Are New Technologies Replacing the Information Produced by Financial Markets?

Laurent Frésard and Marco Zanotti*

Click here for the most recent version.

November 14, 2025

ABSTRACT

Using the staggered adoption of data technologies providing firms with novel insights about their customers, we show that their investment becomes significantly less sensitive to non-fundamental stock price movements after adoption. The effect is consistent with data technologies improving managers' internal information about future demand, thereby reducing their reliance on stock market signals. This "replacement effect" is robust to alternative explanations and reverses under data-privacy restrictions. Our findings suggest that the diffusion of data technologies weakens the informational role of stock prices in guiding real investment decisions.

Key words: Investment, revelatory price efficiency, managerial learning, data technology, price pressure. *JEL classification*: D84, G14, G17, M41

*Universita della Svizzera Italiana (Lugano) and Swiss Finance Institute. Frésard can be reached at laurent.fresard@usi.ch. Zanotti can be reached at marco.zanotti@usi.ch. All errors are the authors' alone.

"*The debates over market efficiency, exciting as they are, would not be important if the stock market did not affect real economic activity*" Morck et al. (1990)

I Introduction

The emergence of new information technologies is profoundly changing how information is gather, processed, and used in financial markets. Technology is making new sources of information available to market participants. Big data, analytics, and advances in artificial intelligence allow them to process this information faster and more efficiently, providing new and timely predictions and insights about firms' fundamentals. A rapidly growing literature studies how the adoption of these data technologies affect investors' information acquisition decisions, investment choices, and the resulting informational efficiency of stock prices (Foucault et al. (2025)). These are important questions because firms (and other decision makers) rely on stock prices as a source of information to guide their decisions.[1] Hence, if they make markets more efficient, data technologies could indirectly improves firms' decisions and bolsters real economic efficiency.

Yet, mirroring investors, firms are also increasingly adopting big data analytics and artificial intelligence to inform their decisions (Baley and Veldkamp (2025)). These tools enable managers to extract more and better information from internal data sources (e.g., data collected on customers, products, employees, or suppliers) and obtain novel insights about their business opportunities.[2] Our contribution is to show that by improving firms' internal information, data technologies *reduce* their reliance on stock prices as an external source of information. Thus, firms' growing use of data technologies lessens the informational role of stock markets in affecting real economic activities. Using the words of Morck et al. (1990), our results suggest that the advent of data technologies, perhaps paradoxically, makes the "debates over market efficiency" less important.

We reach this conclusion by tracking firms' reliance on their stock price for information after they adopt specific technologies that collect and analyze data generated by visitors on

[1] See Bond et al. (2012) and Goldstein (2023) for surveys of the literature.

[2] For instance, Amazon analyzes web-traffic data to "improve their products" Amazon (2025), and Nike analyze similar data to "understand personal preferences" Nike (2025).

their website (e.g., Google Analytics, LiveRamp, or Hotjar). For instance, data analytics tools improve predictions of future demand or costs based on customers' activities (Brynjolfsson and McElheran (2016) or Goldfarb and Tucker (2019)). To understand why data technologies can alter managers' reliance on their stock price and how we can infer this effect from the data, we use a standard investment model in which a manager chooses how much to invest in a growth opportunity. The optimal investment increases in her expectation of the payoff of this opportunity, which depends on her internal information and the firm's stock price, which contains relevant but noisy information about the payoff (e.g., due to liquidity or sentiment shocks). Because her internal information is imprecise and the stock price is informative about the opportunity, the manager optimally conditions her decision on both her internal information and her firm's stock price (Chen et al. (2007)).

The firm's investment is thus sensitive to its stock price, and the strength of this sensitivity increases with the informativeness of the stock price, that is, the precision of the price signal. Building on the idea that data technologies reduce uncertainty about future outcomes (Baley and Veldkamp (2025)), we posit that their use provides the manager with more precise internal information about the growth opportunity. A direct implication is that, all else equal, the manager relies more this information and less on the stock price when she uses technologies to process internal data. The firm's investment becomes *less* sensitive to its stock price, an hypothesis we label the "replacement" effect.

Testing this hypothesis is challenging because the econometrician does not observe the manager's private internal information. If this information overlaps with the information contained in the stock price–which is likely–an OLS regression of the firm's investment on its stock price yields an estimate of the manager's "true" reliance on the stock price that is biased upward (Foucault and Fresard (2014)). Hence, as it is well-known in the literature, the estimated sensitivity could be positive even if the manager ignores the stock market completely. Furthermore, the model predicts that the OLS estimate of the investment-to-price sensitivity should *not* change when the manager uses data- technologies. With more precise internal information, the manager's true reliance on her firm's stock price decreases (the replacement effect). Yet, the bias increases because her private information weight more in her investment decision. Since both effects offset exactly, we cannot infer the replacement effect by regressing firms' investment on their stock price.

We can however infer it by using an instrumental variable for the firm's stock price that is orthogonal to the manager's private information. Similar to Dessaint et al. (2019) we use the noise in the firm's stock price generated by liquidity shocks as an instrument. This source of variation is exogenous to the firm's growth opportunity by design, and unless the manager can perfectly identify the noise in the price, it should be orthogonal to her internal information.[3] The replacement effect implies that the firm's investment should becomes *less* sensitive to the noise component of its stock price when the manager uses data technologies to improve her internal information. This effect can be estimated using an "instrumented" difference-in-differences approach that estimates whether the investment of firms that adopt data technologies becomes less sensitive to their *instrumented* stock price compared to that of non-adoption firms. Notably, although the decision to adopt data technologies may arguably be endogeneous (e.g., firms adopting when they have better growth opportunities), we show that it does not bias our estimation of the replacement effect as long as firms' adoption is unrelated to the noise in their stock prices (which is both plausible and supported by the data).

We implement this approach using a large panel of 8,815 U.S. firms over the period 1980–2023. We identify whether and when firms adopt data technologies using BuiltWith, a company that tracks the installation and removal of various technologies by analyzing webpage code. We focus on 822 specific technologies (or plugins) related to collection of customer data, tracking, and A/B testing and we concentrate on retail-oriented firms because they are the most likely to extract insights about their growth opportunities from data generated on their website.[4] In our sample, adoption of these tools rose from about 20% in 2005 to 40% in 2010, and nearly 95% by 2023. This pattern mirrors the broad diffusion of analytics and tracking technologies over the same period. We show that adopting firms are

[3] Dessaint et al. (2019) show that noise in stock prices influences investment through managerial learning, and that managers cannot perfectly filter out the noise in stock prices.

[4] These tools allow firms to forecast product demand and learn about consumer preferences. For instance, industry reports describe how Adobe Analytics helps online retailers forecast demand ahead of the holiday season (see news.adobe.com/news/[...]online-rising-yoy). Similarly, firms use A/B testing—tools akin to randomized control trials—to compare alternative product pages. By observing which version generates higher engagement or conversion, firms can infer consumer preferences.

more likely to issue sales guidance, and issue more precise sales forecasts, consistent with data technologies improving internal information about future demand.

We measure non-fundamental variation in stock prices using quarterly reinvestment of (cash) dividends by mutual funds. We construct the hypothetical purchases of each firm's stock implied by mutual funds reinvesting the cash received from dividends. These purchases are hypothetical in the sense that they are derived assuming that mutual funds reinvest the cash payments from firms' dividends keeping the distribution of their holdings constant (which we confirm in the data). Like Schmickler and Tremacoldi-Rossi (2023) and Chen (2024) we find that these dividend-induced purchases are associated with positive price pressures that eventually revert (within three months), in line with these purchases representing non-fundamental demand shocks.

Consistent with the replacement effect, the sensitivity of investment to the non-fundamental component of their stock – the "investment-to-noise sensitivity" – decreases for firms after adopt data technologies compared to non-adopting firms. Firms' reliance on their stock price for information decreases by about one half following the adoption of data technologies: a one-standard deviation increase in firms' instrumented stock price is associated with a 7.18% increase in investment before adoption, but 3.43% after adoption.[5] Tracing the investment-to-noise sensitivity in event-time around adoption reveals no difference between adopting and non-adopting firms prior to adoption, and a replacement effect that materializes two years after adoption. The results are robust to stacking the sample to account for heterogeneous treatment effect (Baker et al. (2022)), to account for common variation within industries, and to alternative sample definitions or measurement choices, such as using Total Q from Peters and Taylor (2017). Moreover, falsification tests confirm that the replacement effect is *not* present after firms adopt technologies that are *unrelated* to tracking customer data, which dispel concerns that the replacement effect captures broader changes in firms' web strategy. This effect is also not present for non-retail firms, which are less likely to obtain internal information about future demand from customers' web traffic.

[5]A one–standard deviation increase in the instrumented stock price raises investment by 2.3 percentage points relative to the unconditional mean of 0.32 (7.18%), and by 1.1 percentage points after adoption (3.43%).

Notably, the replacement effect partly reverses for adopting firms that are subject to the California Consumer Privacy Act (CCPA) after 2019. This regulation limits the personal information that firms can collect and process from customers in California. By limiting the scope of data collection, this act lowers the effectiveness of data technologies to improve managers' internal information. The magnitude of the replacement effect drops by about 12.5% for adopting firms become subject to the CCPA.

We consider various alternative explanations for the negative effect of firms' adoption of data technologies on their investment-to-noise sensitivity. First, instead of improving managerial information about their fundamentals, data technologies might help managers to better identify the noise in their stock price. This would reduce their investment-to-noise sensitivity. We confirm that managers can partly detect the noise in their stock price since they trade *against* (i.e., selling shares following dividend-reinvestment price pressure). Yet, inconsistent with a learning about noise channel, their propensity to trade against the noise does *not* change after they install data technologies on their websites.

Second, the use of data technologies may render stock prices less informative for managers, *without* affecting their internal information. This could happen, for instance, if investors have less incentives to produce information about firms using data technologies. In that scenario, managers' reliance on their stock price decreases not because it is replaced by better internal information, but simply because the informational content of stock prices is lower. Consistent with this channel, proxies for the informativeness of firms' stock prices (i.e., their ability to anticipate future outcomes) decreases following the adoption of data technologies. The replacement effect is however present, with similar magnitude, for firms that do not experience any change in the informativeness of their stock price after adopting data technologies.

We also investigate the role of well-known channels that could explain changes in the relation between firms' investment and their stock price. First, we examine whether the adoption of data technologies improves firms' financing conditions (e.g., by making their profits less volatile), which could make their investment less sensitive to their stock price (Baker et al. (2003)). Using various proxies for the cost and availability of external financing, we find no support for such a financing channel. Second, we investigate if the adoption of data technologies increases firms' market power, since this could also create a disconnect between firms' investment and their stock price (Hayashi (1982) or Gutiérrez and Philippon (2017)). We find however no significant change in firms' market share, price-cost margins, or markups after adopting data technologies.

Overall, our findings indicate that data technologies substitute for the informational role of financial markets in firms' investment decisions. At a more aggregate level, we estimate that industries with more intensive adoption of data technologies exhibit a decline in the sensitivity of investment to Q over the past decades. While descriptive, this pattern suggests that the replacement effect may help reconcile part of the growing aggregate disconnect between financial markets and real investment documented by Gutiérrez and Philippon (2017) and Alexander and Eberly (2018).

Related literature. Our results add to, and connect, two distinct strands of research. First, we contribute to the literature analyzing how stock prices affect real decisions through their informational content (see Goldstein (2023) for a recent survey). Morck, Shleifer, and Vishny (1990) first noted that secondary stock markets matter for the real economy only if they influence firms' decisions. When they do, efficient stock markets reveals valuable information to decision markets, and improves capital allocation in the economy, a concept named "revelatory price efficiency" (Bond et al. (2012)). We show that improvements in the quality of managers' internal information obtained from data technologies lessens how strongly their investment respond to their stock price. Although not necessarily impairing real economic efficiency, this replacement effect implies that the adoption of data technologies lowers the informational importance of the stock market as guiding firms' investment, despite the overall increase of stock prices' informativeness documented by Bai et al. (2016) and Davila and Parlatore (2025). In his recent survey, Goldstein (2023) asks "*whether alternative sources of information can replace information from the market.*" Our analysis directly speaks to that question.

Relatedly, our results shed light on what kind of information stock prices reveal to managers. Since data analytics and customer-tracking tools improves internal information pertaining to future demand and customers' preferences, the replacement effect suggests that it is information about these dimensions that managers typically glean from stock prices. These results complement that of Dessaint et al. (2025) who document that, in the context of acquisitions, stock markets reveals primarily information about discount rates.

Second, our paper also adds to the growing literature examining how data technologies affect firms, investors, and financial markets. Recent work finds that the adoption of data technologies affects firms' decisions (e.g., hiring or products), organization, product markets, and performance.[6] A complementary stream of studies indicates that the use of data technologies by investors affect their forecasts, capital collection and allocation, and modifies the efficiency of stock markets (see Foucault et al. (2025) for a recent survey).[7] By showing that data technologies lowers firms' reliance on their stock price as a source of information, our analysis reveals that data technologies modifies the interactions between firms and financial markets. In a related paper, Cao et al. (2023) study one facet of these interactions by showing that firms adjust their disclosure when investors use data technologies to read and analyze firms' regulatory filings. To the best of our knowledge, the replacement effect is novel and suggests that the informational efficiency of financial markets, which is potentially bolstered by data technologies, may be becoming a second-order driver of real economic efficiency.

Our paper further contributes to a growing methodological effort to measure firms' adoption and usage of data technologies. Recent work uses survey data (Brynjolfsson and McElheran (2016) or Acemoglu et al. (2022)), the hiring of specialists in data-related areas (Alekseeva et al. (2021) or Babina et al. (2024a)), the production of patents related to data technologies (Chen et al. (2024) or Mihet et al. (2025a)) or mentions of these technologies in firms' communication (Lu et al. (2024) or Mihet et al. (2025b)). We complement these approaches by focusing on the actual adoption of data technologies from firms' web-

[6]For instance, the use of data technologies improves firms' productivity in certain tasks (Brynjolfsson et al. (2023)), increase sales growth (Babina et al. (2024b)), investment efficiency (Ferracuti et al. (2024)) and innovation (Cockburn et al. (2018)). These technologies also alter firms' internal organizations (Labro et al. (2023) and Babina et al. (2023)), risk (Babina et al. (2024a)), disclosure practices (Blankespoor et al. (2022)), and reshape product markets concentration (Lu et al. (2025)).

[7]For instance, investors' use of data technologies affects their forecasting (Dessaint et al. (2024) and Chi et al. (2025)), their capital collection (Zanotti (2025)) and capital allocation (Abis (2022), Bonelli and Foucault (2025), or Bonelli (2025)). The effects on stock market efficiency are studied by Martin and Nagel (2022), Dou et al. (2025), and Dugast and Foucault (2025).

site. While it captures specific technologies (e.g., data analytics) applied to specific datasets (gathered from customers' traffic activity), our approach enables us to measure firms' adoption precisely (e.g., its exact timing), and plausibly link this adoption to improvements in managers internal information about future demand. This granularity is important to test the replacement effect.

II The replacement hypothesis

To examine how the use of data technologies can replace the information produced by financial markets, we consider a simple model of investment in which a manager relies on internal data and the stock market as sources of information about an investment opportunity. Our objective is to (i) illustrate how the use of data technologies affects firms' reliance of stock market information to make decisions, (ii) underline the challenges to identify this effect empirically, and (iii) justify our identification strategy.

A The investment model

As in Subrahmanyam and Titman (1999), we consider an investment model with two dates, 1 and 2. A firm has a growth opportunity with an uncertain payoff at date 2 equal to

$$G(I,\tilde{\theta}) = \tilde{\theta}I - \frac{I^2}{2} \tag{1}$$

where I is the investment in the opportunity and $\tilde{\theta}$ is the marginal productivity of this investment (the firm's fundamentals). For instance, $\tilde{\theta}$ could represent the future demand for the firm's products or its cost of production. The fundamental $\tilde{\theta}$ is normally distributed, with mean zero (for simplicity) and variance σ^2_θ.

At date 1, the manager chooses the investment that maximizes the expected payoff conditional on her information, Ω_1 , about $\tilde{\theta}$. The optimal investment I^* solves

$$Max_I \mathrm{E}(G(I,\tilde{\theta})|\Omega_1) = \mathrm{E}(\tilde{\theta}|\Omega_1)I - \frac{I^2}{2}, \tag{2}$$

so that,

$$I^*(\Omega_1) = \mathrm{E}(\tilde{\theta}|\Omega_1). \tag{3}$$

Thus, the optimal investment is equal to the manager's expectation of the marginal return on her investment. To form this expectation, the manager has access to two distinct sources of information (signals) at date 1. Her set of signals about $\tilde{\theta}$ is $\Omega_1 = \{s, p\}$ and, conditional on $\tilde{\theta}$, these signals are independent. We describe each in turn.

First, the manager possesses internal data (e.g., data about customers, products, or employees) and processes it to obtain a noisy signal about the fundamentals $\tilde{\theta}$. We denote this signal by $s = \tilde{\theta} + \chi$ where χ is an error term, that is normally distributed with mean zero and variance σ_χ^2. We assume that the manager has imperfect internal information ($\sigma_\chi^2 > 0$), because, otherwise, she would not rely on the stock market for information. We denote the precision of the manager's signal as $\tau_\chi = \frac{1}{\sigma_\chi^2}$. Building on the idea that the manager relies on data technologies to reduces uncertainty about future outcomes (Baley and Veldkamp (2025)), we assume (in reduced form) that the precision of s is higher if she uses data technologies to collect and process data to anticipate $\tilde{\theta}$. For instance, the use of data analytics could improve the manager's prediction of future demand based on customers' data, or her prediction of future costs based on procurement and suppliers' data.

Second, the manager can obtain external information from the firm's stock price, p. We assume that $p = \tilde{\theta} + u$ where u is the noise in the stock price that is normally distributed with mean zero and variance σ_u^2. This noise captures in reduced-form variations in the firm's stock price p that are unrelated to the fundamental $\tilde{\theta}$ (e.g., due to liquidity or sentiment shocks). Hence, the stock price is only partially revealing $\tilde{\theta}$. Similar to Dessaint et al. (2019), we do not model investors' equilibrium trading decisions and market clearing conditions, but we simply assume that p is a noisy signal about $\tilde{\theta}$, whose precision increases with the amount of investors' information about $\tilde{\theta}$. In other words τ_u increases with (unmodeled) investors' information. In the terminology of Bond et al. (2012), τ_u captures the extent of "forecasting price efficiency" (FPE), that is, how accurately the stock price predicts the future value of $\tilde{\theta}$.

The firm's optimal investment given the manager's information at date 1 is

$$I^*(\Omega_1) = E(\tilde{\theta}|\Omega_1) = \frac{\tau_\chi}{\tau_\theta + \tau_\chi + \tau_u} s + \frac{\tau_u}{\tau_\theta + \tau_\chi + \tau_u} p. \tag{4}$$

Equation (4) describes how the manager's signals at date 1 affect her investment decision. For simplicity, let's define the relative weights on each signal as $\alpha = \frac{\tau_\chi}{\tau_\theta+\tau_\chi+\tau_u}$ and $\beta = \frac{\tau_u}{\tau_\theta+\tau_\chi+\tau_u}$ so that the optimal investment is given by $I^*(\Omega_1) = \alpha s + \beta p$. As the manager's internal signal about the fundamental $\tilde{\theta}$ is imperfect ($\tau_\chi < \infty$), she can improve her estimate of $\tilde{\theta}$ by relying on her firm's stock price as a source of information, that is, $\beta > 0$. In other words, the firm's investment is *sensitive* to its stock price, and the strength of this sensitivity is captured by β. Following Bond et al. (2012), β captures "revelatory price efficiency" (RPE). All else equal, this sensitivity increases when the stock price is more informative about $\tilde{\theta}$ since β increases in τ_u (see, for instance, Chen, Goldstein, and Jiang (2007)).

Equation (4) formalizes the idea that the use of data technologies can *replace* the information produced by the stock market for the manager, i.e., reduces RPE. Under our assumption that data technologies improve the precision of the manager's internal signal about $\tilde{\theta}$, Equation (4) implies that, all else equal, the manager relies relatively more on her internal signal when she uses data technologies ($\frac{\partial\alpha}{\partial\tau_\chi} > 0$). Unless the use of these technologies is associated with an improvement of the precision of the firm's stock price signal (τ_u) or reduction of the unconditional uncertainty of the fundamentals ($1/\tau_\theta = \sigma^2_\theta$), an improvement in τ_χ induces the manager to rely *less* on her firm's stock price to invest, that is, $\frac{\partial\beta}{\partial\tau_\chi} < 0$. We obtain the following prediction.

Proposition 1 *(The replacement hypothesis) All else equal, when the manager uses data technologies, her firm's investment is more sensitive to her internal signal s (α is larger), and simultaneously less sensitive to her firm's stock price p (β is lower).*

B The econometrician's problem

Testing Proposition 1 is difficult because the estimation of β is challenging. Unlike the manager in the model, the econometrician does not perfectly observe the private signal s (e.g., Erickson and Whited (2000)), creating an omitted variable problem. To understand the econometrician's inference problem, consider what he realistically observes to infer β from the data: a sample of firms' investment I^* and their stock price p. Using this sample, he might estimate the linear Equation (4) such that:

$$I^* = c + bp + \xi \tag{5}$$

where c is a constant and the normally distributed error term ξ contains the unobserved private signal of the manager, s, and some error ε (i.e., $\xi = \alpha s + \varepsilon$). The coefficient b measures the sensitivity of investment to the stock price. Estimated by Ordinary Least Squares (OLS), the coefficient b^{OLS} is a biased estimate of true sensitivity β. Using Equation (4), it is given by

$$b^{OLS} = \beta + \alpha \frac{Cov(s, p)}{Var(p)}, \tag{6}$$

where $\alpha > 0$ (because $\tau_\chi < \infty$). Since $Cov(s, p) = \tau_\theta^{-1}$ and $Var(p) = \tau_\theta^{-1} + \tau_u^{-1}$, we conclude that $b^{OLS} > \beta$, that is, the OLS estimate is biased upward. The presence of this bias implies that we cannot test the replacement effect by tracking how b^{OLS} changes when the firm adopts data technologies (i.e., when τ_χ increases). Indeed, simple calculations yields $\frac{\partial b^{OLS}}{\partial \tau_\chi} = 0$ (see the Appendix for details), that is, b^{OLS} is insensitive to changes in the precision of the manager's signal (for τ_θ and τ_u fixed). To see why, let's write b^{OLS} as

$$b^{OLS}(\tau_\chi) = \underbrace{\frac{\tau_u}{\tau_\theta + \tau_\chi + \tau_u}}_{\beta(\tau_\chi)} + \underbrace{\frac{\tau_\chi}{\tau_\theta + \tau_\chi + \tau_u} \cdot \frac{\tau_\theta^{-1}}{\tau_\theta^{-1} + \tau_u^{-1}}}_{Bias(\tau_\chi)}. \tag{7}$$

The first term in Equation (7) decreases with τ_χ since the manager relies less on her firm's stock price when the precision of her internal signal improves – the true β decreases. Yet, the second term (i.e., the bias) increases with τ_χ because the managers' unobserved private signal gets more weight in her decision, which increases the bias. When τ_θ and τ_u are fixed, both effects offset exactly.

The model suggests however that the true β (and its sensitivity to τ_χ) can be identified by instrumental variables. To understand the required properties of a prospective instrument, recall that the manager's private signal is $s = \tilde{\theta} + \chi$ and the price is $p = \tilde{\theta} + u$. Thus, any prospective instrument z must be correlated with the firm's stock price p such that $Cov(p, z) \neq 0$ (i.e., satisfies the inclusion restriction). In addition, the instrument must be uncorrelated with the error term ξ in Equation (5) such that $Cov(\xi, z) = Cov(\alpha \cdot s + \varepsilon, z) = 0$ (i.e., satisfies the exclusion restriction). Since the error term ξ contains the unobserved private signal s, z must be uncorrelated with the manager's private signal s, or more precisely, with the part of s that's *orthogonal* to the price p. Put differently, z should capture a portion of the firm's stock price p that does not overlap with the manager's signal s.

The model points to an instrument that captures variation in the firm's stock price p that is unrelated to the fundamentals $\tilde{\theta}$. By construction, variation in p that is related to $\tilde{\theta}$ will be correlated with the manager's private signal s, and therefore violates the exclusion restriction. However, an instrument that only correlates with the noise component of the stock price, u, satisfies the exclusion restriction. Imagine an instrument z that satisfies this condition such that $Cov(u, z) \neq 0$ and $Cov(\tilde{\theta}, z) = 0$. If the econometrician observes z, he can identify the true β by estimating Equation (5) by instrumental variables, such that

$$b^{IV} = \frac{Cov(I^*, z)}{Cov(p, z)} = \frac{Cov(\alpha \cdot s + \beta \cdot p, z)}{Cov(p, z)} = \alpha \cdot \underbrace{\frac{Cov(s, z)}{Cov(p, z)}}_{=0 \text{ since } s \perp z} + \beta \cdot \frac{Cov(p, z)}{Cov(p, z)} = \beta. \tag{8}$$

Similar to Dessaint et al. (2019), the intuition behind z is as follows. If the manager relies on her stock price as a source of information ($\beta \neq 0$), her investment decision is sensitive to variation in p, whether this variation is fundamental ($\tilde{\theta}$) or non-fundamental (u). If the manager cannot perfectly distinguish these components of p (i.e., χ and u are independent), her investment should be *less* sensitive to the component of p that is instrumented by z, i.e., the non-fundamental component. In contrast, if the manager could perfectly distinguish these components (i.e., perfectly filter out the noise in its stock price), her investment decision should be insensitive to the noise component of the stock price, whether she uses data technologies or not.[8]

C Variation in τ_χ and identification

To test Proposition 1, we exploit the adoption of data technologies as a proxy for an increase in τ_χ and estimate whether β decreases when the firm adopts these technologies compared to when it does not (yet). We can perform this test estimating an "instrumented" difference-in-differences specification such that:

$$I^* = c + b_1 p + b_2 \delta_{\tau_\chi} + b_3 p \times \delta_{\tau_\chi} + \xi, \tag{9}$$

[8] See Dessaint et al. (2019) or Yan (2024) for evidence that corporate managers cannot perfectly distinguish fundamental from non-fundamental movements in stock prices.

where δ_{τ_χ} is a binary variable that equals one if the firm adopts data technologies, and zero otherwise, such that $\tau_\chi | \delta_{\tau_\chi} = 1 > \tau_\chi | \delta_{\tau_\chi} = 0$. The replacement effect implies that the firm's reliance of its stock price to decide on investment should be *lower* when it uses data technologies ($\beta_{\delta_{\tau_\chi}=1}$) than when it does not ($\beta_{\delta_{\tau_\chi}=0}$). Therefore, it predicts:

Prediction 1 *(Identifying the replacement effect) When the manager uses data technologies, her firm's investment becomes less sensitive to the non-fundamental component of her firm's stock price p, that is, $b_3^{IV} < 0$.*

In reality, however, the adoption of data technologies is unlikely to be exogenous, and might depend on unobserved factors related to firms' investment decisions (e.g., firms adopting when they have valuable growth opportunities). This potential endogeneity might in fact not be an issue. Because we need to instrument p with z to properly identify β, we estimate eq.(9) by instrumental variables, instrumenting p using z, and the interaction term $p \cdot \delta_{\tau_\chi}$ using the interacted instrument $z \times \delta_{\tau_\chi}$. Hence, as long as the decision to adopt data technologies is *not* correlated with z (e.g., firms adopt when their stock price is more noisy), the coefficient b_3^{IV} is an unbiased estimate of $\beta_{\delta_{\tau_\chi}=1} - \beta_{\delta_{\tau_\chi}=0}$.[9] This condition is important as it adds on the required properties that a prospective instrument needs to satisfy to ensure proper identification.

III Data and methodology

A Sample construction

We use data from BuiltWith to measure firms' adoption of data-related technologies. BuiltWith is an alternative data provider that analyzes websites' source code and searches for specific patterns, such as HTML tags, that identify the presence of technologies. Every website includes different "technologies" (or equivalently "plugins"), which work as building blocks. For instance, a website may use "Google Maps" to display an interactive map (e.g., to show store locations) and/or "Adobe Analytics" to collect and analyze web visitors' data. All of these technologies leave identifiable traces in the source code of a website. Thus, inspecting

[9] Precisely, b_1^{IV} is an unbiased estimate of $\beta_{\delta_{\tau_\chi}=0}$, and $b_1^{IV} + b_3^{IV}$ is an unbiased estimate of $\beta_{\delta_{\tau_\chi}=1}$. Therefore, b_3^{IV} is an unbiased estimate of the difference in β when the firm uses or not data-related tools.

the website's code allows one to observe which plugins are installed in a given point in time. BuiltWith continuously crawls websites to capture the installation (and removal) of various technologies starting January 2000. As of September 2024, BuiltWith covers 15,881 distinct technologies, divided into 34 distinct categories, such as Analytics (e.g., Google Analytics), Security (e.g., reCAPTCHA, which limit access to users suspected of being bots), or Payment (e.g., PayPal).

We focus specifically on the "Analytics" category. We carefully read the description of what these technologies actually do, and retain only 822 unique plugins designed to collect, process, and analyze customers data. These include, for instance, technologies labeled as "User Tracking and Analytics", "Social Media Tracking and Analytics", "A/B Testing", or "Cloud and Customer Relationship". For example, Hotjar is a data analytics plugin that records user interactions to help firms understand how users behave on their website, what they search, and where they spend more time. These plugins allow firms to gather rich information about customers—such as demographics, browsing patterns, and product interest—that managers can use to tailor product design and communication.

[Insert Table I about here]

In Table I we list the leading data technologies, as of December 2023, for our sample of U.S. retail firms. Google Analytics accounts for the lion's share of adoption, with around 95% of retail firms installing it.[10] Other leading plugins include RapLeaf, which allows for detailed user tracking, as well as LiveRamp, which integrates big data across platforms. All technologies in Table I are designed to extract signals from web traffic visits. These features are not limited to the most commonly used data analytics tools by firms in our sample. Appendix Figure C.1 displays the word cloud extracted from the description of all these data technologies. Frequent terms include for instance "customer", "data", "analytics" and "tracking", consistent with their focus on collecting and analyzing customer data.

We merge BuiltWith data with CRSP-Compustat merged dataset (annual data) using firms' websites. We start with all U.S. firms (fic=USA, loc=USA, and curcd=USA) with

[10] Appendix D shows that all the main results are not driven solely by Google Analytics (Table D.5).

common stocks (shrcd 10 or 11) trading on AMEX, NYSE, or Nasdaq (exchcd equal to 1, 2, or 3) for the 1980-2023 period. We keep observations that have (i) non-missing information on total assets, sales, capital expenditures, property, plant and equipment (PP&E), and (ii) total assets and sales greater than $1 million. From the 13,645 distinct firms in this set, 8,815 have a website that is (ultimately) covered by BuiltWith.

In our main test, we concentrate on retail-oriented firms because they are the most likely to extract insights about their growth opportunities from data generated on their website. We identify these firms using the text-based approach of Frésard et al. (2019) linking product vocabularies from the Bureau of Economic Analysis (BEA) input-output tables to firms' 10-K product descriptions. The input-output tables record the purchases of "commodity" outputs (any good or service) by producing "industries" or "end-users" (including exports, personal consumption, or the government). Each commodity comprises a textual description (between 1 to 25 distinct words) that summarizes the nature of the good or service provided. Frésard et al. (2019) identify the specific words associated with the goods or services that are directly purchased by end-users (together with their economic intensity). Contrasting these words with those extracted from firms' 10-K business descriptions, they compute (for each firm-year) a score that measures the intensity with which each firm's products and service flows towards end-users. We define retail-oriented firms as those firms that have a score above 50% (at least once throughout the sample), and discard firms with a score below 50%.[11]

[Insert Figure I about here]

For each firm-year, we define the binary variable $\textsc{Data}_{i,t}$ as equal to one if firm i has installed at least one data technologies on its website in year t, and zero otherwise. This variable is the empirical equivalent to δ_{τ_χ} in eq. (9), and we use it to capture an increase in the precision of the manager's signal (τ_χ). Figure I plots the evolution of the average value of $\textsc{Data}_{i,t}$ by year, tracking the adoption of data technologies. The adoption of data technologies started in 2005, and the adoption rate reached 40% in 2010 and 80% in 2015.

[11] We show later that we obtain similar results if we define retail-oriented firms based on SIC codes associated with "retail" industries.

In 2023, 96% of retail-oriented firms have data technologies installed on their website. Our empirical strategy exploits the heterogeneous adoption across firms and time.

B Non-fundamental shocks to stock prices

We identify non-fundamental variation in stock prices (the empirical analog of the instrument z in the model) using the trading generated by the reinvestment of (cash) dividends received by mutual funds from their portfolio firms. Schmickler and Tremacoldi-Rossi (2023), Chen (2024) and van der Beck (2024) shows that when funds reinvest cash dividends, they generate non-fundamental buying pressure for all stocks in their portfolio. The intuition behind this dividend-induced pressure can be illustrated with the following simple example. Consider a mutual fund holding an equally weighted portfolio consisting of three stocks, A, B, and C. At time t, stock A pays a cash dividend, while B and C do not. The fund receives cash from A and reinvests it proportionally across all holdings, increasing its positions in A, B, and C equally. Although A's dividend payment might reveal information about its fundamentals (e.g., its future performance), the increased demand for B and C is plausibly unrelated to their fundamentals. We rely on this demand pressure to capture variation in stock price that is plausibly unrelated to corporate managers' information about their firm's fundamentals ($\tilde{\theta}$).

Following Schmickler and Tremacoldi-Rossi (2023), we use quarterly holdings of mutual funds from Thomson Reuters and retain all U.S. funds between 1980 and 2023. We consider mutual funds because most have specific benchmarks and are thus more likely to reinvest the cash received from dividend payments in their existing holdings. We obtain mutual funds' asset under management (AuM) from the CRSP Mutual Funds Database, and cash dividends paid by stock from CRSP. Let $d_{i,q}$ denote the cash dividend per share paid by stock i in quarter q, and $S_{j,q^{ex}}(i)$ the number of shares of stock i held by mutual fund j, as of the ex-dividend date of the dividend paid in quarter q.[12] For each mutual fund j, we

[12] Stock dividends are paid to whom is holding the stock at market opening of the ex-dividend date, while the actual payment will be made several days later. In our sample, the dividend payment is made 43 days after the ex-date, on average. We denote the dividend payment quarter as q, and the associated ex-date q^{ex}.

construct the total cash dividend received in quarter q as

$$C_{j,q} = \sum_{i=1}^{N} d_{i,q} \cdot S_{j,q^{ex}}(i), \tag{10}$$

where N is the total number of stocks held by mutual fund j in quarter q. $C_{j,q}$ represents the dollar amount of dividends received by mutual fund j in quarter q. Then, we construct the *hypothetical* reinvestment purchase made by mutual fund j in stock i (in its portfolio) as $C_{j,q} \cdot w_{j,q}(i)$, where $w_{j,q}(i)$ denotes the weight of stock i in mutual fund's j portfolio in quarter q. Finally, we define $HDIT_{i,q}$ as the total *hypothetical dividend-induced trades* made by all mutual funds holding stock i, normalized by total assets ($A_{i,t}$):

$$HDIT_{i,q} = \frac{\sum_{j=1}^{J} C_{j,q} \cdot w_{j,q}(i)}{A_{i,t}}, \tag{11}$$

where J is the total number of mutual funds. To avoid including fundamental information coming from a firm's own dividend, we follow Schmickler and Tremacoldi-Rossi (2023) and construct $HDIT$ for stock i using dividend payments to all funds, but excluding dividends from stock i and from all firms in the same two-digit SIC industry, since they may reveal correlated fundamental information. We obtain $HDIT_{i,t}$ for each stock-year by taking the average of $HDIT_{i,q}$ across quarters.[13]

Mimicking the approach of Edmans et al. (2012), $HDIT$ aggregates the hypothetical number of shares mutual funds purchase in response to receiving cash dividends, as if mutual funds were reinvesting all of the cash inflow from dividends in proportion to their existing portfolio weights. We do not use *actual* trades because they likely reflect fund managers' private information about firms' fundamentals. Instead, we assume that in response to inflows from dividends, mutual funds use all cash inflow to expand their existing holdings proportionally. We confirm in Appendix B that mutual funds do reinvest dividends proportionally to their existing holdings. Furthermore, we find no evidence that mutual funds save part of the cash inflow from dividends as cash holdings (Appendix Table B.1).

By construction, the variation of $HDIT$ for a given firm (in a given year) is entirely driven by the dividends paid by unrelated firms and is thus unlikely to reflect changes in

[13] We obtain similar results when we take the sum of $HDIT_{i,q}$ across quarters.

investors' views about the fundamentals of that firm. In addition, it does not depend on choices made by fund managers about which stocks to buy (or sell) to invest the cash received from dividends. As such, $HDIT$ is a plausible proxy for non-fundamental shocks to stock prices. In support of this claim, Figure II displays the relationship between mutual fund hypothetical dividend-induced trades and stock returns around the (cash) dividend payment. We define an "event" for stock i in month m of year t when $HDIT_{i,m,t}$ is above the tenth percentile of the sample distribution of $HDIT$ in year t. We then estimate a regression of the monthly abnormal returns for stocks affected by these events on event-month dummy variables, and plot the associated coefficients (i.e., the average abnormal return) around the event. Figure II reveals that, on average, stock affected by the dividend-induced buying pressure experience a 2% increase in their price in the month following the event. In line with the non-fundamental nature of this price increase, the effect is temporary and reverts completerly after three months.

[Insert Figures II and C.2 about here]

We verify that $HDIT$ is associated with meaningful non-fundamental price variation throughout our full sample period, rather than only during specific episodes. Appendix Figure C.2 plots the relationship between mutual fund hypothetical dividend-induced trades and stock prices by decade–1980s, 1990s, 2000s, and 2010s– and shows that $HDIT$ triggers transitory price pressures in each period. This evidence confirms that our identification rests on a persistent and robust first-stage across the full 1980–2023 sample. We further show that $HDIT$ is unlikely to capture economy-wide or industry-specific characteristics. Figure III displays the average value of $HDIT$ across firms for each year in our sample and across SIC 2-digits industries. We observe no obvious clustering in any particular industry.

Finally, we verify that $HDIT$ does not explain firms' adoption decisions to adopt data technologies. An important condition for the identification of the replacement effect is that our instrument (z in the model) is not correlated with the (endogenous) decision to adopt data technologies. A priori, there is no obvious reason why the reinvestment of dividends paid by unrelated firms held in the same institutional portfolios should influence a firm's decision to adopt. We nevertheless formally test this condition by regressing the adoption indicator on contemporaneous and lagged values of $HDIT$, including firm and year fixed effects. Table D.1 shows that variation in $HDIT$ does not predict whether or when firms adopt data-related tools. This result supports the condition underlying our identification: non-fundamental shocks to stock prices generated by dividend reinvestments are unrelated to the changes in the precision of managers' internal signals, τ_χ. Therefore, $HDIT$ isolates variation in the non-fundamental component of prices that appears orthogonal to managers' internal information.

[Insert Table D.1 about here]

C Empirical specification

We specify the empirical counterpart of the model's difference-in-differences eq.(9) as follows:

$$I_{i,t} = a_1 Q_{i,t-1} + a_2 \text{DATA}_{i,t-1} + a_3 Q_{i,t-1} \times \text{DATA}_{i,t-1} + \Gamma X_{i,t-1} + \gamma_i + \eta_t + \varepsilon_{i,t}, \tag{12}$$

where $I_{i,t}$ is the ratio of capital expenditure scaled by lagged fixed assets (property, plant, and equipment) in year t for firm i. $Q_{i,t-1}$ is the (normalized) stock price defined as the market value of firm i (total assets minus book value of equity plus the market value of equity) divided by its total assets. The binary variable $\text{DATA}_{i,t-1}$ captures the adoption of data technologies. The vector $X_{i,t-1}$ includes standard control variables in investment models and variables capturing fundamental information about investment opportunities known to managers at the time they decide on investment. Specifically, we control for the one-year lagged values of the natural logarithm of assets ("firm size") and contemporaneous cash flows. In addition, we control for time-invariant firm heterogeneity by including firm fixed effects (γ_i), and aggregate fluctuations by including year fixed effects (η_t). Standard errors are clustered in two ways, by firm and by year. Appendix A describes the construction of all variables. We winsorize all ratios at 1% in each tail to reduce the effect of outliers. Table II presents the descriptive statistics. They are in line with related research.

[Insert Table II about here]

We estimate eq.(12) by instrumental variables, using $HDIT_{i,t-1}$ to instrument $Q_{i,t-1}$ and the interaction between $HDIT_{i,t-1}$ and $\text{DATA}_{i,t-1}$ to instrument $Q_{i,t-1} \times \text{DATA}_{i,t-1}$. The coefficient of interest in equation (12) is a_3. It is the empirical equivalent of b_3 in equation (9). It measures whether the sensitivity of firms' investment to a non-fundamental component of their stock price (i.e., the *instrumented* component of Q) changes after they adopt data technologies. The replacement effect predicts that this "investment-to-noise" sensitivity declines, that is, $a_3 < 0$. Although the adoption of data technologies starts in 2005, we estimate eq.(12) over the period 1980-2023, following Gardner et al. (2024), who recommend long pre-treatment periods in staggered adoption settings.[14]

Similar to Dessaint et al. (2019) our test posits that managers do not perfectly observe the non-fundamental variation in their stock prices triggered by dividend-induced pressure. The econometrician can measure these changes ex post (possibly, a long time after they happened). However, our identification requires that these changes are not perfectly observed by managers when they happen, that is, managers cannot completely filter out the noise in their stock price. If they could, managers' beliefs (and investment decisions) would *not* be influenced by the effect $HDIT$ on stock prices (i.e., α_1 and α_3 in eq.(12) should be equal to 0). We show later (see Section V.A) that, to some extent, managers detect part of the noise in their stock price. Yet, importantly, this detection ability is not affected by the adoption of data technologies.

D Data technologies and managers' information: Validation

The replacement effect rests on the hypothesis that the adoption of data technologies increases the precision of managers' internal signals about their fundamentals (τ_χ). Unfortunately, since we do not observe the precision of managers' signals, we cannot directly test this hypothesis. We provide however two pieces of evidence that support its validity, exploiting managers' guidance.

Managers issue guidance (i.e., forecasts) about their future sales and earnings on a voluntary basis, and are more likely to issue these forecasts when their internal signals differ from those of analysts or other market participants (Coller and Yohn, 1997). We posit that If data technologies improve the precision of managers' internal signal (τ_χ), they should issue guidance more frequently, and their forecasts should be more precise. We collect one-year

[14] We show later that our conclusions are similar if we estimate eq.(12) on a sample that starts in 1990.

ahead sales forecasts issued by managers from I/B/E/S Guidance from 1980 to 2023. We use one-year sales because managers most often release forecasts at that horizon. When forecasts are issued as a range instead of a point forecast, we use the upper bound of the range since it better reflects managerial expectations than the midpoint (Ciconte et al. (2014); Call et al. (2024)). We then define a binary variable equal to one if a firm releases a sales forecasts in a given year, and zero otherwise. We measure the precision of each forecast computing the mean squared forecast error with the respect to the actual realization of sales, normalized by lagged sales.

[Insert Table III about here]

Consistent with τ_χ increasing with data technologies, Table III shows that managers are more likely to issue sales forecasts after they adopt these technologies. Columns (1) and (2) indicate that, after adoption, firms are about 10% more likely to release forecasts about their sales. Columns (3) and (4) further show that the adoption of data technologies is associated with a significant decline in managers' forecast errors. Taken together, results Table III support our hypothesis that data technologies are associated with improved managers' internal information (about future demand).

IV The replacement effect

A Baseline results

To facilitate interpretation, in estimating Eq.(12), we scale all independent variables by their within-firm sample standard deviation prior to estimation, such that coefficients represent the estimated change of investment in response to a one-standard deviation change in each independent variable. We report the estimates of Eq.(12) in Table IV.[15] Estimated by 2SLS, the coefficient on Q is 0.023, with a t-statistic of 7.66. This estimate confirms that, overall,

[15]Because we include firm fixed effect in Eq.(12), the source of variation used to estimate the coefficient is within-firm over time. For that reason, we standardize the independent variables by their within-firm standard deviation Liu and Winegar (2025). In this vein, in the 2SLS estimations, we standardize the *instrumented* variables (i.e., Q and $Q\times$Data) by their within-firm standard deviation. This choice of standardization explains why the standard errors in the 2SLS estimates (column (2)) are smaller than their OLS counterparts.

firms' investment is sensitive to the non-fundamental component of their stock price (e.g., Dessaint et al. (2019)). A one-standard deviation increase in this component is associated with a 2.3 percentage point increase in investment. Compared to the OLS estimation in column (3), this magnitude represents about a quarter of that of the non-instrumented coefficient on Q (0.089 with a t-statistic of 14.83), which captures both the fundamental and non-fundamental component of firms' stock price.

[Insert Table IV about here]

Consistent with the replacement effect (Prediction 1), the coefficient on $Q\times$DATA estimated by 2SLS (i.e., α_3) is negative and statistically different from zero. The point estimate in column (2) is -0.012, with a t-statistic of 4. As predicted, the use of data-related technologies renders firms' investment *less* sensitive to the non-fundamental component of their stock price. After adopting these technologies, a one-standard deviation increase in the non-fundamental component of firms' stock price is associated with a 1.1 (i.e., 0.023-0.012) percentage point increase in investment, or about half the estimated magnitude prior to adoption (compared to non-adopting firms). The results are qualitatively similar (albeit smaller) in column (1) when we exclude the control variables.

Columns (4) and (5) presents the first-stage estimates. Confirming the results presented in Section III.B, and in line with Schmickler and Tremacoldi-Rossi (2023) $HDIT$ is positively and significantly related to Q. The magnitude of this effect is large, since a one-standard deviation increase in $HDIT$ is associated with a 0.652 increase in Q. This increase corresponds to about half of the within-firm standard-deviation of Q in the sample. While the strength of the instrument $HDIT$ is slightly lower after firms adopt data-related tools (i.e., the coefficient on $HDIT\times$DATA is negative), it remains substantial, alleviating weak instrument concerns.[16] This conclusion is further confirmed in column (5) as we observe a strong positive relation between $Q\times$DATA and $HDIT\times$DATA, and looking at the (Kleibergen–Paap) F-statistic of 36.71.

[16]Note that this variation in the instrument strength in the first-stages has no effect on the magnitudes estimated in the second-stage because we standardize the instrumented Q and $Q\times$DATA by their within-firm standard deviation.

[Insert Figure IV about here]

Figure IV plots the evolution of the sensitivity of firms' investment to the non-fundamental component of their stock price prior and after the adoption of data-related tools, compared to non-adopting firms. We set the year prior to adoption as the (excluded) reference period, and replace the binary variable DATA in Eq.(12) by event-time binary variables identifying the years surrounding firms' data-tools' adoption (e.g., $\text{DATA}_{i,t}^{-2} = 1$ if firm i will adopt in two years, and 0 otherwise). In the years preceding adoption, we find no difference in the sensitivity of investment to the non-fundamental component of stock prices between adopting and non-adopting firms. Confirming that our results stem from the adoption of data-related tools, the investment of adopting firms become less sensitive to non-fundamental changes in their stock price in the third year following adoption. While supporting the replacement effect, Figure IV suggests that this effect is not immediate, perhaps because it takes time for firms to harness valuable insights from internal data (i.e., improving the precision of the internal signal, τ_χ) after adopting the new tools.

We conduct further robustness checks. Given firms' staggered adoption of data-related tools, we estimate Eq.(12) using a stacked regression approach. Following Baker et al. (2022), for each year t, we create a "cohort-t" sample, retaining only firms adopting data-related tools in year t as well as firms that never adopt these tools (in any year t) as control firms. We then estimate Eq.(12) on the sample of stacked cohort samples, replacing firm fixed effects by firm×cohort fixed effects to account for the duplicity of control firms in the stacked sample. Table V presents the results, and indicates that changing the estimation approach does not modify our conclusion. We continue to observe a reduction in the sensitivity of firms' investment to the non-fundamental component of their stock price after adoption data-related tools. Further, the economic magnitude of the estimated replacement effect matches that obtained in Table IV.

[Insert Tables V and VI about here]

In addition, we show in Table VI that our conclusions are robust to various modifications of our main specification. In particular, they continue to hold if we control for common variation within SIC 2-digits industries (with industry×year fixed effects), define retail sector firms based on their SIC codes as opposed to the text-based "end users" label, or start the sample in 1990. The results also remain unchanged when we interact the control variables with DATA to capture potential indirect effects of adopting data-related tools.

We further report in the Appendix that the main results in Table IV are robust to alternative measurements of several key variables in our tests. In particular, we obtain similar results if we replace Q by the measure of Total Q that scales firm value by a measure of "total" capital that accounts for intangibles (Table D.2), following Peters and Taylor (2017). Results are unchanged if we scale capital expenditures by total assets instead of fixed assets (Table D.3), or if we consider investments in tangible and intangible assets (Table D.4), using the total investment of Peters and Taylor (2017). Notably, we find no replacement effect for purely intangible investments, that is, this type of investments (e.g., R&D) does not become less sensitive to the non-fundamental component of firms' stock price after adopting data-related technologies. We also show that our conclusions continue to hold when we exclude Google Analytics, the most adopted tool, from the list of data-related technologies (Table D.5).

B Placebo tests: Non-data tools and non-retail firms

To further show that our main results stem from an improvement in the quality of managers' internal signals (τ_χ) due to their use of data-related tools, we perform two falsification tests. First, we replace the adoption of data-related tools with that of tools unrelated to the collection and analysis of customer data. From BuiltWith, we gather data on the adoption of 823 distinct tools related to digital agency (e.g., Wix Studio), registrar (e.g., technologies included in the hosting provider's registration), copyright (e.g., Copyscape), shipping (e.g., Shopify shipping). We restrict our attention to these four categories because they exhibit adoption rates that are similar to data-related tools, and the timing of their adoption has little overlap with that of data-related tools (i.e., <10% overlap). This ensures that the results are (i) unlikely driven by a lack of power, and (ii) not capturing the simultaneous effect of data-related tools.

[Insert Tables VII and VIII about here]

We re-estimate Eq.(12), replacing $\textsc{Data}_{i,t}$ with the adoption of each of these "placebo" technologies and present the results in Table VII. Overall, the adoption of these tools does not modify the sensitivity of firms' investment to the non-fundamental component of their stock price. The coefficients on the interaction between Q and the adoption of these tools are insignificant in three specifications. It is negative and significant for registrar tools, but the estimated magnitude is small. These results dispel the potential concerns that our results are driven by overall changes in firms' web strategy, or the adoption of any type of tools.

Second, we alter our sample construction to include only non-retail firms, defined as those with a value of "end-users" below 50%. We posit that these firms should benefit less from the adoption of data-related tools installed in their website. Since their customers are not directly retail consumers, these firms should gather less information about their growth opportunities from the clients' activities on their website activities. In line with this conjecture, Table VIII shows that the coefficient on the interaction between Q and $\textsc{Data}$ is again negative but not statistically significant. Thus, the sensitivity of these firms investment to the non-fundamental component of their stock prices does not change after adopting data-related tools.

C Regulatory limitations of data access

The replacement effect stems from firms' ability to collect and analyze customers' data to obtain more precise insights about their growth opportunities (τ_χ). Thus, this effect should weaken when, after adopting data-related tools, firms face limitations in the amount and type of data they can collect from website traffic. To test this prediction, we exploit the limitations imposed by the California Consumer Privacy Act (CCPA). This regulation was enacted in 2020 with the objective of protecting Californian customers' privacy. It requires that firms doing business in California disclose to customers what personal data they collect, how they use it, and gives them the option to prohibit the sale of their data to third-parties or request their deletion.[17]

[17]CCPA applies to any firm with business in California that meets at least one of three criteria: (i) annual (gross) revenues of $25 million, (ii) collecting personal information of 50,000 customers or more, or (iii)

To assess whether the replacement effect weakens for firms subject to data limitations, we define $\text{CCPA}_{i,t}$, a binary variable that is equal to one if firm i is incorporated in California and year t is greater than 2019, and zero otherwise.[18] We then augment Eq.(12), interacting all variables with CCPA. In the spirit of a partial treatment "reversal", the coefficient of interest is that on the triple interaction $Q \times \text{DATA} \times \text{CCPA}$. Estimated by 2SLS (i.e., instrumenting Q with $HDIT$), it captures how the sensitivity of investment to the non-fundamental component of stock price changes when firms' using data-related tools face constraints in their data collection and usage.[19] We predict a positive coefficient, consistent with firms relying more on their stock price for information when the quality of their internal information deteriorates due to data limitations.

[Insert Table IX about here]

The results in Table IX support this prediction. Following the enactment of the CCPA, firms that use data-related tools and are subject to the CCPA experience a significant increase in their reliance on their stock price. In column (1), the coefficient on $Q \times \text{DATA} \times \text{CCPA}$ is positive and significant. In terms of magnitude, the estimates in Table IX indicate that, after adopting data-related tools, the sensitivity of firms' investment to the non-fundamental component of their stock price decreases by about 1.6 percentage points, in line with our main results. However, the third row of Table IX indicates that data-usage restrictions following the CCPA's enforcement in 2020 offset this effect by 0.2 percentage points for Californian retail firms using data-related tools. The size of this effect is economically reasonable and consistent with the magnitude of the replacement effect we document.

V Alternative channels and implications.

In this section, we consider various alternative channels that could explain why firms' investment becomes less sensitive to the non-fundamental component of their stock price after

---

having half of sales from selling customers' data.

[18]The law was effective from June 2018, but firms had a "grace period" until July 2020.

[19]In estimating this specification, we estimate four first-stages, for Q, $Q \times \text{DATA}$, $Q \times \text{CCP}$, and $Q \times \text{DATA} \times \text{CCPA}$.

they adopt data-related tools. Overall, we find no consistent empirical support for these channels. We also discuss the implications of the replacement effects in the aggregate.

A Learning about the noise in stock prices

First, we consider the possibility that the reduced sensitivity of firms' investment to the non-fundamental component of their stock price arises not because of the replacement effect, but because the adoption of data-related tools improves managers' ability to detect the noise in their stock price. In that scenario, managers do *not* change their reliance on their stock price for information about fundamentals ($\frac{\partial \beta}{\partial \tau_\chi} = 0$), but can better distinguish the fundamental from the non-fundament component of their stock price. In that case, their investment should be less sensitive to the noise in their stock price, that is, less "complying" with the instrument.

To assess this possibility, we examine the trading behavior of corporate insiders. To the extent that managers can identify the non-fundamental upward price pressure generated by hypothetical dividend-induced trades, they should opportunistically and profitably trade against it (e.g., Khan et al. (2012)). We collect insider trades from the Thomson Reuters insider filings database, and focus on all officers with decision-making authority over the operations of the firm, all board members and beneficial owners of more than 10% of a company's stock. Following Cohen et al. (2012) we consider opportunistic (or non-routine) trades and measures opportunistic insiders' net sales in a given year either as insiders' annual (dollar) sales minus purchases, divided by their firm market capitalization (at the end of the year) or as the net number of shares sold.[20]

[Insert Table X about here]

Table X indicates that the adoption of data-related tools does not affect intensity of insiders' opportunistic trades. The coefficients on DATA are not significant in columns (1)

[20] Similar to Cohen et al. (2012) we require an insider to make at least one trade in each of the three preceding years to define her as either an opportunistic or a routine trader. We define then a routine trader as an insider who placed a trade in the same calendar month for at least three consecutive years. Opportunistic traders are defined as everyone else. We set insiders net sales to zero when firm do not have opportunistic trades in a given year.

through (8). Columns (3) and (4) further indicates that insiders' net selling activity is sensitive to the non-fundamental component of their stock price. Consistent with existing research examining insider trading around downward price pressure generated by large mutual funds outflows (e.g., Khan et al. (2012) or Ali et al. (2011)), insiders also appear to detect part of the the non-fundamental component on their stock price: they sell significantly more shares when their stock price increases due to the buying pressure generated by hypothetical dividend-induced trades. However, the sensitivity of insiders' net sales to the non-fundamental component of their stock price does *not* change after they adopt data-related tools (columns (3) and (7)). For this reason, the negative effect of data-related tools on firms' investment-to-noise sensitivity is unlikely to stem from managers' increased ability to detect the non-fundamental component of their stock price analyzing their customers' data.

B Fundamental uncertainty and the precision of the price signal

Second, recall that in the model firms' reliance on their stock price is given by $\beta = \frac{\tau_u}{\tau_\theta+\tau_\chi+\tau_u}$ (see Eq.(4)). So far, we have attributed the estimated decrease in β to an improvement in the precision of managers' internal information (τ_χ) obtained from the use of data-related tools – the replacement effect. Yet, Eq.(4) indicates that a reduction of β could arise even in the absence of a replacement effect, that is, if $\frac{\partial\beta}{\partial\tau_\chi} = 0$. Indeed, Eq.(4) indicates that a reduction of β could be observed if the adoption of data-related tools coincides with a reduction of the unconditional uncertainty of firms' fundamentals ($1/\tau_\theta = \sigma^2_\theta$) and/or a decrease in the precision of firms' stock price as a signal for managers about their fundamentals (τ_u).

To assess this possibility, we study whether proxies for the the uncertainty of firms' fundamentals and the informativeness of their stock prices change around the adoption of data-related tools. As proxies for $1/\tau_\theta$, we consider the volatility of firms sales, cash flows, and earnings (scaled by assets), computed using three-years rolling windows. As a proxy for τ_u, we consider the jump-ratio measure developed by Weller (2018), which compares the amount of a stock's price movement caused by the announcement of earnings to its total price movement during the earnings period. A higher jump-ratio indicates prices adjust abruptly when earnings are released (they are less informative), while a lower ratio indicates information is incorporated into prices more smoothly and continuously (they are more informative). We also proxy τ_u by measuring the ability of firms' stock price to predict their future earnings, following Bai et al. (2016) (Appendix Table D.6).

[Insert Tables XI about here]

Table XI shows that, all else equal, the adoption of data-related tools is not associated with a lower variance of firms' fundamentals. The coefficients on DATA are insignificant for the sales and cash flow volatility, and marginally significant but positive for earnings volatility. Hence, whereas data-related tools improve managers' signals about their firms' fundamentals (see Table III), they do not alter the unconditional uncertainty of the fundamentals (σ_θ^2). This channel is thus unlikely to explain our main result. In contrast, the adoption of data-related tools is associated with a significant increase in firms' jump-ratio, that is, a *deterioration* of their price informativeness. We find a similar deterioration in column (5) when we look at stock prices' ability to predict firms' next year earnings. Notably, the reduction of price informativeness occurs despite the increased incidence of guidance provided by adopting firms (see Table III), perhaps reflecting a crowding out of private information production by market participants (e.g., Morris and Shin (2002)).

To evaluate whether the effect data-related tools on firms' investment-to-noise sensitivity solely stems from the reduction of τ_u, we focus on firms that do not experience a reduction of τ_u after they adopt data-related tools. We identify those firms computing the contribution of each firm to the average effect of data-related tools on the jump-ratio.[21] We then estimate our main specification on the subsample of firms for which the treatment effect on price informativeness is positive or zero. Table D.7 confirms that our results do not originate solely from a deterioration in price informativeness. Even among firms experiencing an *improvement* in price informativeness (columns (1) and (2)), column (4) shows that the coefficient on $Q\times$DATA estimated by 2SLS remains negative and statistically different from zero. In other words, consistent with the replacement effect, after adoption of data-related

[21] The contribution of each firm to the average treatment effect on jump-ratio are obtained using the influence function method Cook and Weisberg (1982). We estimate the average contribution of each firm (DFBETA) and isolate those firms above the median, i.e., for which the effect of data-related tools on jump-ratio is more negative.

tools firms' investment are less sensitive to non-fundamental variation of their stock price.

C Financing channel

Third, previous work shows that the investment of financially constrained firms is more sensitive to the non-fundamental component of their stock price because their price determines how cheaply they can raise funds (e.g., Baker et al. (2003) or Hoberg and Maksimovic (2015)). An increase in stock price lowers the cost of funds, leading to more investment (the financing channel). Thus, the decrease in firms' investment-to-noise sensitivity following their adoption of data-related tools could reflect reductions of firms' financial constraints if adoption lowers the cost and availability of financing. It might if, for instance, the use of data-related tools lowers information asymmetries as managers provide more credible disclosures (see Table III) or improves the predictability of earnings for firms' outsiders.

To assess whether our results could reflect such a financing channel, we test whether firms' cost of financing drops after they adopt data-related tools. We use the measures of financing constraints developed by Hoberg and Maksimovic (2015) and extended by Linn and Weagley (2023), which are based on textual analysis of firms' 10-Ks. In particular, we use their scores of the intensity of debt-market and equity-market constraints, where a higher score indicates more binding constraints. These scores are available for all firms since 1997, including firms that are not seeking access to financing, and as such, are free of any selection bias. In addition, we rely on managers' perceived cost of capital and hurdle rate developed by Gormsen and Huber (2025) collected from firms earnings call transcripts, available since 2002.

[Insert Tables XII about here]

Overall, we find no evidence supporting a financing channel. Table XII indicates that the access to external finance does not improve after firms adopt data-related tools. While we find no change in the intensity of debt-market constraints, column (2) indicates that adoption is associated with *higher* equity constraint. The coefficient on DATA is positive and significant with the measure of Linn and Weagley (2023), in line with the deterioration of price informativeness (i.e., more information asymmetry) reported in Table XI. The last two columns show no significant changes in managers' perceived cost of capital and hurdle rate after adopting data-related tools.

D Market power channel

Fourth, we examine whether the reduction of firms' investment-to-noise sensitivity could result from increased market power obtained through the use of data-related tools (the market power channel). Recent work suggests that the use of data enables firms to gain market shares and power, and affects product market dynamics (e.g., Babina et al. (2024b). At the same time, existing research argues and shows that market power renders firms' investment less sensitive to their stock prices (e.g.. Gutiérrez and Philippon (2017). This reduced sensitivity could occur because of a lower need for firms with market power to be responsive to demand shocks (reflected in the stock prices) or investment opportunities (because capacity expansion depresses product prices), or because of lower cost of financing due to more stable and predictable cash flows.

[Insert Tables XIII about here]

Inconsistent with this channel, Table XIII shows that the adoption of data-related tools does not lead to increased market power. In columns (1) and (2), we find no effect on firms' market shares (computed at the SIC3 level) or price-cost margins (defined as sales minus cogs, divided by sales, following Nickell (1996)). In column (3), we follow De Loecker et al. (2020) and focus on firms' markups. Contrary to the market power channel, we observe a significant decrease in markups after firms adopt data-related tools.

E Implications for long-run trends in investment

Taken together, our tests provide strong empirical support for the replacement effect: the adoption of data-related technologies improves the precision of firms' internal information and reduces their reliance on their stock price as an external source of information. This results is unlikely to reflect alternative channels. Figure I indicates that, since the mid-2000s, an increasing number of firms have adopted these tools. Thus, our findings imply that, over time, corporate investment should have become less sensitive to stock prices. In this section, we provide suggestive evidence that the replacement effect might partly explain the increasing disconnect between firms' investment and Q documented by Gutiérrez and Philippon (2017) and Alexander and Eberly (2018).

Specifically, we obtain the evolution of the investment-to-price sensitivity by estimating the following specification for each SIC 2-digits industry j:

$$I_{i,t} = d_1^j Q_{i,t-1} + d_2^j Q_{i,t-1} \times (t - t_0) + \Gamma^j X_{i,t-1} + \gamma_i + \eta_t + \varepsilon_{i,t}, \tag{13}$$

where i and t represent firms (operating in indusry j) and years, and t_0 is 1985, such that $t - t_0$ represents an incremental year count after that year (i.e., the period 1980-1985 are benchmark years). Similar to Eq.(12), we include firm and year fixed effect, and control for firms' size and cash flows. The coefficients d_2^j capture the trend in the investment-to-price for each industry. To relate these trends to the adoption of data-related tools across industries, we calculate for each firm the fraction of years for which DATA = 1, and average these fraction by industry.

[Insert Figure V about here]

Figure V plots the estimated trends (the t-statistics of d_2^j) in investment-to-price sensitivity for each industry against their intensity of data-tools adoption. Confirming the growing disconnect between investment and Q documented by Gutiérrez and Philippon (2017) and Alexander and Eberly (2018), the estimates of d_2^j are negative for abot 70% of industries (49 out of 71 industries). Notably, the estimated trends are negatively and significantly (at 10%) related to the intensity of data adoption across industries (coefficient of -8.03 with a t-statistic of 1.90). Overall, the results in Figure V are consistent with the prediction of the replacement effect that, over time, firms using data-related tools rely less on their stock price to decide on investment. Arguably, they could also be consistent with various alternative explanations. A more systematic analysis of the determinants of the aggregate decline in the sensitivity of firms' investment to their stock price is beyond the scope of this paper, but represents an interesting avenue for future research.

VI Conclusions

We show that the diffusion of data technologies changes how firms learn from financial markets. When managers gain more precise internal signals from customer-generated data, they rely less on their stock price to guide real investment. Using non-fundamental variation in prices induced by mutual funds' dividend reinvestment, we document a sizable decline in investment-to-price sensitivity after firms adopt technologies that help them extract information about potential customers. The effect emerges only after adoption of these data technology, and is concentrated among retail-oriented firms—those for which customer data are likely most informative about future product demand. Consistent with data technologies replacing the informational role of financial markets, we find that the effect partially reverts when data-privacy restrictions, such as the California Consumer Privacy Act, constrain firms' ability to collect customer information.

Taken together, our results suggest that improvements in managerial information reduce the role of the stock market as a source of guidance for real decisions. This "replacement effect" does not imply worse investment; rather, it indicates that the source of information used in capital allocation is shifting from market signals towards firms' own datasets. At an aggregate level, this shift may help explain the gradual weakening of the link between investment and market valuations observed in recent decades. These findings open several further questions. For example, data technologies might similarly replace—or instead complement—market signals for other economic actors such as regulators, creditors, customers, or analysts. The answer to this question is not trivial, as managers' *internal* information set is very different from the one of other economic agents.

References

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

Acemoglu, D., Anderson, G. W., Beede, D. N., Buffington, C., Childress, E. E., Dinlersoz, E., Foster, L. S., Goldschlag, N., Haltiwanger, J. C., Kroff, Z., Restrepo, P., and Zolas, N. (2022). Automation and the workforce: A firm-level view from the 2019 annual business survey. Working Paper 30659, National Bureau of Economic Research.

Alekseeva, L., Azar, J., Gine, M., Samila, S., and Taska, B. (2021). The demand for ai skills in the labor market. Labour Economics (Forthcoming).

Alexander, L. and Eberly, J. (2018). Investment hollowing out. *IMF Economic Review*, 66(1):pp. 5–30.

Ali, A., Wei, K., and Zhou, Y. (2011). Insider trading and option grant timing in response to fire sales (and purchases) of stocks by mutual funds. *Journal of Accounting Research*, 49(3):595–632.

Amazon (2025). Privacy cookie policy. Amazon.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2023). *Firm Investments in Artificial Intelligence Technologies and Changes in Workforce Composition*, pages 75–117. University of Chicago Press.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024a). Artificial intelligence and firms' systematic risk. Working Paper, University of Maryland.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024b). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745.

Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122:625–654.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Baker, M., Stein, J., and Wurgler, J. (2003). When does the market matter? stock prices and the investment of equity-dependent firms. *Quarterly Journal of Economics*, 3(118):969–1005.

Baley, I. and Veldkamp, L. (2025). *The Data Economy: Tools and Applications*. Princeton University Press.

Blankespoor, E., Hendricks, B., Piotroski, J., and Synn, C. (2022). Real-time revenue and firm disclosure. *Review of Accounting Studies*, 27:1079–1116.

Bond, P., Edmans, A., and Goldstein, I. (2012). The real effects of financial markets. *Annual Reviews of Financial Economics*, 4:339–360.

Bonelli, M. (2025). Data-driven investors. *The Review of Financial Studies*, page hhaf078.

Bonelli, M. and Foucault, T. (2025). Does big data devalue traditional expertise? evidence from active funds. Working Paper, HEC Paris.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative ai at work. Working Paper 31161, National Bureau of Economic Research.

Brynjolfsson, E. and McElheran, K. (2016). The rapid adoption of data-driven decision-making. *The American Economic Review*, 106(5):133–139.

Call, A. C., Hribar, P., Skinner, D. J., and Volant, D. (2024). Corporate managers' perspectives on forward-looking guidance: Survey evidence. *Journal of Accounting and Economics*.

Cao, S., Jiang, W., Yang, B., and Zhang, A. L. (2023). How to talk when a machine is listening: Corporate disclosure in the age of ai. *The Review of Financial Studies*, 36(9):3603–3642.

Chen, H. (2024). Cash induced demand. *Journal of Financial and Quantitative Analysis*, (53):195–220.

Chen, Q., Goldstein, I., and Jiang, W. (2007). Price informativeness and investment sensitivity to stock prices. *Review of Financial Studies*, 20(3):619–650.

Chen, W., Shi, T., and Srinivasan, S. (2024). The value of ai innovation. Working Paper, Harvard Business School.

Chi, F., Hwang, B.-H., and Zheng, Y. (2025). The use and usefulness of big data in finance: Evidence from financial analysts. *Management Science*, 71(6):4599–4621.

Ciconte, W., Kirk, M., and Tucker, J. W. (2014). Does the midpoint of range earnings forecasts represent managers' expectations? *Review of Accounting Studies*, 19:628–660.

Cockburn, I. M., Henderson, R., and Stern, S. (2018). The impact of artificial intelligence on innovation. Working Paper 24449, National Bureau of Economic Research.

Cohen, L., Malloy, C., and Pomorski, L. (2012). Decoding inside information. *The Journal of Finance*, 67(3):1009–1043.

Coller, M. and Yohn, T. L. (1997). Management forecasts and information asymmetry: an examination of bid-ask spreads. *Journal of Accounting Research*, 35(2):181–191.

Cook, R. D. and Weisberg, S. (1982). Residuals and influence in regression. *Biometrical Journal*, 27(1).

Davila, E. and Parlatore, C. (2025). Identifying price informativeness. Working paper.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications*. *Quarterly Journal of Economics*, 135(2):561–644.

Dessaint, O., Foucault, T., and Frésard, L. (2024). Does alternative data improve financial forecasting? The horizon effect. Journal of Finance, Forthcoming.

Dessaint, O., Foucault, T., Frésard, L., and Matray, A. (2019). Noisy stock prices and corporate investment. *Review of Financial Studies*, 32(7):2625–2672.

Dessaint, O., Gondhi, N., and Peress, J. (2025). Learning about discount rates. Working Paper, INSEAD.

Dou, W. W., Goldstein, I., and Ji, Y. (2025). Ai-powered trading, algorithmic collusion, and price efficiency. Working Paper 34054, National Bureau of Economic Research.

Dugast, J. and Foucault, T. (2025). Equilibrium data mining and data abundance. *The Journal of Finance*, 80(1):211–258.

Edmans, A., Goldtsein, I., and Jiang, W. (2012). The real effects of financial markets: The impact of prices on takeovers. *Journal of Finance*, 67(3):933–971.

Ferracuti, E., Koo, M., Lee, M., and Stubben, S. (2024). Acquisition of customer information and corporate decision making. Working Paper, Hong Kong University Business School.

Foucault, T. and Fresard, L. (2014). Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, 111(3):554–577.

Foucault, T., Gambacorta, L., Jiang, W., and Vives, X. (2025). Barcelona 7: Artificial intelligence in finance. CEPR Press, Paris London.

Frésard, L., Hoberg, G., and Phillips, G. M. (2019). Innovation activities and integration through vertical acquisitions. *The Review of Financial Studies*, 33(7):2937–2976.

Gardner, J., Thakral, N., To, L. T., and Luther, Y. (2024). Two-stage differences in differences. *Working paper*.

Goldfarb, A. and Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1):3–43.

Goldstein, I. (2023). Information in Financial Markets and Its Real Effects. *Review of Finance*, 27:1–32.

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

Gormsen, N. J. and Huber, K. (2025). Corporate discount rates. *American Economic Review*, 115(6):2001–49.

Gutiérrez, G. and Philippon, T. (2017). Investment-less growth: An empirical investigation. *Brookings Papers on Economic Activity*.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica*, 50:215–224.

Hoberg, G. and Maksimovic, V. (2015). Redefining financial constraints: A text-based analysis. *Review of Financial Studies*, 28(5):1312–1352.

Khan, M., Kogan, L., and Serafaim, G. (2012). Mutual fund trading pressure: Firm-level stock price impact and timing of seos. *The Journal of Finance*, 67(4):1371–1395.

Labro, E., Lang, M., and Omartian, J. D. (2023). Predictive analytics and centralization of authority. *Journal of Accounting and Economics*, 75(1):101526.

Linn, M. and Weagley, D. (2023). Uncovering financial constraints. *Journal of Financial and Quantitative Analysis*, page 1–36.

Liu, Z. and Winegar, A. (2025). Economic magnitudes within reason. *Journal of Corporate Finance*, 91.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies*, 25(12):3457–3489.

Lu, Y., Phillips, G., and Yang, J. (2024). The impact of cloud computing and ai on industry dynamics and concentration. *Working paper*.

Lu, Y., Phillips, G., and Yang, J. (2025). The impact of cloud computing and ai on industry concentration and dynamics. Working Paper, Tuck School of Business.

Martin, I. W. and Nagel, S. (2022). Market efficiency in the age of big data. *Journal of Financial Economics*, 145(1):154–177.

Mihet, R., Rishabh, K., and Gomes, O. (2025a). Is it ai or data that drives firm growth? Working Paper, Swiss Finance Institute.

Mihet, R., Rishabh, K., and Jang-Jaccard, J. (2025b). Cyberrisk and ai firms. Working Paper, Swiss Finance Institute.

Morck, R., Shleifer, A., and Vishny, R. (1990). The stock market and investment: Is the market a sideshow? *Brookings Papers on Economic Activity*, 21:157–216.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Nickell, S. J. (1996). Competition and corporate performance. *Journal of Finance*, 104(4):724–746.

Nike (2025). Privacy cookie policy. Nike.

Peters, R. and Taylor, L. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 100:251–272.

Schmickler, S. and Tremacoldi-Rossi, P. (2023). Spillover effects of payouts on asset prices and real investment. Working paper.

van der Beck, P. (2024). Flow-driven esg returns. *Working paper*.

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

Yan, D. (2024). Do private firms (mis)learn from the stock market? *Review of Finance*, 28(5):1483–1511.

Zanotti, M. (2025). Asset (and data) managers. Working Paper, Swiss Finance Institute.

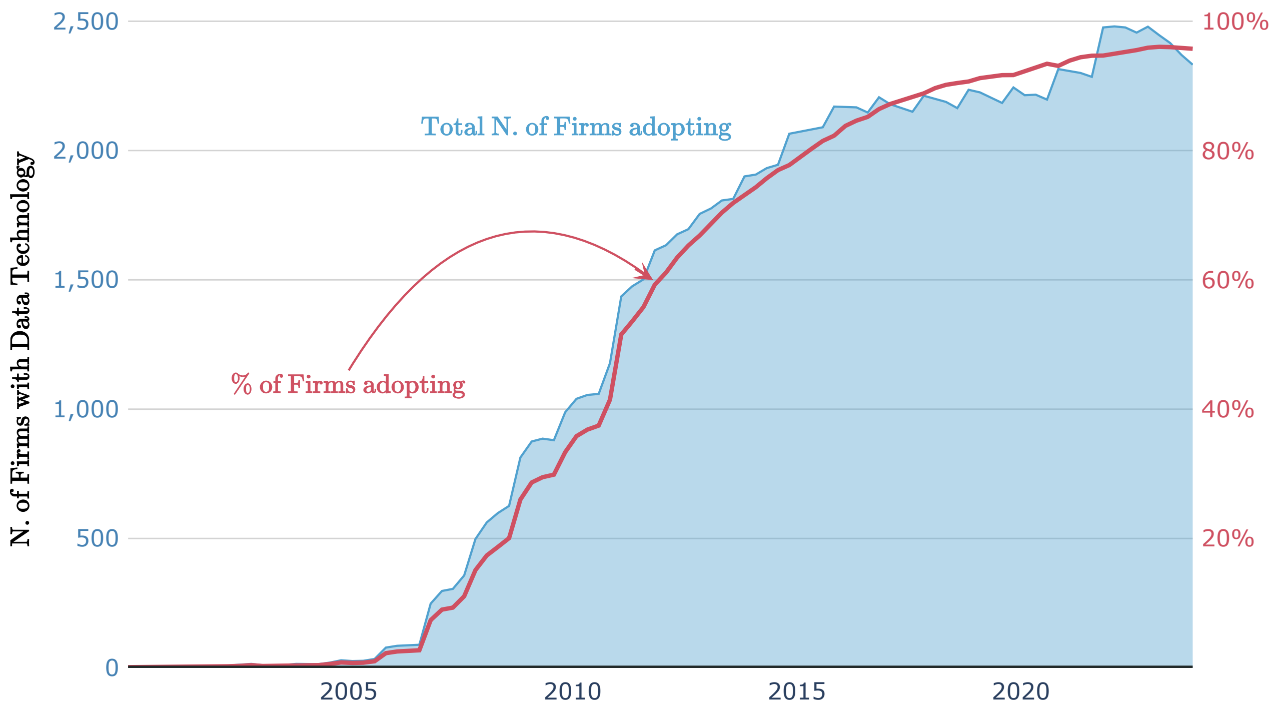


Figure I: Firms and Data Technology adoption. This figure shows the adoption of data technologies in our sample. The data are from BuiltWith, which detects the installation and removal of various technologies by analyzing webpage code. The blue line (left axis) represents the number of funds with at least one data technology in place for each month of the sample period. The red line (right axis) shows the percentage of funds adopting data technologies relative to the total number of funds in a given month.

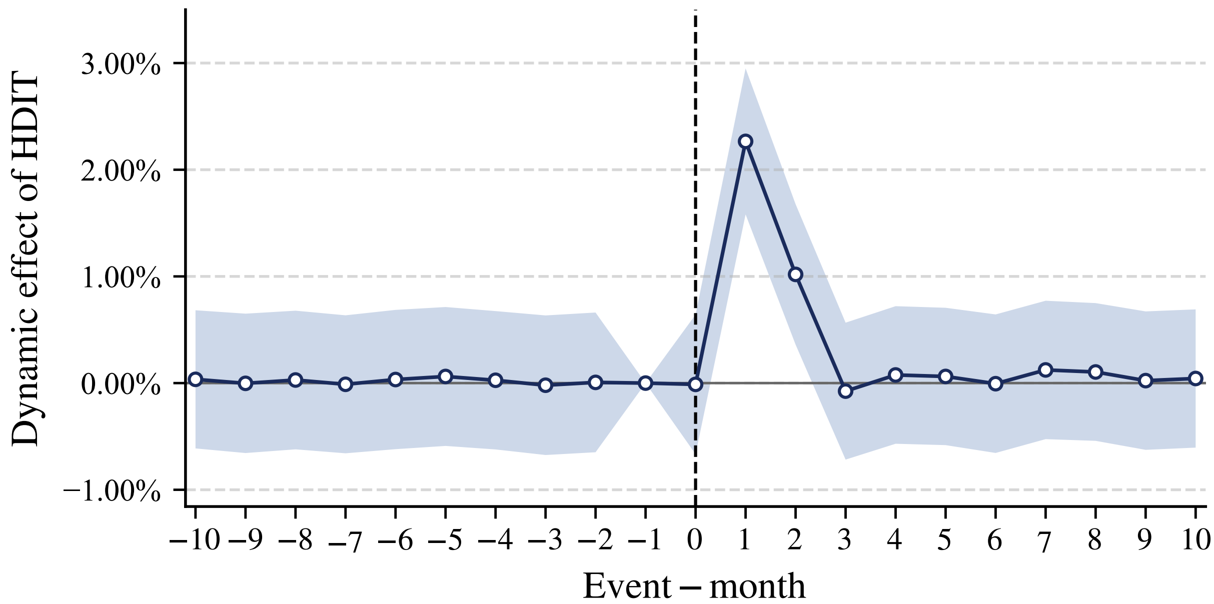
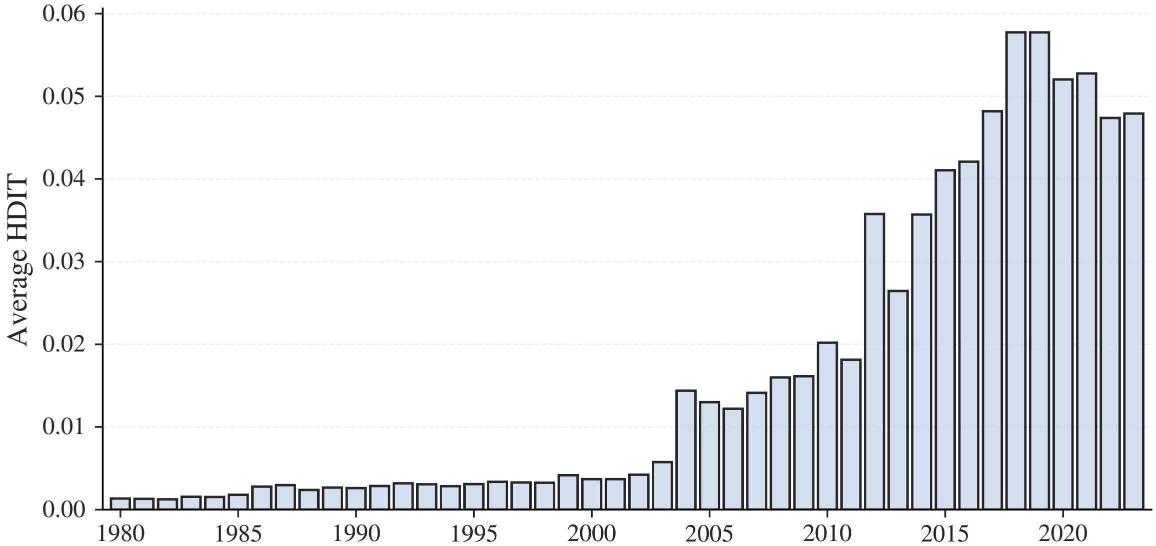
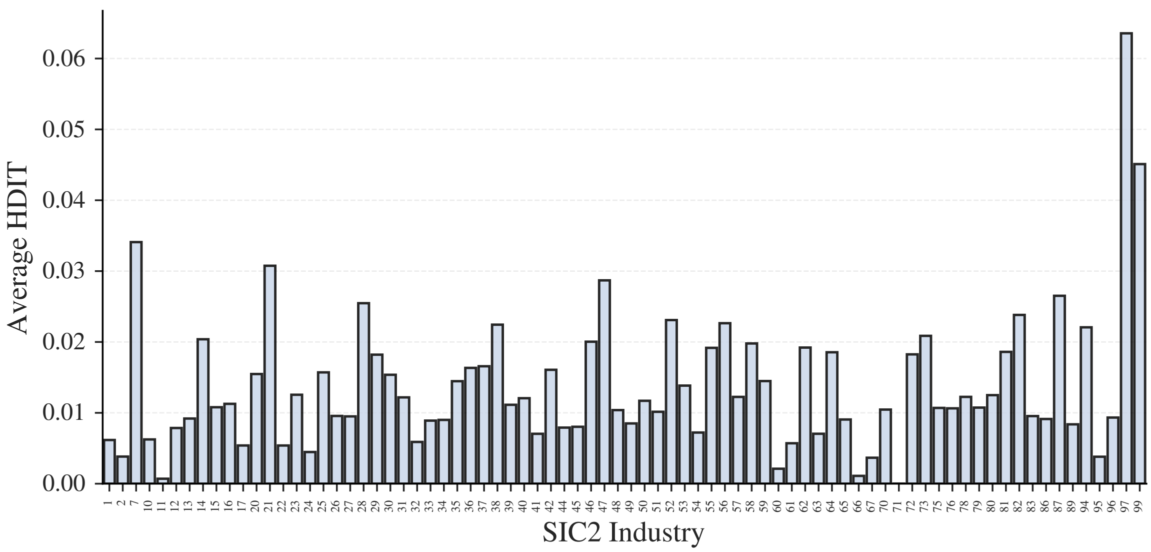


Figure II: Dynamics of the effects of HDIT on prices. This figure plots the monthly average abnormal returns of stocks subject to Hypothetical Dividend Induced Trading (HDIT) price pressure around the event, where an event is defined as a firm-month observation in the top decile of HDIT within each year. For each year in the sample, we take stocks in the top decile of the monthly HDIT distribution, and plot a coefficient regression from a linear regression of stock returns on even-month dummies, with firm×cohort fixed effects. All standard errors are two-way clustered by firm×cohort and day.



A. Average HDIT by year.



B. Average HDIT by SIC2 Industry.

Figure III: Hypothetical Dividend Induced Trading across Time and Industry. This figure plots the distribution of average Hypothetical Dividend Induced Trading (HDIT) by each year in our sample (Panel A), and by each SIC2 industry (Panel B).

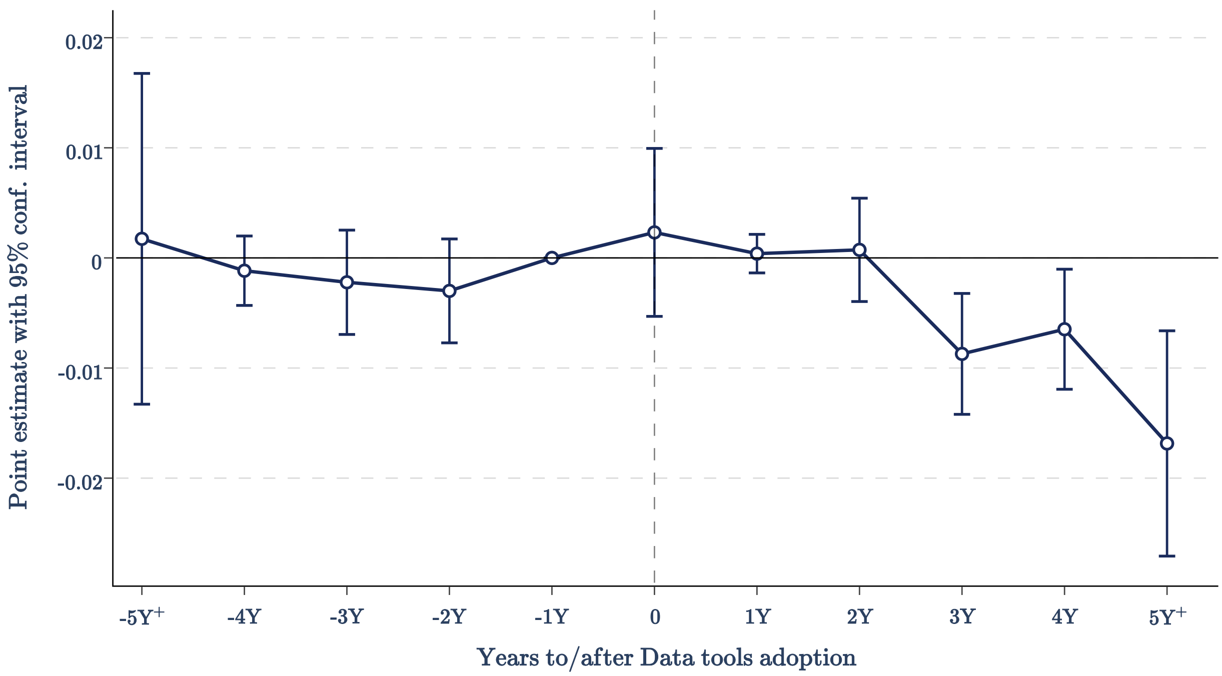


Figure IV: Dynamics of the effects of Data on Managerial Learning. This figure shows results for the specification where Tobin's Q is instrumented by the Dividend Induced Trading (DIT) normalized by Total Assets. Each point represents the estimated coefficient on the interaction with each year before/after data technology adoption. The treatment is a dummy equal to one if a firm i has a data technology in place in year t ($\text{DATA}_{i,t}$). The control variables include a firm's size (*log* assets) in quarter t and contemporaneous cash flows over assets. The regression includes firm and year fixed effects, and the vertical bars represent the 95% confidence interval for the coefficient estimates. The year just before data technology adoption (-1) is the excluded category in the regression and is reported as zero in the figure. The rightmost estimates include all observations after 5 years from the adoption year. The annual sample is from 1980 to 2023.

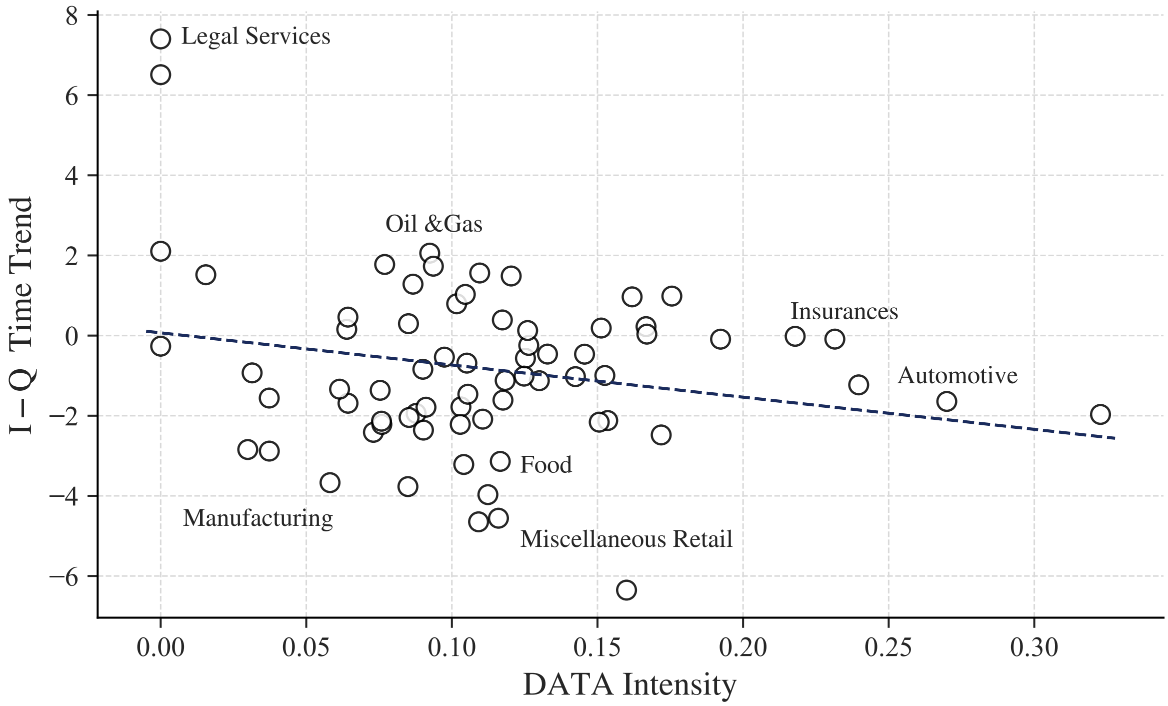


Figure V: Secular Decline in I-Q Sensitivity and Data Technologies. This figure shows the scatter plot between the time-trend in I-Q sensitivity for each industry and the intensity of adoption within industry. For this plot we use the full sample of U.S. publicly listed firms (i.e., we do not restrict to retail-oriented firms.) In particualar, we run the specification $I_{i,t} = \beta_1 Q_{i,t-1} + \beta_2 Q_{i,t-1} \times (t - t_0) + \Gamma X_{i,t-1} + \gamma_i + \eta_t + \varepsilon_{i,t}$, for each two-digit SIC industry, and plot on the y-axis the t-statistic for β_2. The control variables include a firm's size (*log* assets) in quarter t and contemporaneous cash flows over assets. On the x-axis, we plot the average portion of data-adopters in our sample, within each two-digit SIC industry. The full annual sample is from 1980 to 2023.

Data Technology Name	Installation % (in 2023)	Description
Google Analytics	95.43	Users Tracking and Analytics
Google Analytics 4	59.12	Users Tracking and Analytics
Facebook Pixel	51.59	Social Media Tracking and Analytics
LinkedIn Insights	35.97	Social Media Tracking and Analytics
Bing Universal Event Tracking	27.17	Users Tracking and Analytics
Hotjar	24.91	Users Tracking and Analytics
Adobe Analytics	19.07	Users Tracking and Analytics
Twitter Analytics	18.72	Social Media Tracking and Analytics
Google Optimize 360	16.02	A/B Testing
RapLeaf	14.74	Users Tracking
Yahoo Web Analytics	14.69	Users Tracking and Analytics
Hubspot	10.76	Users Tracking and Analytics
Crazy Egg	9.73	Track and Visualize User Interaction
LiveRamp	9.58	Data Connectivity Platform
Yahoo Dot	8.50	Users Tracking and Analytics
Salesforce	8.40	Cloud and Customer Relationship

Table I: Main Data Technologies: This table reports the main data technologies installed on firms' websites, as of December 2023. These plugins are aimed to collect and process website visitors' data. The second column shows the percentage of firms having the technology installed on its website with respect to the total number of funds, as of December 2023. The third column reports a short description of the technology's features.

	Obs.	Mean	Standard Deviation		P5	P25	P50	P75	P95
			sample	within-firm					
CAPEX	103,285	0.32	0.39	0.39	0.03	0.11	0.21	0.38	0.94
Size (*log* Assets)	103,285	5.90	2.15	0.89	2.68	4.27	5.75	7.37	9.74
Tobin's Q	103,285	1.98	1.68	1.26	0.82	1.05	1.38	2.19	5.28
HDIT	103,285	0.02	0.03	0.02	0.00	0.00	0.00	0.02	0.08
Cash Flow	103,285	0.00	0.26	0.18	-0.49	0.01	0.06	0.11	0.20
N. of $Data\ Tech$.	103,285	1.39	3.84	-	0.00	0.00	0.00	1.00	9.00
DATA	103,285	0.25	0.44	-	0.00	0.00	0.00	1.00	1.00

Table II: Summary Statistics Annual: This table reports summary statistics for the full sample. For each variable, the table shows the number of available observations (*obs*,), the mean (*mean*), the sample standard deviation (*sample*), the within-firm standard deviation ($within - firm$), the 5th ($p5$), 25th ($p25$), 50th ($p50$), 75th ($p75$), and the 95th ($p95$) percentiles. CAPEX is CAPX divided by PPENT, Cash Flows and Hypothetical Dividend Induced Trading (HDIT) are normalize by Total Assets. The variable N. of $Data\ Tech$. represents the total number of data technologies installed on the firm's website in a given year, the variable DATA is a dummy equal to 1 if the firm-year observation has at least one data technology installed. All variables are winsorized at the 1% and 99% level. The annual sample is from 1980 to 2023.

	$\mathbb{P}\{\text{Issue Guidance}\}$		Guidance Squared Error	
	(1)	(2)	(3)	(4)
$\textsc{Data}_{i,t}$	0.110***	0.106***	-0.039*	-0.033***
	(0.012)	(0.007)	(0.020)	(0.011)
Controls	×	✓	×	✓
Firm FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Outcome mean	0.111	0.111	0.074	0.074
Outcome SE	0.314	0.314	0.225	0.225
Obs.	116,626	111,713	9,808	9,638
Adjusted R^2	0.176	0.285	0.277	0.287

Table III: Manager Guidance and Data Technologies: This table shows results of panel regression in which the dependent variable is the probability that a manager issue sales guidance (columns (1)-(2)), and the normalized squared error in the manager's guidance (columns (3)-(4)). The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\textsc{Data}_{i,t}$), and controls for firm-year characteristics. The control variables include a (lagged) firm's size (*log* assets) and contemporaneous cash flows over assets. Managers' guidance data are from I/B/E/S. All columns show estimates using difference-in-differences estimator robust to staggered treatment concerns (Gardner et al., 2024). The annual sample is from 1980 to 2023. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX			Q	$Q\times$ DATA
	(1)	(2)	(3)	(4)	(5)
$Q_{i,t-1}$	0.017***	0.023***	0.089***		
	(0.002)	(0.003)	(0.006)		
$Q_{i,t-1}\times$ DATA$_{i,t-1}$	-0.005***	-0.012***	-0.034***		
	(0.001)	(0.003)	(0.009)		
DATA$_{i,t-1}$	0.116***	0.098***	-0.213***	0.069	1.182***
	(0.023)	(0.023)	(0.063)	(0.044)	(0.049)
Size$_{i,t-1}$		-0.045***	-0.058***	-0.363***	-0.071***
		(0.004)	(0.006)	(0.019)	(0.014)
Cash Flow$_{i,t}$		0.032***	0.034***	0.083***	0.014**
		(0.003)	(0.003)	(0.010)	(0.005)
HDIT$_{i,t-1}$				0.652***	0.024*
				(0.052)	(0.013)
HDIT$_{i,t-1}$ $\times$ DATA$_{i,t-1}$				-0.171***	0.360***
				(0.058)	(0.024)
Estimator	2SLS	2SLS	OLS	First-stage	First-stage
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.353	0.353	0.353	1.985	1.985
Outcome SE	0.460	0.460	0.460	1.772	1.772
Obs.	103,285	103,285	125,899	103,285	103,285
Adj. R^2	0.043	0.059	0.287	0.613	0.790
Kleibergen-Paap F-stat	36.53	36.71			

Table IV: Main Results: Managers Learning and Data Technologies: This table shows results of panel regression in which the dependent variable is the one-year-ahead firm's investment. The dependent variable in columns (1) to (3) is CAPEX. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data technology in place in year t (DATA$_{i,t}$), and controls for firm-year characteristics. All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX			Q	$Q\times$ Data
	(1)	(2)	(3)	(4)	(5)
$Q_{i,t-1}$	0.016***	0.021***	0.090***		
	(0.004)	(0.004)	(0.006)		
$Q_{i,t-1}\times$ Data$_{i,t-1}$	-0.004**	-0.008**	-0.035***		
	(0.001)	(0.004)	(0.009)		
Data$_{i,t-1}$	0.108***	0.103***	-0.077	0.171***	1.084***
	(0.024)	(0.024)	(0.059)	(0.062)	(0.048)
Size$_{i,t-1}$		-0.040***	-0.043***	-0.395***	-0.024***
		(0.006)	(0.006)	(0.021)	(0.006)
Cash Flow$_{i,t}$		0.026***	0.027***	0.067***	0.004**
		(0.004)	(0.004)	(0.012)	(0.002)
HDIT$_{i,t-1}$				0.718***	0.016***
				(0.059)	(0.004)
HDIT$_{i,t-1}$ $\times$ Data$_{i,t-1}$				-0.226***	0.345***
				(0.061)	(0.022)
Estimator	2SLS	2SLS	OLS	First-stage	First-stage
Cohort$\times$Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.363	0.363	0.363	1.976	1.976
Outcome SE	0.474	0.474	0.474	1.814	1.814
Obs.	309,468	309,468	354,853	309,468	309,468
Adj. R^2	0.044	0.056	0.253	0.580	0.813
Kleibergen-Paap F-stat	59.107	62.772			

Table V: Robustness of Main Results stacking: This table shows robustness tests for the main results in the stacked sample. For each year in which at least one firm is treated, we maintain firm-year observations of untreated firms as of the event year (i.e., the control group include never treated and not-yet-treated firms) and stack all obsvervations. The specification is equivalent to the one in the main results in Table IV. The dependent variable in columns (1) to (3) is CAPEX. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data technology in place in year t (Data$_{i,t}$), and controls for firm-year characteristics. All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include cohort$\times$firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX							
Robustness:	Industry×Year FE		Retail SIC Codes		Sample starts 1990		Interacted Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Q_{i,t-1}$	0.021***	0.027***	0.015***	0.019***	0.017***	0.021***	0.020***	0.021***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)
$Q_{i,t-1}\times$ $\text{DATA}_{i,t-1}$	-0.007***	-0.015***	-0.006***	-0.012**	-0.004***	-0.008***	-0.005***	-0.004***
	(0.001)	(0.004)	(0.002)	(0.005)	(0.001)	(0.002)	(0.001)	(0.001)
$\text{DATA}_{i,t-1}$	0.094***	0.079***	0.107***	0.096***	0.113***	0.094***	-0.009	-0.061
	(0.019)	(0.019)	(0.033)	(0.033)	(0.023)	(0.024)	(0.063)	(0.061)
$\text{Size}_{i,t-1}$		-0.043***		-0.033***		-0.041***	-0.049***	-0.048***
		(0.004)		(0.005)		(0.005)	(0.006)	(0.005)
Cash $\text{Flow}_{i,t}$		0.030***		0.025***		0.028***	0.033***	0.031***
		(0.003)		(0.003)		(0.003)	(0.003)	(0.003)
Interacted Controls	×	×	×	×	×	×	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	×	×	✓	✓	✓	✓	✓	×
Industry×Time FE	✓	✓	×	×	×	×	×	✓
Outcome mean	0.353	0.353	0.303	0.303	0.349	0.349	0.349	0.349
Outcome SE	0.460	0.460	0.393	0.393	0.454	0.454	0.454	0.454
Obs.	99,753	99,753	52,881	52,881	83,118	83,118	100,205	99,753
Adj. R^2	0.039	0.053	0.038	0.051	0.043	0.056	0.059	0.053
Kleibergen-Paap F-stat	36.622	36.987	40.930	41.066	37.382	37.657	37.698	37.780

Table VI: Robustness of Main Results: This table shows robustness tests for the main results in the paper. The dependent variable is CAPEX across all specifications. Columns (1) and (2) report results including Industry×Year fixed-effects, using SIC 2-digits industry codes. Columns (3) and (4) show results using SIC codes to identify retail sector's firms in our sample (rather than "end-users"). Retail sector's SIC are 3-digits code 080, 090, 200-211, 226-239, 251-254, 259, 265, 267, 283-284, 300-309, 321-323, 363-366, 393-396, 411, 481, 491-493, 503-504, 506, 508, 509, 514-599, 602-606, 611, 623, 641, 701-704, 720-726, 729, 751-753, 760-799, 801-809, 841-843, 849, 922. Columns (5) and (6) show robustness results to using 1990 as starting date for our sample. Columns (7) and (8) report results interacting controls. The table report results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). The regressors are a dummy equal to one if a firm i has a placebo technology in place in year t ($\text{PLACEBO}_{i,t}$). All regressors are standardized by within-firm variation. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

	CAPEX			
Placebo:	Agency	Registrar	Copyright	Shipping
	(1)	(2)	(3)	(4)
$Q_{i,t-1}$	0.016***	0.016***	0.016***	0.016***
	(0.003)	(0.003)	(0.003)	(0.003)
$Q_{i,t-1} \times$ PLACEBO$_{i,t-1}$	-0.003	-0.003*	-0.003	-0.003
	(0.020)	(0.001)	(0.029)	(0.003)
Controls	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Outcome mean	0.354	0.354	0.354	0.354
Outcome SE	0.460	0.460	0.460	0.460
Obs.	103,285	103,285	103,285	103,285
Adj. R^2	0.050	0.050	0.050	0.050
Kleibergen-Paap F-stat	202.29	201.75	196.39	201.22

Table VII: Placebo Technologies: This table shows the results of panel regression in which the dependent variable is the one-year-ahead firm's investment. The dependent variable is CAPEX. The table reports results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). The regressors are a dummy equal to one if a firm i has a placebo technology in place in year t (PLACEBO$_{i,t}$). The placebo technologies are reported above each respective column: column (1) uses Digital Agency technologies as placebo (e.g., Wix Studio), column (2) uses Registrar-specific plugins (e.g., technologies included in the hosting provider's registration), column (3) uses Copyright plugins (e.g., Copyscape), and column (4) Shipping technologies (e.g., Shopify shipping). The control variables (omitted for brevity) include a lagged firm's size (*log* assets) and contemporaneous cash flows over assets. All regressors are standardized by within-firm variation. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX			Q	$Q\times$ Data
	(1)	(2)	(3)	(4)	(5)
$Q_{i,t-1}$	0.016***	0.022***	0.026***		
	(0.004)	(0.005)	(0.006)		
$Q_{i,t-1}\times$ Data$_{i,t-1}$	-0.003	-0.009	-0.014		
	(0.002)	(0.009)	(0.017)		
Data$_{i,t-1}$	0.051	0.034	0.020	0.099*	1.395***
	(0.031)	(0.032)	(0.032)	(0.055)	(0.060)
Size$_{i,t-1}$		-0.038***	-0.037***	-0.193***	-0.042***
		(0.004)	(0.004)	(0.018)	(0.011)
Cash Flow$_{i,t}$		0.022***	0.020***	0.096***	0.024***
		(0.003)	(0.003)	(0.011)	(0.007)
HDIT$_{i,t-1}$				0.461***	0.021
				(0.046)	(0.013)
HDIT$_{i,t-1}$ $\times$ Data$_{i,t-1}$				-0.112**	0.326***
				(0.051)	(0.023)
Estimator	2SLS	2SLS	OLS	First-stage	First-stage
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.303	0.303	0.303	1.799	1.799
Outcome SE	0.372	0.372	0.372	1.530	1.530
Obs.	43,547	42,103	41,601	45,142	45,142
Adj. R^2	0.061	0.081	0.072	0.657	0.826
Kleibergen-Paap F-stat	28.702	28.499			

Table VIII: Placebo Tests with "end-users" below 50%: This table shows placebo tests on the sample of firms with "end-users" below 50% (i.e., non-retail firms). The dependent variable in columns (1) to (3) is CAPEX. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Total Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data tool in place in year t (Data$_{i,t}$), and controls for firm-year characteristics. All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX		
	(1)	(2)	(3)
$Q_{i,t-1}$	0.023***	0.025***	0.026***
	(0.002)	(0.002)	(0.002)
$Q_{i,t-1}\times$ DATA$_{i,t-1}$	-0.020***	-0.016***	-0.019***
	(0.004)	(0.003)	(0.003)
$Q_{i,t-1}\times$ DATA$_{i,t-1}\times$ CCPA$_{i,t-1}$	0.002***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)
DATA$_{i,t-1}\times$ CCPA$_{i,t-1}$	0.061***	0.032***	0.029*
	(0.010)	(0.013)	(0.017)
$Q_{i,t-1}\times$ CCPA$_{i,t-1}$	-0.047***	-0.046***	-0.034***
	(0.003)	(0.004)	(0.006)
DATA$_{i,t-1}$	0.061***	0.054***	-0.106*
	(0.012)	(0.012)	(0.053)
Controls	×	✓	✓
Interacted Controls	×	×	✓
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
Outcome mean	0.354	0.354	0.354
Outcome SE	0.460	0.460	0.460
Obs.	97,383	97,383	97,383
Adj. R^2	0.045	0.061	0.061
Kleibergen-Paap F-stat	63.26	70.47	76.35

Table IX: Managers Learning and Data Technologies after CCPA: This table shows results of panel regression in which the dependent variable is the one-year-ahead firm's investment. All columns reports results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). The dependent variable is CAPEX. The regressors are a dummy equal to one if a firm i has a data technology in place in year t (DATA$_{i,t}$), a dummy equal to one if firm i is subject to the California Consumer Privacy Act (CCPA) in year t (CCPA$_{i,t}$), and controls for firm-year characteristics (omitted for brevity). The CCPA became enforceable on January 1, 2020. The control variables include a (lagged) firm's size (*log* assets) and contemporaneous cash flows over assets. Column (3) includes control variables interacted with DATA and CCPA in all possible combination. All regressors are standardized by within-firm variation. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by state and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	Corporate Insiders' Net Sales				Corporate Insiders' Returns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Q_{i,t}$			0.034***				-0.029**	
			(0.010)				(0.012)	
$Q_{i,t}\times$ $\text{DATA}_{i,t}$			-0.016				0.005	
			(0.010)				(0.008)	
$\text{DATA}_{i,t}$	-0.031	-0.011	0.046	-0.002	0.040	0.033	-0.033	-0.014
	(0.023)	(0.011)	(0.042)	(0.016)	(0.028)	(0.022)	(0.054)	(0.015)
$\text{HDIT}_{i,t}$				0.021***				-0.026***
				(0.007)				(0.007)
$\text{HDIT}_{i,t}\times$ $\text{DATA}_{i,t}$				-0.007				0.015
				(0.010)				(0.010)
$\text{Size}_{i,t-1}$		0.016***	0.002	0.009		-0.009***	0.007	0.009
		(0.004)	(0.007)	(0.006)		(0.003)	(0.009)	(0.006)
Cash $\text{Flow}_{i,t}$		0.029***	0.030***	0.029***		-0.026***	-0.028***	-0.028***
		(0.003)	(0.005)	(0.005)		(0.003)	(0.006)	(0.006)
Estimator	Staggered DiD		2SLS	OLS	Staggered DiD		2SLS	OLS
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Outcome mean	-0.010	-0.010	-0.010	-0.010	0.073	0.073	0.073	0.073
Outcome SE	1.016	1.016	1.016	1.016	0.441	0.441	0.441	0.441
Obs.	96,391	96,391	96,391	96,391	34,701	34,701	34,701	34,701
Adj. R^2	0.001	0.000	0.002	0.098	0.003	0.002	0.016	0.144
Kleibergen-Paap F-stat			41.778				34.725	

Table X: Corporate Insiders' contemporaneous Net Sales and Returns on Insider Trades: This table shows results of panel regression in which the dependent variable is the total (nonroutine) net sales of corporate insiders (columns (1) to (4)) and 12-month excess return of corporate insiders for firm i in year t. Nonroutine trades follows definition in Cohen et al. (2012), while insiders include officers with decision-making authority, all board members and beneficial owners of more than 10% of a company's stock. Control variables include a (lagged) firm's size (*log* assets) and contemporaneous cash flows over assets. Regressors are standardized by within-firm variation. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. Standard errors are two-way clustered by state and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	*vol*(Sales)	*vol*(Cash Flow)	*vol*(Earnings)	Jump Ratio (Weller, 2018)	
	(1)	(2)	(3)	(4)	(5)
$\textsc{Data}_{i,t}$	-0.001	0.001	0.010*	0.042***	0.038**
	(0.005)	(0.003)	(0.006)	(0.011)	(0.006)
Controls	✓	✓	✓	×	✓
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.145	0.070	0.071	0.383	0.383
Outcome SE	0.181	0.125	0.313	0.339	0.339
Obs.	91,527	101,181	97,436	82,534	80,378
Adjusted R^2	0.408	0.626	0.607	0.176	0.178

Table XI: **Firm Volatility and Price Informativeness**: This table shows results of panel regression in which the dependent variables is a measure of firm volatility using 3-years rolling window (columns (1) to (3)) or stock price informativeness (columns (4) and (5)). The measures of firm volatility are the volatility of sales over total assets (column (1)), the volatility of cash flow over total assets (column (2)), and the volatility of earnings over total assets (column (3)). Columns (4) and (5) use the stock price jump ratio around earning announcements (Weller, 2018). A higher jump ratio is associated with less informative stock prices. The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\textsc{Data}_{i,t}$), and controls for firm-year characteristics. The control variables include a (lagged) firm's size (*log* assets), Tobin's Q, and contemporaneous cash flows over assets. All columns show estimates using difference-in-differences estimator robust to staggered treatment concerns Gardner et al. (2024). The annual sample is from 1980 to 2023. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Panel A: Equity Constraints						
Financial Constraints:	Equity Constraints		Debt Constraints		Cost of Capital	Hurdle Rate
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{DATA}_{i,t}$	0.002	0.056***	-0.002	-0.010	0.000	0.000
	(0.003)	(0.022)	(0.001)	(0.011)	(0.000)	(0.000)
Measure from:	HM (2015)	LW (2023)	HM (2015)	LW (2023	Gormsen and Huber (2025)	
Controls	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Outcome mean	-0.006	-0.018	-0.001	0.018	0.091	0.118
Outcome SE	0.093	0.669	0.056	0.603	0.014	0.017
Obs.	38,624	81,402	36,622	81,401	29,456	29,455
Adjusted R^2	0.629	0.705	0.453	0.569	0.899	0.837

Table XII: Financial Constraints and Data Technologies: This table shows results of panel regression in which the dependent variables is a measure of financial constraints. Columns (1) to (4) use text-based measured developed in Hoberg and Maksimovic (2015) and Linn and Weagley (2023). Columns (1) and (2) refer to equity financial constraints, while columns (3) and (4) to debt constraints. In columns (1) and (3) we use measures from Hoberg and Maksimovic (2015) (HM (2015)) and columns (2) and (4) from Linn and Weagley (2023) (LW (2023)). In columns (5) and (6) the dependent variables is a measure of managers' perceived cost of capital from Gormsen and Huber (2025) (column (5)) and managers' perceived hurdle rate (columns (6)). The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\text{DATA}_{i,t}$), and controls for firm-year characteristics. The control variables include a (lagged) firm's size (*log* assets), Tobin's Q, and contemporaneous cash flows over assets. All columns show estimates using difference-in-differences estimator robust to staggered treatment concerns (Gardner et al., 2024). All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	Market Share	Price Cost Margin	DeLoecker et al. (2020)
	(1)	(2)	(3)
$\text{DATA}_{i,t}$	-0.009	-0.003	-0.344***
	(0.006)	(0.032)	(0.094)
Controls	✓	✓	✓
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
Outcome mean	0.087	0.253	1.658
Outcome SE	0.210	0.663	1.428
Obs.	103,285	103,285	103,285
Adjusted R^2	0.670	0.604	0.580

Table XIII: Market Power and Data Technologies: This table shows results of panel regression in which the dependent variables is a measure associated with market power. Columns (1)-(2) use a firm-year market share, computed as the share of sales with respect to the total sales of the SIC-3 digits industry. Columns (3)-(4) use the firm-year measure from De Loecker et al. (2020), while columns (5)-(6) use the firm-year price cost margin computed as $(sales - cogs)/sales$. The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\text{DATA}_{i,t}$), and controls for firm-year characteristics. The control variables include a (lagged) firm's size (*log* assets), Tobin's Q, and contemporaneous cash flows over assets. All columns show estimates using difference-in-differences estimator robust to staggered treatment concerns (Gardner et al., 2024). The annual sample is from 1980 to 2023. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Appendix

A Data Appendix

In this Appendix we describe the main dataset construction procedure. We use four main data sources: (i) Compustat Annual, (ii) CRSP Stock File Data, (iii) CRSP Mutual Fund Holdings, and (iv) BuiltWith for websites' technology installation/removal dates.

We construct the annual panel of U.S. public firms by merging Compustat Fundamentals Annual and CRSP through the standard CRSP–Compustat link (CCM). We filter the sample including only firms listed on the NYSE, NASDAQ, or AMEX (exchange codes 1–3) and classified as common shares (share codes 10 or 11). In our main text, we focus on retail-oriented firms because they are the most likely to extract signals about their growth opportunities from customer data from their website. We identify these firms using the text-based approach of Frésard et al. (2019) linking product vocabularies from the Bureau of Economic Analysis (BEA) input-output tables to firms' 10-K product descriptions. Frésard et al. (2019) identify the specific words associated with firms' goods or services that are directly purchased by end-users (together with their economic intensity). Then, they compute a score that measures the intensity with which each firm's products and service flows towards end-users. In our main sample, we keep only firms with a end-users score above 50% at least once throughout the sample (i.e., retail-oriented firms). Finally, we remove firms with negative or missing total assets and restrict the sample to observations after January 1980 to align with the availability of mutual fund holdings data.

To merge BuiltWith firms' website technologies with the main CRSP/Compustat panel, we rely on each firm's website URL from Compustat. We standardize URLs by removing prefixes and suffixes (e.g., "www.", "http://", "/index.html") and convert all domains to lowercase. For firms that change domains over time, we record the first and last year during which each website is active. Then, we merge each firm-year observation with the corresponding website technologies from BuiltWith. The final annual panel contains 103,285 firm-year observations, with 8,815 unique firms.

We construct $HDIT$ using CRSP Mutual Funds holdings (s12) at the quarterly frequency. The procedure follows Schmickler and Tremacoldi-Rossi (2023). We use CRSP stock data at the daily frequency to align dividend ex- and payment dates with the corresponding fund-stock holdings.

We start from all U.S. mutual funds in the CRSP Mutual Funds database between 1980 and 2023. Using the s12 quarterly holdings, we retain equity positions (share codes 10–11, exchange codes 1–3) and merge them with fund identifiers via MF_LINK tables provided by WRDS. We then compute portfolio weights each quarter using the fund's total assets under management in equity.

Daily dividend information are from CRSP Distribution. We keep only cash dividends (distribution codes 1000–1399) and non-missing payment–ex-date pairs. We remove duplicate and require the payment date to be after the dividend announcement date (we drop 47 observations with this filter, about 0.01% of total stock dividend payment observations). For each dividend, we identify the trading day before the ex-date as the relevant date determining ownership. We obtain the dollar dividend amount by multiplying the dividend per share by shares outstanding on that date.

We assume holdings are constant within reporting quarter. For each stock-fund-day dividend event, we compute the dollar inflow to fund j from stock i in day d as:

$$\text{payFlow}_{i,j,d} = divamt_{i,d} \cdot q^{ex}_{i,j,d}, \tag{A.1}$$

where $divamt$ is the dollar dividend paid by stock i in day d, and $q^{ex}_{i,j,d}$ is the fraction of outstanding shares of i held by fund j on the ex-dividend date. We then aggregate these "dividend-induced" inflows across all dividend-paying stocks in the fund's portfolio to obtain the total dividend inflow received by the fund at each payment day. We then compute the dividend-induced inflow into each stock as

$$\text{inflow}_{i,j,d} = w_{i,j,d} \cdot \text{payFlow}_{i,j,d}, \tag{A.2}$$

with $w_{i,j,d}$ being the portfolio weight of stock i in fund j at that date (assuming constant holdings within the quarter).

We then aggregate across all funds holding stock i on each day to obtain the total hypothetical dividend-induced trading ($HDIT$). To remove any fundamental information, we exclude from $HDIT_{i,d}$ dividends paid by the stock itself and by firms in the same two-digit SIC industry. We then aggregate to the annual level using the average across the four quarters of year t, results are unchanged summing within year.

B Dividend Reinvestment

The instrumental variable we use to identify plausibly non-fundamental demand shocks is the Hypothetical Dividend Induced Trading ($HDIT$) from mutual funds reinvestment of dividends. As detailed in Section B, the underline assumption for this source of variation in stock prices is that mutual funds *reinvest* the cash they receive from dividends. We find empirical support for this assumption. First, we follow the approach used to validate similar assumptions in the literature (e.g., Lou (2012)). Specifically, we run the following pooled regression on our sample of mutual fund-stock-quarter holdings:

$$\frac{\Delta Shares_{j,q}(i)}{Shares_{j,q-1}(i)} = \alpha_j + \alpha_q(i) + \beta \, \frac{C_{j,q}}{AUM_{j,q}} + \varepsilon_{j,q}(i), \tag{B.1}$$

where $Shares_{j,q}(n)$ denotes the number of shares held by mutual fund j in quarter q, for stock i, $C_{j,q}$ is the total dividends payment received by fund j in quarter q (eq. 10), and $AUM_{j,q}$ is a fund's total assets under management. If mutual funds reinvest dividends proportionally to their existing holdings, we should find $\beta = 1$. If they do not reinvest any of the cash received from dividends, we should expect $\beta = 0$. We follow van der Beck (2024) in excluding large trades that exceed 100% of previous holdings ($\Delta Shares_{j,q}(i) < 1$) and excluding fund holdings below 0.1% of total assets under management. Columns (1) and (2) in Table B.1 show the results in our sample. The estimates for β are close to 1, suggesting that mutual funds reinvest dividends proportionally in their existing assets. On average, when funds in our sample receive an inflow from dividends, they scale up their existing portfolio positions proportionally.

Second, we directly test whether mutual funds' cash holdings change after receiving dividends. If funds exert a certain degree of discretion in reinvesting dividend payments, they might use dividends to increase their cash buffer. To test this, we regress fund-quarter cash holdings on dividend repayments:

$$\frac{Cash_{j,q}}{AUM_{j,q}} = \alpha_j + \alpha_q + \delta \frac{C_{j,q}}{AUM_{j,q}} + \epsilon_{j,q}(i), \tag{B.2}$$

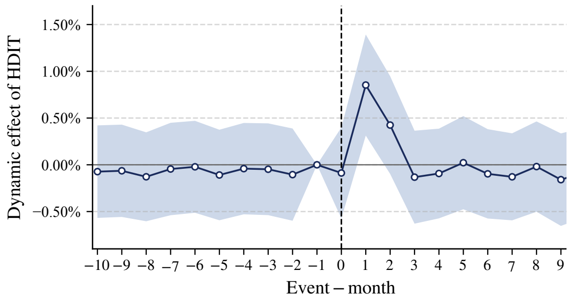
where $Cash_{j,q}$ denotes the cash holdings of fund j in quarter q. If mutual funds use dividend repayments to increase their cash buffer, we should find $\delta = 1$. On the other hand, if, on average, funds do not maintain dividend repayments as cash, we should expect $\delta = 0$. Columns (3) and (4) in Table B.1 report the results. On average, fund managers do not increase their cash holdings when receiving dividend repayments. This evidence corroborates results in other papers using a similar approach to identify non-fundamental demand shocks.

	$\Delta Shares_{j,q}(i)/Shares_{j,q-1}(i)$		$Cash_{j,q}/AUM_{j,q}$	
	(1)	(2)	(3)	(4)
$C_{j,q}/AUM_{j,q}$	0.952***	0.952***	0.000	0.001
	(0.257)	(0.257)	(0.000)	(0.000)
$Flow_{j,q}/AUM_{j,q}$		0.000		0.008***
		(0.000)		(0.001)
Firm FE	✓	✓	✓	✓
Time FE	×	×	✓	✓
Stock×Time FE	✓	✓		
Outcome mean	0.219	0.219	0.034	0.034
Outcome SE	0.478	0.478	0.107	0.107
Obs.	23,441,516	23,441,516	1,867,721	1,867,721
Adj. R^2	0.515	0.515	0.573	0.573

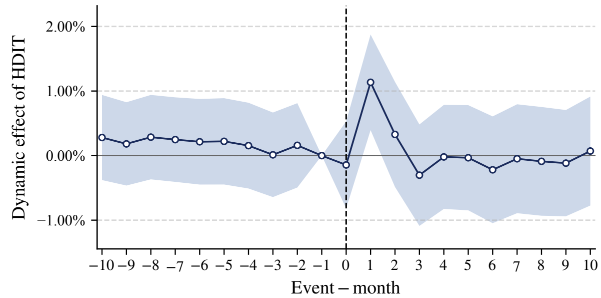
Table B.1: Dividend Reinvestment: This table shows empirical support for the assumption that mutual funds reinvest dividend payments. Column (1) reports results for the fund-stock-quarter regression in equation B.1. Column (2) further controls for mutual fund flows. Columns (3) and (4) show results for the fund-quarter regression in equation B.2. The quarterly sample is from 1980 to 2023. Standard errors (in parentheses) in columns (1) and (2) are two-way clustered by fund and stock-quarter, while standard errors in columns (3) and (4) are two-way clustered by fund and quarter. *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

C Appendix Figures

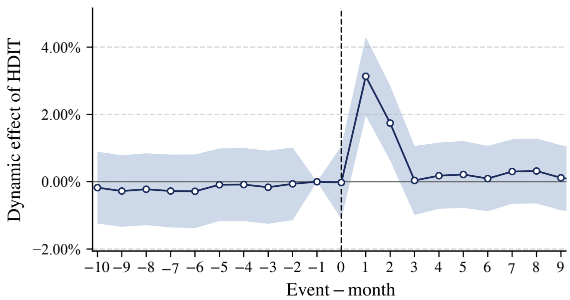
Figure C.1: Word Cloud of Data Technologies' Descriptions. The figure shows the word cloud of data plugins' features in our full sample. For each data technology, we obtain a short description of what the technology allows to do (e.g., ipstack: "*provides IP to geolocation APIs and global IP database services*") and plot the word cloud where larger fonts indicate a higher word frequency.



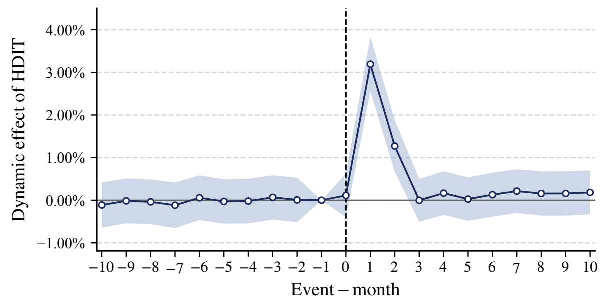
A. HDIT Effect in the 1980s.



B. HDIT Effect in the 1990s.



C. HDIT Effect in the 2000s.



D. HDIT Effect in the 2010s.

Figure C.2: Dynamics of the effects of HDIT on prices.. This figure plots the monthly average abnormal returns of stocks subject to Hypothetical Dividend Induced Trading (HDIT) price pressure around the event-month, where an event is defined as a firm-month observation in the top decile of HDIT within-year distribution. For each month in the sample, we take stocks in the top decile of the within-year HDIT distribution, and plot a coefficient regression from a linear regression of stock returns on even-month dummies, with firm×cohort fixed effects. We plot the results for years between 1980 to 1989 (Panel A), from 1990 to 1999 (Panel B), from 2000 to 2009 (Panel C), and from 2010 to 2023 (Panel B). All standard errors are two-way clustered by firm and day.

D Appendix Tables

	$\mathbb{P}\{\text{Adoption}\}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$log(\text{Mkt Cap})_{i,t}$	0.004*	0.002	0.001			
	(0.002)	(0.002)	(0.002)			
Inst. Ownership$_{i,t}$	0.035***	0.029**	0.027**			
	(0.013)	(0.014)	(0.012)			
Size$_{i,t}$	-0.001	0.000	0.002			
	(0.002)	(0.002)	(0.002)			
Age$_{i,t}$	0.001	0.000	0.000			
	(0.001)	(0.001)	(0.001)			
HDIT$_{i,t}$		0.004	0.003			
		(0.003)	(0.002)			
Peers' Adoption$_{i,t}$			0.919***			
			(0.024)			
$log(\text{Mkt Cap})_{i,t-1}$				0.003	0.001	0.001
				(0.002)	(0.002)	(0.002)
Inst. Ownership$_{i,t-1}$				0.029**	0.026*	0.027**
				(0.012)	(0.013)	(0.012)
Size$_{i,t-1}$				-0.001	0.000	0.002
				(0.002)	(0.002)	(0.002)
Age$_{i,t-1}$				0.001	0.000	0.000
				(0.001)	(0.001)	(0.001)
HDIT$_{i,t-1}$					0.003	0.001
					(0.003)	(0.002)
Peers' Adoption$_{i,t-1}$						0.696***
						(0.037)
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Obs.	127,518	110,972	110,972	115,780	101,030	101,030
Adj. R^2	0.818	0.817	0.840	0.815	0.813	0.826

Table D.1: Probability of adopting a data-related tool: This table shows results of panel regression in which the dependent variable is the probability that a firm adopts a data-related tool. The regressors are the *log* market capitalization (*log*Mkt Cap), the share of total shares outstanding held by institutional investors (Inst. Ownership), a firm's size (*log* assets), age, Hypothetical Dividend Induced Trading ($HDIT$), and the share of peers with a data-related tool installed (Peers' Adoption within SIC 3-digits). The annual sample is from 1980 to 2023. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX			Total$Q$	Total$Q\times$ DATA
	(1)	(2)	(3)	(4)	(5)
Total$Q_{i,t-1}$	0.023***	0.028***	0.118***		
	(0.003)	(0.004)	(0.005)		
Total$Q_{i,t-1}\times$ DATA$_{i,t-1}$	-0.010***	-0.014***	-0.038***		
	(0.002)	(0.005)	(0.010)		
DATA$_{i,t-1}$	0.063***	0.047***	-0.284***	0.079	0.336***
	(0.015)	(0.015)	(0.063)	(0.087)	(0.035)
Size$_{i,t-1}$		-0.056***	-0.074***	-0.138***	-0.012**
		(0.004)	(0.005)	(0.035)	(0.005)
Cash Flow$_{i,t}$		0.026***	0.031***	0.204***	0.013***
		(0.003)	(0.003)	(0.016)	(0.004)
HDIT$_{i,t-1}$				0.606***	0.014
				(0.073)	(0.008)
HDIT$_{i,t-1}$ $\times$ DATA$_{i,t-1}$				-0.087	0.169***
				(0.084)	(0.019)
Estimator	2SLS	2SLS	OLS	First-stage	First-stage
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.354	0.354	0.354	1.777	1.777
Outcome SE	0.460	0.460	0.460	3.786	3.786
Obs.	92,384	92,421	112,513	92,384	92,384
Adj. R^2	0.051	0.073	0.303	0.524	0.426
Kleibergen-Paap F-stat	16.425	16.144			

Table D.2: Robustness of Main Results using Total Q: This table shows robustness tests for the main results using Total Q (Peters and Taylor, 2017). The dependent variable in columns (1) to (3) is CAPEX. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Total Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data tool in place in year t (DATA$_{i,t}$), and controls for firm-year characteristics. All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPX/Total Assets			
	(1)	(2)	(3)	(4)
$Q_{i,t-1}$	0.003***	0.005***	0.012***	0.011***
	(0.000)	(0.001)	(0.001)	(0.001)
$Q_{i,t-1} \times$ $\text{DATA}_{i,t-1}$	-0.001***	-0.003***	-0.007***	-0.007***
	(0.000)	(0.001)	(0.001)	(0.001)
$\text{DATA}_{i,t-1}$	0.019***	0.016***	0.009***	-0.028***
	(0.004)	(0.004)	(0.002)	(0.008)
$\text{Size}_{i,t-1}$		-0.008***		-0.010***
		(0.001)		(0.001)
Cash $\text{Flow}_{i,t}$		0.004***		0.004***
		(0.000)		(0.001)
Estimator	2SLS	2SLS	OLS	OLS
Firm FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Outcome mean	0.060	0.060	0.060	0.060
Outcome SE	0.079	0.079	0.079	0.079
Obs.	100,205	101,629	128,216	122,905
Adj. R^2	0.041	0.061	0.528	0.545
Kleibergen-Paap F-stat	36.531	37.511		

Table D.3: Robustness using different measure of investment: This table shows robustness tests for the main results in the paper, using CAPX over Total Assets as measure of investment. The dependent variable is CAPX over Total Assets across all specifications. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Total Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) and (4) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data tool in place in year t ($\text{DATA}_{i,t}$), and controls for firm-year characteristics. All regressors are standardized by within-firm variation. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	Intangible Investment					
	(1)	(2)	(3)	(4)	(5)	(6)
$Q_{i,t-1}$	0.007***	0.003***	0.003***			
	(0.000)	(0.001)	(0.001)			
$Q_{i,t-1} \times$ $\textsc{Data}_{i,t-1}$	-0.002***	0.000	-0.001			
	(0.001)	(0.000)	(0.001)			
Total$Q_{i,t-1}$				0.005***	0.003***	0.004***
				(0.000)	(0.001)	(0.001)
Total$Q_{i,t-1} \times$ $\textsc{Data}_{i,t-1}$				-0.001*	0.000	-0.001
				(0.001)	(0.000)	(0.001)
Estimator	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Controls	✓	×	✓	✓	×	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Outcome mean	0.221	0.221	0.221	0.221	0.221	0.221
Outcome SE	0.102	0.102	0.102	0.102	0.102	0.102
Obs.	55,676	47,808	46,634	55,423	47,580	46,418
Adj. R^2	0.601	0.015	0.046	0.610	0.035	0.065
Kleibergen-Paap F-stat		56.55	53.90		17.016	16.893

Table D.4: Placebo on R&D and Intangible Investment: This table shows the results of panel regression in which the dependent variable is the one-year-ahead intangible investment, where intangible investment is defined following Peters and Taylor (2017) as R&D + 0.30×SG&A divided by Intangible Capital. Column (1) and (4) show results for the OLS estimates, while columns (2), (3), (5), and (6) reports results for the instrumental variable (IV) estimates where Tobin's Q and TotalQ are instrumented by the Hypothetical Dividend Induced Trading (HDIT). The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\textsc{Data}_{i,t}$), and controls for firm-year characteristics. The control variables (omitted for brevity) include a lagged firm's size (*log* assets) and contemporaneous cash flows over assets. All regressors are standardized by within-firm variation. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	CAPEX			Q	$Q\times$ DATA
	(1)	(2)	(3)	(4)	(5)
$Q_{i,t-1}$	0.013***	0.017***	0.085***		
	(0.002)	(0.002)	(0.005)		
$Q_{i,t-1}\times$ DATA$_{i,t-1}$	-0.005***	-0.007***	-0.037***		
	(0.001)	(0.001)	(0.009)		
DATA$_{i,t-1}$	0.102***	0.096***	-0.112*	0.085*	1.110***
	(0.021)	(0.019)	(0.060)	(0.046)	(0.046)
Size$_{i,t-1}$		-0.048***	-0.051***	-0.359***	-0.029***
		(0.004)	(0.005)	(0.019)	(0.007)
Cash Flow$_{i,t}$		0.033***	0.032***	0.084***	0.007***
		(0.003)	(0.002)	(0.011)	(0.002)
HDIT$_{i,t-1}$				0.566***	0.020***
				(0.031)	(0.004)
HDIT$_{i,t-1}$ $\times$ DATA$_{i,t-1}$				-0.090**	0.316***
				(0.037)	(0.027)
Estimator	2SLS	2SLS	OLS	First-stage	First-stage
Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.353	0.353	0.353	1.985	1.985
Outcome SE	0.460	0.460	0.460	1.772	1.772
Obs.	100,206	100,206	121,055	100,206	100,206
Adj. R^2	0.037	0.056	0.286	0.611	0.814
Kleibergen-Paap F-stat	92.115	87.764			

Table D.5: Robustness of Main Results excluding Google Analytics: This table shows robustness excluding Google Analytics from the set of data-related tools. The dependent variable is the one-year-ahead firm's investment. The dependent variable in columns (1) to (3) is CAPEX. Columns (1) and (2) reports results for the instrumental variable (IV) estimates, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (3) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data technology in place in year t (DATA$_{i,t}$) excluding Google Analytics, and controls for firm-year characteristics. All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

	$E_{i,t+1}/A_{i,t}$		$E_{i,t+3}/A_{i,t}$		$E_{i,t+5}/A_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$log(\text{Mkt Cap}/A)_{i,t}$	0.017***	0.016***	0.007**	0.005	0.019***	0.017***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
$log(\text{Mkt Cap}/A)_{i,t} \times$ $\text{DATA}_{i,t}$	-0.005**	-0.007**	-0.003	-0.002	-0.001	0.006
	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Controls	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Time FE	✓	×	✓	×	✓	×
Industry×Time FE	×	✓	×	✓	×	✓
Outcome mean	0.028	0.028	0.055	0.055	0.084	0.084
Outcome SE	0.222	0.222	0.267	0.267	0.314	0.314
Obs.	142,111	140,199	116,962	115,400	96,767	95,467
Adj. R^2	0.814	0.817	0.627	0.635	0.562	0.571

Table D.6: Price Informativeness and data-related tools (Bai et al., 2016): This table shows results of panel regression in which the dependent variable is the h-years-ahead firm's earnings over current assets, for $h = \{1, 3, 5\}$. The regressors are a the *log* market capitalization over assets, and an interaction term with a dummy equal to one if a firm i has a data technology in place in year t ($\text{DATA}_{i,t}$), and controls for firm-year characteristics (omitted for brevity). The coefficient on the *log* market capitalization over assets captures price informativeness (Bai et al., 2016). The control variables include a firm's current earnings over assets, size (*log* assets), and cash flows over assets in quarter t. Industry×Time fixed effects use SIC 2-digits industry codes. The annual sample is from 1980 to 2023. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.

Dependent variable:	Jump Ratio (Weller, 2018)		CAPEX		
	(1)	(2)	(3)	(4)	(5)
$Q_{i,t-1}$		0.001	0.022***	0.026***	0.096***
		(0.002)	(0.005)	(0.006)	(0.007)
$Q_{i,t-1}\times$ $\text{D}_{\text{ATA}_{i,t-1}}$			-0.012***	-0.014**	-0.043***
			(0.004)	(0.007)	(0.010)
$\text{D}_{\text{ATA}_{i,t-1}}$	-0.263***	-0.276***	0.129***	0.117***	-0.037
	(0.011)	(0.011)	(0.034)	(0.033)	(0.079)
$\text{Size}_{i,t-1}$		0.039***		-0.039***	-0.040***
		(0.006)		(0.007)	(0.007)
Cash $\text{Flow}_{i,t}$		0.090***		0.026***	0.025***
		(0.015)		(0.005)	(0.004)
Estimator	Stacked DiD		2SLS	2SLS	OLS
Cohort×Firm FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Outcome mean	0.361	0.361	0.368	0.368	0.368
Outcome SE	0.336	0.336	0.478	0.478	0.478
Obs.	146,301	141,969	152,834	152,834	176,597
Adj. R^2	0.198	0.208	0.059	0.070	0.270
Kleibergen-Paap F-stat			50.996	54.582	

Table D.7: Robustness of Main Results on subsample with positive effect on Price Informativeness: This table shows robustness of the main results for a subsample in which the adoption of data-related tools is associated with an increase in measures of price informativeness Weller (2018). We identify the subsample for this robustness using influence functions for the effect of data-related tools on the Weller (2018) jump ratio. We compute the influence of each firm on the estimates (DFBETA) and keep only firms above the median, i.e., we isolate firms for which the effect of data-related tools on price informativeness is positive. Columns (1)-(2) show the effect of data technology on stock price jump ratio around earning announcements (Weller, 2018). A higher jump ratio is associated with less informative stock prices. Columns (3) and (4) reports results for the instrumental variable (IV) estimates in the subsample, where Tobin's Q is instrumented by the Hypothetical Dividend Induced Trading (HDIT). Column (5) reports the OLS results. The regressors are a dummy equal to one if a firm i has a data technology in place in year t ($\text{D}_{\text{ATA}_{i,t}}$) excluding Google Analytics, and controls for firm-year characteristics. All estimates are in the stacked sample (Gormley and Matsa, 2011). All regressors are standardized by within-firm variation. Columns (4) and (5) report the first-stage results. The annual sample is from 1980 to 2023. All columns include firm and year fixed-effects. All standard errors are two-way clustered by firm and year (in parentheses). *, **, and *** denote statistical significance at the 10%, 5% and 1% respectively.