Displaced by Big Data?
Evidence from Active Fund Managers

Maxime Bonelli and Thierry Foucault*

November 5, 2024

ABSTRACT

Alternative data provides new signals for active fund managers, but requires specific skills to leverage. Managers lacking these skills could experience a decline in their ability to outperform, unless their expertise produces information distinct from that in alternative data. Consistent with the former, we find that the release of satellite data tracking firms' parking lots significantly reduces fund managers' stock-picking abilities in covered stocks. This effect is stronger for funds leveraging traditional expertise, like industry specialization or geographic proximity, leading them to divest from covered stocks. Our findings suggest that alternative data can reshape the determinants of active funds' performance.

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

JEL classification: G11, G14, G23.

*Bonelli is affiliated with London Business School and can be reached at mbonelli@london.edu. Foucault is affiliated with HEC Paris can be reached at foucault@hec.fr. We thank Laurent Fresard, Joachim Grammig, Jillian Grennan, Nicholas Hirschey, Augustin Landier, Marcel Müller, Marcus Painter, Nikolai Roussanov, Simon Straumann, Russ Wermers and Tianhao Yao, as well as conference and seminar participants at the NBER Big Data and Securities Markets Fall 2023, the European Winter Finance Conference 2024, the DGF conference 2024, the SGF conference 2024, the 2024 Western Finance Association Conference (WFA), the 2024 European Finance Association Conference, Chinese University of Hong Kong, Bocconi University, Hong Kong City University, London Business School, National University of Singapore, Toulouse Business School, UC Irvine, University of Cologne, and University of Macau. 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 asset management industry is experiencing a significant transformation due to the advent of big data and artificial intelligence (AI).[1] Active asset managers produce information about future asset returns with the aim of delivering superior performance. Traditionally, they have done so by relying on expertise derived from specialized industry knowledge, networking, or location advantages, among other factors. Alternative data, such as social media, point-of-sale data, sensors, and satellite imagery, offers new opportunities to gain more precise signals for active fund managers. However, leveraging these data requires new skills and significant investments.[2] As a result, alternative datasets are effectively used by only a small fraction of active fund managers, mostly those using a quantitative approach ("quant funds"). What is the effect of alternative data on other fund managers' performance, that is, those lacking the skills and money to use these data? This question is crucial given the size of the active asset management industry and the considerable interest in understanding the determinants of active fund managers' performance.[3]

The answer is not straightforward. It depends on whether the signals provided by

[1] See for instance "Make way for the robot stock pickers", Financial Times, June 26, 2016, "Fund managers deny AI threatens jobs", Financial Times, August 14, 2017 and "At BlackRock, Machines are rising over managers to pick stocks", The New-York Times, March 28, 2017.

[2] For instance, respondents to a survey of 43 asset managers by Greenwich Associates highlight the costs of alternative datasets and the lack of expertise to process these datasets as major roadblocks to the use of alternative data for investment (see Alternative Data for Alpha). See also the 2024 employer survey results of the CFA institute in which only 7% of respondents indicate that there is no need of upskilling in data analytics of their workforce.

[3] According to the 2024 Investment Company Fact Book, assets managed by active mutual funds totaled \$14 Trillions in the U.S. alone in 2023.

alternative data and those produced by traditional expertise are about the same dimensions of firms' fundamentals (e.g., future sales) or about different dimensions (e.g., future sales vs. investment in intangibles).[4] In both cases, alternative data should enhance stock price informativeness (see Section II for details).[5] However, in the first case, this reduces the expected excess return on private information for "traditional managers" (that is, fund managers relying on traditional sources of information) while in the second case, it could increase it. Indeed, in the latter scenario, the improvement in price informativeness allows traditional managers to better leverage their own signals because it reduces their uncertainty about the dimension of asset payoff on which they have no direct information.[6] Thus, ultimately, whether the introduction of alternative data about a stock reduces or increases traditional fund managers' performance in this stock is an empirical question that, to our knowledge, has not yet been studied.

Our contribution is to fill this gap. To this end, we use the staggered introduction of new alternative data for retailers' stock, namely daily store-level car counts from satellite imagery by RS Metrics (an alternative data provider). This shock is well-suited for our tests because satellite imagery is one important type of alternative data, used by fund managers.[7] Moreover, it generates data that has predictive power for firms' future cash-flows or stock returns (see Zhu, 2019; Kang et al., 2021; Katona et al., 2023) and it enhances price informativeness (Zhu, 2019; Mukherjee et al., 2021). Last, extracting information from the store-level data considered in our tests requires quantitative skills (as the dataset contains several millions of data points).

Our sample comprises "covered" stocks, i.e., stocks for which store-level car counts from

[4]For instance, traditional fund managers might have unique expertise in assessing the effect of investments in intangibles on firm value because doing so requires judgement. In line with this possibility, Cao et al. (2024) find that humans' forecasts outperform machines' forecasts (generated using techniques from artificial intelligence) when institutional knowledge is important, e.g., involving intangible assets and financial distress.

[5]See Zhu (2019), Mukherjee et al. (2021), Grennan and Michaely (2021) for evidence that the availability of new alternative data enhances price informativeness.

[6]This insight is due to Goldstein and Yang (2015). See Section II for further details.

[7] See, for instance, "How satellite imagery is helping hedge funds outperform", International Banker, June 26, 2020; "Stock Picks From Space", The Atlantic, May 2019; and Blackrock's podcast available at "How geospatial data can inform investment decisions" .

satellite imagery become available over our sample period (2009-2020), and control stocks, i.e., stocks for which such data is not available. Moreover, we observe portfolio holdings of about 4,000 different active mutual funds, at the beginning of each quarter. With this data, in each quarter, we measure a fund's stock picking ability in a given stock by *"Picking"*, the product between (i) the stock's subsequent idiosyncratic return at various horizons and (ii) the deviation of its weight in the fund's portfolio from its weight in the market portfolio.[8] Intuitively, *"Picking"* captures a fund's ability to tilt its stock holding in the direction of future return adjusted for systematic risk. Our main tests study the evolution of funds' stock picking ability in treated stocks after these stocks become covered using a difference-in-differences specification, with stock fixed effects and fund-quarter fixed effects (which control for changes in variables affecting the overall stock picking ability of a fund).

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

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

[8] This measure is similar to that used by Kacperczyk et al. (2014) to evaluate the stock picking ability of active mutual funds. The key difference is that our measure is at the fund-stock level while their is at the fund level. See also Grinblatt and Titman (1993a), Daniel et al. (1997) and Jiang and Zheng (2018) for related measures of a fund's skill.

to control stocks begins only *after* the stock becomes covered, and reaches its lowest point approximately one year after coverage initiation.

Our baseline tests measure the average drop in stock picking ability across all funds holding covered stocks in our sample, that is, those who don't buy the data considered in our tests and those who do (we do not observe which funds buy the data from RS Metrics). We then test our additional hypotheses regarding the heterogeneity of this drop across funds in our sample (see Section II). First, we show that this decline is stronger for funds that are more likely to rely on traditional methods to obtain their signals, namely industry expertise (Kacperczyk et al., 2005) or geographical location (Coval and Moskowitz, 2001) as a source of information. For example, we find that the decline in stock picking ability is approximately four times larger for funds located in the same area as the covered stocks' headquarters or major point of sales, highlighting the substantial impact of alternative data on funds that traditionally depend on local informational advantage. Second, as predicted, we show that funds with a higher stock picking ability before the shock considered in our tests experience a stronger decline in their stock picking ability, even after controlling for reversals in funds' performance.

Thus, the availability of new alternative data reduces traditional experts' return on their private information, consistent with the scenario in which the information available in alternative data and produced by experts are substitutes. In a second step, we study how experts respond to this threat to their expertise. First, we show that funds reduce the weight of covered stocks in their portfolios, and this effect is stronger for funds with initially high stock picking ability, industry expertise, or geographical advantage. Some funds even entirely divest from covered stocks: After coverage, the number of funds holding a covered stock drops by about 20%, relative to control stocks.

Furthermore, fund managers reallocate their portfolios to "peer stocks" in which they can exploit their expertise (based on industry or geography) but which are never covered by the data provider considered in our tests. Specifically, we show that the number of funds holding a stock increases by 10% after one of its geographical or industry peer becomes covered. Overall, these findings suggest that fund managers adapt to the rise of alternative data by moving their investments to stocks that are not covered by alternative data.[9] As a result, the way information is produced could become increasingly different between stocks covered by alternative data (which tends to be large capitalization stocks) and stocks that are not.

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

Our paper contributes to the growing literature on alternative data in financial markets. Many papers show that alternative data (e.g., social media data, geolocation data, employee satisfaction data, or satellite images) contain information on firms' fundamentals and stock returns.[10] Fewer papers focus on the effect of alternative data on various types of participants in financial markets (e.g., security analysts, short-sellers, retail investors). Chi et al. (2023) finds that equity analysts mentioning the use of alternative data in their reports have smaller forecasting errors, while Dessaint et al. (2024) finds that the availability of new alternative data (from social media) make short-term (long-term) analysts' forecasts more

[9]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.

[10]See, for instance, Chen et al. (2014) (social media data), Froot et al. (2017) (consumer devices such as mobile phones), or Green et al. (2019) (employee satisfaction data). Dessaint et al. (2024) lists 26 academic papers showing that different types of alternative data has predictive power for future returns (see Table A.1 in their online appendix).

(less) informative. More related to our study, Katona et al. (2023) demonstrate that the introduction of new alternative data increases information asymmetry between sophisticated and unsophisticated investors. Specifically, after a stock become covered by an alternative data provider (RS Metrics), short-selling activity ahead of negative earnings surprises for this stock rises, disadvantaging retail investors. Our findings further show that the availability of alternative data can also negatively impact sophisticated investors, such as active mutual fund managers. Kang (2022) finds that there is a decline of information demand by institutional investors, via the participation to syndicated loans, when borrowers' stocks is covered by satellite imagery. However, this study does not examine the impact of this decline on investors' stock picking abilities (the decline may, in fact, reflect a shift in institutional demand towards alternative data). To our knowledge, our paper is the first to directly investigate this effect.[11]

Our paper is also related to the large academic literature on active fund managers' performance. Fama and French (2010) finds that the average U.S. equity fund manager has no skill (zero alpha on average). However, there is heterogeneity among active mutual managers and some generate significant and persistent abnormal returns for their clients (see Wermers, 2000; Kosowski et al., 2006; Kacperczyk et al., 2014; Jiang and Zheng, 2018; Berk and van Binsbergen, 2015; Barras et al., 2022). One important question is how fund managers obtain superior performance. One possibility is that they trade on private information, as found by Kacperczyk and Seru (2007) and Hendershott et al. (2015),[12] and the literature has identified a few characteristics associated with a fund manager's ability to outperform: education (Chevalier and Ellison, 1999), geographical proximity to firms (Coval and Moskowitz, 2001), educational network (Cohen et al., 2008), and industry expertise (Kacperczyk et al., 2005; Kempf et al., 2017). On this front, our findings show that the rise of alternative data reduces

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

[12]Another possibility is that they offer low cost exposure to systematic factors to investors (Gerakos et al., 2021) or that they receive a compensation for liquidity provision (Anand et al., 2020).

the profitability of traditional methods (such as industry expertise or geographical location) to obtain private information for active mutual managers.

Finally, some papers have used satellite imagery data to explore the extent to which fund managers and security analysts rely on local information. For instance, Kang et al. (2021) use daily store-level car counts from satellite imagery (before it becomes commercially available for funds) to show that fund managers' stock holdings are sensitive to local information, highlighting its importance as a source of expertise. Using similar data, Gerken and Painter (2022) find that equity analysts' forecasting errors for retailers are influenced by local parking lot traffic. Unlike these studies, we exploit satellite imagery data as a shock to the availability of alternative data on covered stocks and focus on how this affects the performance of active mutual fund managers.

The rest of the paper is organized as follows. In Section II, we develop our testable hypotheses regarding the effect of alternative data on the stock picking ability of traditional fund managers. We present the data that we use to test these hypotheses in Section III and the findings of our main tests in Section IV. In Section V, we study how funds reallocate their portfolios after a stock becomes by alternative data. Finally, in Section VI, we show that the the informativeness of a stock price increases after the stock becomes covered by the alternative provider considered in our paper. Section VII concludes.

II Hypotheses Development

Our goal in this paper is to study empirically how alternative data affects the performance of fund managers who do not use this data. To develop the testable hypotheses considered in our tests, we consider in the Internet Appendix (Section I.1) a standard noisy rational expectations model with two types of active asset managers: "experts" and "data miners."[13] Both types receive informative signals about the payoff v of a risky asset (a stock) and trade this asset with noise traders. However, experts and data miners use different technologies

[13] The model is in the tradition of Grossman and Stiglitz (1980) and more specifically related to Dugast and Foucault (2024).

to produce these signals. The precision of experts' signals is fixed (e.g., by education or industry expertise developed over time) and equal to τ_{ex} (index "ex" designates experts). In contrast, data miners can increase the precision of their signals, τ_{dm} (index "dm" designates data miners) when new datasets become available (provided they buy those) because they have the skills to extract information from these datasets.[14] Since our contribution is not theoretical, here we just outline the main implications of the theory for our tests and leave the details of the formal analysis to the internet appendix.

Let $s(\tau_j)$ be the signal of a fund manager with type $j \in \{dm, ex\}$ and let $w(s(\tau_j))$ be the weight of the stock in the fund manager's portfolio (the fraction of her AUM invested in this stock) given this signal. The fund manager's realized excess return on her position in the stock is:

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

where R^e is the excess return on the stock. An informative signal enables a fund manager to tilt her holdings in the stock in the direction of its future excess return, so that $w(s(\tau_j))$ and R^e co-vary positively. Thus, an informed manager's expected excess return, that is, $\bar{R}(\tau_j) \equiv \mathrm{E}(R(s(\tau_j))$ is strictly positive if her signal is informative. In equilibrium, $\bar{R}(\tau_j)$ increases with the precision of the fund manager's signal and decreases with the informativeness of the stock price (see eq.(IA.12) in the Internet appendix).[15] Thus, $\bar{R}(\tau_j)$ is a measure of a fund's *stock picking ability* (everything else equal, the higher is τ_j, the higher the average return of the fund in the stock).

The average stock picking ability *across* experts and data miners, $\bar{R}^A$, is (by definition):

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

where μ is the fraction of capital controlled by data miners. The average stock picking ability across funds, $\bar{R}^A$, is always strictly positive in equilibrium because active asset managers

[14] We assume for simplicity that all fund managers in a given group have signals of the same precision. This assumption can be relaxed without affecting the conclusions.

[15] As is standard, the informativeness of the stock price is defined as the inverse of the variance of the asset payoff conditional on the price of the asset $(Var(v \mid p)^{-1})$.

make trading profits at the expense of noise traders. In other words, trading is not a zero sum game among active asset managers because of the presence of noise traders.

An increase in the precision of data miners' signals improves price informativeness and, as a result, reduces experts' average stock picking ability ($\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}} < 0$; see eq.(IA.13) in the Internet Appendix). As a result, for μ not too large, an increase in the precision of data miners' signals reduces the average stock picking ability of active fund managers as a group ($\frac{\partial \bar{R}^A}{\partial \tau_{dm}} < 0$) because the drop in the average performance of experts as a group more than offsets the increase in the average performance of data miners as a group.[16] We interpret the availability of new alternative data for a stock as a positive shock to τ_{dm}.[17] This yields our first testable hypothesis.

H.1: The introduction of alternative data for a stock reduces the average stock picking ability of fund managers holding this stock.

A second implication is that the drop in the average stock picking ability of an expert should be stronger than the drop in the average stock picking of all funds (that is, $\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}} < \frac{\partial \bar{R}^A}{\partial \tau_{dm}}$) since the performance of data miners increases when new alternative data becomes available ($\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} > 0$ for μ not too large). This yields our second testable hypothesis.

H.2: The introduction of alternative data for a stock has a stronger negative effects for experts than for the average fund.

Finally, we show in the Internet Appendix that the drop in the stock picking ability of experts following the introduction of alternative data is stronger when the precision of their signal, and therefore their stock picking ability before the shock, is higher (that is,

[16] The fraction of capital allocated to data miners does not need to be very small for this to happen (e.g., we give a numerical example in the internet appendix in which $\mu = 40\%$). In practice, quant funds still represent a relative small fraction (about 7% of mutual funds; see Abis, 2022) of total assets under management by active funds. We assume that all data miners use new alternative data. If instead only a fraction of data miners experience an increase in the precision of their signal (e.g., because only a few buy the new data), the effect is even stronger (as if μ was smaller) because the data miners who do not buy the data are also negatively affected.

[17] Our testable implications do not depend on whether or not data miners have a signal of higher precision than experts. Thus, we do not need assumptions on the ranking of τ_{dm} and τ_{ex} before and after the availability of new alternative data.

$\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{ex} \partial \tau_{dm}} < 0$). This yields our third testable hypothesis.

H.3: The introduction of alternative data for a stock has a stronger negative effect for funds with a high stock picking ability before the introduction of alternative data.

In the model we use to derive these testable hypotheses, data miners and experts receive signals about v, that is about the same dimension of uncertainty for the asset fundamental. In Section I.1.3 of the online appendix, we consider an alternative formulation in which $v = v_{dm} + v_{ex}$. Experts receive a signal about v_{ex} and data miners receive a signal about v_{dm} if alternative data is available about the asset. That is, experts and data miners produce information on different determinants of the asset fundamental, v.[18] This version of the model is similar to Goldstein and Yang (2015).

In this case, the availability of alternative data about v_{dm} triggers an improvement in price informativeness, as in the baseline case. However, in contrast to the baseline case, it can increase the expected excess return of experts and the average expected excess return of all fund managers ($\bar{R}^A$). Indeed, an increase of the mass of traders informed about one dimension of uncertainty of the asset payoff, say, v_{dm}, can increase the value of being informed about the other dimension, v_{ex}. The reason is that the stock price becomes more informative about v_{dm} for experts, which reduces their uncertainty about the asset payoff and enables them to trade more aggressively on their information about v_{ex}. When this "uncertainty reduction" effect (using the terminology of Goldstein and Yang, 2015) is strong enough, experts' expected excess return increases when data miners have access to information about v_{dm} relative to the case in which they don't. Moreover, even when this is not the case, the uncertainty reduction effect dampens the negative effect of an increase in price informativeness on experts' expected excess returns so that the average expected excess returns across all funds increases following the availability of alternative data on v_{dm}. In sum, under this scenario, Hypotheses H.1 and H.2 can be reversed. As explained below, our empirical findings are not consistent with this

[18] For instance, experts may have unique abilities to assess the potential of the development of new products for a firm or long-run implications of a new strategy while alternative data could be useful to gauge existing demand for a firm's products.

alternative scenario.

III Data and Measurements

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

A Data

Our first data set is from the Center for Research on Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database. This database provides comprehensive information about funds, such as their returns and size (total net assets). We focus our analysis on US domestic equity funds from January 2009 to December 2020, for which the holdings data (described below) are most complete and reliable.[19] 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.[20] To address the possibility of incubation bias, we further exclude observations if the year of the observation is prior to the reported fund starting year or if the name of the fund is missing. From this database, we also obtain the contact information of each fund to determine its geographical location.

Our second data set is from the CRSP Mutual Fund Holdings database. This database provides stock holdings of mutual fund portfolios, collected both from mandatory SEC reports by mutual funds and voluntary reports. We use portfolio holdings disclosed by funds at the end of each quarter.[21] We further filter our sample by excluding funds that do not hold more than 80% of equities as well as funds that in the previous month had less than $5 million of assets under management (as in Kacperczyk et al., 2014). Our final sample features 3,962 funds holding 9,781 distinct stocks.

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

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

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

For the companies held by mutual funds, we gather monthly stock returns from CRSP and accounting variables 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.[22]

Our last data set comes from Remote Sensing (RS) Metrics (see `https://RSMetrics.com/`). RS Metrics is the first data provider to have commercialized store-level parking lot traffic data for U.S. retail firms based on satellite imagery (see Figure A.1 in the Appendix for examples) and its clients include asset management firms. RS Metrics track parking lots for more than 65,000 retail store locations around the United States. Each store is monitored multiple times in a given month, which enables RS Metrics to provide daily store-level information (using various techniques to analyze images) about parking lot capacity and utilization, across major U.S retailers (e.g., Walmart, Home Depot, Best Buy, Starbucks, Tractor Supply Co., etc.).[23]

We obtain from RS Metrics their historical dataset on parking lot traffic, which includes the exact dates at which it starts collecting 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). For each retailer in our sample, we use this date as shock to the availability of alternative data about this retailer. We refer to this date as the date at which RS metrics "initiates coverage"

[22] 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.

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

of the retailer's stock, and we say that a stock is "covered" if at some point during our sample period RS Metrics initiates coverage of this stock. We do not observe the date at which some funds start buying data from RS Metrics but press reports suggest that some firms began using RS Metrics' data as early as 2010, (see https://www.cnbc.com/2010/08/16/new-big-brother-marketmoving-satellite-images.html), that is, soon after the RS Metrics initiates coverage of the first stock in our sample.[24] Figure A.2 and Table B.1 in the Appendix show the number of new coverage initiations in each quarter during our sample period and the different industries (NAICS sectors) of the stocks covered by RS Metrics, respectively.

The number of alternative datasets (credit card data, email receipts, geolocation data, etc.) relevant for a stock has been steadily growing over time (see Dessaint et al., 2024). Thus, some stocks in our sample might be covered by other types of alternative data providers before RS Metrics initiates coverage of these stocks. Therefore, what we measure is the marginal effect of introducing a new alternative data set for a stock on the stock-picking ability of funds holding this stock. This effect must be weaker than the total effect of all alternative data about a stock, which works against finding any effect if parking lot traffic data contains no or little incremental information relative to other types of alternative data.

B Measuring Funds' Stock Picking Ability

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

$$Picking^{h}_{f,i,t} = 100 \times (w^{f}_{i,t} - w^{m}_{i,t})(R^{i}_{t+h} - \beta_{i,t} R^{m}_{t+h}), \tag{3}$$

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

[24] RS Metrics typically makes its data available via its "Traffic Signals™" product. A fund subscribing to "Traffic Signals" gets access to data about all stocks covered by RS Metrics at the time of the purchase.

different horizons in our empirical analysis, namely $h = 1$, 2, 3 and 4 quarters.

A fund's stock picking ability in stock i, $Picking^h_{f,i,t}$ is a proxy for $R(s) = w(s) \times R^e_i$ in eq.(1), where $w(s)$ is measured by $(w^f_{i,t} - w^m_{i,t})$ and $R^e_i = (R^i_{t+h} - \beta_{i,t}R^m_{t+h})$. It should be positive on average if $(w^f_{i,t} - w^m_{i,t})$ and $(R^i_{t+h} - \beta_{i,t}R^m_{t+h})$ are positively correlated, that if fund f tilts its position in stock i relative to a passive benchmark (the market portfolio) in the direction of its future idiosyncratic return, suggesting that the fund exploits useful signals. This measure is closely related to that used by Kacperczyk et al. (2014) (see also Jiang and Zheng, 2018). The main difference is that we measure picking at the fund-stock level while Kacperczyk et al. (2014) measures it at the fund level (by summing the stock-fund picking ability measure across stocks held by each fund).[25]

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

$$Trading^h_{f,i,t} = 100 \times (w^f_{i,t} - w^f_{i,t-4})(R^i_{t+h} - \beta_{i,t}R^m_{t+h}), \tag{4}$$

where $w^f_{i,t-4}$ is the weight of stock i in fund f in quarter $t-4$.[26] This measure captures the ability of fund f to change its holdings in a given stock in the direction of subsequent excess returns. In contrast to the two other measures, it is based on trades (change in funds' holdings in a given stock) rather than on the deviation of the fund's holdings relative to a benchmark portfolio. Our results with these alternative measures are qualitatively similar to those obtained with $Picking^h_{f,i,t}$ (reported in Section IV). Thus, we only report them in the

[25] Kacperczyk et al. (2014) also introduces another measure ("*Timing*") of a funds' ability to anticipate a stock's systematic return ($\beta_{i,t}R^m_{t+h}$ in our notations). The alternative data considered in our tests are unlikely to affect a fund's "*Timing*" ability since (i) signals about a retailers' sales are likely to be very noisy signals of market returns and (ii) parking lots counts are not obviously related to a retailers' *beta*s. Consistent with this, we show in Section I.5 of the Internet Appendix that the coverage of a stock by RS Metrics has no effect on funds' timing ability.

[26] 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).

Internet Appendix (see Section I.4). For brevity, we do not systematically recall this point when discussing the results regarding $Picking^h_{f,i,t}$.

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

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.[27]

[Insert Table I about here]

[27] 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 hypotheses because (i) satellite traffic data about a firm contain information useful to forecast future sales and stock returns for this firm (see Zhu, 2019; Katona et al., 2023, and Section I.3 in the Internet Appendix for evidence), and (ii) analyzing these data to extract meaningful signals likely requires data science expertise. Notably, the RS Metrics data set tracks parking lots for more than 65,000 retail store locations around the United States, yielding several millions of historical data points to process. Furthermore, there is anecdotal evidence that some funds (most likely quant funds) use these data (see Footnote 7). Thus, as explained in Section II, coverage of a stock by RS Metrics should allow these funds to obtain higher precision signals about this stock and therefore *reduce* other funds' stock picking ability in this stock (in particular, those who rely solely on traditional human expertise to obtain their signals).

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

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

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

The coefficient β measures the extent to which a fund's stock-picking ability in a given stock is affected by the RS metrics' initiation of coverage for this stock. Our first hypothesis (H.1) implies that $\beta < 0$. One concern is that a fund's stock picking ability might vary over time due to unobserved factors, not necessarily related to alternative data. For example, a fund might decide to beef up its team of analysts. If such decision affects a fund's stock picking ability and is correlated with the availability of alternative data, it would bias our estimate of the effect of RS Metrics' coverage initiation on a fund's stock picking ability, β. The inclusion of fund×quarter fixed effect in our specification (eq.(5)) addresses this problem. Indeed, it controls for all unobserved variables that might affect a fund stock picking ability in a given quarter. Given the fund×quarter fixed effects, β captures the change in a fund's picking ability after a stock becomes covered by RS Metrics relative to stocks in the *same* fund's portfolio not covered by RS Metrics.

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

[Insert Table II about here]

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

Consistent with hypothesis H.1, β is always significantly negative for all horizons in all specifications (even after controlling for stock and fund-quarter fixed effects in columns (2), (5), (8) and (11)). Thus, funds stock picking ability declines after RS Metrics start covering these stocks. The magnitude of this decline is economically significant. For instance, consider the four-quarter horizon where the mean of "Picking" is zero, the standard deviation is 0.361, and only 5% of the realizations exceed 0.4 p.p. (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 p.p. after coverage initiation, which is nearly one-third of its standard deviation. In columns (3), (6), (9) and (12), we include fund×stock fixed effects. Results are unchanged with this approach.

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

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

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 coverage initiaton for this stock ($k \in \{-12, -11, ..., 15, 16+\}$). 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 I about here]

Figure I plots the estimates of the β_ks' in eq.(6) when Picking is measured at the yearly horizon (dashed lines correspond to 95% confidence intervals). We observe that these estimates are not significantly different from zero before coverage and become significantly negative only after coverage starts. Thus, the evolution of a fund's Picking ability is not different for control and covered stocks *before* coverage initiation, indicating no observable "pre-treatment" trend. The drop in funds' Picking ability for covered stocks only after coverage initiation supports our interpretation that coverage initiation is the cause, and not the consequence, of this drop.

Notably, the decline in Picking in covered stocks starts rapidly after coverage starts, reaching its lowest point approximately one year later. Specifically, within four quarters following coverage initiation, there is a significant drop of over 0.1 p.p. in "Picking." While we do not observe the exact timing of data acquisition and processing by market participants, this finding suggests that some (e.g., quant funds) swiftly exploit RS Metrics' data after they become available, thereby quickly reducing other funds Picking's ability. Although there is a slight rebound in funds' picking skills for covered stocks after the initial decline, the effect remains negative over time after coverage begins. Moreover, Picking does not revert to pre-coverage level even after four years, indicating a persistent and enduring impact of coverage initiation on Picking.

B Cross-Sectional Heterogeneity

B.1 Experts are more Negatively Affected (H.2)

The tests in the previous section show that the availability of new alternative data for a stock has a negative effect on active funds' average (across all funds) stock picking ability in this stock ($\bar{R}$ in eq.(2)). However, as explained in Section II, this negative effect should be more pronounced for traditional experts, that is, asset managers relying on traditional methods to obtain their signals. In this section we test whether this is the case.

The literature suggests two ways in which active mutual funds have traditionally obtained private information: (i) industry expertise (Kacperczyk et al., 2005), and (ii) local information (see Coval and Moskowitz, 2001; Bae et al., 2008; Chen et al., 2022; Cicero et al., 2023). In particular, funds that are located in proximity to headquarters or the stores of firms covered by RS Metrics are better able to benefit from networking opportunities with managers of these firms or to directly observe sales by these firms. For instance, Kang et al. (2021) show that institutional investors adjust their holdings of retailers stocks in response to sales of stores in their geographical area (local stores) and that these trades are profitable. Funds that rely on these sources of expertise to generate their signals should not see a decline in the quality of their signals following RS metrics coverage. However, this shock should reduce their ability to profitably trade on their signals according to our hypothesis H.2. (see Section II).

Industry Expertise

To study the role of industry expertise, we proceed as follows. In the spirit of Kacperczyk et al. (2005), 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 (Table B.1 in the Appendix).[28]

We then estimate the following specification:

$$
\begin{aligned}
Picking^h_{f,i,t} = \beta\{Covered \times Post\}_{i,t} \\
+ \beta_{expert}\{Covered \times Post\}_{i,t} \times IndustryExpert_{f,i} \\
+ \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t},
\end{aligned}
\tag{7}
$$

where $IndustryExpert_{f,i}$ is a dummy variable equal to 1 either if fund f is a"Sector Fund" or if it is an "Industry Specialist" for the industry of stock i. $\alpha_{f\times i}$ are fund×stock fixed effects. Other variables are the same as in eq.(5) and standard errors are double-clustered

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

at the fund and stock levels. The coefficient of interest is β_{expert}, which measures the effect of coverage initiation of a given stock on industry experts' stock picking ability, above and beyond the baseline average effect (measured by β) for all funds holding this stock.

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

More important, 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). Thus, funds with high stock selection ability, in this case resulting from their industry specialization, are more negatively affected by the availability of new alternative data for a stock, in line with hypothesis H.2.

Geographical Location

To measure location-based expertise, we identify, for each covered stock, the metropolitan statistical areas (MSAs) corresponding to (i) the firm's headquarters (sourced from Compustat) and (ii) the location where the highest number of the firm's parking lots are located (the stock's "primary MSA"). Then, we categorize a fund as "Local" for a specific stock if it is located in the same MSA as either (i) the firm's headquarters or (ii) the stock's primary MSA.[29] We do not exclusively rely on firms' headquarters for identifying local funds because, among the retailers covered by RS Metrics in our sample, very few share the MSA of their headquarters with funds. Specifically, only 0.10% of observations in our sample correspond to funds categorized as "Local" based on the MSA of a covered firm's headquarters. In contrast, this percentage increases to 6% when "Local" is defined using both the firm's headquarters and the primary location of its parking lots.[30] We then estimate the following specification:

$$\begin{aligned} Picking^{h}_{f,i,t} = {} & \beta\{Covered \times Post\}_{i,t} \\ & + \beta_{local}\{Covered \times Post\}_{i,t} \times Local_{f,i} \\ & + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{8}$$

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 IV about here]

Table IV reports estimates of eq.(8). In columns (1), (3), (5), and (7), we estimate this equation without fund-stock fixed effects but we include stock fixed effects and the interaction variable $Covered_{i,t} \times Local_{f,i}$ equal to 1 if fund f is classified as "Local." The coefficient on this variable is positive and significant. Thus, mutual funds located in areas where covered companies have the highest number of their stores or their headquarters have a higher stock picking ability prior to coverage initiation. More important for our purpose, in all specifications, the coefficient, β_{local}, is negative and significant. Thus, following coverage initiation, local funds experience a more pronounced decline in their stock picking ability

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

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

than other funds holding covered stocks (for which we also find a drop in stock picking ability; β is significantly negative in all specifications).[31] 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 that of other funds.

B.2 High-Skilled Funds are more Negatively Affected (H.3)

Our third hypothesis (H.3) is that the availability of new alternative data for a stock should have a stronger negative effect on the stock picking ability of funds that show a high stock picking ability before the data is available. To test this hypothesis, we estimate:

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

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. (7) and standard errors are double-clustered at the fund and stock levels. In estimating eq. (9), we exclude observations for funds that began holding covered stocks after coverage initiation since, by definition, $HighPickingPre_{f,i}$ cannot be observed for these funds.

Again, we include fund×stock fixed effects, $\alpha_{f\times i}$, to make sure that the coefficients β and β_{high} are estimated by comparing picking abilities before versus after coverage initiation for the *same* fund and stock. Because $HighPickingPre_{f,i}$ is constant over time for a given fund-stock, the interaction term $Covered_i \times HighPickingPre_{f,i}$ is absorbed by the fund-stock

[31] This finding is different from Kang et al. (2021), who use satellite imagery data of retailers' parking lots before they become commercially available to show that some funds rely on local information to make investment decisions. Indeed, this finding does not per se imply that local information becomes less valuable when a stock becomes covered by a satellite imagery data provider, which is what Table IV shows.

fixed effects.

[Insert Table V about here]

Table V presents the estimates of eq.(9). 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.

A potential concern is that the above results may be due to reversals in mutual funds' stock picking abilities.[32] 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:

$$\begin{aligned} Picking^h_{f,i,t} = & \sum_{k=-12,k\neq -1}^{16+} \beta_k \{Covered \times Quarter(k)\}_{i,t} \\ & + \sum_{k=-12,k\neq -1}^{16+} \beta_{high,k} \{Covered \times Quarter(k)\}_{i,t} \times HighPickingPre_{f,i} \\ & + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{10}$$

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 II about here]

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

Panel A and B of Figure II plot the estimates of respectively the β_ks' and the $\beta_{high,k}$s' for each quarter in eq.(10) 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 -0.3 p.p, and does not revert to pre-coverage level even after four years.

Even though Panels A and B of Figure II depict no discernible pre-trend in picking abilities before coverage initiation, funds with abnormally high picking ability levels might experience a subsequent reversal, i.e., a relative decline post-coverage due to *Picking* reverting to its mean level. To further address this concern, we re-estimate equations (5) and (9) with controls for the lagged value of "Picking". If our previous results are just mechanically due to reversals in funds' performance, the lagged value of picking should subsume the effect of coverage initiation on funds with high picking abilities pre-coverage.

The estimation results are presented in Appendix Table B.2. Across all specifications, we observe a negative coefficient on the lagged value of Picking ("Past Picking"), between -0.065 (for Picking 1-Q) and -0.207 (for Picking 4-Q). Thus, there is relative reversal in picking abilities at the *fund-stock* level. For instance, assuming lagged Picking 1-Q is 0.19 p.p. (its 95th percentile), this translates to a reduction of current Picking 1-Q by -0.065 $\times$ 0.19 = -0.01 p.p.. More important for our purposes, the coefficients on the interaction term "Covered $\times$ Post" and on the triple interaction "Covered $\times$ Post $\times$ High Picking Pre" remain significantly negative, and their magnitudes align with the estimates in Tables II and V. Thus our findings are unlikely to be driven by reversals in funds' stock picking abilities.

B.3 Is the Effect Weaker for Quant Funds?

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

As explained previously, we conjecture that quant funds are more likely to buy alternative data as they have the skills required to use them. Identifying these funds in our sample is not straightforward as we do not directly observe a fund's type. To overcome this issue, we build proxies for quant funds, exploiting the text of mutual fund prospectuses, as explained in Section I.6 of the Internet Appendix. We search for specific keywords, such as "quantitative stock selection", in the strategy section of fund prospectuses which provides information regarding funds' investment process.[33] We then estimate a specification similar to eq.(7), interacting "Covered × Post" with various proxies for quant funds. Internet Appendix Table IA.8 presents estimation results of these tests, and shows that the coefficients on these triple interaction terms are positive and statistically significant in most specifications. Thus, as predicted, coverage initiation has a weaker negative effect on funds that are more likely to be able to use the data considered in our tests. Our inference however is limited by our inability to directly observe funds that buy the data and by the fact we only focus on mutual funds. Indeed, we expect sophisticated investors out of our sample, such as hedge funds, to also use alternative data, and therefore to benefit after coverage initiation.

C Robustness and Additional Tests

C.1 Selection of Covered Stocks and Matching

One concern is that there are time-varying variables that both affect RS Metrics's coverage decision and funds' stock picking ability in covered stocks differently than for control stocks.

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

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

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

Interestingly, the number of mutual funds holding the stock (a measure of potential demand) and funds' lagged stock picking ability do not predict coverage initiation (see Columns (4) and (5)). This observation, together with the results from the previous section, mitigates the possibility of reverse causality (RS Metrics initiating coverage in response to a strong demand by funds because of a decline in the quality of their information). Lastly, measures of stock price informativeness and analysts' forecasts errors similarly do not appear to be predictors for coverage initiation (see Column 6).[34] Again, this suggests that RS Metrics' decision to cover a stock is not driven by a decline in the informativeness of a firm's stock price.

In sum, the only robust predictor of coverage for a stock is market capitalization, maybe

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

because data providers anticipate higher demand for large-cap stocks, allowing them to better amortize the fixed costs of cleaning up and preparing the data.[35] Another potential factor is that RS Metrics may have prioritized firms with more parking lots to ensure better data accuracy, which tends to correlate with larger market capitalizations.

As a robustness test, we re-estimate our specifications (5), (7), (8) and (9), with a matched group of control stocks. Specifically, one year prior to the initiation of coverage, we match each covered stock with five non-covered stocks based on industry (NAICS 2-digit sector) and market capitalization. The results with this matching approach are presented in the Internet Appendix (Section I.7) and confirm those discussed above.

We also conduct supplementary robustness tests of our cross-sectional heterogeneity results. Specifically, we re-estimate regressions (7), (8) and (9) adding stock-quarter fixed effects. These fixed effects capture time-varying fluctuations in picking abilities that are common across all funds holding a specific stock. As such, they absorb the interaction term $Covered \times Post$ and control for any time-varying variables at the stock level that may affect both RS Metrics's decision to cover a stock and funds' picking abilities (e.g., changes in the stock's factor exposures). This specification still enables us to estimate the coefficient on the triple-interaction terms between $Covered \times Post$ and each measure of expertise. The estimation results are presented in Table IA.13 (Internet Appendix) and remain consistent with the results discussed earlier.

C.2 Cohort Analysis

RS Metrics is the first data vendor starting to sell parking lot traffic data. A competitor, Orbital Insight, introduced a similar product in 2014,[36] and began distributing its parking lot data through Bloomberg terminals in 2018, potentially making satellite data more accessible

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

[36] See https://www.globenewswire.com/en/Orbital-Insight-Expands-U-S-Retail-Traffic-Product-to-More-Than-100-Retailers.html.

for integration into trading strategies.[37] It is therefore possible that the effect of RS Metrics coverage on funds' stock picking ability declines after 2014 and 2018 simply because, in some cases, stocks covered by RS Metrics are already covered by Orbital Insight.

To address this point, we estimate multiple specifications focusing on different cohorts of covered stocks based on their coverage initiation dates. Specifically, we use a "stacked" difference-in-differences approach, as in Gormley and Matsa (2011) and Cengiz et al. (2019), which is robust to treatment heterogeneity across groups and periods when treatment is staggered. This method involves creating separate cohorts for each coverage initiation date, with distinct time windows before and after the treatment. These cohorts are then "stacked" into a panel for estimation. For each treatment date, we construct a 10-year window of treated and control observations (5 years before and 5 years after the coverage initiation), using stocks that are never covered as controls. We then estimate our main specification (5) on this stacked cohort panel, allowing fixed effects to vary by cohort.

The results are reported in the Internet Appendix (Table IA.14). They show that across all specifications and cohorts—whether including treatment years up to 2013 (pre-Orbital Insight), 2015, or 2017 (pre-Bloomberg)—the results remain consistent. This suggests that the effects we capture are unlikely to be driven by the introduction of Orbital Insight's parking lot data or its subsequent integration into Bloomberg terminals. Additionally, we re-estimate our main specification using only data from the pre-2018 period. The results, reported Table IA.15 (Internet Appendix), remain qualitatively similar, supporting the robustness of our findings.

C.3 Short-Term Trading

One limitation of our empirical analysis is that we cannot observe funds' within-quarter trades. If funds execute short-term trades based on parking lot car counts (e.g., a few days before firms' earnings) and close their positions before quarter-end, the CRSP quarterly holdings data will not reflect these trades. Thus, one concern might be that, while we observe

[37]See https://www.bloomberg.com/news/articles/2019-05-15/orbital-insight-uses-satellite-imagery-to-interrogate-the-world

a decline in funds' stock picking ability over various horizons (from one to four quarters), their short-term trading ability may still improve. To address this concern, we conduct an additional analysis using the return gap measure from Kacperczyk et al. (2008). The return gap reflects the difference between a fund's reported returns (capturing both short-term trades and unobserved actions that occur between holding disclosure dates) and the returns implied by its disclosed quarterly holdings. Our findings (reported in Table IA.16 of the Internet Appendix) indicate that there is no significant change for the return gap of of funds holding covered stocks before and after the availability of satellite data. It is therefore unlikely that the (unobserved) short-term trading ability of mutual funds between quarterly disclosure dates improves after a stock becomes covered.

C.4 Decomposing Picking Skills

In our empirical analysis, a fund's picking ability in a given stock is the product of (i) the fund's tilt towards the stock (relative to the market) and (ii) the stock's abnormal return (see equation (3)). This measure captures how a fund's holding of each stock, relative to the market, comoves with the idiosyncratic component of the stock return. Indeed, using eq.((3)), we have

$$\mathbb{E}(Picking^h_{f,i,t}) = 100\times\Big\{\mathbb{E}(w^f_{i,t} - w^m_{i,t})\mathbb{E}(R^i_{t+h} - \beta_{i,t}R^m_{t+h}) + \mathrm{Cov}(w^f_{i,t} - w^m_{i,t}, R^i_{t+h} - \beta_{i,t}R^m_{t+h})\Big\} \tag{11}$$

Thus, the drop in the average stock picking ability of funds in covered stocks can stem from a drop in $\mathrm{Cov}(w^f_{i,t} - w^m_{i,t}, R^i_{t+h} - \beta_{i,t}R^m_{t+h})$ or a drop in $\mathbb{E}(w^f_{i,t} - w^m_{i,t})\mathbb{E}(R^i_{t+h} - \beta_{i,t}R^m_{t+h})$ or both. We expect the first effect to drive the findings since the mechanism presented in Section II does not imply that the release of alternative data should affect the average abnormal return of a stock. To check this point, we estimate specification (5), using either the fund's tilt towards the stock relative to the market, the stock's return, the stock's beta or the stock's abnormal return as dependent variable. Results are reported in Table IA.17 of the Internet Appendix and, as expected, there is no significant change in these variables after a stock becomes covered by RS Metrics. Thus, the drop in Picking documented in the previous section stems from a drop in funds' ability to tilt their holdings in a stock in the direction of future abnormal returns ($\text{Cov}(w^f_{i,t} - w^m_{i,t}, R^i_{t+h} - \beta_{i,t} R^m_{t+h})$).

V Adaptation

In line with our hypotheses, the availability of new alternative data for a stock reduces the value of traditional expertise in this stock. In this section, we study how fund managers adapt to this evolution. We do not expect them to immediately or necessarily switch to other occupations. Instead, they could adapt by reducing the capital allocated to the stocks in which their expertise is less valuable and shift it to stocks in which this expertise remains valuable. We examine whether this is the case in Sections V.A and V.B, respectively.

A Divestment

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

We then estimate specification (5) using as dependent variable the (negative) natural logarithm of the stock rank in the fund portfolio (so that a higher value of the dependent variable corresponds to a greater investment in a stock).[38] Results are presented in columns (1) and (2) of Table VI. After coverage initiation, covered stocks experience a 9% decrease in their rank, that is, a fall of 11 places (since the average fund in our sample holds 123 stocks). We emphasize that the regression in column (2) incorporates fund-stock fixed effects, ensuring that the coefficient on "Covered × Post" is estimated by comparing a given fund's holdings

[38]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".

in covered stocks after and before 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 including triple interactions as in equations (7), (8) and (9), using the negative of the logarithm of the stock rank as the dependent variable. Table VII presents the results. Consistent with our previous results, we find that funds classified as superior stock pickers before coverage initiation, industry specialists, and "local" funds divest to a larger extent from covered stocks after coverage initiation.

[Insert Table VII about Here]

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

[Insert Figure III about Here]

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

B Assets Reallocation

We now examine whether, after coverage, funds redirect their investments towards uncovered stocks in which the value of their expertise is valuable. To test this conjecture, we study, around coverage initiations, the evolution of the number of funds holding non-covered stocks in the same industry or same 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 highest number of parking lots are in the same Metropolitan Statistical Area (MSA) as the focal stock's headquarters.[39]

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

[Insert Table VIII about Here]

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

This last finding suggests that coverage of a stock might, indirectly, affect its peers even though they are not covered. To rule out the possibility that such spillover effects from covered stocks to non-covered stocks contaminate our main results, we re-estimate our main specifications regarding the effect of coverage on funds' stock picking (those reported in Table II), omitting non-covered peer stocks from the set of control stocks. Our findings with this approach are unchanged (see Table IA.18 in the Internet Appendix).

VI Mechanism: The Role of Price Informativeness

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

A Measuring Price Informativeness

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

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

B Empirical Findings

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

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

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, that is, NAICS sectors where RS Metrics provides satellite imagery data for at least one company (cf., Appendix Table B.1).

[Insert Table IX about here]

Table IX reports estimates of eq.(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. (2023) find no statistically significant effect of the introduction of satellite data on stock price informativeness. They conjecture that this result could be due to the fact the effect of satellite data on price informativeness takes time to materialize and provide evidence consistent with this conjecture. The difference between their findings and ours may reflect the fact that our sample is not exactly the same (in particular, our sample ends in 2020 while their ends in 2017) and that our measure of stock price informativeness is not defined exactly in the same way. In any case, the novelty of our paper is not to study the effect of alternative data on price informativeness but its effect on funds' stock picking ability.

C Better Private Signals or Better Public Signals?

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

Nevertheless, to study whether the public information channel plays a role, we study whether sell-side equity analysts' earnings forecasts become more accurate after a stock becomes covered since equity analysts are an important source of public information in stock markets (see, for instance, Lee and So, 2017; Chen et al., 2020). For this, we use data from I/B/E/S to calculate standardized analysts' earning forecast errors for each stock in industries covered by RS Metrics. For every quarter, we compute these errors by taking the absolute difference between the actual earnings and the average of the latest analyst forecasts, divided by the standard deviation of those forecasts. We compute such forecasting errors for the subsequent quarter, for the second, third, and fourth quarter ahead, as well as for the forthcoming year. We then estimate eq. (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 Appendix Table B.4, reveal no significant coefficients on the interaction term between *Covered* and *Post*. Thus, there is no significant reduction in analysts' forecasts errors after a stock becomes covered.[40] Thus, the improvement in price informativeness for covered stocks is unlikely to be due to an improvement in the quality of analysts' forecasts for these stocks. Of course, the availability of new alternative data could enhance the accuracy of public signals for investors via other sources than sell-side analysts' forecasts (e.g., managerial disclosure).

VII Conclusion

This paper shows that the availability of satellite imagery data tracking retailer firms' parking lots reduces the average stock picking ability of active mutual fund managers in stocks covered by this data. This decline is particularly pronounced for funds that heavily rely on traditional sources of expertise, such as industry specialization and geographical location. This finding is predicted by a standard model of trading on private information in which (i) the availability of new alternative data for a stock is a positive shock to the precision of signals about this stock for funds who can use alternative data and (ii) the signals of these funds are on the same dimension of fundamentals as the signals produced by traditional fund managers.

We obtained these results by examining the introduction of one type of alternative data (derived from satellite imagery) for a limited subset of stocks. A natural question is whether similar results would hold for other types of alternative data. The theory suggests that this is not necessarily the case for alternative datasets providing insights into aspects of stock fundamentals distinct from those traditionally analyzed by fund managers.[41] The introduction of such datasets could, in theory, enhance the expected performance of traditional

[40]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. (2023) 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. (2023), only 3% of analysts' reports in their data mention the use of data based on satellite imagery as a source of information.

[41]For example, as shown by Bonelli (2023) in the context of startup financing, data-driven venture capitalists could lack the "human" expertise to identify highly innovative startups with rare major successes.

fund managers (see Section II). Future research could therefore focus on testing whether the effects of alternative datasets on traditional fund managers' expected returns depend on the types of fundamental information they provide.

Our findings also indicate that satellite imagery coverage of a stock is associated, on average, with a decrease in active mutual fund holdings of covered stocks and a shift by fund managers towards stocks in which they retain expertise but that lack alternative data coverage. This adaptation by traditional fund managers in response to the rise of alternative data suggests a potential segmentation in information production: stocks covered by alternative data could increasingly rely on quant funds for information generation, while stocks without such coverage could continue to depend on traditional fund managers leveraging judgment and expertise. Investigating whether this segmentation develops, as well as its implications for asset prices and capital allocation, represents a promising direction for future research.

References

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

Anand, A., Jotikasthira, C., and Venkataraman, K. (2020). Mutual Fund Trading Style and Bond Market Fragility. The Review of Financial Studies, 34(6):2993–3044.

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

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.

Bonelli, M. (2023). Data-driven investors.

Cao, S., Jiang, W., Wang, J., and Yang, B. (2024). From man vs. machine to man + machine: The art and ai of stock analyses. Journal of Financial Economics, 160:103910.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics, 134(3):1405–1454.

Chen, H., De, P., Hu, Y., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. Review of Financial Studies, (27):1367–1403.

Chen, H., Qu, Y., Shen, T., Wang, Q., and Xu, D. X. (2022). The geography of information acquisition. Journal of Financial and Quantitative Analysis, 57(6):2251–2285.

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. (2023). The use and usefulness of big data in finance. Working paper, Cornell University.

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

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

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

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

Dessaint, O., Foucault, T., and Fresard, L. (2024). Does alternative data improve financial forecasting? the horizon effect. Journal of Finance, 79:2237–2287.

Dugast, J. and Foucault, T. (2024). Equilibrium data mining and data abundance. Forthcoming Journal of Finance.

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.

Froot, K., Kang, N., Ozik, G., and Sadka, R. (2017). What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? Journal of Financial Economics, 125:143–162.

Gerakos, J., Linnainmaa, J. T., and Morse, A. (2021). Asset managers: Institutional performance and factor exposures. Journal of Finance, 76(4):2035–2075.

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.

Goldstein, I. and Yang, L. (2015). Information diversity and complementarities in trading and information acquisition. The Journal of Finance, 70(4):1723–1765.

Gormley, T. A. and Matsa, D. A. (2011). Growing out of trouble? corporate responses to liability risk. The Review of Financial Studies, 24(8):2781–2821.

Green, T. C., Huang, R., Wen, Q., and Zhou, D. (2019). Crowdsourced employer reviews and stock returns. Journal of Financial Economics, 134:236–251.

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

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

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

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

Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. American Economic Review, 70:393–408.

Hendershott, T., Livdan, D., and Schürhoff, N. (2015). Are institutions informed about news? Journal of Financial Economics, 117(2):249–287.

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

Kacperczyk, M. and Seru, A. (2007). Fund managers use of public information: New evidence on managerial skills. Journal of Finance, 62:485–528.

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., Sialm, C., and Zheng, L. (2008). Unobserved actions of mutual funds. The Review of Financial Studies, 21(6):2379–2416.

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

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

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

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

Kempf, E., Manconi, A., and Spalt, O. G. (2017). Learning by doing: The value of experience and the origins of skill for mutual fund managers. Available at SSRN 2124896.

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.

Mukherjee, A., Panayotov, G., and Shon, J. (2021). Eye in the sky: Private satellites and government macro data. Journal of Financial Economics, 141(1):234–254.

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

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

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

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

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

Tables

Table I: Sample Descriptive Statistics

This table presents descriptive statistics for the main variables employed in our main analysis. The sample includes 3,962 funds and 9,781 distinct stocks between 2009 and 2020. In Panel B, Market Cap., Book to Market, Assets, Leverage, price informativeness measures ($ACAR$ and CAR measures), and analysts' earning forecast measures ("Std. FE") are reported for the sample of stocks used in our regressions studying price informativeness in Section VI, i.e., stocks in the industries covered by RS Metrics (cf. Appendix Table B.1).

Panel A: Fund-Quarter level

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

Panel B: Stock-Quarter level

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

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

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

Table II: Alternative Data and Stock Picking Skills

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

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

Table III: Heterogeneous Effect based on Industry Expertise

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

Panel A: Industry Specialists

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

Panel B: Sector Funds

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

Table IV: Heterogeneous Effect based on Geographical Location

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

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

Table V: Heterogeneous Effects based on Skills before Coverage Initiation

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

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

Table VI: Divestment from Covered Stocks

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

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

Table VII: Divestment of Experts from Covered Stocks

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

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

Table VIII: Investment in Peers of Covered Stocks

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

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

Table IX: Alternative Data and Price Informativeness

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

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

Figures

Figure I: Alternative Data and Fund Picking Abilities

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

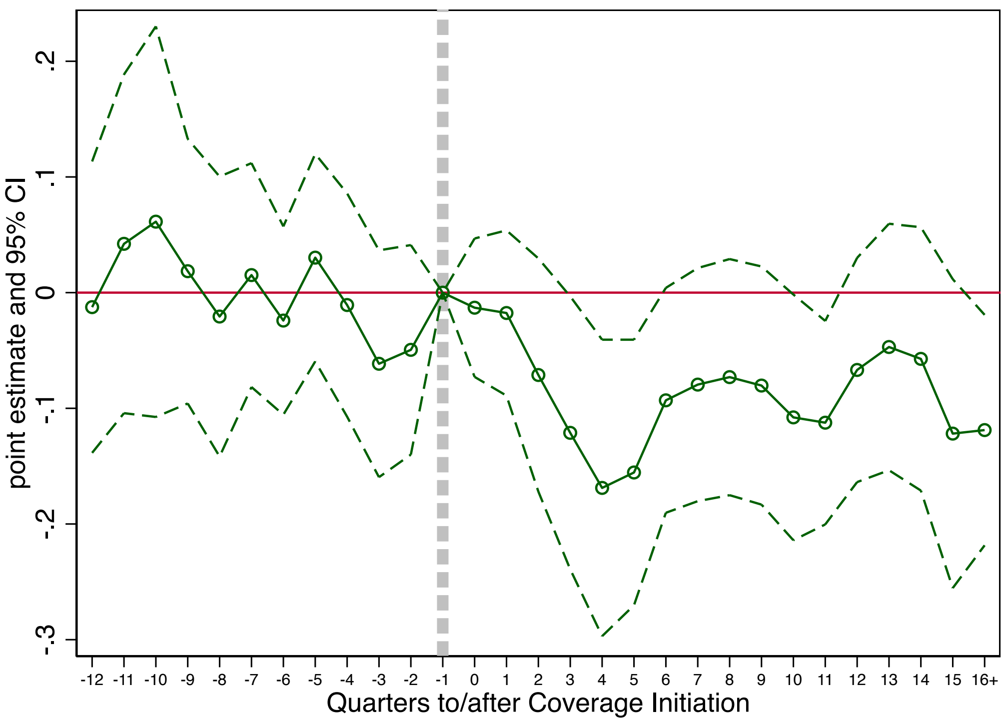
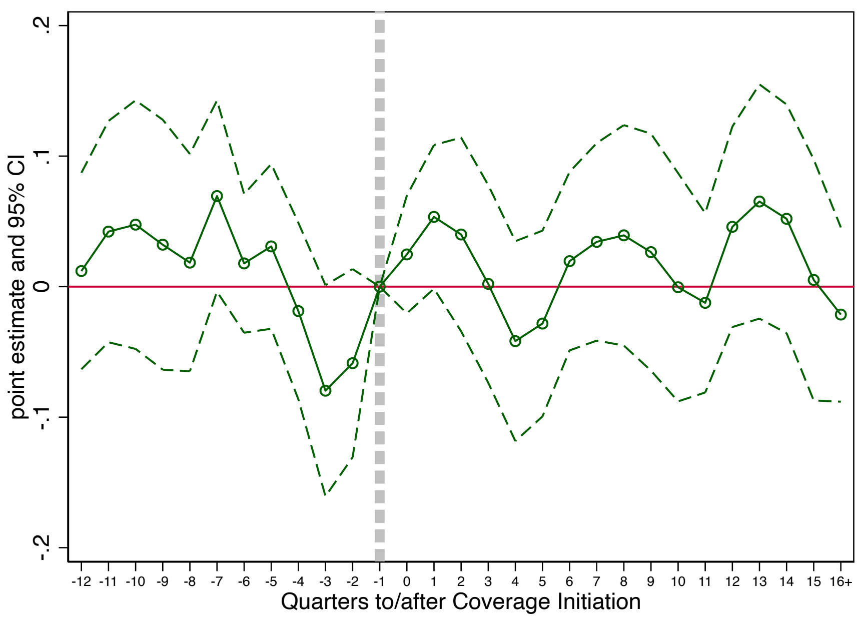


Figure II: Fund Picking Abilities depending on Skills before Coverage Initiation

This figure reports the dynamic effect of the release of alternative data on fund picking abilities. The specification corresponds to equation (10). 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

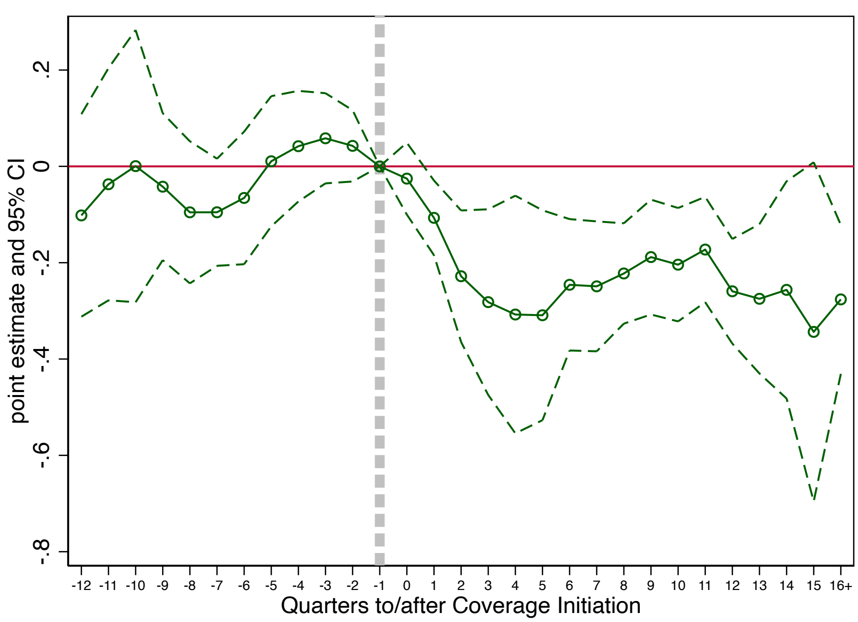
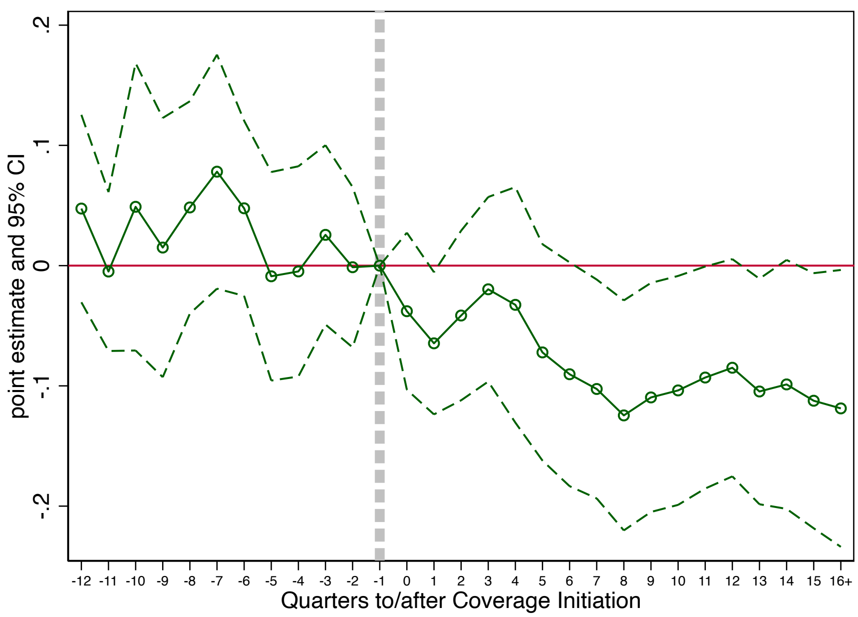


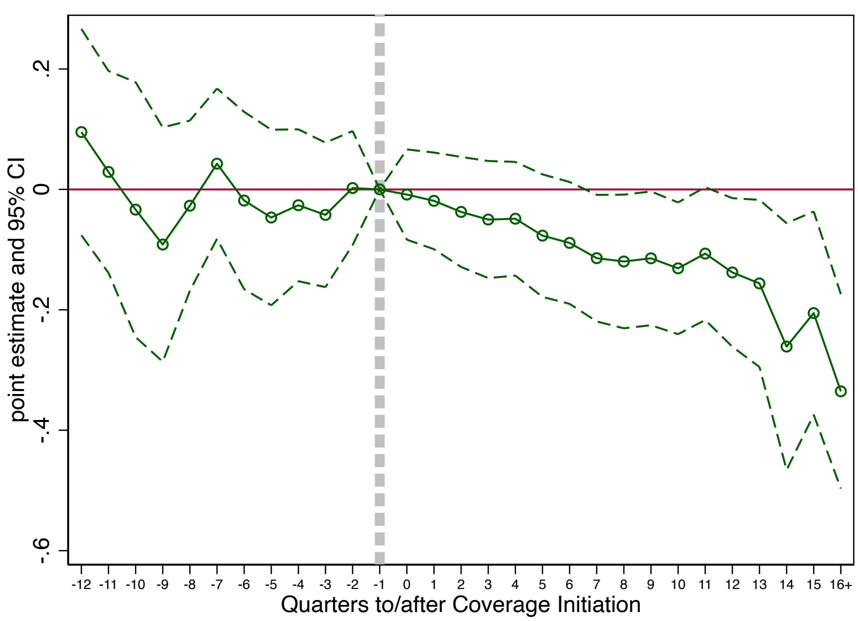
Figure III: Alternative Data and Fund Divestment

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.RS Metrics.com/trafficsignals/retail/monitoring.

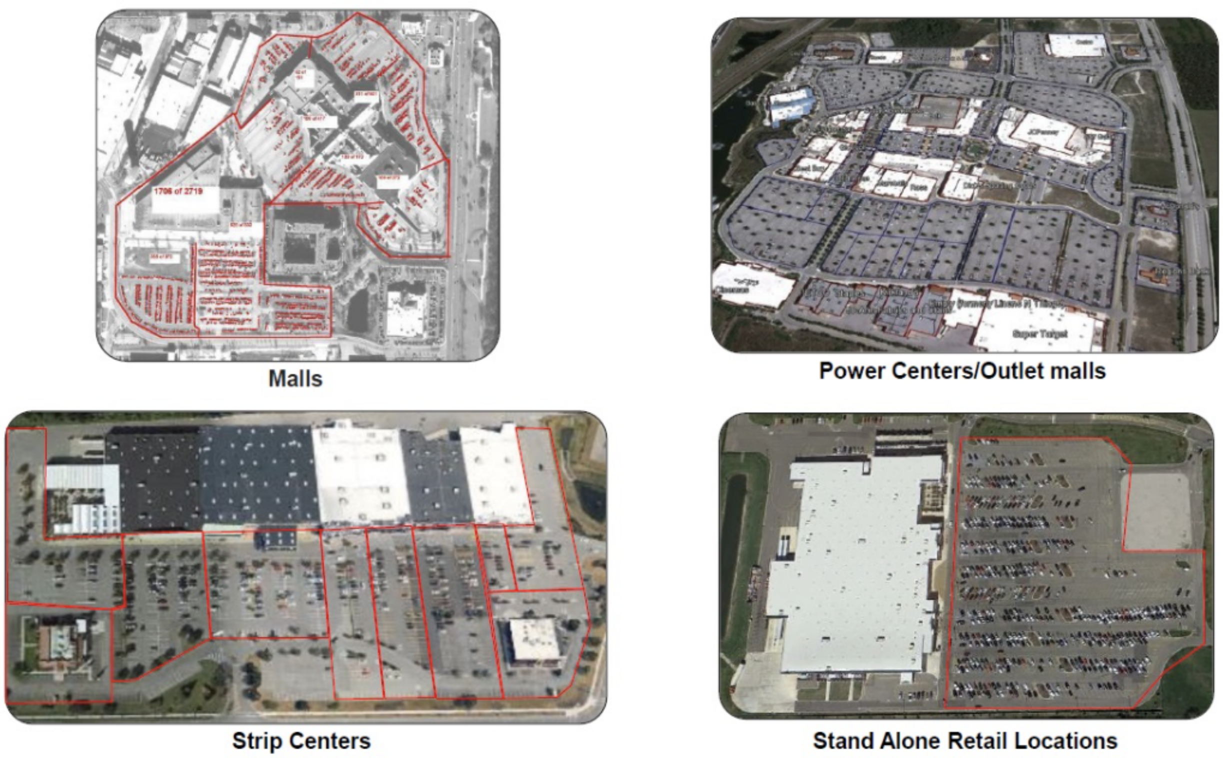
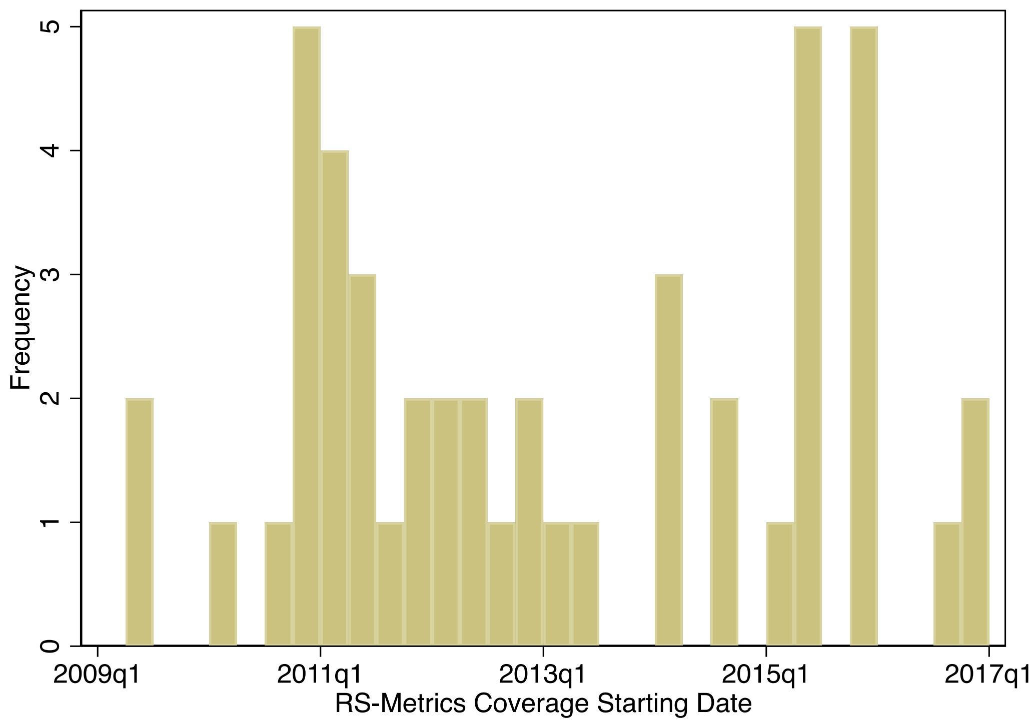


Figure A.2: Number of new Stocks Covered by Satellite Data over Time

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.



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: Controlling for Reversal in Picking Abilities

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

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

Table B.3: Stock-level Characteristics and Coverage Initiation

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

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

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

The table presents the results of our study on the impact of alternative data on analysts' earnings forecast errors. Regressions are estimated at the stock-quarter level. The dependent variable is standardized analysts' forecasting error calculated at different horizons: for the subsequent quarter (column 1), for the second (column 2), third (column 3), and fourth quarter (column 4) ahead, as well as for the forthcoming year (column 5). We compute the standardized analysts' earnings forecast errors as the absolute value of the corresponding actual earnings minus the average of the most recent analyst forecasts, divided by the standard deviation of those forecasts. *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. $Covered \times Post$ 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	-0.059	-0.199	-0.065	0.192	-0.154
	(0.191)	(0.206)	(0.255)	(0.266)	(0.266)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	44,197	44,541	42,355	39,352	11,021
R^2	0.14	0.12	0.15	0.16	0.23

Online Appendix for

Displaced by Big Data? Evidence from Active Fund Managers

(not intended for publication)

Maxime Bonelli and Thierry Foucault

November 5, 2024

Contents

I.1 Hypothesis Development: Theory **2**

I.1.1 Framework . . . 2

I.1.2 Testable Hypotheses . . . 5

I.1.3 Experts and Data Miners Produce Information on Different Dimensions . . . 9

I.2 Stock Picking and Alpha **17**

I.3 Forecasting Sales using Satellite Imagery Data **19**

I.4 Alternative Measures of Stock Picking Ability **21**

I.5 Timing Ability and Alternative Data **27**

I.6 Proxies for Quant Funds **29**

I.7 Stock Picking using a Matched Sample of Control Stocks **33**

I.8 Robustness Test: Adding Stock-Quarter Fixed Effects **38**

I.9 Stock Picking across Cohorts **40**

I.10 Stock Picking before 2018 **42**

I.11 Analysis of Funds' Return Gap **44**

I.12 Decomposing Picking **45**

I.13 Stock Picking Excluding Industry and Geographical Peers **46**

References **48**

I.1 Hypothesis Development: Theory

In this section, we formalize the mechanism that leads to the implications tested in the paper using a simplified version of Dugast and Foucault (2024). We present the model in Section I.1.1. We then derive the implications (hypotheses H.1, H.2, and H.3 in the text) in Section I.1.2. Finally, in Section I.1.3, we consider a different formulation in which data miners and experts have access to signals about different dimensions of uncertainty about the asset payoff. We show that Hypotheses H.1 and H.2 can be reversed in this case.

I.1.1 Framework

We consider the market for a risky asset whose payoff, v, is realized at date 2. This payoff is normally distributed with mean zero and variance σ_v^2. At date 1, fund managers receive signals about this payoff and can invest in the risky asset and a risk free asset, whose rate of return is normalized to zero. Specifically, fund manager j with type k receives the signal:

$$s_j(\tau_k) = v + \tau_k^{-1/2}\varepsilon_j, \tag{IA.1}$$

where τ_k depends on the type of the fund manager (see below). The noise in fund managers' signals (ε_j) is normally distributed with mean zero and variance σ_v^2 and is independent across managers. The higher is τ_k, the higher is the precision ("quality") of fund manager j's signal.

There is a continuum of fund managers (of mass 1) with two types: (i) "Experts" and (ii) "Data Miners". We denote by μ (resp., $(1-\mu)$) the fraction of data miners (resp., experts) and use subscripts "*dm*" and "*ex*" to refer to "data miners" and "experts", respectively. Each fund manager is endowed with the same amount of capital, W_0 (that is, each fund has the same size). Moreover, we assume that all experts have a signal of precision τ_{ex} while all data miners have a signal of precision τ_{dm}. We do not make any assumption on the ranking of τ_{ex} and τ_{dm}.

Data miners have the skills (ability) to process large datasets while experts don't. Thus, data miners can improve the precision of their signals when new alternative datasets become available while experts cannot. In other words, the availability of a new alternative dataset raises τ_{dm} while leaving unchanged τ_{ex} (we discuss the case in which only a fraction of data miners get access to the new dataset at the end of the next section).

After observing her signal, each fund manager chooses a trading strategy, i.e., a demand schedule, $x(s_j(\tau_k), p)$, where, p, is the price of the risky asset. Thus, in period 2, fund manager j with type $k \in \{dm, ex\}$ returns to her client:

$$W_{j,k} = W_0 + x(s_j(\tau_k), p), p)(v - p). \tag{IA.2}$$

Fund managers trade with noise traders and risk-neutral market makers. The noise traders' aggregate demand is price-inelastic and denoted by η, where $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$ (η is independent of v and errors' in fund managers' signals). Market-makers observe the aggregate demand for the asset, $D(p) = \int_0^\mu x(s_j(\tau_{dm}), p)dj + \int_\mu^1 x(s_j(\tau_{ex}), p)dj + \eta$ and post a price such that they obtain zero expected profits. Thus, the equilibrium price, p^*, is equal to their expectation of the asset payoff conditional on the aggregate demand for the asset:

$$p^*(D) = \mathsf{E}\left[v \,|D(p^*) = D\right]. \tag{IA.3}$$

Investors have a CARA utility function (with risk aversion ρ) and fund managers invest in their clients' best interest. Thus, each fund manager j of type $k \in \{dm, ex\}$ chooses her portfolio at date 1 to solve:

$$Max_x \, \mathsf{E}\left[-\exp(-\rho(W_0 + x(s_j(\tau_k), p)(v - p))\right], \tag{IA.4}$$

The equilibrium of the model is a pair $(x^*(., p), p^*(D)$ such that

1. The trading strategy $x^*(s(\tau), p)$ maximizes the expected utility of a fund manager receiving the signal $s(\tau)$ with precision τ, given that other fund managers follow this strategy and that market makers' price is given by $p^*(D)$.

2. The asset price, $p^*(D)$, satisfies $p^* = \mathsf{E}\left[v \,|D^*(p^*) = D\right]$ where $D^*(p^*) = \int_0^\mu x^*(s_j(\tau_{dm}), p)dj+$

$\int_{\mu}^{1} x^*(s_j(\tau_{ex}), p)dj + \eta$.

In sum, we consider a noisy rational expectations model with two types of informed investors who differ in the precision of their signals. This set-up is a special case of Dugast and Foucault (2024) who consider a more general framework. First, they allow the precision of signals within a given group of fund managers to be heterogeneous and it endogenizes the distribution of the precision of data miners' signals using a search model. This enables them to make predictions about how the availability of new data affects the entire distribution of data miners' signals precision and performance. As these predictions are not our focus, we simplify their model by assuming that all fund managers within a given group have signals of the same precision. Second, they endogenize the allocation of capital to each type of fund managers by investors. This enables them to make predictions about the long run effects of the growth of alternative data on the capital allocated to each type of fund manager. As these predictions are not our focus here, we just take this allocation as given.

Thus, we can derive the equilibrium of the market for the risky asset as in Dugast and Foucault (2024), replacing the average precision of experts' signals in their model (that they denote $\bar{\tau}_{ex}$) by τ_{ex} and the average precision of data miners' signals (that they denote $\bar{\tau}_{dm}$) by τ_{dm}. This yields the following characterization of the equilibrium. As it follows directly from Proposition 1 in Dugast and Foucault (2024), we do not reprove it here.

Proposition 1 *(Dugast and Foucault (2024)): Let $\bar{\tau} = \mu\tau_{dm} + (1 - \tau)\tau_{ex}$ be the average precision of signals across fund managers. In equilibrium, a fund manager's demand for the risky asset is*

$$x^*(s(\tau), p) = \frac{\mathsf{E}[v|s(\tau), p] - p}{\rho \, \mathsf{Var}[v|s(\tau), p]} = \beta(\tau) \left(s(\tau) - p\right), \tag{IA.5}$$

where $\beta(\tau) = \frac{\tau}{\rho\sigma_v^2}$ and the equilibrium price of the asset is

$$p^*(D) = \mathsf{E}[v|D^*(p) = D] = \lambda(\bar{\tau})\xi, \tag{IA.6}$$

where $\xi \equiv v + \rho\sigma_v^2\bar{\tau}^{-1}\eta$ *and*

$$\lambda(\bar{\tau}) \equiv \frac{\bar{\tau}^2}{\bar{\tau}^2 + \rho^2\sigma_v^2\sigma_\eta^2}. \tag{IA.7}$$

I.1.2 Testable Hypotheses

Using the previous proposition, we can derive the implications that we test in the paper (the analysis again closely follows Dugast and Foucault, 2024). Price informativeness, $\mathcal{I}(\bar{\tau})$, can be measured by the inverse of the residual uncertainty about the asset payoff conditional on its price. From Proposition 1, it is direct that:

$$\mathcal{I}(\bar{\tau}) \equiv \mathsf{Var}[v \mid p^*]^{-1} = \frac{1}{\sigma_v^2} + \frac{\bar{\tau}^2}{\rho^2\sigma_v^4\sigma_\eta^2}. \tag{IA.8}$$

Thus, the asset price is more informative when the average quality of fund managers' predictors, $\bar{\tau}$, increases.

Now consider the case in which data miners have access to a new alternative data. We assume that this raises the precision of their signal, τ_{dm} and therefore $\bar{\tau}$. As a result price informativeness increases. Intuitively, the reason is that an increase in τ_{dm} induces data miners to take larger position for a given deviation between their signal and the asset price (see eq.(IA.5)). As a result, the aggregate demand for the risky asset ($D(p^*)$) becomes more informative and therefore the price is more informative.

We denote by $R(s_j(\tau_k))$ the gross excess return (i.e., net of the return on the risk free asset) of fund manager j with type $k \in \{ex, dm\}$ on her position in the risky asset:

$$R(s_j(\tau_k)) \equiv \frac{x^*(s_j(\tau_k), p^*) \times (v - p^*)}{W_0} = w(s_j(\tau_k))R^e, \tag{IA.9}$$

where

$$w(s_j(\tau_k)) = x^*(s_j(\tau_k), p^*)p^*/W_0, \tag{IA.10}$$

is the weight of the risky asset in the portfolio of the fund manager and $R^e = v/p^* - 1$ is the risky asset return in excess of the return on the risk free asset. This weight increases with

$s_j(\tau_k)$ and, holding $s_j(\tau_k)$ constant, it increases in τ_k in absolute value.

From eq.(IA.5), we obtain

$$x^*(s_j(\tau_k), p^*) = \frac{1}{\rho\sigma_v^2}\left(\tau_k(v - p^*) + \tau_k^{1/2}\varepsilon_j\right). \tag{IA.11}$$

Thus, the *expected* gross excess return of a fund manager with predictor's quality τ_k is

$$\bar{R}(\tau_k) \equiv \mathsf{E}(R(s_j(\tau_k))) = \frac{\tau}{W_0\rho\sigma_v^2}\mathsf{Var}[v - p^*] = \frac{\tau_k}{W_0\rho\sigma_v^2\mathcal{I}(\bar{\tau})}, \tag{IA.12}$$

where the second equality follows from the independence between $(v - p^*)$ and ε_j and the last equality follows from the fact that $p^* = \mathsf{E}(v \mid p^*)$ so that $\mathsf{Var}[v - p^*] = \mathsf{Var}[v \mid p^*] = \mathcal{I}^{-1}$.

Hence, the expected gross asset return of a fund manager in the risky asset (her "stock picking ability") is only determined by her type, increases with the quality of her signal and decreases with price informativeness. It follows that, when the precision of data miners' signal increases, $\bar{R}(\tau_{ex})$ decreases since $\mathcal{I}(\bar{\tau})$ increases (see eq.(IA.8)). Indeed:

$$\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}} = -\tau_{ex}\frac{\frac{\partial \mathcal{I}}{\partial \tau_{dm}}}{\mathcal{I}^2} < 0. \tag{IA.13}$$

In contrast, the effect of an increase in the precision of data miners' signal on their expected gross return, $\bar{R}(\tau_{dm})$, is ambiguous since it increases both the numerator and the denominator of eq.(IA.12). However, for μ small enough, the effect is always positive. Indeed, using eq.(IA.8), eq.(IA.12) and the definition of $\bar{\tau}$, one obtains:

$$\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} = \frac{1}{W_0\rho\sigma_v^4\mathcal{I}(\bar{\tau})^2}\left(1 + \frac{1}{\rho^2\sigma_v^2\sigma_\eta^2}((1-\mu)^2\tau_{ex}^2 - \tau_{dm}^2\mu(2-\mu))\right) \tag{IA.14}$$

Let $A = \frac{\tau_{ex}^2 + \rho^2\sigma_v^2\sigma_\eta^2}{\tau_{ex}^2 + \tau_{dm}^2}$. It is immediate that if $A > 1$ then $\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} > 0$ for all μ and if $A \leq 1$, $\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} > 0$ iff $0 < \mu < 1 - \sqrt{1 - A}$. Thus, for μ small enough, $\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}}$. Figure IA.1 below provides a numerical example for two different values of μ (respectively, 10% and 40%).

Now consider the average expected gross return of all fund managers, denoted $\bar{R}^A$. By definition, this is:

$$\bar{R}^A \equiv (1-\mu)\bar{R}(\tau_{ex}) + \mu \bar{R}(\tau_{dm}). \tag{IA.15}$$

We have:

$$\frac{\partial \bar{R}^A}{\partial \tau_{dm}} = (1-\mu)\underbrace{\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}}}_{<0} + \mu \frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} \tag{IA.16}$$

For μ small enough, the sign of $\frac{\partial \bar{R}^A}{\partial \tau_{dm}}$ is the same as $\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}}$ and is therefore negative even if $\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} > 0$. This delivers of our first testable hypothesis (H.1). Moreover, in this case:

$$\frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}} < \frac{\partial \bar{R}^A}{\partial \tau_{dm}}, \tag{IA.17}$$

since $\frac{\partial \bar{R}(\tau_{dm})}{\partial \tau_{dm}} > 0$. This is our second testable hypothesis. Last, Eq.(IA.13) shows that the drop in performance is larger when experts' signals have a higher precision since $\mid \frac{\partial \bar{R}(\tau_{ex})}{\partial \tau_{dm}} \mid$ is higher when τ_{ex} is higher, other things equal. Thus, if we had experts with signals of different precisions (extending the model in this case is straightforward), the drop in performance following an increase in data miners' signal precision would be larger for experts with signals of higher precision.[1] This delivers our third testable hypothesis (H.3).

Figure IA.1 illustrates H.1 (black curve) and H.2 (red curve) numerically for two different values of μ (10% and 40%).

Remarks:

1. **Information Acquisition by Data Miners.** In the previous derivations, we have implicitly assumed that when new alternative data becomes available, it is bought by all data miners. One can easily endogenize this decision and obtain three possible situations: (i) All data miners buy the data because the cost of the data is small enough, in which case the previous analysis is unchanged, (ii) nobody buys the data

[1] For instance, suppose we have 2 types of experts with signals of precision τ_{ex}^H (type H) and τ_{ex}^L (type L) such that $\tau_{ex}^L < \tau_{ex}^H$. Moreover the mass of experts with type H in the population of experts is λ_H. Defining, $\bar{\tau} = \mu\tau_{dm} + (1-\mu)(\lambda_H^{ex}\tau_{ex}^H + \lambda_L \tau_{ex}^L)$, the rest of the analysis is identical and one gets from Eq.(IA.13) that $\frac{\partial \bar{R}(\tau_{ex}^H)}{\partial \tau_{dm}} < \frac{\partial \bar{R}(\tau_{ex}^L)}{\partial \tau_{dm}} < 0$.

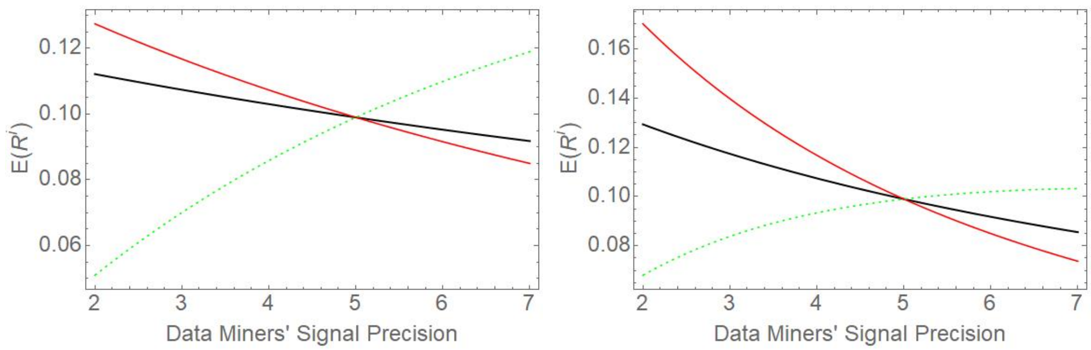


Figure IA.1: Left Panel: $\mu = 0.1$. Other parameter values: $\sigma_v^2 = \sigma_\eta^2 = 1$, $\rho = 0.5$, $\tau_{ex} = 5$, $W_0 = 1$. **Right Panel:** $\mu = 0.4$. Other parameter values are as in the left panel. In each case the figure shows the effect of a variation in the precision of data miners' signals (τ_{dm}) on (i) experts' gross expected return, $\bar{R}(\tau_{ex})$ (red line), (ii) data miners' gross expected return, $\bar{R}(\tau_{dm})$ (green dashed line) and (iii) fund managers' average gross return, $\bar{R}^A$ (plain black line).

because the cost of the data is too high (in which case there is no effect), and (iii) an interior case in which only a fraction α of data miners buy the data. In this last case, the analysis is slightly modified since one must now distinguish between three types of fund managers: (a) experts (fraction $(1-\mu)$), (b) data miners who buy the data (fraction $\alpha\mu$) and (c) data miners who do not buy the data (fraction $(1-\alpha)\mu)$). In this case, those purchasing the data exerts a negative externality on experts (as in the baseline case in which data miners buy the data) and the data miners not buying the data. The gross expected return of those buying the data must improve so that, net of the cost of the data, their expected utility is just equal to the expected utility of data miners not buying the data. Predictions are therefore identical to the baseline case and even stronger (as data miners not buying the data are negatively affected).

2. **Liquidity** Using Proposition 1 , note that $D(p^*)(\frac{\bar{\tau}}{\rho\sigma_v^2})^{-1} + p^* = \xi$. Thus, using eq.(IA.6), we can write $p^*(D)$ as $p^*(D) = \hat{\lambda}D$ where $\hat{\lambda} = \frac{\lambda}{1-\lambda}(\frac{\bar{\tau}}{\rho\sigma_v^2})^{-1} = \frac{\bar{\tau}}{\sigma_\eta^2}$, where the last equality follows from the definition of λ. Thus, $\hat{\lambda}$ measures the sensitivity of the asset price to the order flow D. Thus, it is a measure of illiquidity. When new data becomes available, τ_{dm} increases and therefore $\bar{\tau}$ increases. Thus, according to the model, illiquidity should increase, as found in Katona et al. (2023).

I.1.3 Experts and Data Miners Produce Information on Different Dimensions

Now we assume that the fundamental v has two dimensions, namely $v = v_{dm} + v_{ex}$ where v_j is normally distributed with mean zero and variance σ_j^2 for $j \in \{ex, dm\}$. We assume that only experts can produce information about v_{ex} and only data miners can produce information about v_{dm}, provided that alternative data about the asset is available. This formulation captures the idea that experts and data miners have expertise in producing different types of information. This version of the model is similar to Goldstein and Yang (2015). As in their paper, we assume that speculators specialized in one dimension perfectly observe the realization of the fundamental on this dimension (e.g., experts perfectly observe v_{ex}). One difference with their approach is that, in our setting, the price is set by risk neutral market makers and satisfies therefore $IA.3$. A second, minor, difference is that we allow the variance of v_{dm} and v_{ex} to be different.

We first solve the model assuming that alternative data is available. Solving the model in this case is more complex than in the baseline case. The reason is that fund managers of type j can use the asset price and their observation of v_j to filter out of the price a signal about v_{-j}, the component of the asset payoff they do not observe. This possibility gives them an edge over the dealers, as if they had an imperfect signal about v_{-j}. We proceed using the standard "guess and verify" procedure to solve for linear rational expectations equilibria. That is, we conjecture that the demand for the asset of fund managers of type $j \in \{ex, dm\}$ is:

$$x_j(v_j, p) = \alpha_j v_j - \beta_j p, \tag{IA.18}$$

and that the equilibrium price is:

$$p^* = \lambda v_{dm} + \gamma v_{ex} + \kappa \eta. \tag{IA.19}$$

We then solve for each fund manager's optimal demand given the conjectured price and for the equilibrium price given fund managers' conjectured demands. This approach yields a system of equations for the coefficients $(\alpha_{dm}, \alpha_{ex}, \beta_{dm}, \beta_{ex}, \lambda, \gamma, \kappa)$ whose solution is the rational expectations equilibrium of the model.

First consider experts' demand function. As they know v_{ex}, experts can extract from the price a signal $z_{ex} = \frac{p^* - \gamma v_{ex}}{\lambda} = v_{dm} + \frac{\kappa}{\lambda}\eta$ about v_{dm}. Thus, experts' expectation of the asset payoff condotional on the asset price and their information about v_{ex} is:

$$\mathsf{E}\left[v \left| p^*, v_{ex}\right.\right] = \mathsf{E}\left[v \left| z_{ex}, v_{ex}\right.\right] = v_{ex} + \mathsf{E}\left[v_{dm} \left| z_{ex}\right.\right] \tag{IA.20}$$

Writing the first order condition of an expert's expected utility maximization problem, we obtain that his demand function is:

$$x^*(v_{ex}, p) = \frac{\mathsf{E}\left[v \left| p^*, v_{ex}\right.\right] - p}{\rho \, \mathsf{Var}[v|p^*, v_{ex}]} = \frac{v_{ex} + \mathsf{E}\left[v_{dm} \left| z_{ex}\right.\right] - p}{\rho \, \mathsf{Var}[v_{dm}|z_{ex}]}. \tag{IA.21}$$

Calculations yield:

$$\mathsf{E}\left[v_{dm} \left| z_{ex}\right.\right] = \frac{\tau_\eta}{\tau_\eta + (\frac{\kappa}{\lambda})^2 \tau_{dm}} z_{ex}, \tag{IA.22}$$

and

$$\mathsf{Var}[v_{dm}|z_{ex}] = \frac{1}{\tau_{dm} + (\frac{\lambda}{\kappa})^2 \tau_\eta}, \tag{IA.23}$$

where $\tau_\eta = (\sigma_\eta^2)^{-1}$ and $\tau_j = (\sigma_j^2)^{-1}$. Substituting these expressions in eq.(IA.21) and using the fact that $z_{ex} = \frac{p^* - \gamma v_{dm}}{\lambda}$, $x^*(v_{ex}, p)$ is as conjectured iff:

$$\alpha_{ex} = \rho^{-1}(\tau_{dm} + \frac{\lambda \tau_\eta}{\kappa^2}(\lambda - \gamma)), \tag{IA.24}$$

and

$$\beta_{ex} = \rho^{-1}(\tau_{dm} - \frac{\lambda \tau_\eta}{\kappa^2}(1 - \lambda)). \tag{IA.25}$$

Following the same steps, we obtain:

$$\alpha_{dm} = \rho^{-1}(\tau_{ex} + \frac{\gamma \tau_\eta}{\kappa^2}(\gamma - \lambda)), \tag{IA.26}$$

and

$$\beta_{dm} = \rho^{-1}(\tau_{ex} - \frac{\gamma \tau_\eta}{\kappa^2}(1 - \gamma)). \tag{IA.27}$$

The aggregate demand is $D(p) = \mu x_{dm}(v_{dm}, p) + (1-\mu)x_{ex}(v_{ex}, p) + \eta$. Using eq.(IA.18), the aggregate demand is therefore informationally equivalent to $z_D = D(p) + (\mu\alpha_{dm} + (1 - \mu)\alpha_{ex})p = \mu\alpha_{dm}v_{dm} + (1-\mu)\alpha_{ex}v_{ex} + \eta$. Thus,

$$p^*(D) = \mathsf{E}[v|D^*(p) = D] = \mathsf{E}[v|z_D = \mu\alpha_{dm}v_{dm} + (1-\mu)\alpha_{ex}v_{ex} + \eta], \tag{IA.28}$$

Computing the last conditional expectation, we deduce that p^* is is consistent with our conjecture if and only if

$$\lambda = \frac{((\mu\alpha_{dm})^2\sigma_{dm}^2 + \mu(1-\mu)\alpha_{dm}\alpha_{ex}\sigma_{ex}^2}{(\mu\alpha_{dm})^2\sigma_{dm}^2 + ((1-\mu)\alpha_{ex})^2\sigma_{ex}^2 + \sigma_\eta^2}, \tag{IA.29}$$

$$\gamma = \frac{((1-\mu)\alpha_{ex})^2\sigma_{ex}^2 + \mu(1-\mu)\alpha_{dm}\alpha_{ex}\sigma_{dm}^2}{(\mu\alpha_{dm})^2\sigma_{dm}^2 + ((1-\mu)\alpha_{ex})^2\sigma_{ex}^2 + \sigma_\eta^2}, \tag{IA.30}$$

$$\kappa = \frac{\mu\alpha_{dm}\sigma_{dm}^2 + (1-\mu)\alpha_{ex}\sigma_{ex}^2}{(\mu\alpha_{dm})^2\sigma_{dm}^2 + ((1-\mu)\alpha_{ex})^2\sigma_{ex}^2 + \sigma_\eta^2}. \tag{IA.31}$$

Eq.(IA.24) to (IA.31) form a system of 7 equations with 7 unknowns $\alpha_{dm}, \alpha_{ex}, \beta_{dm}, \beta_{ex}, \lambda, \gamma, \kappa$.

Observe that:

$$\frac{\gamma}{\kappa} = (1-\mu)\alpha_{ex}, \tag{IA.32}$$

and

$$\frac{\lambda}{\kappa} = \mu\alpha_{dm}. \tag{IA.33}$$

Thus, using the expression for α_{dm} (eq.(IA.26)), we have:

$$\frac{\lambda}{\kappa} = \mu\rho^{-1}(\tau_{ex} + \tau_\eta((\frac{\gamma}{\kappa})^2 - \frac{\lambda}{\kappa}\frac{\gamma}{\kappa})). \tag{IA.34}$$

Let $q \equiv \frac{\lambda}{\kappa}$ and $y \equiv \frac{\gamma}{\kappa}$. From the previous equation, we deduce:

$$q(y) = \frac{\mu(\tau_{ex} + \tau_\eta y^2)}{\rho + \mu\tau_\eta y}. \tag{IA.35}$$

Then, substituting $q(y)$ and α_{ex} by their expressions in eq.(IA.32), we deduce that y must solve:

$$y = \frac{(1-\mu)}{\rho}[\tau_{dm} + \tau_\eta q(y)(q(y) - y)] \tag{IA.36}$$

There is no closed-form solution for this equation. However, it can be solved numerically. Once the solution is obtained one can derive $q = \frac{\lambda}{\kappa}$, and therefore α_{ex}, α_{dm} and κ (using eq.(IA.32), eq.(IA.33 and eq.(IA.31)). This yields λ and γ (via eq.(IA.29) and eq.(IA.30) again) and therefore β_{ex} and β_{dm}.

Similar to the baseline case, we define the excess return of a fund manager with type j (the fund manager's stock picking ability) as:

$$R_j \equiv \frac{x^*(v_j, p^*) \times (v - p^*)}{W_0}. \tag{IA.37}$$

Thus, the *expected* gross excess return of a fund manager with type $j \in \{ex, dm\}$ is

$$\bar{R}_j \equiv \mathsf{E}(R_j) = \frac{Cov(x^*(v_j, p^*), v - p^*)}{W_0}, \tag{IA.38}$$

where the last equality follows from $\mathsf{E}(p^*(D)) = \mathsf{E}(\mathsf{E}[v|D^*(p) = D]) = \mathsf{E}(v) = 0$. Using the expression for $x^*(v_j, p^*)$ (eq.(IA.18)), we deduce after some algebra:

$$\bar{R}_{ex} = \frac{(1-\gamma)(\alpha_{ex} - \beta_{ex}\gamma)\sigma_{ex}^2 - \beta_{ex}\lambda(1-\lambda)\sigma_{dm}^2 + \kappa^2\beta_{ex}\sigma_\eta^2}{W_0}. \tag{IA.39}$$

Similarly:

$$\bar{R}_{dm} = \frac{(1-\lambda)(\alpha_{dm} - \beta_{dm}\gamma)\sigma_{dm}^2 - \beta_{dm}\gamma(1-\gamma)\sigma_{dex}^2 + \kappa^2\beta_{dm}\sigma_\eta^2}{W_0}. \tag{IA.40}$$

We also compute the informativeness of the equilibrium price as:

$$\mathsf{Var}[v \mid p^*]^{-1} = \frac{\lambda^2\sigma_{dm}^2 + \gamma^2\sigma_{ex}^2 + \kappa^2\sigma_\eta^2}{(\lambda-\gamma)^2\sigma_{ex}^2\sigma_{dm}^2 + \kappa^2\sigma_\eta^2(\sigma_{ex}^2 + \sigma_{dm}^2)} \tag{IA.41}$$

We can proceed exactly in the same way when no alternative data is available about v_{dm}. In fact in this case, everything is as if $\lambda = 0$. One can therefore obtain the equilibrium by setting $\lambda = 0$ in eq.(IA.24) and (IA.25). It follows that with no alternative data on v_{dm}, we have:

$$\alpha_{ex}^{no} = \beta_{ex}^{no} = \tau_{dm}\rho^{-1}, \tag{IA.42}$$

where superscript "no" indicates the absence of alternative data. Proceeding as in the previous case, we obtain that the expected excess return of an expert in this case is:

$$\bar{R}_{ex}^{no} = \frac{\alpha_{ex}^{no}(1-\gamma)^2\sigma_{ex}^2 + \kappa^2\beta_{ex}\sigma_\eta^2}{W_0}. \tag{IA.43}$$

The informativeness of the equilibrium price in this case is:

$$\mathsf{Var}^{no}[v \mid p^*]^{-1} = \frac{\gamma^2\sigma_{ex}^2 + \kappa^2\sigma_\eta^2}{\gamma^2\sigma_{ex}^2\sigma_{dm}^2 + \kappa^2\sigma_\eta^2(\sigma_{ex}^2 + \sigma_{dm}^2)} \tag{IA.44}$$

We can then analyze the effect of the availability of alternative data about v_{dm} by comparing the expected excess returns of the different types of fund managers and price informativeness with and without alternative data, holding all other parameters constant. Importantly, in this comparison, we hold μ constant. That is, we do not assume that alternative data reduces the mass of experts relative to data miners (from 1 to $(1-\mu)$). Rather we assumes that once alternative data about the asset becomes available, it enables the sector of fund managers with ability to use these data to start trading on v_{dm}.

We define the change in experts' expected excess return following the availability of alternative data about v_{dm} as:

$$\Delta \bar{R}_{ex} = \bar{R}_{ex} - \bar{R}_{ex}^{no}. \tag{IA.45}$$

Similarly, we can define the change in data miners' expected excess return as $\Delta \bar{R}_{dm} = \bar{R}_{dm} - \bar{R}_{dm}^{no} = \bar{R}_{dm}$ (since $\bar{R}_{dm}^{no} = 0$) and the average change in fund manager's expected excess return as a group as:

$$\Delta \bar{R}^{A} = \mu \bar{R}_{dm} + (1-\mu)\bar{R}_{ex} - (1-\mu)\bar{R}_{ex}^{no}. \tag{IA.46}$$

In the baseline case considered in the previous section, the availability of alternative data always reduces both the expected excess returns of experts and fund managers as a group, that is, in the baseline case $\Delta \bar{R}_{ex} < 0$ and $\Delta \bar{R}^{A} < 0$. As shown in Figure IA.2, this is not necessarily the case when experts and data miners produce information on different dimensions of the asset payoff.

First, for μ large enough, experts' expected excess return increases when data miners get access to alternative data (see the L.H.S panel of Figure IA.2).[2] The intuition is as follows. Ceteris paribus, as explained by Goldstein and Yang (2015), an increase in the trading intensity of some speculators (say data miners) on signals about one dimension of uncertainty for the asset payoff has two opposite effects on the trading intensity of speculators trading on signals about the other dimension (say experts). First, the uncertainty about the asset payoff for experts decreases because they can extract information about v_{dm} from the price. Second, holding the price constant, experts infer from a strong realization of their

[2]This effect is not due to a decrease in the mass of experts since, as previously explained, this mass is held constant with and without alternative data.

signal that data miners must have received a low signal. This "inference augmentation effect" (Goldstein and Yang (2015)'s terminology) works to reduce experts' responsiveness to their own signal and thus plays in a direction opposite to the uncertainty reduction effect. The first effect increases the value of their own information for experts (and therefore their expected excess return) while the second decreases it.[3] As μ increases, the uncertainty reduction effect increases and ends up dominating the inference augmentation effect. For this reason, $\Delta\bar{R}_{ex}$ is non monotonic in μ and becomes positive for μ large enough (the exact threshold depends on other parameter values).

The L.H.S panel of Figure IA.2 also shows that the net effect of alternative data on experts' expected excess return is always quite small because the uncertainty reduction effect and the inference augmentation effect tend to neutralize each other. For this reason, the average change in fund managers' expected excess returns (averaged across all managers) is always positive (see the R.H.S panel in Figure IA.2).

In sum, this analysis shows that the direction of the effects of alternative data on the expected excess returns of experts (those who cannot use alternative data) and the average excess return of all active fund managers crucially depends on whether experts and data miners produce information on the same dimension of the fundamental (as in the baseline case) or different dimensions (as assumed in this section). Our tests support the implications of the first scenario and therefore reject the second one, at least for the type of alternative data considered in our tests.

Importantly, these effects arise despite the fact that price informativeness is higher when alternative data is available than when it is not as Figure IA.3 shows for the same parameter values as those in Figure IA.2.

[3] In other words, as shown by Goldstein and Yang (2015)), the uncertainty reduction effect is a source of complementarities between speculators informed on different dimensions of uncertainty.

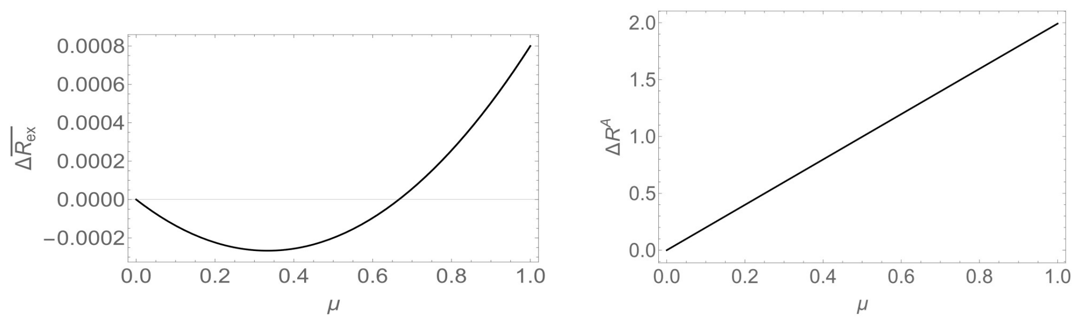


Figure IA.2: Left Panel: Change in experts' expected excess return ($\Delta\bar{R}_{ex}$) when alternative data becomes available. **Right Panel:** Change in average excess returns across all fund managers ($\Delta\bar{R}^A$) when alternative data becomes available. Parameter values: $\sigma_{dm}^2 = \sigma_{ex}^2 = 1$, $\sigma_\eta^2 = 1000$, $\rho = 0.5$, $W_0 = 1$

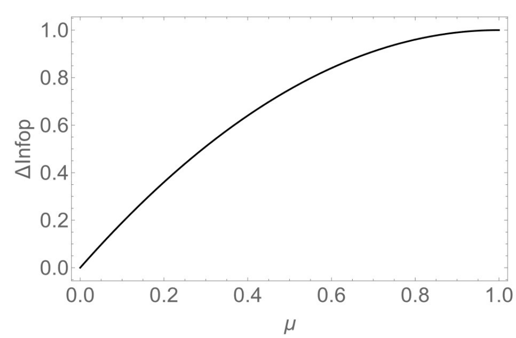


Figure IA.3: Change in the informativeness of the asset price with and without alternative data: $\Delta Infop = \mathsf{Var}[v \mid p^*]^{-1} - \mathsf{Var}^{no}[v \mid p^*]^{-1}$ where $\mathsf{Var}[v \mid p^*]^{-1}$ is given in eq.(IA.41) and $\mathsf{Var}^{no}[v \mid p^*]^{-1}$ is given in eq.(IA.44). Parameter values: $\sigma_{dm}^2 = \sigma_{ex}^2 = 1$, $\sigma_\eta^2 = 1000$, $\rho = 0.5$, $W_0 = 1$. $\Delta Infop$ is multiplied by 10^4 for readibility.

I.2 Stock Picking and Alpha

In this section, we explain how our stock-level measure of a fund's stock picking ability is related to its "alpha" (average abnormal performance). Let $\omega_{if,t}$ the weight of stock i in fund f's portfolio at the beginning of period t and let R_{ft} be the rate of return on fund f's portfolio from date t to date $t+1$. Last, we denote by R_{0t} the rate of return on the risk free asset over the same period. We have:

$$R_{ft} = R_{0t} + \sum_{i=1}^{N} \omega_{if,t}(R_{it} - R_{0t}), \tag{IA.47}$$

where R_{it} is the return on stock i and N is the number of stocks. Now, let the CAPM abnormal return on stock i from date t to date $t+1$ be:

$$\epsilon_{it} = R_{it} - R_{0t} - \beta_i(R_{mt} - R_{0t}), \tag{IA.48}$$

where R_{mt} is the market return and β_i the beta of stock i. Using this equation, we can rewrite eq.(IA.47) as:

$$R_{ft} = R_{0t} + \sum_{i=1}^{N} \omega_{if,t}\epsilon_{it} + \beta_f(R_{mt} - R_{0t}), \tag{IA.49}$$

where $\beta_f = \sum_{i=1}^{N} \omega_{if,t}\beta_i$ is the beta of fund f. Hence, the abnormal return of fund f is:

$$\epsilon_{ft} \equiv R_{ft} - R_{0t} - \beta_f(R_{mt} - R_{0t}) = \sum_{i=1}^{N} \omega_{if,t}\epsilon_{it} = \sum_{i=1}^{N}(\omega_{if,t} - \omega_{im,t})\epsilon_{it}, \tag{IA.50}$$

where $\omega_{im,t}$ is the weight of stock i in the market portfolio and the last equality follows from the fact that $\epsilon_{mt} = \sum_{i=1}^{N} \omega_{im,t}\epsilon_{it} = 0$ since $\beta_m = 1$. We deduce that fund f's alpha (its expected abnormal return) is

$$\alpha_f = \mathrm{E}(\epsilon_{ft}) = \sum_{i=1}^{N} E((\omega_{if,t} - \omega_{im,t})\epsilon_{it}). \tag{IA.51}$$

Thus, the alpha of fund f is sum of its expected stock picking ability in each stock.

I.3 Forecasting Sales using Satellite Imagery Data

We validate the relevance of satellite imagery of parking lots for forecasting retailer sales. Specifically, for firms covered by RS Metrics, we test whether variations in parking lot filling rates derived from RS Metrics data are predictive of firm sales growth. Following Katona et al. (2023), we compute year-on-year growth in same-store parking lot utilization (we focus on year-on-year growth rather than sequential growth due to seasonal effects in retailer performance). For each retailer-quarter, we sum up across individual store locations with available year-on-year satellite coverage to obtain the aggregate parking lot traffic $Cars_{i,t}$, and the aggregate parking lot space $Spaces_{i,t}$. We calculate the firm-level parking lot fill rate as the ratio of aggregate parking lot traffic divided by aggregate parking lot space $FillingRate_{i,t} = Cars_{i,t}/Spaces_{i,t}$. We then compute our main predictive variable of interest as the (standardized) year-on-year growth in same-store parking lot fill rates $FillingRateGrowth_{i,t} = (FillingRate_{i,t} - FillingRate_{i,t-4})/FillingRate_{i,t-4}$. From Compustat, we obtain each firm's quarterly sales and we compute year-on-year growth in sales as $SalesGrowth_{i,t} = (Sales_{i,t} - Sales_{i,t-4})/Sales_{i,t-4}$. In Appendix Table IA.1, we present estimates of the following regression:

$$SalesGrowth_{i,t} = \beta FillingRate_{i,t} + \gamma SalesGrowth_{i,t-1} + \theta_i + \delta_{q\times t} + \epsilon_{i,t}, \qquad \text{(IA.52)}$$

where θ_i and $\delta_{q\times t}$ are respectively firm fixed effects and firm size quintile × year-quarter fixed effects.

Table IA.1 confirms that parking lot filling rates data reported by RS Metrics are a relevant predictor of retailer sales performance.

Table IA.1: Forecasting Sales using Satellite Imagery Data

This table provides evidence that growth in parking lot filling rates obtained from satellite imagery data predicts growth retailer sales. Regressions are estimated at the stock-quarter level. The dependent variable is the year-on-year growth in sales. The main dependent variable is the (standardized) firm-level growth in parking lot fill rate. The sample includes the firms covered by RS Metrics. Standard errors are adjusted for heteroskedasticity. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Sales Growth					
	(1)	(2)	(3)	(4)	(5)	(6)
Filling Rate Growth	0.013***	0.005***	0.012***	0.007***	0.015***	0.008***
	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Sales Growth (Past Quarter)		0.799***		0.648***		0.631***
		(0.042)		(0.070)		(0.069)
Year-Quarter FE	No	No	Yes	Yes	No	No
Stock FE	No	No	Yes	Yes	Yes	Yes
Size Quintile × Year-Quarter FE	No	No	No	No	Yes	Yes
Observations	758	758	758	758	758	758
R^2	0.02	0.68	0.52	0.73	0.61	0.77

I.4 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 IA.2: Alternative Measures of Skills

This table reproduces our results on the effect of the release of alternative data on fund picking abilities presented in Table II, but uses alternative measures of fund skills as dependent variables: fund picking abilities measured using the composition of the S&P 500 index (Panel A) and stock trading abilities 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 Abilities

	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***	-0.041***	-0.037**	-0.052***	-0.066***	-0.058*	-0.080***	-0.104***	-0.096**
	(0.006)	(0.007)	(0.008)	(0.012)	(0.015)	(0.019)	(0.020)	(0.025)	(0.031)	(0.028)	(0.035)	(0.046)
Covered	0.024***			0.047***			0.074***			0.105***		
	(0.005)			(0.010)			(0.016)			(0.022)		
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	0.08	0.13	0.34	0.08	0.14	0.38

Panel B: Trading Abilities

	Trading 1-Q			Trading 2-Q			Trading 3-Q			Trading 4-Q		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Covered × Post	-0.013***	-0.015***	-0.014***	-0.026***	-0.028***	-0.022**	-0.042***	-0.048***	-0.035**	-0.064***	-0.076***	-0.061***
	(0.004)	(0.005)	(0.005)	(0.008)	(0.010)	(0.009)	(0.012)	(0.015)	(0.014)	(0.017)	(0.021)	(0.020)
Covered	0.016***			0.034***			0.055***			0.080***		
	(0.004)			(0.008)			(0.012)			(0.017)		
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	0.05	0.07	0.20	0.05	0.08	0.28	0.05	0.09	0.33	0.05	0.09	0.36

Table IA.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 abilities measured using the composition of the S&P 500 index (Panel A) and stock trading abilities 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 Abilities

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.003	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	0.34	0.38

Panel B: Trading Abilities

	Trading 1-Q	Trading 2-Q	Trading 3-Q	Trading 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.004	0.017**	0.022*	0.013
	(0.004)	(0.008)	(0.012)	(0.014)
Covered × Post × High Trading Pre	-0.043***	-0.090***	-0.130***	-0.166***
	(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 IA.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 abilities measured using the composition of the S&P 500 index (Panel A) and stock trading abilities 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	-0.138**	-0.159***	-0.343**	-0.414***	-0.569**	-0.703***	-0.856**	-1.074***
	(0.065)	(0.054)	(0.164)	(0.140)	(0.272)	(0.233)	(0.402)	(0.354)
Covered × Industry Specialist	0.104***		0.247***		0.393***		0.560***	
	(0.035)		(0.084)		(0.131)		(0.177)	
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	0.34	0.14	0.38

Panel B: Trading Abilities

	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	-0.089**	-0.018	-0.163***	-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*		0.168**		0.233**	
	(0.021)		(0.056)		(0.084)		(0.101)	
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 IA.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 abilities measured using the composition of the S&P 500 index (Panel A) and stock trading abilities 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)	(0.026)
Covered × Post × Sector Fund	-0.102	-0.131**	-0.253	-0.338**	-0.423	-0.576**	-0.628*	-0.878**
	(0.063)	(0.053)	(0.157)	(0.140)	(0.258)	(0.233)	(0.380)	(0.356)
Covered × Sector Fund	0.088**		0.211**		0.337**		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	0.34	0.14	0.38

Panel B: Trading Abilities

	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 IA.6: Alternative Measures of Skills and Geographical Location

This table reproduces our results on the differential impact of alternative data on fund picking abilities depending on fund location in Table V, but uses alternative measures of fund skills as dependent variables: fund picking abilities measured using the composition of the S&P 500 index (Panel A) and stock trading abilities 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***	-0.047*	-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**	-0.055***	-0.075**	-0.114**	-0.113*	-0.178**	-0.162*	-0.267**
	(0.015)	(0.019)	(0.036)	(0.054)	(0.058)	(0.085)	(0.083)	(0.123)
Covered × Local	0.027**		0.054**		0.079**		0.112**	
	(0.011)		(0.024)		(0.037)		(0.053)	
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	9972897	9972897	9681448	9681448	9383519	9383519	9083331	9083331
R^2	0.10	0.21	0.12	0.29	0.14	0.34	0.14	0.39

Panel B: Trading Abilities

	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**	-0.073***	-0.055***
	(0.005)	(0.005)	(0.011)	(0.010)	(0.016)	(0.014)	(0.021)	(0.020)
Covered × Post × Local	-0.011*	-0.017**	-0.020*	-0.030**	-0.028	-0.042*	-0.037	-0.062*
	(0.006)	(0.007)	(0.012)	(0.014)	(0.019)	(0.024)	(0.028)	(0.034)
Covered × Local	0.011**		0.022**		0.027		0.031	
	(0.005)		(0.011)		(0.017)		(0.024)	
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	9972897	9972897	9681448	9681448	9383519	9383519	9083331	9083331
R^2	0.07	0.21	0.08	0.29	0.09	0.34	0.10	0.37

I.5 Timing Ability and Alternative Data

The alternative data considered in our tests are unlikely to affect a fund's ability to anticipate a stock's systematic return, since (i) signals about a retailers' sales are likely to be very noisy signals of market returns and (ii) parking lots counts are not obviously related to a retailers' βs. To test this, we introduce an alternative measure "Timing" (similar to Kacperczyk et al., 2014), defined as follows for fund f in stock i at horizon h in quarter t:

$$Timing^h_{f,i,t} = 100 \times (w^f_{i,t} - w^m_{i,t})(\beta_{i,t} R^m_{t+h}), \tag{IA.53}$$

where $w^f_{i,t}$ is the fraction of fund f's assets held in stock i at the end of quarter t, $w^m_{i,t}$ is the fraction of total market capitalization in stock i (its weight in the "market portfolio") at the end of quarter t, R^m_{t+h} is the return of the stock market over the following h quarters, and $\beta_{i,t}$ is the beta of stock i with the market (computed using daily returns over the last 252 days). As in our main analysis, we consider four different horizons, namely $h = 1$, 2, 3 and 4 quarters.

We estimate our main specification (equation (5) in the main text) using "*Timing*" as dependent variable. Table IA.7 presents the estimation results and shows that the coverage initiation of a stock by RS Metrics has no effect on funds' timing ability.

Table IA.7: Alternative Data and Timing Abilities

This table presents our results on the effect of the release of alternative data on fund timing abilities. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Timing* calculated at different horizons ranging from one quarter to one year, and is defined in equation (IA.53). *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.

	Timing 1-Q			Timing 2-Q			Timing 3-Q			Timing 4-Q		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Covered × Post	-0.005	-0.004	-0.006	-0.012	-0.009	-0.012	-0.018	-0.013	-0.017	-0.021	-0.013	-0.019
	(0.006)	(0.005)	(0.006)	(0.011)	(0.009)	(0.010)	(0.015)	(0.012)	(0.014)	(0.020)	(0.014)	(0.017)
Covered	0.002			0.003			0.005			0.005		
	(0.007)			(0.012)			(0.017)			(0.022)		
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	1.29e+07	1.29e+07	1.29e+07	1.26e+07	1.26e+07	1.26e+07	1.23e+07	1.23e+07	1.23e+07	1.20e+07	1.20e+07	1.20e+07
R^2	0.60	0.61	0.67	0.59	0.61	0.70	0.58	0.60	0.72	0.56	0.59	0.74

I.6 Proxies for Quant Funds

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

We conjecture that quant funds are more likely to buy alternative data as they have the skills required to use them. Identifying these funds in our sample is not straightforward as we do not directly observe a fund's type. In addition, quant funds might be sophisticated investors out of our sample, such as hedge funds. To build a proxy for quant funds in our sample, we rely on the text of fund prospectuses that mutual funds must submit to the SEC at least once a year. In its prospectus, a fund provides information on its strategy, risks, fees, and performance. In particular, in the strategy section, funds provide information regarding their investment process. Thus, we search for specific keywords in this section of fund prospectuses to identify funds that are more likely to be quant.[4]

We obtain fund prospectuses (Form N-1A) from the SEC EDGAR (Electronic Data Gathering, Analysis, and Retrieval) system. Focusing on the Principal Investment Strategy (PIS) section, the most informative about fund investment process, we extract this section and merge it with our fund holding dataset using tickers. We follow standard text cleaning procedures to clean the prospectus text. We only keep the English words in the prospectuses by removing numbers, symbols and special characters. We also remove all the stop words. In addition, we stem each word to its root using the Porter stemmer algorithm (e.g.

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

'mathematic', 'mathematics', ... = 'mathemat').

We then search for various combinations of keywords in fund prospectuses. We do not rely solely on a generic term like "quantitative" to avoid potential false positives. Indeed, about 30% of fund prospectuses mention this term, sometimes in a way that is not highly indicative of a quantitative investment process.[5] Instead, we focus on more specific keywords likely to be relevant in our context. First we search for "quantitative stock selection" (stemmed to "quantit stock select") and create a fund-level dummy variable, "Quantit Stock Select", equal to one if the fund ever mentions this term in its prospectus. We further identify funds combining this term with others like "proprietary" (stemmed to "proprietari") and "rank", suggested to be highly informative about quant funds by Abis (2022), resulting in two additional dummy variables: "Quantit Stock Select & Proprietari" and "Quantit Stock Select & Rank".[6]

We then estimate a specification similar to eq.(9), interacting "Covered × Post" with either "Quantit Stock Select", "Quantit Stock Select & Proprietari" or "Quantit Stock Select & Rank". We present estimates of these tests in Table IA.8. Columns (1), (4), (7), and (10) show that the coefficient on the triple interaction term "Covered × Post × Quantit Stock Select" is positive, although not statistically significant in all specifications. In Columns (2)-(3), (5)-(6), and (11)-(12), we find that the coefficients on the triple interaction term "Covered × Post × Quantit Stock Select & Proprietari" as well as "Covered × Post × Quantit Stock Select & Rank" are both positive and statistically significant. The magnitude of these coefficients is similar to that of the coefficient on the interaction term "Covered × Post".

Thus, as predicted, coverage initiation has a weaker negative effect on funds that are more likely to be able to use the data considered in our tests. Our inference however is limited by our inability to directly observe funds that buy the data and by the fact we only

[5] For example, a fund mentions in its prospectus *"The subadviser will exercise judgment to determine ESG best practices based on its long standing experience managing ESG investment strategies [...]. Leadership may be assessed both quantitatively and qualitatively, through the subadviser."*

[6] Abis (2022) also finds that the word "model" is very informative about the fund being a quant fund. In our sample all prospectuses mentioning "quantitative stock selection" also mention "model".

focus on mutual funds. Indeed, we expect sophisticated investors out of our sample, such as hedge funds, to also exploit and trade on alternative data, and therefore to benefit after coverage initiation.

Table IA.8: 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. $Covered \times Post$ 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 variables "Quantit Stock Select", "Quantit Stock Select & Proprietari" and "Quantit Stock Select & Rank", equal to one if the fund ever mentions the corresponding terms in its prospectus. The regressions in the table do not include the picking abilities for funds for which we cannot identify a prospectus in the SEC EDGAR system. 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	-0.021**	-0.021**	-0.021**	-0.039**	-0.039**	-0.039**	-0.060*	-0.060*	-0.060*	-0.099**	-0.099**	-0.099**
	(0.009)	(0.009)	(0.009)	(0.020)	(0.020)	(0.020)	(0.032)	(0.032)	(0.032)	(0.047)	(0.047)	(0.047)
Covered × Post × Quantit Stock Select	0.012*			0.031*			0.045			0.078		
	(0.007)			(0.018)			(0.031)			(0.048)		
Covered × Post × Quantit Stock Select & Proprietari		0.015**			0.042***			0.061**			0.104**	
		(0.007)			(0.014)			(0.026)			(0.042)	
Covered × Post × Quantit Stock Select & Rank			0.014**			0.041***			0.061**			0.107**
			(0.006)			(0.013)			(0.026)			(0.042)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8638837	8638837	8638837	8380015	8380015	8380015	8116899	8116899	8116899	7852414	7852414	7852414
R^2	0.20	0.20	0.20	0.28	0.28	0.28	0.33	0.33	0.33	0.38	0.38	0.38

I.7 Stock Picking using a Matched Sample of Control Stocks

Table IA.9: 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 abilities. 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* × *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	740,301	740,301	740,301	723,093	723,093	723,093	705,773	705,773	705,773	688,352	688,352	688,352
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

Table IA.10: Heterogeneous Effect based on Industry Expertise using a Matched Sample of Control Stocks

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, 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. 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	-0.014***	-0.015***	-0.027***	-0.023**	-0.039***	-0.029*	-0.059***	-0.049**
	(0.005)	(0.006)	(0.010)	(0.011)	(0.015)	(0.017)	(0.021)	(0.024)
Covered × Post × Industry Specialist	-0.144***	-0.150***	-0.338***	-0.372***	-0.540***	-0.608***	-0.791***	-0.910***
	(0.047)	(0.038)	(0.116)	(0.093)	(0.192)	(0.155)	(0.291)	(0.245)
Covered × Industry Specialist	0.093***		0.207***		0.315***		0.432***	
	(0.025)		(0.058)		(0.091)		(0.126)	
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	740,301	740,301	723,093	723,093	705,773	705,773	688,352	688,352
R^2	0.33	0.40	0.34	0.46	0.36	0.51	0.37	0.55

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	-0.014***	-0.014***	-0.028***	-0.022**	-0.039***	-0.028*	-0.060***	-0.047**
	(0.005)	(0.005)	(0.009)	(0.011)	(0.015)	(0.016)	(0.021)	(0.023)
Covered × Post × Sector Fund	-0.116***	-0.134***	-0.262**	-0.322***	-0.417**	-0.524***	-0.600**	-0.782***
	(0.043)	(0.033)	(0.107)	(0.086)	(0.174)	(0.144)	(0.260)	(0.226)
Covered × Sector Fund	0.074***		0.168***		0.264***		0.368***	
	(0.026)		(0.060)		(0.091)		(0.123)	
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	740,301	740,301	723,093	723,093	705,773	705,773	688,352	688,352
R^2	0.33	0.40	0.34	0.46	0.36	0.51	0.37	0.55

Table IA.11: Heterogeneous Effect based on Geographical Location using a Matched Sample of Control Stocks

The table presents the results of our study on the differential impact of alternative data on fund picking abilities depending on fund location. Regressions are estimated at the fund-stock-quarter level, 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. The table presents estimation results of specifications that include interactions with a dummy variable, "Local", which equals one if the fund is located in the same MSA as either (i) the firm's headquarters or (ii) the stock's primary MSA based on parking lots, as identified through satellite imagery data (the MSA where the highest number of the firm's parking lots are located). The regressions in the table do not include the picking skills for funds for which we are unable to obtain the official address from the CRSP database. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.016***	-0.018***	-0.032***	-0.032**	-0.046**	-0.044**	-0.069***	-0.071**
	(0.005)	(0.006)	(0.012)	(0.013)	(0.018)	(0.020)	(0.025)	(0.029)
Covered × Post × Local	-0.028***	-0.044***	-0.056**	-0.090***	-0.088**	-0.139***	-0.126**	-0.208***
	(0.011)	(0.013)	(0.023)	(0.034)	(0.035)	(0.051)	(0.051)	(0.076)
Covered × Local	0.021**		0.041***		0.062***		0.084***	
	(0.008)		(0.015)		(0.023)		(0.032)	
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	557,134	557,134	544,481	544,481	531,705	531,705	518,814	518,814
R^2	0.33	0.41	0.35	0.46	0.36	0.52	0.37	0.55

Table IA.12: Heterogeneous Effects based on Skills before Coverage Initiation using a Matched Sample of Control Stocks

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, 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. The table presents estimation results of specifications that include interactions with a dummy variable, "High Picking Pre", which equals one if the stock is covered by RS Metrics and the fund has a picking ability above the median for that stock before the release of satellite data imagery. The regressions in the table do not include the picking skills for stocks covered by RS Metrics for funds that start holding the stock after the release of satellite data imagery. In other words, we only analyze the effect of alternative data on funds that had a certain level of stock-picking ability before the satellite data imagery was released. All picking skills for uncovered stocks are included. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.003	-0.001	0.008	0.002
	(0.005)	(0.011)	(0.015)	(0.020)
Covered × Post × High Picking Pre	-0.043***	-0.090***	-0.149***	-0.208***
	(0.009)	(0.018)	(0.032)	(0.049)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	599,805	587,808	575,682	563,237
R^2	0.42	0.47	0.52	0.55

I.8 Robustness Test: Adding Stock-Quarter Fixed Effects

Table IA.13: Heterogeneity across Funds when Including Stock-Quarter Fixed Effects

This table reproduces our main results on the heterogeneous effect across funds of the release of alternative data on picking abilities. Regressions are estimated at the fund-stock-quarter level as in the main text but add stock-quarter fixed effects. 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)	(13)	(14)	(15)	(16)
Covered × Post × High Picking Pre	-0.056***				-0.115***				-0.179***				-0.241***			
	(0.010)				(0.024)				(0.042)				(0.063)			
Covered × Post × Industry Specialist		-0.116**				-0.311**				-0.537**				-0.837**		
		(0.054)				(0.152)				(0.258)				(0.401)		
Covered × Post × Sector Fund			-0.090**				-0.233*				-0.400*				-0.620*	
			(0.046)				(0.131)				(0.224)				(0.347)	
Covered × Post × Local				-0.011				-0.029				-0.041				-0.059
				(0.009)				(0.019)				(0.027)				(0.039)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1.27e+07	1.28e+07	1.28e+07	9972770	1.23e+07	1.24e+07	1.24e+07	9681321	1.19e+07	1.20e+07	1.20e+07	9383392	1.15e+07	1.16e+07	1.16e+07	9083207
R^2	0.49	0.49	0.49	0.49	0.55	0.55	0.55	0.54	0.58	0.58	0.58	0.58	0.61	0.61	0.61	0.61

I.9 Stock Picking across Cohorts

Table IA.14: Alternative Data and Stock Picking Skills across Coverage Cohorts

This table presents our main results on the effect of the release of alternative data on fund picking abilities, estimating the effect across different cohorts of covered stocks. 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. Each column corresponds to a "stacked regression" approach. For each treatment date, we create a cohort of treated stocks and never-covered stocks, and we restrict the sample to a window of 5 years pre- and post-coverage date. We create a similar sample for each treatment date and we then "stack" the samples into one dataset, creating a variable that identifies the event (i.e., "cohort") each observation belongs to. Columns (1), (4), (7) and (10) stack the cohorts with treatment year up to 2013. Columns (2), (5), (8) and (11) stack the cohorts with treatment year up to 2015. Columns (3), (6), (9) and (12) stack the cohorts with treatment year up to 2017. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q			Picking 2-Q			Picking 3-Q			Picking 4-Q		
	$\leq$ 2013	$\leq$ 2015	$\leq$ 2017	$\leq$ 2013	$\leq$ 2015	$\leq$ 2017	$\leq$ 2013	$\leq$ 2015	$\leq$ 2017	$\leq$ 2013	$\leq$ 2015	$\leq$ 2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Covered × Post	-0.020**	-0.025***	-0.020***	-0.032*	-0.049***	-0.040***	-0.055*	-0.078***	-0.064**	-0.090**	-0.102***	-0.102***
	(0.009)	(0.007)	(0.007)	(0.019)	(0.015)	(0.015)	(0.030)	(0.026)	(0.025)	(0.041)	(0.036)	(0.036)
Fund × Year-Quarter × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock × Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1.10e+08	1.67e+08	1.99e+08	1.09e+08	1.66e+08	1.96e+08	1.08e+08	1.64e+08	1.93e+08	1.07e+08	1.89e+08	1.89e+08
R^2	0.09	0.09	0.09	0.11	0.11	0.12	0.12	0.13	0.13	0.14	0.15	0.15

I.10 Stock Picking before 2018

Table IA.15: Alternative Data and Stock Picking Skills before 2018

This table presents our main results on the effect of the release of alternative data on fund picking abilities, restricting our sample to the pre-2018 period, i.e., before the retail traffic product by Orbital Insight (the main competitor of RS Metrics) were distributed via the Bloomberg terminal. 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***	-0.024***	-0.023***	-0.036***	-0.043***	-0.036**	-0.055***	-0.064***	-0.050*	-0.082***	-0.099***	-0.086**
	(0.006)	(0.007)	(0.008)	(0.013)	(0.015)	(0.017)	(0.021)	(0.025)	(0.028)	(0.030)	(0.034)	(0.041)
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	9529922	9529922	9529922	9452755	9452755	9452755	9368098	9368098	9368098	9277784	9277784	9277784
R^2	0.07	0.09	0.20	0.07	0.11	0.28	0.07	0.12	0.34	0.06	0.14	0.39

I.11 Analysis of Funds' Return Gap

Table IA.16: Alternative Data and Return Gap

This table presents our results on the effect of the release of alternative data on funds' return gap when holding covered stocks. Regressions are estimated at the fund-month level. The dependent variable is the fund's return gap (Kacperczyk et al., 2008) in percentage points, calculated as the difference between the net investor return and the net holdings return. The return gap captures the funds' unobserved actions. "Fraction of Assets in Covered Stocks" is the fraction (between 0 and 1) of the fund's total net assets invested in stocks that are at some point covered by RS-Metrics. "Fraction of Assets in Covered Stocks Pre-Coverage" is the fraction of the fund's total net assets invested in stocks that will be but are not yet covered by RS-Metrics. "Fraction of Assets in Covered Stocks Post-Coverage" is the fraction (between 0 and 1) of the fund's total net assets invested in stocks that are currently covered by RS-Metrics. Standard errors are double-clustered at the fund and year-month levels.

	Return Gap (%)			
	(1)	(2)	(3)	(4)
Fraction of Assets in Covered Stocks	-0.268	-0.247		
	(0.288)	(0.286)		
Fraction of Assets in Covered Stocks Pre-Coverage			-1.164	-1.120
			(1.174)	(1.178)
Fraction of Assets in Covered Stocks Post-Coverage			-0.092	-0.076
			(0.355)	(0.354)
Log(TNA)		-0.015		-0.015
		(0.018)		(0.018)
Log(Fund Age)		-0.092		-0.092
		(0.065)		(0.065)
Year-Month FE	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes
Observations	314,374	314,374	314,374	314,374
R^2	0.09	0.09	0.09	0.09

I.12 Decomposing Picking

Table IA.17: Alternative Data and Each Component of Picking

This table presents our results on the effect of the release of alternative data on each component of fund picking abilities. In columns (1) and (2), regressions are estimated at the fund-stock-quarter level and the dependent variable is the fund's tilt towards the stock relative to the market. This is defined as $w_{i,t}^{f} - w_{i,t}^{m}$, 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 stock i (its weight in the "market portfolio") at the end of quarter t. In columns (3) to (5), the regressions are estimated at the stock-quarter level. The dependent variable is the stock's quarterly return in column (3), the stock's beta with respect to the market in column (4), and the stock's quarterly abnormal return (i.e., the difference between the stock's quarterly return and the product of its beta and the market's quarterly return) in column (5). *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 in columns (1) and (2), and are clustered at the stock level in columns (3) to (5).

	$w^f - w^m$		Stock Return	β	Stock Abnormal Return
	(1)	(2)	(3)	(4)	(5)
Covered × Post	-0.001*	-0.001	0.004	0.048	-0.006
	(0.001)	(0.001)	(0.011)	(0.037)	(0.010)
Fund × Year-Quarter FE	Yes	Yes	No	No	No
Stock FE	Yes	No	Yes	Yes	Yes
Fund × Stock FE	No	Yes	No	No	No
Year-Quarter FE	No	No	Yes	Yes	Yes
Observations	1.32e+07	1.32e+07	220,712	220,712	220,712
R^2	0.58	0.84	0.20	0.62	0.09

I.13 Stock Picking Excluding Industry and Geographical Peers

Table IA.18: 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 abilities. 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 largest number 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**	-0.055***	-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***			0.048***			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	4194735	4194735	4194735	4059453	4059453	4059453	3921939	3921939	3921939	3782795	3782795	3782795
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

References

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

Dugast, J. and Foucault, T. (2024). Equilibrium data mining and data abundance. Forthcoming Journal of Finance.

Goldstein, I. and Yang, L. (2015). Information diversity and complementarities in trading and information acquisition. *The Journal of Finance*, 70(4):1723–1765.

Kacperczyk, M., Sialm, C., and Zheng, L. (2008). Unobserved actions of mutual funds. *The Review of Financial Studies*, 21(6):2379–2416.

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

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