Online Appendix for

Does Big Data Devalue Traditional Expertise? Evidence from Active Funds

(not intended for publication)

Maxime Bonelli and Thierry Foucault

Contents

I.1. Hypothesis Development: Theory

In this section, we present the framework that generates the implications tested in the paper. We present the baseline model in Section I.1.1, derive its equilibrium in Section I.1.2 and its testable implications (hypotheses H.1, H.2, and H.3 in the text) in Section I.1.3. The baseline model is similar to Dugast and Foucault (2025) and uses some of the results derived in this paper.

I.1.1. *Framework*

We consider the market for a risky asset whose payoff, v, is realized at date 2. The payoff of the risky asset is normally distributed with mean zero and variance σ_v^2.

Fund Managers. 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.

At date 1, fund managers generate signals about the payoff of the risky asset 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 $k \in \{ex, dm\}$ receives the signal:

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

where $\tau_k(\iota_j) = \tau_k + \iota_j\Delta_k$ depends on the type $k \in \{ex, dm\}$ of the fund manager and ι_j is determined by whether or not fund manager j purchased new data available at date 0: $\iota_j = 1$ if the manager has purchased new data at date 0 and $\iota_j = 0$ otherwise (more on this 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 $\tau_k(\iota_j)$, the higher is the precision ("quality") of fund manager j's signal.

New Data. At date 0, fund managers have the possibility to buy new data about the asset at cost C. If manager j of type $k \in \{ex, dm\}$ buys the data then $\iota_j = 1$ and the precision of her signal at date 1 is:

$$\tau_k(1) = \tau_k + \Delta_k. \tag{IA.2}$$

If instead the manager decides not to buy the data, then $\iota_j = 0$ and the precision of her signal at date 1 is $\tau_k(0) = \tau_k$. Thus, new data enables fund managers to improve the precision of their signal (τ_k is the precision in the absence of new data for a fund manager with type k).

If they buy data, data miners can improve the precision of their signals by a larger margin than experts. Specifically we assume that: $\Delta_{ex} = \kappa\Delta_{dm}$, where $0 \leq \kappa < 1$. This assumption reflects that data miners (quants) have better infrastructure (e.g., storage and computing capacities) and teams (data analysts) to combine large datasets and process these data quantitatively. As a result, they extract more precise signals from the same data.

In the absence of data, we assume that data miners' signals are less precise than experts' signals. That is, $\tau_{dm} \leq \tau_{ex}$. This assumption is not necessary for our implications, but it facilitates the exposition and the derivation of our results. Moreover, it captures a plausible scenario in which the availability of alternative data enables data miners to catch up with experts.

Given these assumptions, we show in Section I.1.2 that data miners have a higher willingness to pay than experts for the data. As a result, in equilibrium, there are always more data miners buying data than experts, holding C constant. To simplify the exposition, we focus on the case in which it is optimal for all data miners to buy the data. As shown below (see (IA.21)), this is the case when $C \leq C_1(\alpha, \Delta_{dm}, \kappa)$, where α is the fraction of experts buying data. Thus, in the rest of the analysis, we assume C satisfies this condition and set $\iota = 1$ for data miners.

Trading. After observing her signal, each fund manager chooses a trading strategy, i.e., a demand schedule, $x(s_j(\tau_k(\iota)_j), p)$, where, p, is the price of the risky asset. Thus, at date

2, fund manager j with type $k \in \{dm, ex\}$ and data choice ι_j returns to her client:

$$W_{j,k} = W_0 + x(s_j(\tau_k(\iota_j)), p)(v - p). \tag{IA.3}$$

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, $D(p) = \int_0^\mu x(s_j(\tau_{dm}(1)), p)dj + \int_\mu^1 x(s_j(\tau_{ex}(\iota_j), p)dj + \eta$ 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:

$$p^*(D) = \mathsf{E}\left[v \,|D(p^*) = D\right]. \tag{IA.4}$$

Fund managers' objective function. 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 trading strategy, $x(s_j(\tau_k(\iota_j)), p)$, and her decision to buy data or not, $\iota_j \in \{0, 1\}$, to maximize her client's expected utility. Each fund manager's problem is dynamic because she first decides whether to buy data at date 0, and then, at date 1, she chooses her demand for the risky asset after observing her signal.

This dynamic optimization problem can be solved by backward induction. That is, at date 1, fund manager j chooses her demand of the risky asset, $x^*(s_j, p)$, to solve:[1]

$$\max_x \mathsf{E}\left[-\exp\big(-\rho\big(W_0 - \iota_j \times C + x(v-p)\big)\big) \,\big|\, s_j(\tau_k(\iota_j)),\, p = p^*(D)\right], \tag{IA.5}$$

and, at date 0, she makes her decision to buy data or not, ι_j, to maximize the unconditional expected utility:

$$\max_{\iota_j \in \{0,1\}} \mathsf{E}(\big[-\exp\big(-\rho\big(W_0 - \iota_j \times C + x^*(s_j, p)(v-p)\big)\big)\big]). \tag{IA.6}$$

[1]As there is a continuum of fund managers, each is price-taker.

In choosing her portfolio, each fund manager accounts for the information in the equilibrium price, in addition to her own signal.

I.1.2. *Equilibrium*

To solve for equilibrium decisions in the model, we proceed by backward induction. We first characterize the market equilibrium at date 1, taking as given the mass of experts, α, who purchase data at date 0. We then solve for α. At every step, we closely follow the derivations in Dugast and Foucault (2025), since the framework we consider is an adaptation of theirs, specifically tailored to study how the availability of new data affects fund managers' performance.

B.1 Equilibrium at Date 1.

At date 1, for a fixed fraction α of experts buying the data, a noisy rational expectations equilibrium of the market is a vector $(x^*(\cdot,\cdot), p^*(D))$ such that:

1. $x^*(s(\tau), p)$ maximizes (IA.5) for a fund manager receiving the signal $s(\tau)$ with precision τ, given that other fund managers's demand for the risky asset is given by $x^*(\cdot,\cdot)$ and that market makers' price is given by $p^*(D)$.

2. The price of the risky asset, $p^*(D)$, satisfies $p^* = \mathsf{E}\left[v \,|D^*(p^*) = D\right]$ where $D^*(p^*) = \int_0^\mu x^*(s_j(\tau_{dm}(1)), p)dj + \int_\mu^1 x^*(s_j(\tau_{ex}(\iota_j)), p)dj + \eta$.

We denote the average precision of signals across fund managers by $\bar{\tau}(\alpha, \Delta_{dm}, \kappa)$. It is given by:

$$\begin{aligned}\bar{\tau}(\alpha, \Delta_{dm}, \kappa) &= \mu\tau_{dm}(1) + (1-\mu)\alpha\tau_{ex}(1) + (1-\mu)(1-\alpha)\tau_{ex}(0) \\ &= \mu(\tau_{dm} + \Delta_{dm}) + (1-\mu)\tau_{ex} + (1-\mu)\alpha\kappa\Delta_{dm},\end{aligned} \tag{IA.7}$$

where the second equality follows from the definition of $\tau_k(\iota_j)$. To avoid cluttering notation, we will sometimes denote $\bar{\tau}(\alpha, \Delta_{dm}, \kappa)$ simply by $\bar{\tau}$, unless doing so reduces clarity.

We can derive the equilibrium of the market for the risky asset as in Dugast and Foucault (2025), replacing the average precision of fund managers' signals (that they denote $\bar{\tau}$) by

$\bar{\tau}(\alpha, \Delta_{dm}, \kappa)$. This yields the following characterization of the equilibrium. As it follows directly from Proposition 1 in Dugast and Foucault (2025), we do not reprove it here.

Proposition IA.1 *(Dugast and Foucault (2025)): In equilibrium, the demand for the risky asset of manager receiving a signal of precision τ is*

$$x^*(s(\tau), p) = \frac{\mathsf{E}[v|s(\tau), p] - p}{\rho\,\mathsf{Var}[v|s(\tau), p]} = \frac{\tau\,(s(\tau) - p)}{\rho\sigma_v^2}, \tag{IA.8}$$

and the equilibrium price of the asset is

$$p^*(D) = \mathsf{E}[v|D^*(p) = D] = \left(\frac{\bar{\tau}(\alpha, \Delta_{dm}, \kappa)^2}{\bar{\tau}(\alpha, \Delta_{dm}, \kappa)^2 + \rho^2\sigma_v^2\sigma_\eta^2}\right)\xi, \tag{IA.9}$$

where $\xi \equiv v + \rho\sigma_v^2\bar{\tau}(\alpha, \Delta_{dm}, \kappa)^{-1}\eta$.

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 IA.1, it is direct that:

$$\mathcal{I}(\bar{\tau}) \equiv \mathsf{Var}[v \mid p^*]^{-1} = \frac{1}{\sigma_v^2} + \frac{\bar{\tau}(\alpha, \Delta_{dm}, \kappa)^2}{\rho^2\sigma_v^4\sigma_\eta^2}. \tag{IA.10}$$

From (IA.7) and (IA.10), we deduce that

$$\frac{\partial \mathcal{I}}{\partial \Delta_{dm}} = \frac{2(\mu + (1-\mu)\kappa\alpha)\bar{\tau}(\alpha, \Delta_{dm}, \kappa)}{\rho^2\sigma_v^4\sigma_\eta^2} > 0. \tag{IA.11}$$

Thus, the availability of new data about the risky asset enhances price informativeness because it increases the precision of the signals used by fund managers buying the data. This effect drives many of our implications.

We denote by $R_k(s_j(\tau_k(\iota_j)))$ 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:

$$R_k(s_j(\tau_k(\iota_j))) \equiv \frac{x^*(s_j(\tau_k(\iota_j)), p^*) \times (v - p^*)}{W_0} = w(s_j(\tau_k(\iota_j)))R^e, \tag{IA.12}$$

where $R^e = v/p^* - 1$ is the risky asset return in excess of the risk free rate [2] and

$$w(s_j(\tau_k(\iota_j))) = \frac{x^*(s_j(\tau_k(\iota_j)), p^*)p^*}{W_0}, \tag{IA.13}$$

is the weight of the risky asset in the portfolio of the fund manager. This weight increases with $s_j(\tau_k(\iota_j))$ and, holding $s_j(\tau_k(\iota_j))$ constant, it increases in $\tau_k(\iota_j)$ in absolute value. Intuitively, fund managers take larger bets on their signals when the precision of these signals is higher.

From (IA.8), we obtain

$$x^*(s_j(\tau_k(\iota_j)), p^*) = \frac{1}{\rho\sigma_v^2}\left(\tau_k(\iota_j)(v - p^*) + \tau_k(\iota_j)^{1/2}\varepsilon_j\right). \tag{IA.14}$$

We deduce that the *expected* gross excess return of fund manager j of type k with predictor's quality $\tau_k(\iota_j)$ is

$$\bar{R}_k(\tau_k(\iota_j)) := \mathbb{E}(R(s_j(\tau_k))) = \frac{\tau_k(\iota_j)}{W_0\rho\sigma_v^2}\mathsf{Var}[v - p^*] = \frac{\tau_k(\iota_j)}{W_0\rho\sigma_v^2\mathcal{I}(\bar{\tau})}, \quad k \in \{ex, dm\}, \tag{IA.15}$$

where the second equality follows from substituting (IA.13) in (IA.12) and the independence between $(v - p^*)$ and ε_j. The last equality follows from the fact that $p^* = \mathsf{E}(v \mid p^*)$ so that $\mathsf{Var}[v - p^*] = \mathsf{Var}[v \mid p^*] = \mathcal{I}^{-1}$.

In line with intuition, the expected gross asset return of a fund manager in the risky asset (her "stock picking ability") increases with the precision of her signal ($\tau_k(\iota_j)$) and decreases with price informativeness.

B.2 Equilibrium Decisions to Buy Data at date 0.

In this section, we show that experts have a lower willingness to buy data than data miners. This leads to the existence of an equilibrium in which no experts buy the data while all data miners do.

Willingness to pay for data. Following the same steps as those in Dugast and Foucault (2025) for their Lemma 2, we obtain that the ex-ante expected utility of a fund manager j

[2] The risk free rate is set to zero, as is standard in noisy rational expectations models.

who makes decision $\iota_j \in \{0,1\}$, when (i) a fraction α of experts buy the data and (ii) all data miners buy the data, is

$$\begin{aligned} g(\tau_k + \iota_j \Delta_k; \alpha) &\equiv \mathsf{E}\big[-\exp\big(-\rho\big(W_0 - \iota_j \times C + x^*(s,p)\big(v - p^*(D)\big)\big)\big)\big] \\ &= -\exp(\rho(\iota_j \times C - W_0)) \left(1 + \frac{\tau_k(\iota_j)}{\sigma_v^2 \mathcal{I}(\bar{\tau}(\alpha, \Delta_{dm}, \kappa))}\right)^{-\frac{1}{2}}, \\ &= -\exp(\rho(\iota_j \times C - W_0)) \left(1 + \rho W_0 \bar{R}_k(\tau_k(\iota))\right)^{-\frac{1}{2}}, \quad \text{for } k \in \{ex, dm\} \text{ and } \iota_j \in \{0,1\}. \end{aligned} \tag{IA.16}$$

In line with intuition, a fund manager's expected utility increases with her expected excess return $\bar{R}_k(\tau_k(\iota_j))$. Moreover, if a fund manager buys the data then she obtains a larger expected excess return than if she does not, holding fixed other fund managers' decisions. Indeed, holding α fixed, a fund manager optimally buys the data iff:

$$g(\tau_k + \Delta_k; \alpha) > g(\tau_k; \alpha), \tag{IA.17}$$

which, using eq.(IA.16), is equivalent to:

$$\left(\frac{1 + \rho W_0 \bar{R}_k(\tau_k(1))}{1 + \rho W_0 \bar{R}_k(\tau_k(0))}\right)^{\frac{1}{2}} > exp(2\rho C) > 1, \tag{IA.18}$$

which requires:

$$\bar{R}_k(\tau_k(1)) \geq \bar{R}_k(\tau_k(0)) \quad for \quad k \in \{dm, ex\}. \tag{IA.19}$$

For a given α, all data miners are better off buying data if and only if (IA.17) holds for $k = dm$. Using (IA.16), we deduce after some algebra that this is the case if and only if

$$C \leq C_1(\alpha, \Delta_{dm}, \kappa), \tag{IA.20}$$

where

$$C_1(\alpha, \Delta_{dm}, \kappa) \equiv \frac{1}{2\rho} ln\left(1 + \frac{\Delta_{dm}}{\tau_{dm} + \rho\sigma_v^2 \mathcal{I}(\bar{\tau}(\alpha, \Delta_{dm}, \kappa))}\right). \tag{IA.21}$$

Observe that $C_1(\alpha, \Delta_{dm}, \kappa)$ is the largest amount that the data miners are willing to pay for the data when all data miners buy the data and when a fraction α of experts buy the data. Hence, we refer to $C_1(\alpha, \Delta_{dm}, \kappa)$ as data miners' willingness to pay for the data.

Now consider experts. When an expert expects a fraction α of other experts to buy the data, he is better off buying data if and only if

$$g(\tau_{ex}; \alpha) \leq g(\tau_{ex} + \Delta_{ex}; \alpha).$$

Using (IA.16), we deduce after some algebra that this condition is equivalent to

$$C \leq C_0(\alpha, \Delta_{dm}, \kappa), \tag{IA.22}$$

where

$$C_0(\alpha, \Delta_{dm}, \kappa) \equiv \frac{1}{2\rho} ln(1 + \frac{\kappa \Delta_{dm}}{\tau_{ex} + \rho \sigma_v^2 \mathcal{I}(\bar{\tau}(\alpha, \Delta_{dm}, \kappa))}). \tag{IA.23}$$

We refer to $C_0(\alpha, \Delta_{dm}, \kappa)$ as experts' willingness to pay for the data since this is the maximum amount that each expert is willing to pay to buy the data when she expects a fraction α of experts and all data miners to buy the data.

Observe that $C_0(\alpha, \Delta_{dm}, \kappa) < C_1(\alpha, \Delta_{dm}, \kappa)$ because $\kappa < 1$ and $\tau_{ex} > \tau_{dm}$. Thus, data miners always have a higher willingness to pay for the data than experts because (a) they process the data more efficiently ($\kappa < 1$) and (b) they start with signals of lower precision ($\tau_{ex} > \tau_{dm}$).

Equilibrium. Fund managers' willingness to pay for the data depends on their expectations about other fund managers' decisions to buy it, because the fraction purchasing the data affects price informativeness and thus performance, depending on whether a given manager buys the data or not. We therefore consider a Nash equilibrium in which each fund manager's decision at date 0 is a best response to her belief about the mass α^* of experts buying the data. This mass is consistent with a Nash equilibrium if and only if, when fund managers expect all data miners to buy the data and a fraction α^* of experts to buy the data then (i) a fraction α^* of experts find it optimal to buy the data, (ii) a fraction $(1-\alpha^*)$ of experts find it optimal not to buy the data and (iii) all data miners find optimal to buy the data.

There are two cases.

Case 1. Let define $\bar{C}_0(\Delta_{dm}, \kappa) = C_0(0, \Delta_{dm}, \kappa)$ and $\bar{C}_1(\Delta_{dm}, \kappa) = C_1(0, \Delta_{dm}, \kappa)$. Case 1 obtains when:

$$\bar{C}_0(\Delta_{dm}, \kappa) \leq C \leq \bar{C}_1(\Delta_{dm}, \kappa), \tag{IA.24}$$

In this case, the unique equilibrium is such that $\alpha^* = 0$. Indeed, as $C > C_0(0, \Delta_{dm}, \kappa)$, no expert find optimal to buy the data given that they expect no expert to buy the data and all data miners to buy the data. Moreover, as $C \leq C_1(0, \Delta_{dm}, \kappa)$, all data miners find optimal to buy the data given that $\alpha^* = 0$. Thus, when (IA.24) holds, $\alpha^* = 0$ is a Nash equilibrium.

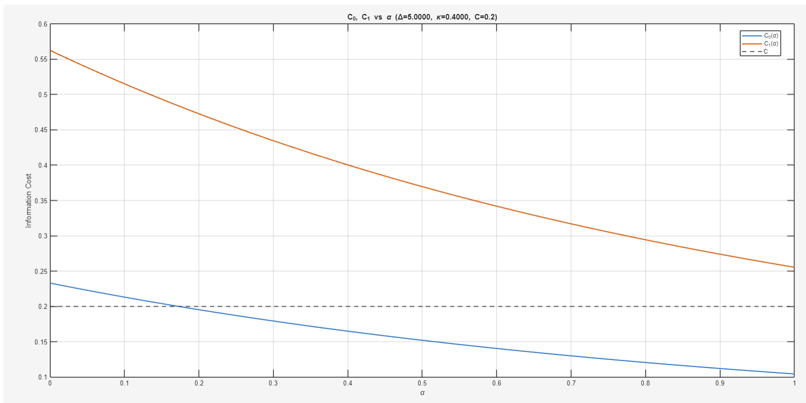
Observe that if κ decreases, then $C_0(0, \Delta_{dm}, \kappa)$ decreases, and therefore the range of values for C such that no experts buy the data becomes larger. When $\kappa = 0$, $C_0(0, \Delta_{dm}, \kappa) = 0$, and therefore experts never buy the data for any parameter values. Thus, there is always a value of κ small enough such that Case 1 arises, no matter how small the cost of data is.

Case 2. When $C < C_0(0, \Delta_{dm}, \kappa)$, we obtain Case 2. In Case 2, the equilibrium is such that $0 < \alpha^* < 1$ and solves:

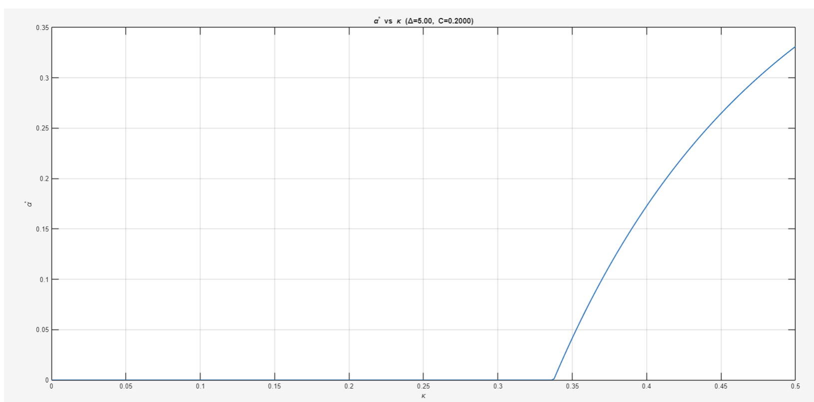
$$C_0(\alpha^*, \Delta_{dm}, \kappa) = C. \tag{IA.25}$$

Indeed, when this condition is satisfied for $\alpha = \alpha^*$, all experts are indifferent between buying and not buying the data, and so a fraction $\alpha^* \in (0, 1)$ buying the data is an equilibrium. This equilibrium is unique because $C_0(\alpha, \Delta_{dm}, \kappa)$ decreases with α. Hence, the equation $C_0(\alpha^*, \Delta_{dm}, \kappa) = C$ has a unique solution in $(0, 1)$ when such a solution exists. If it does not, it must be either because $C > C_0(0, \Delta_{dm}, \kappa)$ (Case 1) or because C is so small that the previous equation has no solution strictly smaller than 1. In the latter case, the equilibrium is $\alpha^* = 1$. It never arises, however, for κ small enough, since $C_0(0, \Delta_{dm}, \kappa)$ goes to zero when κ goes to zero. Figure IA.1 provides a summary of the results in this section using a numerical example.

Figure IA.1: Parameter values: $C = 0.2$ (Panel B), $\mu = 0.1$, $\rho = 0.5$, $\kappa = 0.4$ (Panel A), $\Delta = 5$, $\sigma_v = 1.5$, $\sigma_\eta = 1$, $\tau_{dm} = 1$, $\tau_{ex} = 2$, $W = 1$.



Panel A: $C_0(\alpha, \Delta_{dm}, \kappa)$ (blue curve) and $C_1(\alpha, \Delta_{dm}, \kappa)$ (orange curve) as a function of α. When $C < C_0(0, \Delta_{dm}, \kappa)$, the equilibrium is such that some experts buy the data and all data miners do. For instance, for the parameter values in this example, if $C = 0.2$ (dashed line), $\alpha^* \approx 0.17$.



Panel B: Effect of κ on the equilibrium fraction of experts who buy the data. For κ larger enough, $\alpha^* > 0$ and increases with κ.

I.1.3. *Testable Hypotheses*

In this section, we compare the expected excess return of experts and data miners in the economy in which no new data is available and the economy in which new data is available. The equilibrium at date 1 of the economy in which no new data is available is obtained by simply setting $\Delta_{dm} = 0$ in the characterization of the equilibrium at date 1 (and therefore in (IA.7) and (IA.15)). Indeed, in this case, data miners' signals have the same precision as the precision of their signals in the absence of new data.

We first focus on the case in which $C_0(0, \Delta_{dm}, \kappa) \leq C \leq C_1(0, \Delta_{dm}, \kappa)$ so that when data are available, all data miners buy the data and no experts buy the data (see the previous section). We discuss other cases at the end of the section.

Consider experts first. We denote by $R_{ex}^{diff}(\iota; \alpha, \Delta_{dm}, \kappa)$, the difference between an expert's expected gross excess return when data is available and it is not, given its decision $\iota \in \{0, 1\}$ to buy the data or not when data is available. When $\alpha = 0$ and $\iota = 0$, we deduce from the previous observations and (IA.15) that:

$$\begin{aligned} R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa) &\equiv R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa) - R_{dm}^{diff}(0; 0, 0, \kappa) \\ &= \frac{\tau_{ex}(0)}{W_0 \rho \sigma_v^2} \left(\frac{1}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))} - \frac{1}{\mathcal{I}(\bar{\tau}(0, 0, \kappa))} \right). \end{aligned} \tag{IA.26}$$

As price informativeness ($\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))$) increases with Δ_{dm}, $R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa) < 0$. Thus, experts who do not buy data are worse off when new data becomes available.

Now consider data miners. We denote the difference between a data miner's expected gross excess return when data is available and it is not by $R_{dm}^{diff}(\iota; \alpha, \Delta_{dm}, \kappa)$. When $\alpha = 0$, we deduce from the previous observations and (IA.15) that:

$$\begin{aligned} R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) &\equiv R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) - R_{ex}^{diff}(1; 0, 0, \kappa) \\ &= \frac{1}{W_0 \rho \sigma_v^2} \left(\frac{\tau_{dm} + \Delta_{dm}}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))} - \frac{\tau_{dm}}{\mathcal{I}(\bar{\tau}(0, 0, \kappa))} \right). \end{aligned} \tag{IA.27}$$

The effect of new data on data miners' performance is less clear-cut than for experts. On the one hand, it increases the precision of their signals. Other things equal, this increase their expected gross excess return relative to the case in which data is not available. On the other hand, it increases price informativeness, which reduces their expected gross excess return, as for experts, so the net effect is ambiguous.[3] However, there is always a level of τ_{dm} small

[3]It may seem puzzling that data miners are not unambiguously better off since they all buy the data when $C \leq C_1(\alpha, \Delta_{dm}, \kappa)$. However, the decision to buy data does not hinge on a comparison of their performance in the market with and without data, but on a comparison of their performance when they buy the data and when they don't, *conditional on all other data miners buying the data.* It could be that all data miners would be collectively better off committing not to buy new data and yet, if this commitment cannot be enforced, the only equilibrium is that they all buy the data and end up strictly worse off collectively. See Biais et al.

enough such that $\Delta R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) > 0$.[4]

Empirically we cannot clearly identify who are the data miners and who are the experts. Thus, our main tests focus on the change in the expected gross excess of the "average" fund. That is, our main test is about the effect of new data on the cross-sectional average gross excess returns of all funds in our sample. In our framework, for a given α, this average is:

$$\bar{R}^A(\alpha, \Delta_m, \kappa) := (1-\mu)(1-\alpha)\bar{R}(\tau_{ex}(0)) + (1-\mu)\alpha\bar{R}(\tau_{ex}(1)) + \mu\bar{R}(\tau_{dm}(1)). \qquad \text{(IA.28)}$$

Hence, for $\alpha = 0$, the effect of the availability of new data on average funds' performance is given by:

$$\bar{R}^{A,diff}(\alpha, \Delta_m, \kappa) = \mu R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) + (1-\mu) R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa), \qquad \text{(IA.29)}$$

where $\bar{R}^{A,diff}(\alpha, \Delta_m, \kappa)$ is the difference in average funds' performance when new data is available and when it is not.

Using (IA.26) and (IA.27), we obtain:

$$\bar{R}^{A,diff}(0, \Delta_m, \kappa) = \frac{1}{W_0 \rho \sigma_v^2}\left(\frac{\mu \Delta_{dm}}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))} + \bar{\tau}(0,0,\kappa)\left(\frac{1}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))} - \frac{1}{\mathcal{I}(\bar{\tau}(0,0,\kappa))}\right)\right). \qquad \text{(IA.30)}$$

After some algebra, we deduce that $\bar{R}^{A,diff}(0, \Delta_m, \kappa) < 0$ if and only if:

$$\mu < \bar{\tau}(0,0,\kappa)\left(\frac{2\bar{\tau}(0,0,\kappa) + \mu\Delta_{dm}}{\rho^2\sigma_v^2\sigma_\eta^2 + \bar{\tau}(0,0,\kappa)}\right). \qquad \text{(IA.31)}$$

Calculations show that a sufficient condition for the R.H.S. to be larger than 1 is that

(2015) for a similar possibility.

[4] This follows from the fact the negative term in (IA.27) goes to zero when τ_{dm} goes to zero while the positive term remains strictly positive since $\Delta_{dm} > 0$.

$\mu < \frac{\min\{\tau_{dm},\tau_{ex}\}}{\rho^2\sigma_v^2\sigma_\eta^2+\tau_{dm}}$ or, as $\tau_{ex} \geq \tau_{dm}$, $\mu < \bar{\mu}(\tau_{dm})$ where:[5]

$$\bar{\mu}(\tau_{dm}) = \frac{\tau_{dm}}{\rho^2\sigma_v^2\sigma_\eta^2 + \tau_{dm}}. \tag{IA.32}$$

This condition is sufficient for the decline in experts' expected gross excess return to be large enough to more than offset the possible increase in data miners' expected gross excess return. This delivers our first testable hypothesis (H.1) given in the main text.

Our second prediction is that the average decline in experts' gross expected return is larger than that for the average fund manager. Intuitively, the reason is that they are more negatively affected than the data miners since the latter improve the precision of their signals by buying new data. This is obvious when $R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) > 0$ but less so when $R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) < 0$, which, as explained above, can happen for some parameter values. Thus, to formally derive the implication we compute $\bar{R}^{A,diff}(0, \Delta_m, \kappa) - R_{ex}^{diff}(1; 0, \Delta_{dm}, \kappa)$. Using (IA.26) and (IA.30), we obtain

$$\bar{R}^{A,diff}(0, \Delta_m, \kappa) - R_{ex}(1; 0, \Delta_{dm}, \kappa) = \mu(R_{dm}^{diff}(1; 0, \Delta_{dm}, \kappa) - R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa)), \tag{IA.33}$$

which is (using (IA.26) and (IA.27)):

$$\begin{aligned} \bar{R}^{A,diff}(0, \Delta_m, \kappa) - R_{ex}(1; 0, \Delta_{dm}, \kappa) = \mu\Bigg(& \frac{\Delta_{dm}}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))} \\ & + (\tau_{ex} - \tau_{dm})\left(\frac{1}{\mathcal{I}(\bar{\tau}(0,0,\kappa))} - \frac{1}{\mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))}\right)\Bigg) \\ > 0. & \end{aligned} \tag{IA.34}$$

where the last inequality comes from the fact that $\mathcal{I}(\bar{\tau}(0,0,\kappa)) < \mathcal{I}(\bar{\tau}(0, \Delta_{dm}, \kappa))$ because price informativeness increases with Δ_{dm} and the assumption $\tau_{ex} \geq \tau_{dm}$. This yields our

[5] Indeed, Condition (IA.31) is satisfied if $\mu < \frac{2\bar{\tau}(0,0,\kappa)^2}{\rho^2\sigma_v^2\sigma_\eta^2+\bar{\tau}(0,0,\kappa)}$ since $\Delta_{dm} > 0$. The R.H.S. of this inequality increases with $\bar{\tau}(0,0,\kappa)$. Thus, it is always satisfied if it is satisfied for the smallest value of $\bar{\tau}(0,0,\kappa)$, which is $\min\{\tau_{dm}, \tau_{ex}\} = \tau_{dm}$.

second testable hypothesis (H.2. in the main text).

Last, (IA.26) shows that the drop in performance is larger when experts' signals have a higher precision since $\mid R_{ex}^{diff}(0; 0, \Delta_{dm}, \kappa) \mid$ is higher when τ_{ex} is higher, other things equal. Thus, if we had experts with signals of different precisions τ_{ex} (extending the model in this case is straightforward), the drop in performance when new data becomes available would be larger for experts with signals of higher precision.[6] This delivers our third testable hypothesis (H.3 in the main text).

To facilitate the exposition, so far, we have derived these implications when $C_0(0, \Delta_{dm}, \kappa) \leq C \leq C_1(0, \Delta_{dm}, \kappa)$. Hence, when data are available, all data miners buy the data and no experts buy the data. However, these implications also hold even when a fraction of experts buy the data. Indeed, experts who do not buy the data experience a decline in their expected gross excess returns for exactly the same reason as that when no expert buys the data. Moreover this drop can be large enough to generate a drop in the average expected gross excess returns across all fund managers, again via the same mechanism as that highlighted when $\alpha^* = 0$. In fact this effect tends to be stronger because price informativeness increases by a larger amount when some experts buy the data. Hence, experts who do not buy the data experience a larger drop in their expected gross excess return while fund managers who buy the data experience a smaller increase in their expected gross return.[7]

Figure IA.2 provides a numerical example. It shows the change in experts' expected gross excess returns ($R_{ex}^{diff}(0; \alpha, \Delta_{dm}, \kappa)$; blue bar) and data miners' expected excess returns ($R_{dm}^{diff}(1; \alpha, \Delta_{dm}, \kappa)$; red bar) when new data becomes available as well as the change in the average expected excess returns (orange bar) across all fund managers for two different values of κ and two different values of μ. In each case, one of the value of κ or μ is such that some

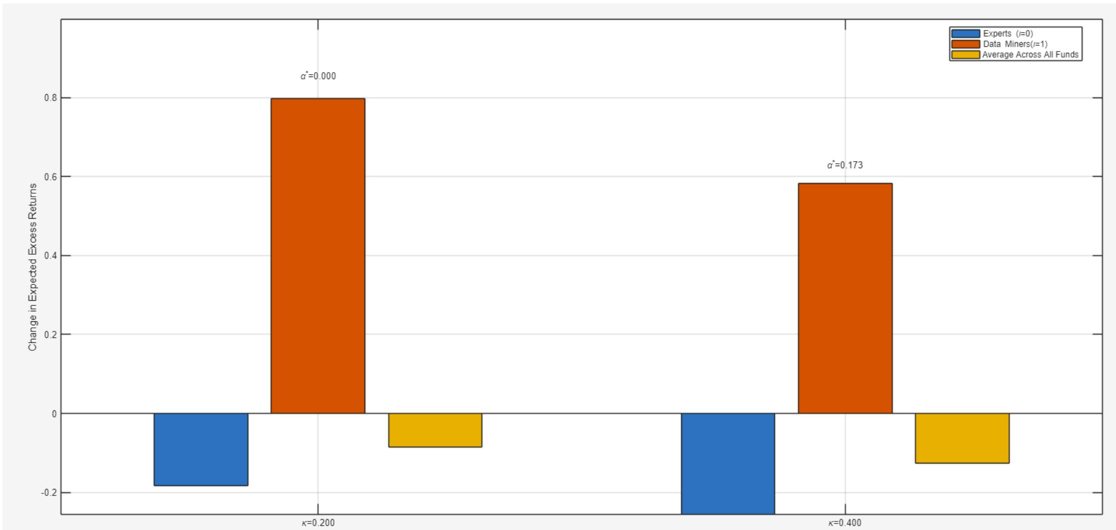
[6] For instance, suppose we have 2 types of experts with signals of precision τ_{ex}^H (type H) and τ_{ex}^L (type L) such that $\tau_{ex}^L < \tau_{ex}^H$. Moreover the mass of experts with type H in the population of experts is λ_H. Defining, $\bar{\tau}(0, \Delta_{dm}, \kappa)) = \mu(\tau_{dm} + \Delta_{dm}) + (1 - \mu)(\lambda_H^{ex}\tau_{ex}^H + \lambda_L \tau_{ex}^L)$, the rest of the analysis is identical and one gets from (IA.26) that experts with type H experience a larger drop in their expected gross excess return than experts with type L.

[7] This clearly holds when we keep all parameter constant, except C because this parameter has no direct effect on price informativeness or fund managers' expected gross returns. This is less clear cut otherwise as a change in parameters affect both α^* and directly price informativeness and fund managers' performance.

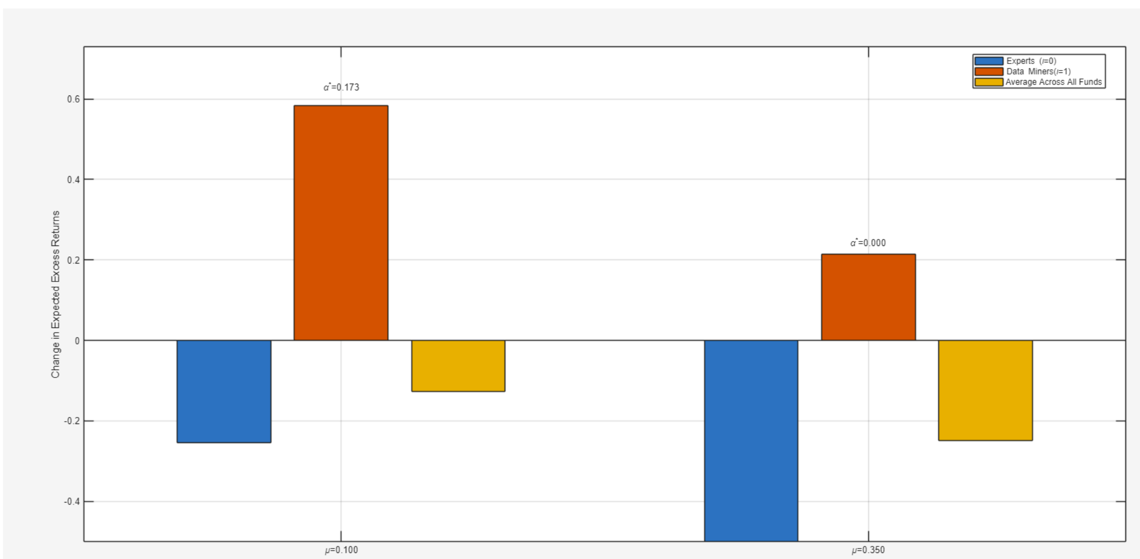
experts buy data ($\alpha^* > 0$) in equilibrium.

For instance, consider the effect of κ (Panel A of Figure IA.2). For $\kappa = 0.2$, the equilibrium is such that no expert buys the data while all data miners do. Consistent with our testable implications, experts' expected gross returns drop when data becomes available and the average expected gross returns across all funds drop as well, even though data miners' expected gross returns increase. For $\kappa = 0.4$, the same patterns are observed, even though in this case some experts buy data ($\alpha^* = 17.3\%$). In fact the effects are more pronounced because experts who do not buy the data experience an even larger drop in their expected gross return (because the price becomes even more informative).

Last consider the case in which $C > C_1(0, \Delta_{dm}, \kappa)$. In this case, in equilibrium no experts buy the data and only a fraction of data miners buy the data. Again, predictions are unchanged. The only difference is that some data miners experience a drop in their expected gross returns, as experts do, since they do not buy the data.



Panel A: The figure shows, for two different values of κ (resp., 20% and 40%), the differences in expected gross returns for fund managers with and without new data, for data miners (red), experts (blue), and the average across both types (orange). The equilibrium fraction of experts buying the data (α^*) is shown at the top of each case.



Panel B: The figure shows, for two different values of μ (resp., 10% and 35%), the differences in expected gross returns for fund managers with and without new data, for data miners (red), experts (blue), and the average across both types (orange). The equilibrium fraction of experts buying the data (α^*) is shown at the top of each case.

Figure IA.2: Parameter values: $C = 0.2$, $\mu = 0.1$ (Panel A), $\rho = 0.5$, $\kappa = 0.4$ (Panel B), $\Delta = 5$, $\sigma_v = 1.5$, $\sigma_\eta = 1$, $\tau_{dm} = 1$, $\tau_{ex} = 2$, $W = 1$.

I.2. Dynamics of Coverage and Covered stocks by Industry

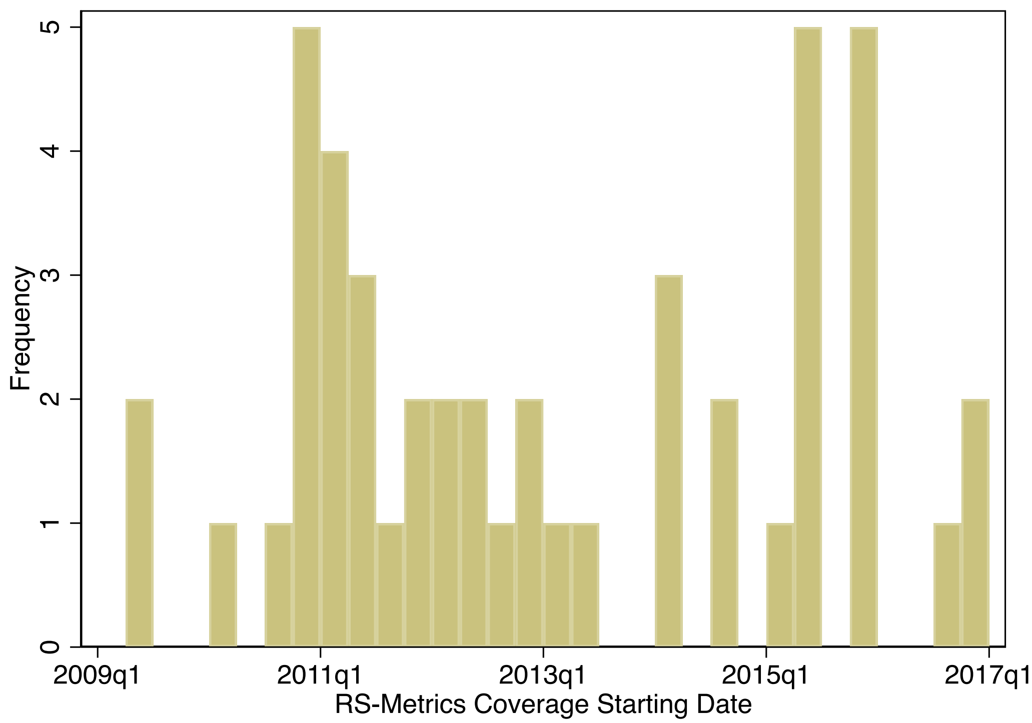
Table IA.1: Covered Stocks in each Industry

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

Figure IA.3: 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.



I.3. Number of Fund-Stock Pairs contributing to Identification

In this section, we provide detailed summary statistics on the number of observations contributing to identification in our estimation sample. These statistics are summarized in Internet Appendix Table IA.2.

- Estimation of Equation (6): The identification sample includes 2,335 distinct funds (out of 3,962 in the full sample) that hold at least one treated stock both before and after coverage. On average, 209 funds hold a treated stock in the quarter it becomes covered. We identify 9,451 fund-*treated* stock pairs that span both the pre- and post-coverage periods of the 48 treated stocks. These pairs contribute a total of 153,090 fund-stock-quarter observations, with 53,151 pre-coverage and 99,939 post-coverage. On average, each fund-treated stock pair spans 6 quarters pre-coverage and 11 quarters post-coverage. In addition, the estimation also uses the set of fund-*control* stock pairs—stocks not covered by RS Metrics but held by the same funds as the treated stocks in both the pre- and post-coverage periods. This set comprises 318,942 fund-stock pairs, corresponding to 6,010,737 fund-stock-quarter observations. Taken together, the total number of fund-stock pairs contributing to the estimation of the coefficient β in Equation (6) is 328,393, corresponding to over 6.1 million fund-stock-quarter observations (out of 12.8 million observations in our overall sample).

- Estimation of Equation (8): The sample used to estimate the coefficient β is the same as in Equation (6). However, the estimation of the triple-interaction coefficient β_{expert} relies only on the subset of fund-stock pairs in which the fund is classified as an Industry Specialist or a Sector Fund. These subsamples are by definition smaller. For Industry Specialists (funds with more than 75% of assets in covered industries), the estimation sample for β_{expert} includes 5,642 fund-stock pairs, corresponding to 133,304 fund-stock-quarter observations. On average, in this subsample, each fund-treated stock pair spans 13 quarters pre-coverage and 20 quarters post-coverage. For Sector Funds, the estimation sample for β_{expert} includes 7,235 fund-stock pairs, corresponding to 165,570 fund-stock-quarter observations. On average, in this subsample, each fund-treated stock pair spans 11 quarters pre-coverage and 17 quarters post-coverage.

- Estimation of Equation (9): The sample used to estimate specification (9) includes all the funds for which we are able to obtain the official address from the CRSP database.[8] The total number of fund-stock pairs contributing to the estimation of the coefficient β is 325,907, corresponding to 6.1 million fund-stock-quarter observations. The estimation of the triple-interaction coefficient β_{local} relies only on the subset of fund-stock pairs in which the fund is classified as "Local" for one of the treated stocks. This subsample includes 120,088 fund-stock pairs, corresponding to 2.2 million fund-stock-quarter observations. On average, in this subsample, each fund-treated stock pair spans 5 quarters pre-coverage and 10 quarters post-coverage.

- Estimation of Equation (10): The sample used to estimate this specification does not include the picking skills for stocks covered by RS Metrics for funds that start holding the stock post-coverage (because for the fund-covered stock pairs we need to estimate a level of stock-picking ability before coverage initiation). The total number of fund-stock pairs contributing to the estimation of the coefficient β is 328,383, corresponding to 6.1 million fund-stock-quarter observations. The estimation of the triple-interaction coefficient β_{high} relies only on the subset of fund-stock pairs in which the fund is classified as High Picking Pre-coverage for one of the treated stocks. This subsample includes 302,003 distinct fund-stock pairs, corresponding to 5.7 million fund-stock-quarter observations. On average, in this subsample, each fund-treated stock pair spans 8 quarters pre-coverage and 16 quarters post-coverage.

[8] We note that, in this updated version of the manuscript, the number of observations in the table using fund address locations is higher than in the initial submission. This is because we re-downloaded the fund address data from CRSP, which now contains more complete address information. The improved coverage increases the number of matched fund observations, but does not affect any of our results or their interpretation. We include this note for full transparency.

Table IA.2: Identification Sample and Estimated Coefficients

This table reports, for each main specification, summary statistics on the identifying variation for the corresponding coefficient. We report the number of fund–stock pairs spanning both pre- and post-coverage periods, the total number of fund–stock–quarter observations, and the average number of quarters before and after coverage initiation per treated pair.

Specification	Coefficient	Fund–Stock Pairs	Observations	Avg. Pre-Qtrs	Avg. Post-Qtrs
Main Spec. (Eq. (6))	β	328,393	6,163,827	6	11
Industry Specialists (Eq. (8))	β_{expert}	5,642	133,304	13	20
Sector Funds (Eq. (8))	β_{expert}	7,235	165,570	11	17
Local Funds (Eq. (9))	β_{local}	120,088	2,214,227	5	10
High Picking Pre (Eq. (10))	β_{high}	302,003	5,717,413	8	16

I.4. 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. (2025), 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.3, we present estimates of the following regression:

$$SalesGrowth_{i,t} = \beta FillingRate_{i,t} + \gamma SalesGrowth_{i,t-1} + \theta_i + \delta_{q\times t} + \epsilon_{i,t}, \qquad \text{(IA.35)}$$

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

Table IA.3 confirms that parking lot filling rates data reported by RS Metrics are a relevant predictor of retailer sales performance.

Table IA.3: 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.5. Alternative Measures of Stock Picking Ability

In this section, we show that our main 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), (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) and (iii) one in which we use the Fama-French 3-factor model to compute idiosyncratic returns. In the later case, we define stock picking ability of fund f in stock i at horizon h in quarter t as:

$$Picking_{f,i,t}^{h} = 100 \times (w_{i,t}^{f} - w_{i,t}^{m})(R_{t+h}^{ie} - \beta_{i,t}^{m} R_{t+h}^{me} - \beta_{i,t}^{SMB} R_{t+h}^{SMB} - \beta_{i,t}^{HML} R_{t+h}^{HML}), \tag{IA.36}$$

where $w_{i,t}^{f}$ is the fraction of fund f's assets invested in stock i at the end of quarter t, and $w_{i,t}^{m}$ is the corresponding market portfolio weight. R_{t+h}^{ie} and R_{t+h}^{me} are the stock's and the market's excess return over the risk-free rate over the next h quarters, respectively. R_{t+h}^{SMB} and R_{t+h}^{HML} denote the returns on the size (SMB) and value (HML) factors, respectively, obtained from Ken French's data library. $\beta_{i,t}^{m}$, $\beta_{i,t}^{SMB}$, and $\beta_{i,t}^{HML}$ are the corresponding betas estimated using daily stock returns over the preceding 252 trading days.

The tests corresponding to approaches (i) and (ii) are reported in Tables IA.4 to IA.8. The tests using the approach (iii) (i.e., Fama-French 3-factor model to compute idiosyncratic returns) are reported in tables IA.9 to IA.13.

Table IA.4: Alternative Benchmark Portfolios

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)
Covered × Post	-0.021***	-0.021**	-0.041***	-0.038**	-0.067***	-0.059*	-0.105***	-0.097**
	(0.007)	(0.008)	(0.015)	(0.019)	(0.025)	(0.031)	(0.036)	(0.046)
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***	-0.014***	-0.028***	-0.022**	-0.048***	-0.035**	-0.077***	-0.061***
	(0.005)	(0.005)	(0.010)	(0.009)	(0.015)	(0.014)	(0.021)	(0.020)
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 Benchmark Portfolios 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 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 Abilities

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.002	0.012	0.025*	0.019
	(0.005)	(0.009)	(0.015)	(0.018)
Covered × Post × High Picking Pre	-0.054***	-0.114***	-0.185***	-0.255***
	(0.013)	(0.031)	(0.054)	(0.081)
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	0.016**	0.022*	0.012
	(0.004)	(0.008)	(0.012)	(0.014)
Covered × Post × High Trading Pre	-0.043***	-0.090***	-0.130***	-0.165***
	(0.006)	(0.012)	(0.018)	(0.023)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.27e+07	1.23e+07	1.19e+07	1.15e+07
R^2	0.20	0.28	0.33	0.36

Table IA.6: Alternative Benchmark Portfolios 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 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 Skills

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.015**	-0.021*	-0.031*	-0.056**
	(0.006)	(0.011)	(0.018)	(0.026)
Covered × Post × Industry Specialist	-0.163***	-0.421***	-0.714***	-1.090***
	(0.055)	(0.143)	(0.236)	(0.360)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+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	-0.015***	-0.023**	-0.034**	-0.058***
	(0.005)	(0.009)	(0.014)	(0.020)
Covered × Post × Industry Specialist	0.029***	0.021	-0.015	-0.073
	(0.006)	(0.014)	(0.028)	(0.047)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.20	0.28	0.33	0.36

Table IA.7: Alternative Benchmark Portfolios 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 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 Skills

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.015**	-0.021*	-0.031*	-0.056**
	(0.006)	(0.011)	(0.018)	(0.026)
Covered × Post × Sector Fund	-0.133**	-0.344**	-0.585**	-0.891**
	(0.054)	(0.142)	(0.237)	(0.362)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+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	-0.015***	-0.023**	-0.034**	-0.058***
	(0.005)	(0.009)	(0.014)	(0.019)
Covered × Post × Sector Fund	0.025***	0.017	-0.019	-0.068*
	(0.006)	(0.012)	(0.024)	(0.041)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.20	0.28	0.33	0.36

Table IA.8: Alternative Benchmark Portfolios 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 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)
Covered × Post	-0.017**	-0.030*	-0.048*	-0.080**
	(0.007)	(0.016)	(0.026)	(0.037)
Covered × Post × Local	-0.057***	-0.117**	-0.181**	-0.270**
	(0.019)	(0.054)	(0.086)	(0.124)
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

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.012**	-0.019*	-0.032**	-0.055***
	(0.005)	(0.010)	(0.014)	(0.020)
Covered × Post × Local	-0.018**	-0.032**	-0.043*	-0.063*
	(0.008)	(0.015)	(0.024)	(0.034)
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.9: 3-Factor Adjusted 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 an alternative measure of fund picking skills as dependent variable: the product of (i) the fund excess weight in the stock relative to the market portfolio and (ii) the stock's abnormal return over the following 1, 2, 3 or 4 quarters according to the Fama-French 3-factor model (see Eq. (IA.36)).

	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*	-0.028*	-0.026	-0.049**	-0.042	-0.081**	-0.076*
	(0.007)	(0.008)	(0.015)	(0.017)	(0.025)	(0.030)	(0.036)	(0.043)
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	0.08	0.17	0.09	0.25	0.10	0.30	0.11	0.35

Table IA.10: 3-Factor Adjusted 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 V, but uses an alternative measure of fund picking skills as dependent variable: the product of (i) the fund excess weight in the stock relative to the market portfolio and (ii) the stock's abnormal return over the following 1, 2, 3 or 4 quarters according to the Fama-French 3-factor model (see Eq. (IA.36)).

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.006	0.018*	0.026	0.019
	(0.005)	(0.010)	(0.016)	(0.020)
Covered × Post × High Picking Pre	-0.049***	-0.101***	-0.158***	-0.220***
	(0.012)	(0.027)	(0.047)	(0.071)
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	0.17	0.25	0.30	0.35

Table IA.11: 3-Factor Adjusted 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 III, but uses an alternative measure of fund picking skills as dependent variable: the product of (i) the fund excess weight in the stock relative to the market portfolio and (ii) the stock's abnormal return over the following 1, 2, 3 or 4 quarters according to the Fama-French 3-factor model (see Eq. (IA.36)).

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.010	-0.013	-0.019	-0.040
	(0.006)	(0.012)	(0.020)	(0.027)
Covered × Post × Industry Specialist	-0.139***	-0.339***	-0.604***	-0.927***
	(0.047)	(0.116)	(0.201)	(0.309)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.17	0.25	0.30	0.35

Table IA.12: 3-Factor Adjusted 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 III, but uses an alternative measure of fund picking skills as dependent variable: the product of (i) the fund excess weight in the stock relative to the market portfolio and (ii) the stock's abnormal return over the following 1, 2, 3 or 4 quarters according to the Fama-French 3-factor model (see Eq. (IA.36)).

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.010	-0.013	-0.019	-0.041
	(0.006)	(0.012)	(0.020)	(0.027)
Covered × Post × Sector Fund	-0.113**	-0.275**	-0.493**	-0.753**
	(0.047)	(0.117)	(0.204)	(0.314)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.17	0.25	0.30	0.35

Table IA.13: 3-Factor Adjusted 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 IV, but uses an alternative measure of fund picking skills as dependent variable: the product of (i) the fund excess weight in the stock relative to the market portfolio and (ii) the stock's abnormal return over the following 1, 2, 3 or 4 quarters according to the Fama-French 3-factor model (see Eq. (IA.36)).

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.011	-0.019	-0.032	-0.059
	(0.007)	(0.015)	(0.025)	(0.036)
Covered × Post × Local	-0.050***	-0.098**	-0.157**	-0.241**
	(0.016)	(0.042)	(0.072)	(0.104)
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	0.17	0.25	0.30	0.35

I.6. Scaled Picking

In this section, we investigate whether the larger estimated effects for industry specialists are driven by mechanical differences in portfolio construction. As industry specialists concentrate their investments in a smaller set of stocks, their portfolio weights—and therefore the scale of the *Picking* measure—are mechanically larger. To address this concern, we re-estimate our baseline specifications after scaling *Picking* by the average stock weight in the fund's portfolio in each quarter. This normalization controls for differences in portfolio concentration and yields a measure that is closer to the return generated per unit of capital allocated. The next table presents results when we re-estimate eq.(8) after scaling $Picking^{h}_{f,i,t}$ by the average stock weight in fund f's portfolio in each quarter.

Table IA.14: Heterogeneous Effect based on Industry Expertise with Picking Scaled By Portfolio Weight

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 scaled by the fund's average stock weight in the corresponding quarter to account for differences in portfolio concentration across funds. *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. 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. We note that the terms "Industry Specialist" and "Sector Fund" are constant at the fund level. 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)
Covered × Post	-2.036**	-3.231**	-4.889*	-8.075**
	(0.877)	(1.621)	(2.631)	(3.642)
Covered × Post × Industry Specialist	-7.131**	-19.361***	-32.981***	-49.908***
	(2.958)	(7.431)	(12.449)	(18.883)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.13	0.20	0.26	0.30

Panel B: Sector Funds

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-2.043**	-3.252**	-4.904*	-8.088**
	(0.881)	(1.625)	(2.639)	(3.655)
Covered × Post × Sector Fund	-5.709**	-15.448**	-26.691**	-40.581**
	(2.879)	(7.513)	(12.528)	(18.968)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.13	0.20	0.26	0.30

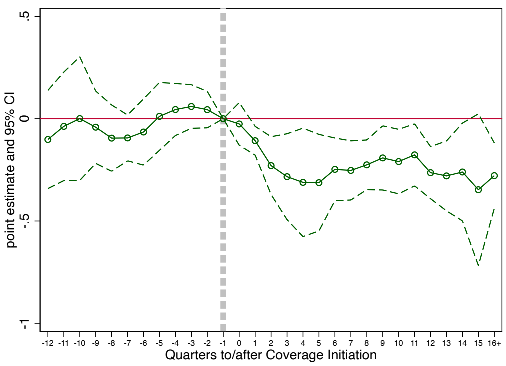
I.7. Dynamics of the Impact of Coverage Initiation

Figure IA.4 reports estimates of specifications similar to (11) in the text, but replacing $HighPickingPref_{f,i}$ by $IndustryExpert_f$ (Panels B and C in Figure IA.4) or $Local_{f,i}$ (Panel D). In each case, we just report the dynamics on the coefficient of interest, namely the coefficient on the triple interaction terms. In Panel A, we just reproduce Panel B of Figure II in the text. In Panels E to H of Figure IA.4, we report estimates of the β_ks in a specification similar to (7), replacing $Picking$ by stocks' excess weights $(w_{i,t}^f - w_{i,t}^m)$, return, β, and stock abnormal return as in Table A.3 in the paper.

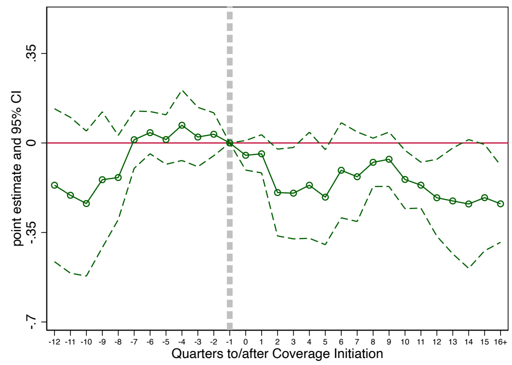
Figure IA.4: Dynamic Effects of Alternative Data across Specifications

This figure reports the dynamic effects of the release of alternative data on fund picking abilities (Picking 4-Q) across multiple specifications in Panels A, B, C and D, as well as on funds' excess weight, stock's quarterly return, beta and abnormal quarterly returns in Panels E, F, G and H, respectively. Each panel plots the coefficients on the interaction between "Covered" and event-time indicators. Circles represent point estimates and dashed lines denote 95% confidence intervals. Standard errors are double-clustered at the fund and stock levels in Panels A, B, C, D and E, and clustered at the stock level in Panels F, G and H.

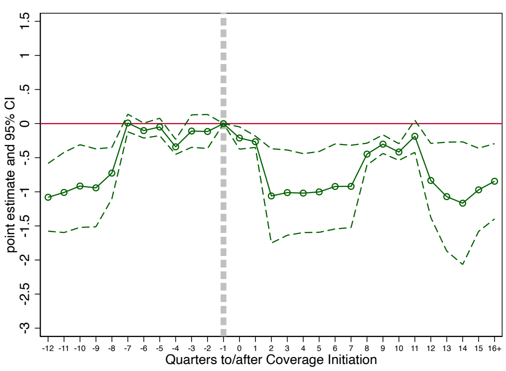
Panel A: Superior Stock Pickers Pre-Coverage



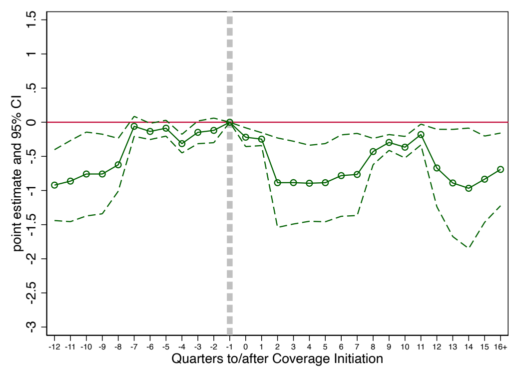
Panel B: Local Funds



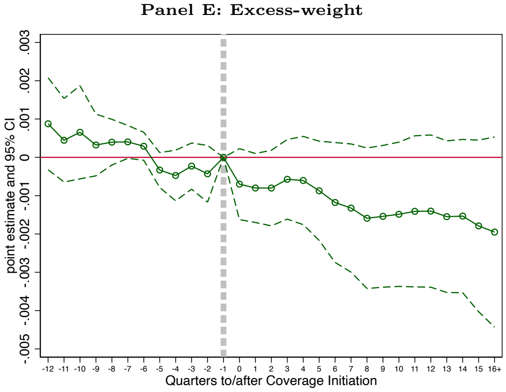
Panel C: Industry Specialists



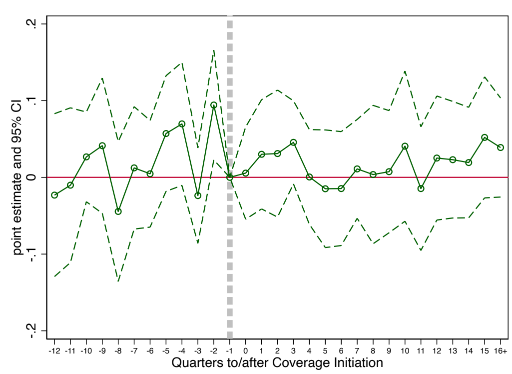
Panel D: Sector Funds



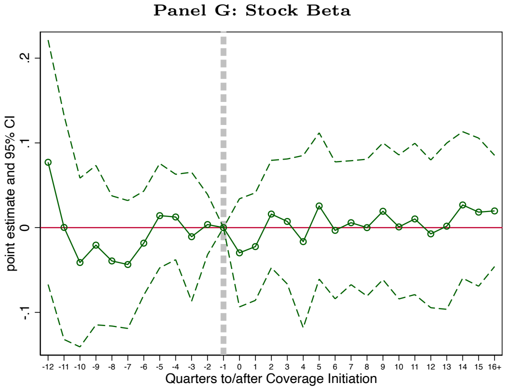
Panel E: Excess-weight



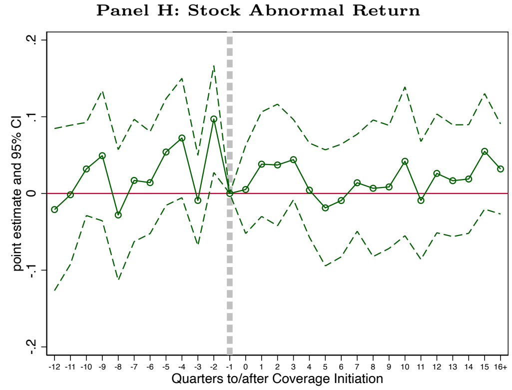
Panel F: Stock Return



Panel G: Stock Beta



Panel H: Stock Abnormal Return



I.8. 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:

$$Timing^{h}_{f,i,t} = 100 \times (w^{f}_{i,t} - w^{m}_{i,t})(\beta_{i,t} R^{me}_{t+h}), \qquad \text{(IA.37)}$$

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^{me}_{t+h} is the excess 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 (6) in the main text) using "*Timing*" as dependent variable. Table IA.15 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.15: 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.37). *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.

	Timing 1-Q		Timing 2-Q		Timing 3-Q		Timing 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.004	-0.006	-0.009	-0.012	-0.013	-0.017	-0.013	-0.019
	(0.005)	(0.006)	(0.009)	(0.010)	(0.012)	(0.014)	(0.014)	(0.017)
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	1.29e+07	1.29e+07	1.26e+07	1.26e+07	1.23e+07	1.23e+07	1.20e+07	1.20e+07
R^2	0.61	0.67	0.61	0.70	0.60	0.72	0.59	0.74

I.9. Value Added

Table IA.16: Alternative Data and Value Added

This table presents our results on the effect of the release of alternative data on fund value added (in $ millions). Regressions are estimated at the fund-stock-quarter level. The dependent variable is the product of the fund's AUM and *Picking* calculated at different horizons ranging from one quarter to one year (defined in equation (4)). $Covered \times Post$ 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.

	VA 1-Q		VA 2-Q		VA 3-Q		VA 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.182**	-0.254**	-0.334*	-0.479**	-0.529*	-0.730*	-0.858**	-1.197**
	(0.086)	(0.109)	(0.191)	(0.243)	(0.297)	(0.384)	(0.429)	(0.564)
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	0.03	0.10	0.03	0.15	0.04	0.20	0.04	0.24

I.10. Leave-one-stock or one-fund Out Analysis

Influential Stocks. As discussed in Section IV.B, we conduct a leave-one-stock-out analysis to assess whether our results are driven by a small number of highly influential (i.e., frequently held) stocks. To this end, we identify the 15 treated stocks (out of 48) that contribute most to identification—i.e., those with the largest number of fund–stock–quarter observations—which together account for more than 50% of all treated observations. We then re-estimate Equation (6), as well as Equations (8), (9), and (10), 15 times, each time excluding one of these stocks.

Panels A–D of Figure IA.5 report the distribution of t-statistics across these regressions. In all cases, the t-statistics remain negative and statistically significant at conventional levels. These results indicate that our findings are not driven by a small subset of heavily held treated stocks. Indeed, even when excluding each of the most influential stocks in turn, we continue to find that alternative data coverage reduces mutual funds' stock-picking performance, with larger effects among funds relying on traditional sources of expertise (e.g., industry specialization or geographic proximity) and among funds with high pre-coverage skill.

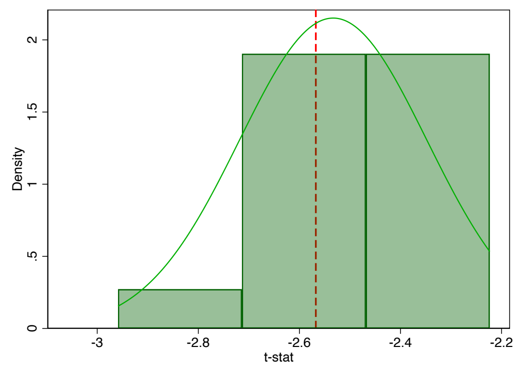
Influential Funds. Similarly, as discussed in Section IV.A, we conduct a leave-one-fund-out analysis to assess whether our results are driven by a small number of funds with large exposures to treated stocks. We identify the 15 funds with the highest average dollar investment in treated stocks over the sample period, with average holdings ranging from \$15 million to \$60 million. We then re-estimate Equation (6), as well as Equations (8), (9), and (10), 15 times, each time excluding one of these funds.

Panels A–D of Figure IA.6 report the distribution of t-statistics across these regressions. As in the leave-one-stock-out analysis, the t-statistics remain negative and statistically significant at conventional levels. These results indicate that our findings are not driven by a small subset of heavily invested funds. Even when excluding each of the largest investors in treated stocks in turn, we continue to find that alternative data coverage reduces mutual funds' stock-picking performance, with larger effects among funds relying on traditional forms of expertise.

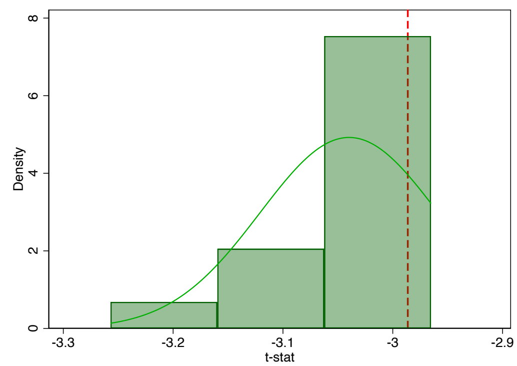
Figure IA.5: Leave-one-stock-out Analysis

This figure reports the results of a leave-one-stock-out analysis. We identify the 15 treated stocks (out of 48) that contribute the most fund–stock–quarter observations and re-estimate each main specification 15 times, each time excluding one of these top-15 treated stocks from the estimation sample. Each panel shows the distribution of the resulting t-statistics for the coefficient of interest across these leave-one-out regressions using Picking 1-Q as the dependent variable. Panel A shows the t-statistics for the baseline effect (β) in regression (6). Panels B and C report the t-statistics for β_{expert} in regression (8), using two definitions of expertise: Panel B defines experts as funds with more than 75% of their portfolio in covered industries (industry specialists), while Panel C uses the CRSP sector-fund classification. Panel D presents the t-statistics for β_{local} in regression (9), which tests for heterogeneous effects by geographical proximity. Panel E reports the t-statistics for β_{high} in regression (10), which tests for heterogeneity by pre-coverage stock-picking skill. The red dashed lines correspond to the t-statistics in the main paper using the full sample.

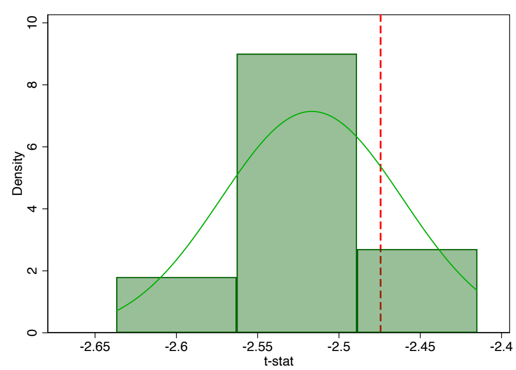
Panel A: Average Effect



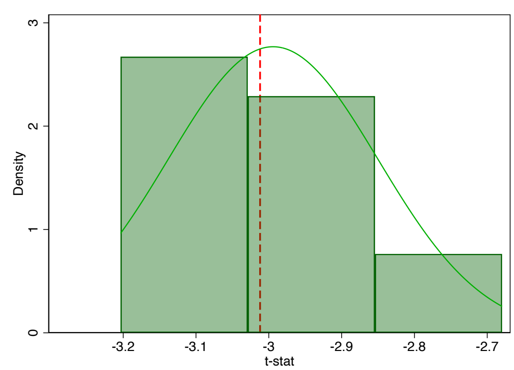
Panel B: Industry Specialists



Panel C: Sector Funds



Panel D: Local Funds



Panel E: Above-median Picking Funds

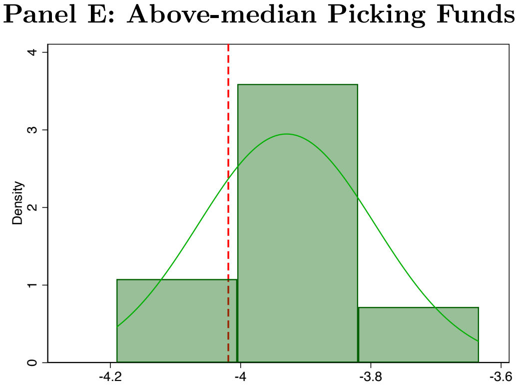
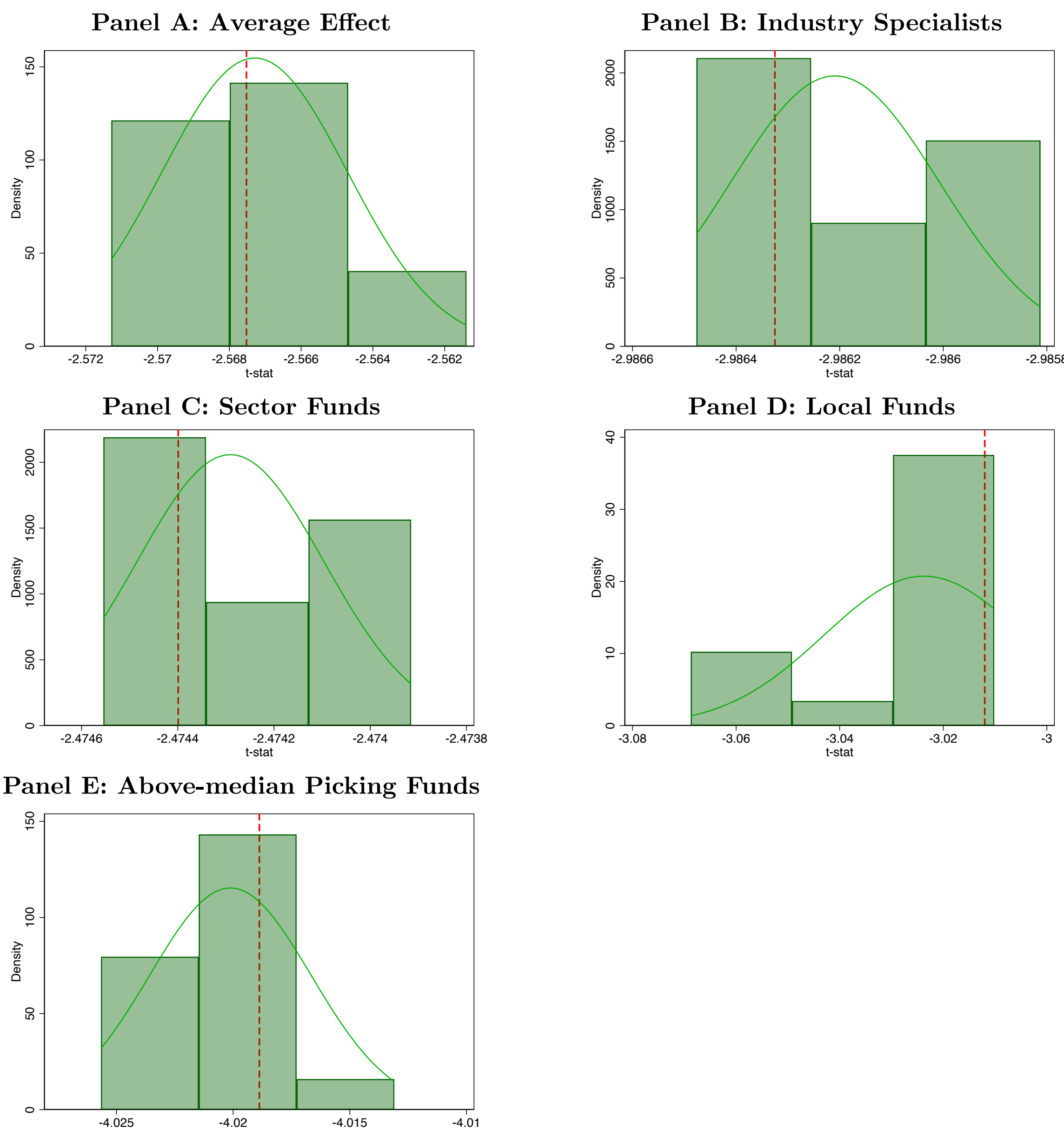


Figure IA.6: Leave-one-fund-out Analysis

This figure presents the results of a leave-one-fund-out analysis. We identify the 15 funds with the highest average dollar investment in treated stocks over the sample period and re-estimate each main specification 15 times, each time excluding one of these funds from the estimation sample. Each panel displays the distribution of t-statistics for the coefficient of interest across these leave-one-out regressions, using *Picking* 1-Q as the dependent variable. Panel A shows the t-statistics for the average treatment effect (β) in regression (6). Panels B and C show the t-statistics for β_{expert} in regression (8), using two definitions of expertise: Panel B defines experts as funds with more than 75% of assets in covered industries (industry specialists), and Panel C uses the CRSP sector-fund classification. Panel D shows the t-statistics for β_{local} in regression (9), testing for heterogeneous effects by geographical proximity. Panel E reports the t-statistics for β_{high} in regression (10), capturing heterogeneity by pre-coverage stock-picking skill. The red dashed lines correspond to the t-statistics in the main paper using the full sample.



I.11. Proxies for Quant Funds

Our findings are consistent with the hypothesis that coverage of a stock by 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 do not buy the data because they lack the expertise required to exploit alternative data. In contrast, those who have this expertise and buy the data should experience an increase or a weaker decline in their stock picking ability when new alternative data becomes available for a stock (see the discussion following eq.(IA.27) and Figure IA.2 in Section I.1). Accordingly, if we zoom in on the funds that are more likely to be data-driven, we should observe a weaker negative, even possibly positive, effect of RS Metrics coverage on their stock picking ability.

We conjecture that quant funds are more likely to buy new data because their strategies are data-driven and they have the data infrastructure required to efficiently use the data. Identifying these funds in our sample is not straightforward because 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. Hence, 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.[9]

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

[9] 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.

words. In addition, we stem each word to its root using the Porter stemmer algorithm (e.g. '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.[10] 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".[11]

We then estimate a specification similar to eq.(8), 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.17. 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

[10] 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."*

[11] 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".

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 exploit and trade on alternative data, and therefore to benefit after coverage initiation.

Table IA.17: 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 (4). $Covered \times Post$ 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. We note that the single terms "Quantit Stock Select", "Quantit Stock Select & Proprietari" and "Quantit Stock Select & Rank" are constant at the fund level and are therefore absorbed by the fixed effects in all specifications. Their interactions with "Covered" are absorbed by the fund-stock fixed effects in all specifications. 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	-0.022**	-0.022**	-0.022**	-0.040**	-0.040**	-0.040**	-0.062*	-0.062*	-0.062*	-0.101**	-0.101**	-0.101**
	(0.009)	(0.009)	(0.009)	(0.020)	(0.020)	(0.020)	(0.032)	(0.032)	(0.032)	(0.048)	(0.048)	(0.048)
Covered × Post × Quantit Stock Select	0.012*			0.031*			0.045			0.079		
	(0.007)			(0.019)			(0.031)			(0.048)		
Covered × Post × Quantit Stock Select & Proprietari		0.015**			0.042***			0.061**			0.105**	
		(0.007)			(0.014)			(0.026)			(0.042)	
Covered × Post × Quantit Stock Select & Rank			0.014**			0.042***			0.062**			0.108**
			(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	8636822	8636822	8636822	8378020	8378020	8378020	8114917	8114917	8114917	7850449	7850449	7850449
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.12. Stock Picking using a Matched Sample of Control Stocks

In this section, we present estimation results based on a matching procedure between treated and control stocks. Specifically, we compute a distance using all firm-level variables included in the selection regressions in Table A.2, namely market capitalization, assets, book-to-market, past stock return, earnings, idiosyncratic volatility, analyst forecast error, analyst coverage, inclusion in the S&P500 Index, number of institutional investors in each category holding the stock, average picking ability, jump ratio, ACAR and bid-ask spread. To construct the matched sample, we compute the Euclidean distance between firms, defined as the square root of the sum of squared differences across standardized variables. This distance metric captures differences across firms along all these dimensions simultaneously. For each treated firm, we identify the five closest control firms in the same industry based on this distance measured one year before coverage initiation.

The results using this matched sample remain fully consistent with our main findings: the decline in mutual fund stock-picking performance for covered firms remains statistically and economically significant, and the magnitude of the effect is similar to that obtained in the full-sample difference-in-differences analysis. The results are reported below in Tables IA.18, IA.19, IA.20 and IA.21.

Table IA.18: 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 smallest Euclidean distance across the full set of firm characteristics listed in Internet Appendix Table A.2. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). $Covered \times Post$ 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)
Covered × Post	-0.020***	-0.023***	-0.038***	-0.041**	-0.056**	-0.057**	-0.083***	-0.090**
	(0.006)	(0.008)	(0.014)	(0.017)	(0.022)	(0.028)	(0.031)	(0.041)
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	644,940	644,940	630,348	630,348	615,641	615,641	600,962	600,962
R^2	0.34	0.41	0.35	0.47	0.37	0.52	0.38	0.56

Table IA.19: 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 smallest Euclidean distance across the full set of firm characteristics listed in Internet Appendix Table A.2. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). $Covered \times Post$ 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 IA.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)
Covered × Post	-0.015**	-0.021*	-0.026	-0.043*
	(0.006)	(0.012)	(0.018)	(0.026)
Covered × Post × Industry Specialist	-0.202***	-0.513***	-0.846***	-1.288***
	(0.068)	(0.160)	(0.261)	(0.401)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	644,940	630,348	615,641	600,962
R^2	0.41	0.47	0.53	0.57

Panel B: Sector Funds

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.015**	-0.020*	-0.024	-0.040
	(0.006)	(0.012)	(0.017)	(0.025)
Covered × Post × Sector Fund	-0.179***	-0.443***	-0.724***	-1.097***
	(0.058)	(0.141)	(0.232)	(0.358)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	644,940	630,348	615,641	600,962
R^2	0.41	0.47	0.53	0.57

Table IA.20: 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 smallest Euclidean distance across the full set of firm characteristics listed in Internet Appendix Table A.2. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). $Covered \times Post$ 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)
Covered × Post	-0.019***	-0.032**	-0.045*	-0.071**
	(0.007)	(0.015)	(0.023)	(0.034)
Covered × Post × Local	-0.049***	-0.113**	-0.183**	-0.271**
	(0.017)	(0.045)	(0.073)	(0.110)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	559,151	546,591	533,927	521,285
R^2	0.42	0.48	0.54	0.57

Table IA.21: 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 smallest Euclidean distance across the full set of firm characteristics listed in Internet Appendix Table A.2. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). *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. 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	-0.004	-0.002	0.009	0.004
	(0.005)	(0.013)	(0.018)	(0.024)
Covered × Post × High Picking Pre	-0.048***	-0.099***	-0.164***	-0.232***
	(0.011)	(0.024)	(0.042)	(0.064)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	504,442	495,061	485,548	475,845
R^2	0.43	0.48	0.53	0.57

I.13. Placebo Analysis

In this section, we implement a placebo analysis in which we treat the closest matched control firms (not the true covered firms) as if they were the ones receiving coverage. Specifically, for each treated stock we take the five closest matched control firm (as defined in our distance matching procedure discussed in Section I.12) and assign to each of them the actual coverage initiation date of the corresponding treated stock. We then re-estimate the regression specifications using these "placebo-treated" firms as the treatment group and the remainder of the sample as controls (excluding the true treated firms).

The estimation results are presented in Tables IA.22, IA.23, IA.24 and IA.25). Across all specifications, we find no significant placebo effects: the coefficients on $Placebo\ Covered \times Post$ interactions are small in magnitude and statistically insignificant. These results provide reassurance that our main findings are not driven by confounding factors that affect firms similar to the treated ones but independent of actual coverage by RS Metrics.

Table IA.22: Stock Picking Skills in a Placebo Sample of Matched Control Firms

This table presents placebo tests using as treated closed matched firms instead of the true ones. Regressions are estimated at the fund-stock-quarter level, but include only the observations corresponding to non-covered stocks. We consider as treated the five closest matched control firms (not the true covered firms) as if they were the ones receiving coverage. Specifically, for each treated stock we take the five closest matched control firm (as defined in our matching procedure presented in Internet Appendix I.12) and assign to each of them the actual coverage initiation date of the corresponding treated stock. We then re-estimate the difference-in-differences specifications using these "placebo-treated" firms as the treatment group and the remainder of the sample (excluding the true treated firms) as controls. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). *Placebo Covered* × *Post* is a dummy equal to one after the placebo coverage initiation date 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)
Placebo Covered × Post	-0.001	0.002	-0.002	0.003	-0.004	0.007	-0.007	0.011
	(0.009)	(0.009)	(0.020)	(0.019)	(0.033)	(0.033)	(0.046)	(0.045)
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	1.25e+07	1.25e+07	1.21e+07	1.21e+07	1.17e+07	1.17e+07	1.13e+07	1.13e+07
R^2	0.10	0.20	0.12	0.28	0.14	0.34	0.15	0.39

Table IA.23: Heterogeneous Effect based on Industry Expertise in a Placebo Sample of Matched Control Firms

This table presents placebo tests using as treated closed matched firms instead of the true ones. Regressions are estimated at the fund-stock-quarter level, but include only the observations corresponding to non-covered stocks. We consider as treated the five closest matched control firms (not the true covered firms) as if they were the ones receiving coverage. Specifically, for each treated stock we take the five closest matched control firms (as defined in our matching procedure presented in Internet Appendix I.12) and assign to each of them the actual coverage initiation date of the corresponding treated stock. We then re-estimate the difference-in-differences specifications using these "placebo-treated" firms as the treatment group and the remainder of the sample (excluding the true treated firms) as controls. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). *Placebo Covered* × *Post* is a dummy equal to one after the placebo coverage initiation date 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 IA.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)
Placebo Covered × Post	-0.001	-0.004	-0.004	-0.006
	(0.009)	(0.017)	(0.031)	(0.041)
Placebo Covered × Post × Industry Specialist	0.056	0.124	0.180	0.280
	(0.042)	(0.105)	(0.177)	(0.263)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.25e+07	1.21e+07	1.17e+07	1.13e+07
R^2	0.20	0.28	0.34	0.39

Panel B: Sector Funds

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Placebo Covered × Post	-0.003	-0.008	-0.009	-0.013
	(0.008)	(0.016)	(0.028)	(0.037)
Placebo Covered × Post × Sector Fund	0.061*	0.134*	0.200	0.294
	(0.031)	(0.079)	(0.132)	(0.197)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.25e+07	1.21e+07	1.17e+07	1.13e+07
R^2	0.20	0.28	0.34	0.39

Table IA.24: Heterogeneous Effect based on Geographical Location in a Placebo Sample of Matched Control Firms

This table presents placebo tests using as treated closed matched firms instead of the true ones. Regressions are estimated at the fund-stock-quarter level, but include only the observations corresponding to non-covered stocks. We consider as treated the five closest matched control firms (not the true covered firms) as if they were the ones receiving coverage. Specifically, for each treated stock we take the five closest matched control firms (as defined in our matching procedure presented in Internet Appendix I.12) and assign to each of them the actual coverage initiation date of the corresponding treated stock. We then re-estimate the difference-in-differences specifications using these "placebo-treated" firms as the treatment group and the remainder of the sample (excluding the true treated firms) as controls. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). $Placebo\ Covered \times Post$ is a dummy equal to one after the placebo coverage initiation date 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 the firm's headquarters. 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)
Placebo Covered × Post	0.001	0.001	0.002	0.005
	(0.009)	(0.018)	(0.031)	(0.043)
Placebo Covered × Post × Local	0.016	0.027	0.053	0.067
	(0.015)	(0.030)	(0.053)	(0.072)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.24e+07	1.20e+07	1.16e+07	1.12e+07
R^2	0.20	0.28	0.34	0.39

Table IA.25: Heterogeneous Effects based on Skills before Coverage Initiation in a Placebo Sample of Matched Control Firms

This table presents placebo tests using as treated closed matched firms instead of the true ones. Regressions are estimated at the fund-stock-quarter level, but include only the observations corresponding to non-covered stocks. We consider as treated the five closest matched control firms (not the true covered firms) as if they were the ones receiving coverage. Specifically, for each treated stock we take the five closest matched control firms (as defined in our matching procedure presented in Internet Appendix I.12) and assign to each of them the actual coverage initiation date of the corresponding treated stock. We then re-estimate the difference-in-differences specifications using these "placebo-treated" firms as the treatment group and the remainder of the sample (excluding the true treated firms) as controls. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (4). $Placebo\ Covered \times Post$ is a dummy equal to one after the placebo coverage initiation date 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 fund has a picking ability above the median for the treated stock before the coverage initiation date. 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)
Placebo Covered × Post	-0.001	-0.001	0.003	0.004
	(0.021)	(0.041)	(0.070)	(0.101)
Placebo Covered × Post × High Picking Pre	0.004	0.004	0.004	0.007
	(0.015)	(0.029)	(0.049)	(0.074)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.25e+07	1.21e+07	1.17e+07	1.13e+07
R^2	0.20	0.28	0.34	0.39

I.14. Funds' Investment Universe

In this section, we re-estimate our main specifications by including all fund–stock–quarter observations in which the stock belongs to the fund's investment universe, defined as all stocks held by the fund in the current or previous 11 quarters (following Koijen and Yogo, 2019). Thus, in this expanded sample, we include both positive and zero weights, effectively considering zero weights as active deviations from the market portfolio. In Tables IA.26 to IA.29, we re-estimate all specifications considered in Tables II to V in the paper.

Table IA.26: Alternative Data and Stock Picking Skills using Funds' Investment Universe Sample

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 and the sample includes all fund-stock-quarter observations where the stock falls within the fund's investment universe. A fund universe is defined following Koijen and Yogo (2019) as the set of stocks held in the current or any of the previous 11 quarters (i.e., potentially includes zero weight observations). The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year. *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)
Covered × Post	-0.016***	-0.017***	-0.031***	-0.033***	-0.049***	-0.051**	-0.075***	-0.079***
	(0.004)	(0.005)	(0.010)	(0.012)	(0.017)	(0.020)	(0.024)	(0.029)
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	2.15e+07	2.15e+07	2.08e+07	2.08e+07	2.01e+07	2.01e+07	1.94e+07	1.94e+07
R^2	0.06	0.13	0.07	0.20	0.08	0.24	0.09	0.28

Table IA.27: Heterogeneous Effect based on Industry Expertise using Funds' Investment Universe Sample

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 and the sample includes all fund-stock-quarter observations where the stock falls within the fund's investment universe. A fund universe is defined following Koijen and Yogo (2019) as the set of stocks held in the current or any of the previous 11 quarters (i.e., potentially includes zero weight observations). The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year. *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. 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. We note that the terms "Industry Specialist" and "Sector Fund" are constant at the fund level. 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)
Covered × Post	-0.016***	-0.027***	-0.040**	-0.063***
	(0.005)	(0.009)	(0.015)	(0.021)
Covered × Post × Industry Specialist	-0.157***	-0.401***	-0.684***	-1.056***
	(0.059)	(0.154)	(0.255)	(0.388)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	2.15e+07	2.08e+07	2.01e+07	1.94e+07
R^2	0.14	0.20	0.24	0.28

Panel B: Sector Funds

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.016***	-0.027***	-0.040***	-0.064***
	(0.005)	(0.009)	(0.015)	(0.021)
Covered × Post × Sector Fund	-0.109*	-0.283*	-0.486*	-0.749**
	(0.058)	(0.149)	(0.250)	(0.382)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	2.15e+07	2.08e+07	2.01e+07	1.94e+07
R^2	0.14	0.20	0.24	0.28

Table IA.28: Heterogeneous Effect based on Geographical Location using Funds' Investment Universe Sample

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 and the sample includes all fund-stock-quarter observations where the stock falls within the fund's investment universe. A fund universe is defined following Koijen and Yogo (2019) as the set of stocks held in the current or any of the previous 11 quarters (i.e., potentially includes zero weight observations). The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year. *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 covered firm's headquarters or (ii) the covered 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). We note that the term *Local* is constant at the fund-stock level. 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)
Covered × Post	-0.018***	-0.033***	-0.052***	-0.082***
	(0.005)	(0.012)	(0.019)	(0.028)
Covered × Post × Local	-0.041**	-0.080*	-0.120*	-0.180*
	(0.016)	(0.043)	(0.070)	(0.103)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	2.14e+07	2.07e+07	2.00e+07	1.93e+07
R^2	0.14	0.20	0.24	0.28

Table IA.29: Heterogeneous Effects based on Skills before Coverage Initiation using Funds' Investment Universe Sample

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 and the sample includes all fund-stock-quarter observations where the stock falls within the fund's investment universe. A fund universe is defined following Koijen and Yogo (2019) as the set of stocks held in the current or any of the previous 11 quarters (i.e., potentially includes zero weight observations). The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year. $Covered \times Post$ 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. We note that the term $HighPickingPre$ is constant at the fund-stock level. 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	0.004	0.011	0.001
	(0.005)	(0.010)	(0.016)	(0.019)
Covered × Post × High Picking Pre	-0.059***	-0.125***	-0.201***	-0.278***
	(0.014)	(0.032)	(0.056)	(0.085)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	2.11e+07	2.04e+07	1.98e+07	1.91e+07
R^2	0.14	0.20	0.24	0.28

I.15. Robustness Test: Adding Stock-Quarter Fixed Effects

Table IA.30: 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 (4). $Covered \times Post$ 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)	(13)	(14)	(15)	(16)
Covered × Post × High Picking Pre	-0.056***				-0.115***				-0.180***				-0.244***			
	(0.010)				(0.024)				(0.043)				(0.064)			
Covered × Post × Industry Specialist		-0.119**				-0.317**				-0.546**				-0.849**		
		(0.055)				(0.154)				(0.262)				(0.407)		
Covered × Post × Sector Fund			-0.092**				-0.237*				-0.407*				-0.629*	
			(0.047)				(0.133)				(0.227)				(0.353)	
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	1.27e+07	1.23e+07	1.24e+07	1.24e+07	1.23e+07	1.19e+07	1.20e+07	1.20e+07	1.19e+07	1.15e+07	1.16e+07	1.16e+07	1.15e+07
R^2	0.49	0.49	0.49	0.49	0.55	0.55	0.55	0.55	0.58	0.58	0.58	0.58	0.61	0.61	0.61	0.61

I.16. Clustering of Standard Errors by Time

In our specifications at the fund-stock-quarter level, we implement a double clustering of standard errors at the fund and stock levels. In this section, we re-estimate regressions considered in Tables II to VIII when clustering standard errors by time: we use triple clustering at the fund, stock, and year–quarter levels whenever the unit of observation is fund–stock–quarter, and, when the unit is stock–quarter, we use double clustering at the stock and year–quarter levels. Under this more conservative approach, our results remain robust. Tables IA.31 to IA.37 below show that the effects of interest remain statistically significant with this approach, even though, as expected, significance is sometimes weaker.

Table IA.31: Alternative Data and Stock Picking Skills with Standard Errors Clustered by Time

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. *Covered* × *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Standard errors are triple-clustered at the fund, stock and year-quarter levels.

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.022**	-0.021**	-0.042**	-0.038*	-0.068**	-0.059*	-0.107**	-0.099*
	(0.010)	(0.010)	(0.020)	(0.022)	(0.031)	(0.035)	(0.043)	(0.050)
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 IA.32: Heterogeneous Effect based on Industry Expertise with Standard Errors Clustered by Time

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. $Covered \times Post$ 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 IA.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. We note that the terms "Industry Specialist" and "Sector Fund" are constant at the fund level. Standard errors are triple-clustered at the fund, stock and year-quarter levels.

Panel A: Industry Specialists

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.015**	-0.022	-0.031	-0.057*
	(0.007)	(0.013)	(0.021)	(0.029)
Covered × Post × Industry Specialist	-0.165***	-0.426***	-0.723***	-1.103***
	(0.055)	(0.144)	(0.238)	(0.361)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.20	0.28	0.34	0.39

Panel B: Sector Funds

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.015**	-0.022	-0.031	-0.057*
	(0.007)	(0.013)	(0.021)	(0.029)
Covered × Post × Sector Fund	-0.135**	-0.348**	-0.592**	-0.901**
	(0.054)	(0.145)	(0.240)	(0.365)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.24e+07	1.20e+07	1.16e+07
R^2	0.20	0.28	0.34	0.39

Table IA.33: Heterogeneous Effect based on Geographical Location with Standard Errors Clustered by Time

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. *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 covered firm's headquarters or (ii) the covered 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). We note that the term *Local* is constant at the fund-stock level. 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 triple-clustered at the fund, stock and year-quarter levels.

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	-0.017*	-0.031*	-0.049*	-0.082*
	(0.009)	(0.018)	(0.029)	(0.041)
Covered × Post × Local	-0.058***	-0.119**	-0.184**	-0.274**
	(0.017)	(0.053)	(0.084)	(0.123)
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 IA.34: Heterogeneous Effects based on Skills before Coverage Initiation with Standard Errors Clustered by Time

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. $Covered \times Post$ 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. We note that the term $HighPickingPre$ is constant at the fund-stock level. 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 triple-clustered at the fund, stock and year-quarter 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
	(0.006)	(0.012)	(0.017)	(0.020)
Covered × Post × High Picking Pre	-0.054***	-0.114***	-0.185***	-0.257***
	(0.016)	(0.033)	(0.057)	(0.083)
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 IA.35: Divestment from Covered Stocks with Standard Errors Clustered by Time

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. $Covered \times Post$ is a dummy equal to one after RS Metrics initiates coverage of the stock. Standard errors are triple-clustered at the fund, stock and year-quarter levels in columns (1) and (2), and double-clustered at the stock and year-quarter levels 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***
	(0.041)	(0.042)	(0.066)	(0.069)
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
Observations	1.28e+07	1.28e+07	229,291	101,786
R^2	0.72	0.87	0.89	0.87

Table IA.36: Divestment of Experts from Covered Stocks with Standard Errors Clustered by Time

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. $Covered \times Post$ 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. We note that "High Picking Pre" is constant at the fund-stock stock level. 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. 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. We note that the terms "Industry Specialist" and "Sector Fund" are constant at the fund level. In column (4), "Local" is a dummy variable that equals one if the fund is located in the same MSA as either (i) the covered firm's headquarters or (ii) the covered 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). We note that "Local" is constant at the fund-stock stock level. Standard errors are triple-clustered at the fund, stock and year-quarter levels.

	Stock Rank			
	(1)	(2)	(3)	(4)
Covered × Post	0.011	-0.062*	-0.061	-0.061
	(0.052)	(0.037)	(0.037)	(0.037)
Covered × Post × High Picking Pre	-0.196***			
	(0.057)			
Covered × Post × Industry Specialist		-0.416**		
		(0.180)		
Covered × Post × Sector Fund			-0.363**	
			(0.157)	
Covered × Post × Local				-0.146
				(0.133)
Log(Holding Duration)	0.154***	0.154***	0.154***	0.154***
	(0.004)	(0.004)	(0.004)	(0.004)
Stock Return Past Month	0.555***	0.557***	0.557***	0.557***
	(0.034)	(0.034)	(0.034)	(0.034)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.26e+07	1.27e+07	1.27e+07	1.26e+07
R^2	0.88	0.88	0.88	0.88

Table IA.37: Investment in Peers of Covered Stocks with Standard Errors Clustered by Time

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 double-clustered at the stock and year-quarter levels.

	Nb. Funds Holding the Stock		
	(1)	(2)	(3)
Industry Peer Covered × Post	0.112***		0.110***
	(0.032)		(0.032)
Local Peer Covered × Post		0.085***	0.083***
		(0.028)	(0.028)
Year-Quarter FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Observations	227,323	227,323	227,323
R^2	0.89	0.89	0.89

I.17. Stacked Difference-in-Differences and Cohort Analysis

Table IA.38: 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 (4). $Covered \times Post$ 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		
	≤ 2013	≤ 2015	≤ 2017	≤ 2013	≤ 2015	≤ 2017	≤ 2013	≤ 2015	≤ 2017	≤ 2013	≤ 2015	≤ 2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Covered × Post	-0.020**	-0.025***	-0.021***	-0.033*	-0.049***	-0.040***	-0.055*	-0.079***	-0.065**	-0.090**	-0.123***	-0.103***
	(0.009)	(0.007)	(0.007)	(0.019)	(0.015)	(0.015)	(0.030)	(0.026)	(0.025)	(0.041)	(0.037)	(0.037)
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	1.61e+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	0.14	0.15

I.18. Additional Evidence Using Alternative Data from Datarade

In this section, we provide additional evidence to assess the external validity of our main findings and to address the concern that our results may be driven by the relatively small number of stocks covered by RS Metrics. Specifically, we conduct a complementary analysis using alternative datasets identified on the Datarade platform (https://datarade.ai), a digital marketplace that aggregates a wide range of commercially available alternative data products.

We focus on datasets classified under the *Commerce* and *Foot Traffic* categories, where similar data such as credit card transactions or geospatial activity measures are most likely to appear. For each dataset in these categories, we manually inspect the dataset descriptions to determine whether specific publicly traded firms (by name or ticker) are explicitly mentioned as covered.[12] When a dataset description clearly lists identifiable firms, we treat those firms as "covered" by that dataset as of the release date provided on Datarade.

Using this procedure, we identify 27 distinct covered firms, of which three overlap with the RS Metrics sample. We then construct two treatment variables, *Commerce Covered* and *FootTraffic Covered*, and interact them with a post indicator based on the dataset release date. We estimate specifications analogous to our main difference-in-differences model (Eq. (6)) using these alternative treatment variables.

The results are reported in Table IA.39. Across specifications, we find that fund stock-picking ability declines following coverage by these alternative datasets, consistent with our main findings using RS Metrics.

We emphasize that this test is necessarily less precise than our main analysis due to the approximate nature of the coverage dates and firm identification. Indeed, for this test,

[12] For example, the dataset *"Satellite US Construction Materials Dataset"* (available at https://datarade.ai/data-products/satellite-us-construction-materials-dataset-package-cemex-v-space-know) explicitly states that it covers Cemex (CX), Martin Marietta (MLM), and Vulcan (VMC), and provides six years of historical data.

we cannot determine whether a treated firm in the sample was covered from the earliest historical date available mentioned on Datarade or added later. Thus, our coverage timing is necessarily an approximation. However, the consistency of the results across distinct categories of alternative data provides additional support for the robustness and generality of our findings.

Table IA.39: Alternative Data Coverage from Commerce and Foot Traffic Datasets and Stock Picking Skills

This table presents additional evidence on the effect of alternative data coverage on fund stock-picking abilities using firms identified as covered by alternative datasets listed on the Datarade data marketplace. 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 defined in equation (4). *Commerce Covered × Post* and *FootTraffic Covered × Post* are dummies equal to one after the release date of a dataset listed on Datarade in the Commerce or Foot Traffic category, respectively, for firms explicitly mentioned as covered in the dataset description. Coverage status and timing are constructed based on publicly available dataset descriptions and release dates on Datarade and therefore represent an approximation of actual firm-level coverage initiation. 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)
Commerce Covered × Post	-0.050**	-0.059***	-0.094**	-0.120***	-0.148***	-0.176***	-0.176***	-0.193**
	(0.021)	(0.022)	(0.047)	(0.045)	(0.050)	(0.055)	(0.063)	(0.078)
FootTraffic Covered × Post	-0.031***	-0.067***	-0.020**	-0.054***	-0.052***	-0.050***	-0.313***	-0.235***
	(0.004)	(0.004)	(0.008)	(0.008)	(0.013)	(0.013)	(0.013)	(0.015)
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	1.25e+07	1.25e+07	1.21e+07	1.21e+07	1.17e+07	1.17e+07	1.13e+07	1.13e+07
R^2	0.10	0.20	0.12	0.28	0.14	0.34	0.15	0.39

I.19. Stock Picking before 2018

Table IA.40: 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 (4). *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)
Covered × Post	-0.024***	-0.023***	-0.043***	-0.037**	-0.064***	-0.051*	-0.100***	-0.087**
	(0.007)	(0.008)	(0.015)	(0.017)	(0.025)	(0.029)	(0.034)	(0.041)
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	9527206	9527206	9450045	9450045	9365427	9365427	9275134	9275134
R^2	0.09	0.20	0.11	0.28	0.12	0.34	0.14	0.39

I.20. Analysis of Funds' Return Gap

Table IA.41: 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.21. Divestment: Funds' Weights Relative to Market Weights

Table IA.42: Divestment of Experts from Covered Stocks: Fund Weights Relative to Market Weights

The table presents the results of our study on the impact of alternative data on fund holdings relative to market weights, depending on fund expertise. We investigate this by examining the investment-size rank of covered stocks in portfolios of funds managed by experts, calculated based on the fund's weight in the stock relative to the market weight. Regressions are estimated at the fund-stock-quarter level, and the dependent variable is the natural logarithm of the stock rank (relative to market) 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. $Covered \times Post$ 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. We note that "High Picking Pre" is constant at the fund-stock stock level. 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 IA.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. We note that the terms "Industry Specialist" and "Sector Fund" are constant at the fund level. In column (4), "Local" is a dummy variable that equals one if the fund is located in the same MSA as either (i) the covered firm's headquarters or (ii) the covered 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). We note that "Local" is constant at the fund-stock stock level. Our specifications include two control variables: the logarithm of the holding duration of the stock by the fund (reset to zero when the stocks is no longer part of the fund portfolio) and the stock return over the past month. Standard errors are double-clustered at the fund and stock levels.

	Stock Rank				
	(1)	(2)	(3)	(4)	(5)
Covered × Post	-0.074**	0.008	-0.057*	-0.056*	-0.055*
	(0.035)	(0.047)	(0.032)	(0.032)	(0.033)
Covered × Post × High Picking Pre		-0.178***			
		(0.067)			
Covered × Post × Industry Specialist			-0.425**		
			(0.183)		
Covered × Post × Sector Fund				-0.374**	
				(0.159)	
Covered × Post × Local					-0.167
					(0.142)
Log(Holding Duration)	0.160***	0.160***	0.160***	0.160***	0.160***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Stock Return Past Month	0.543***	0.541***	0.543***	0.543***	0.542***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	1.27e+07	1.26e+07	1.27e+07	1.27e+07	1.26e+07
R^2	0.85	0.85	0.85	0.85	0.85

I.22. Stock Picking Excluding Industry and Geographical Peers

Table IA.43: 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 (4). *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)
Covered × Post	-0.020***	-0.017***	-0.039***	-0.028**	-0.063***	-0.044**	-0.101***	-0.077***
	(0.006)	(0.006)	(0.012)	(0.012)	(0.019)	(0.019)	(0.027)	(0.027)
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	4194803	4194803	4059525	4059525	3922008	3922008	3782866	3782866
R^2	0.17	0.27	0.20	0.37	0.22	0.43	0.23	0.49

I.23. Fund-Level Alpha and Exposure to Covered Stocks

In this section, we complement our stock-level analysis with a fund-level test to assess whether the reallocation of capital documented in Section V.B offsets the loss of informational advantage in covered stocks. In particular, we examine whether funds that are more exposed to the universe of firms covered by RS Metrics experience changes in their overall performance as coverage expands.

Conducting this analysis at the fund level presents two challenges. First, coverage by RS Metrics expands gradually across stocks and over time, implying that funds are differentially exposed to coverage at any given point. Second, portfolio reallocation decisions following coverage initiation make contemporaneous exposure endogenous, as funds may actively reduce their holdings of covered stocks. To address these issues, we construct a predetermined measure of fund exposure to covered firms that is not mechanically affected by post-coverage rebalancing.

Let U denote the set of stocks that are covered by RS Metrics at some point during our sample period, and let T_s denote the coverage initiation date for stock $s \in U$. For each fund f and month t, we define fund exposure as:

$$FundExposure_{f,t} = \sum_{s \in U} \left(w_{f,s,t-1} \cdot \mathbf{1}[t < T_s] + w_{f,s,T_s-1} \cdot \mathbf{1}[t \geq T_s] \right), \tag{IA.38}$$

where $w_{f,s,t}$ denotes the weight of stock s in fund f's portfolio at time t. For periods prior to coverage ($t < T_s$), exposure is based on lagged portfolio weights. For periods after coverage ($t \geq T_s$), we fix the weight at its last pre-coverage value, w_{f,s,T_s-1}. This construction ensures that the exposure measure is predetermined with respect to post-coverage returns and is not mechanically affected by rebalancing decisions.

We then estimate the following specification at the fund-month level:

$$\alpha_{f,t} = \beta \, FundExposure_{f,t} + \gamma_f + \delta_t + \varepsilon_{f,t}, \tag{IA.39}$$

where $\alpha_{f,t}$ denotes fund f's risk-adjusted return (gross of fees) in month t, γ_f are fund fixed effects, and δ_t are time fixed effects. Identification therefore comes from within-fund variation over time in exposure to the set of covered stocks as coverage expands.

Fund-level alpha is computed as:

$$\alpha_{f,t} = R_t^f - \sum_{i=1}^{F} \beta_{f,t}^i R_t^i, \tag{IA.40}$$

where R_t^f is the fund's return in excess of the risk-free rate, $\beta_{f,t}^i$ are factor loadings estimated using daily returns over a rolling 252-day window, and R_t^i are factor returns. We consider three factor models: (i) CAPM, (ii) Fama–French three-factor, and (iii) Fama–French three-factor augmented with momentum.

Table IA.44 reports the results. Across all specifications, the coefficient on $FundExposure_{f,t}$ is negative and statistically significant. This implies that funds with greater exposure to the set of covered firms experience a larger decline in alpha as coverage expands. In economic terms, the results indicate that the reallocation of capital toward uncovered stocks does not fully offset the loss in informational advantage in covered stocks.

These findings are consistent with the interpretation that funds reallocate capital toward "second-best" investment opportunities following coverage initiation. While the reallocation documented in Section V.B suggests that funds shift capital to stocks where their expertise remains valuable, the fund-level results indicate that these alternative opportunities do not fully compensate for the erosion of alpha in covered stocks.

We emphasize that this fund-level analysis is, by construction, less cleanly identified than our primary fund-stock-level analysis. Even with the predetermined exposure measure, residual concerns remain regarding unobserved differences across funds that may correlate with both exposure and time-varying performance. For this reason, we view these results as complementary evidence. Our primary identification strategy relies on the fund-stock-level *Picking* measure, which allows us to isolate changes in stock-level informational advantage while controlling for time-varying fund characteristics through fund–time fixed effects.

Table IA.44: Fund-Level Alpha and Exposure to Covered Stocks

This table reports estimates of equation (IA.39). The dependent variable is monthly fund alpha. Alpha is measured gross of fees and calculated under three alternative factor models: CAPM (Panel A), Fama-French three-factor model (Panel B), and Fama-French three-factor model augmented with momentum (Panel C). The main independent variable, $FundExposure_{f,t}$, is defined in equation (IA.38). For each fund and month, it sums lagged portfolio weights in stocks that will be covered in the future and freezes portfolio weights at their last pre-coverage level once coverage begins. All specifications include fund fixed effects and year-month fixed effects. Columns (2) and (4) additionally control for the logarithm of lagged fund total net assets (TNA). Columns (3) and (4) further include fund-style × year-month fixed effects to account for time-varying performance differences across CRSP investment styles. Standard errors are clustered at the fund level.

Panel A: CAPM

	Fund Alpha			
	(1)	(2)	(3)	(4)
Fund Exposure	-0.030***	-0.027***	-0.038***	-0.035***
	(0.006)	(0.006)	(0.006)	(0.006)
Log(TNA)		-0.001***		-0.001***
		(0.000)		(0.000)
Fund FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	No	No
Style × Year-Month FE	No	No	Yes	Yes
Observations	206,260	206,260	206,260	206,260
R^2	0.16	0.16	0.26	0.26

Panel B: Fama-French 3-factor Model

	Fund Alpha			
	(1)	(2)	(3)	(4)
Fund Exposure	-0.013**	-0.010*	-0.023***	-0.020***
	(0.006)	(0.006)	(0.005)	(0.005)
Log(TNA)		-0.002***		-0.001***
		(0.000)		(0.000)
Fund FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	No	No
Style × Year-Month FE	No	No	Yes	Yes
Observations	206,260	206,260	206,260	206,260
R^2	0.15	0.15	0.23	0.23

Panel C: Fama-French 3-factor Model with Momentum

	Fund Alpha			
	(1)	(2)	(3)	(4)
Fund Exposure	-0.018***	-0.016***	-0.023***	-0.021***
	(0.005)	(0.005)	(0.005)	(0.005)
Log(TNA)		-0.001***		-0.001***
		(0.000)		(0.000)
Fund FE	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	No	No
Style × Year-Month FE	No	No	Yes	Yes
Observations	206,260	206,260	206,260	206,260
R^2	0.13	0.13	0.20	0.20

References

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

Biais, B., Foucault, T., and Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics*, 116:292–313.

Dugast, J. and Foucault, T. (2025). Equilibrium data mining and data abundance. *The Journal of Finance*, pages 211–258.

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. O., Patatoukas, P. N., and Zeng, J. (2025). On the capital market consequences of big data: Evidence from outer space. *Journal of Financial and Quantitative Analysis*, 60(2):551–579.

Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.