Breaking the Data Chain: The Ripple Effect of Data Sharing Restrictions on Financial Markets*

Simona Abis† Bo Bian‡ Huan Tang§

April 30, 2026

Abstract

Privacy regulation is typically studied through its effects on firms and consumers, while its implications for financial markets remain unexplored. We use Apple's App Tracking Transparency (ATT) to examine whether privacy-driven data sharing restrictions spill over to capital markets by reducing the precision of widely used signals. After ATT, analysts relying more heavily on such data become less accurate, and mutual funds shift attention away from affected stocks. Consequently, firms more exposed to these market participants exhibit weaker price efficiency. Our findings reveal a new fragility in financial markets as surveillance-like alternative data becomes increasingly vulnerable to future regulatory disruption.

Keywords: Alternative Data, Mobile Apps, Data Sharing, Privacy Regulation, Information Friction, Analyst Forecast, Mutual Fund

JEL Codes: G12, G14, G18 G23, G29, O16

*We thank Yakov Amihud, Jennie Bai, Maxime Bonelli, Andrea Buffa, Sean Cao, Tony Cookson, Zhi Da, Olivier Dessaint, Mengqiao Du, Vivian Fang, Thierry Foucault, Itay Goldstein, Clifton Green, Jillian Grennan, Allen Hu, Jiekun Huang, Byoung-Hyoun Hwang, Ryan Israelsen, Wei Jiang, Yukun Liu, Kai Li, Connie Mao, Ye Mao, Yang Song, Gordon Phillips, Laura Veldkamp, Lulu Wang, Ting Xu, Gloria Yu, Xiaoyun Yu, Xin Zheng, and seminar/conference participants at Indiana University, University of Colorado Boulder, ITAM, NBER Summer Institute (Big Data and High-Performance Computing for Financial Economics), Penn/NYU Conference on Law and Finance, University of Kentucky Finance Conference, and UBC Summer Conference for their helpful comments and suggestions. Bo Bian acknowledges financial support from the Social Sciences and Humanities Research Council of Canada (Grant Number: 435-2025-0700). Simona Abis gratefully acknowledges financial support provided by Leeds School of Business. We thank Quinn Danielson, Rohan Joseph, Yiming Ma, and Bolin Xu for excellent research assistance.

†University of Colorado Boulder, Leeds School of Business, simona.abis@colorado.edu

‡University of British Columbia, Sauder School of Business, bo.bian@sauder.ubc.ca.

§University of Pennsylvania, the Wharton School, huan.ht.tang@gmail.com

1 Introduction

The modern economy is increasingly reliant on personal data. Firms collect, aggregate, and exchange granular consumer data to inform marketing, product customization, and competitive strategy. Much of its commercial value derives from linking data across sources and contexts to construct comprehensive behavioral profiles.[1] These practices have attracted growing regulatory scrutiny, resulting in interventions such as the General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA), alongside platform-level policies curtailing data collection, linkage, and sharing. The existing literature has studied this tension extensively, focusing on effects on firms' advertising effectiveness and revenue, and on consumers' privacy and welfare. What this debate has largely overlooked is that the resulting data products have also become critical inputs to financial markets, shaping analysts' forecasts, fund managers' trades, and ultimately asset prices. Whether privacy interventions spill over into price discovery and capital allocation remains an open and consequential question, with direct implications for privacy policy design and for how financial institutions manage alternative-data provenance and quality risks.

We study this question through the lens of mobile-generated alternative data, a widely used input to financial forecasts.[2] These data provide a useful laboratory for causal analysis: Apple's 2021 App Tracking Transparency (ATT) policy introduced an unexpected, ecosystem-wide privacy shock that disrupted the cross-app linkages underpinning their informativeness about future fundamentals, without fully eliminating mobile-generated data. We show that this privacy-driven policy spilled over to financial markets: it increased forecast errors among market participants that relied on those signals and weakened price efficiency.

ATT replaced default cross-app tracking with an explicit opt-in regime. This sharply limited access to device-specific identifiers that had previously allowed firms to link behavior across apps, websites, and datasets. The disruption was large: iOS devices account for over

[1] Mobile devices play a central role in enabling this linkage, by connecting a single user's purchasing behavior, location patterns, and digital engagement across apps, platforms, and time.

[2] Many forms of alternative data, including app usage, geolocation, mobile transaction records, advertising engagement, and real-time customer reviews, are collected through mobile app activity.

50% of the U.S. smartphone market (comScore, 2024), yet only 4–18% of users opted in to tracking (Balasubramanian, 2022). The economic importance of this shock comes from the role of cross-app linkages in signal construction. These linkages allow firms, data vendors, and market participants to turn fragmented consumer activity into leading indicators of demand and firm performance. When they weaken, the informational value of these datasets falls: targeting may become less precise, tracked foot-traffic panels less representative, and observed digital activity less tightly connected to economic outcomes. This mechanism is particularly intuitive in app traffic. Before ATT, for instance, McDonald's app downloads could be informative about sales because the firm could identify and target users with recent indicators of purchase intent, such as browsing competitors' menus on DoorDash or Uber Eats, passing near a McDonald's location, or engaging with late-night snack content on Instagram, and convert that traffic into app activity and ultimately purchases.[3] After ATT, downloads may still occur, but a larger share reflects untargeted traffic and weaker purchase intent. As a result, app traffic becomes a noisier signal of subsequent sales and profits.

Our empirical analysis combines detailed app-level activity data with firm fundamentals, sell-side analyst forecasts, and mutual fund portfolio holdings to trace how such an upstream data-sharing policy propagates through financial markets.

We begin by validating that ATT weakened the informational content of mobile-generated signals for firm fundamentals. We focus on one of the most direct mobile signals: app traffic, measured by new app downloads over a quarter. Prior to ATT, downloads positively predicted sales growth and unexpected earnings; after ATT, this predictive relation weakens substantially. One concern is that this loss of predictability merely reflects ATT's impact on firms' underlying real performance. While ATT may well affect fundamentals, a shift in their level or volatility would not mechanically eliminate the slope linking downloads to subsequent sales and earnings. Instead, we find that app traffic no longer forecasts future fundamentals, consistent with it becoming less informative about firm performance. More importantly,

[3]McDonald's discloses in its app privacy labels that it collects data to track users across apps and websites owned by other companies. See https://apps.apple.com/us/app/mcdonalds/id922103212.

in the analyses that follow, we isolate the impact of ATT on information production by exploiting within-firm-time variation across sell-side analysts and mutual funds with varying reliance on mobile-generated signals, thereby differencing out ATT's direct effect on firm fundamentals.

Having established that ATT reduced the informational content of mobile-generated signals, we examine how market participants respond when widely used signals lose precision. Agents with limited analytical capacity, whether constrained by cognition or budgets, must redirect scarce resources. Sell-side analysts, whose coverage is likely sticky in the short run, face the multitasking trade-off of Dessaint et al. (2024): they can shift effort across forecast horizons but not easily across stocks, so noisier data should reduce forecast accuracy for affected firms unless switching costs are negligible. In contrast, in the rational-inattention model of Kacperczyk et al. (2016), portfolio managers flexibly adjust portfolio allocations, reallocating attention, and thus capital, to stocks with more precise signals. Studying both groups lets us separate two relevant margins for prices: information production conditional on attention and attention allocation across stocks. Accordingly, ATT should worsen information production for exposed firms by reducing analysts' forecast accuracy and inducing funds to reallocate efforts away from those stocks. We test these predictions by studying changes in analyst forecast accuracy and mutual-fund stock picking around ATT.

We begin with sell-side analysts, as we directly observe their forecast histories and information sources. Following the above prediction, for a given forecast horizon, we analyze the *within-stock-quarter* forecast gap between analysts with different reliance on mobile signals.

We proxy for analyst specialization on mobile data by searching reports for mentions of app-related terms or providers (e.g., "daily active users", or "App Annie") and define *keyword intensity* as the share of an analyst's reports containing such references. For each earnings forecast we calculate the analyst's error relative to same-quarter consensus and include firm-quarter fixed effects, removing any direct effect of ATT on firm fundamentals. Hence, a post-ATT widening of the forecast error gap between high- and low-specialization analysts isolates the extra forecast noise faced by the analysts most dependent on the impaired signal.

We show that, pre-ATT analysts most reliant on mobile data posted smaller forecast errors than their non-specialist peers, while these relative errors significantly increase post-ATT. This change is not preceded by a detectable pre-trend, supporting a causal interpretation. The results are robust to adding analyst-by-firm fixed effects, suggesting that they do not reflect specialists sorting into more volatile firms. The same specialists show no such shift when covering placebo firms unlikely to rely on app traffic, further ruling out analyst-level confounds. Interestingly, specialized analysts end up, on average, less accurate than non-specialists, suggesting that their pre-ATT advantage turns into a post-ATT disadvantage.

This aggregate reversal masks important differences in how specialists respond to the shock. We sort them into switchers, those whose app-related keyword usage falls by more than the median after ATT, and non-switchers, comprising the remainder. Switchers, who evidently face lower switching costs, reallocate attention away from mobile metrics and merely surrender their former edge, posting errors similar to those of non-specialists. Non-switchers, likely burdened by higher switching costs, keep scarce effort tied to a less informative signal and become less accurate than analysts who never used mobile data.

Moreover, the deterioration in forecast accuracy is strongest at short horizons, becoming statistically insignificant beyond three quarters, consistent with mobile-generated alternative data being primarily informative about near-term outcomes (Dessaint et al., 2024).

An immediate question is whether these forecast differences matter for prices, since analysts do not trade themselves. To address this, we ask whether the market reacts differently to analyst reports that feature app-usage metrics. If investors view a signal as informative, reports that emphasize it should elicit larger price moves; if the signal's quality deteriorates, that extra reaction should fade. Consistent with this pattern, before ATT, reports containing more app keywords triggered significantly larger cumulative abnormal returns, suggesting that investors regarded the underlying signal as especially reliable. After ATT, the premium disappears, indicating that the market no longer values the now-noisier signal. The loss of this "app-keyword premium" mirrors the post-ATT decline in specialists' forecast accuracy and suggests that the deterioration in mobile-data-based research did propagate to prices.

We then turn to mutual funds, which let us observe directly actors who trade and thus move prices. This setting complements the analyst tests by showing how portfolio considerations shape the transmission of degraded information to prices. The trade-off is lower transparency: we do not observe funds' information sources or forecasts, so both specialization on mobile signals and forecasting ability must be inferred from holdings.

We first test whether fund trades respond to app-traffic signals and use the same framework to construct a fund-level measure of specialization. Using monthly holdings, we isolate trade-induced changes in portfolio weights and relate buy-side changes to lagged abnormal downloads. Prior to ATT, fund buying is positively related to past abnormal downloads; after ATT, this relation becomes statistically insignificant, consistent with funds recognizing the deterioration in signal quality and reducing its use in trading. We then estimate this relation separately for each fund in the pre-ATT period, with month fixed effects, and take the within-fixed-effects adjusted R^2 as signal reliance.[4] Our *specialization* measure ranks funds based on this reliance, scaled by the fund's pre-ATT average portfolio share invested in app-exposed stocks, thereby capturing both how strongly a fund's trades load on mobile signals and the share of its portfolio over which those signals are relevant.

We then test whether funds that relied more on mobile signals before ATT suffered a larger decline in comparative picking skill on affected stocks. For each fund-month-stock observation, we multiply the stock's benchmark-adjusted active weight by its idiosyncratic return over the next 1-18 months. This measure captures the covariance between active weights and future residual returns, with higher values indicating better selection.[5] We classify stocks with an associated app as *exposed* to ATT, and all others as *unexposed.* Since managers allocate limited capacity across these two groups, our object of interest is the gap between average exposed and unexposed picking ability, which captures where forecasting

[4]Our measure is closest to those of Sheng et al. (2024) and Kacperczyk et al. (2008).

[5]See Kacperczyk et al. (2014) and Bonelli and Foucault (2023) for related measures.

effort is most effective. We estimate a triple-difference specification comparing pre- versus post-ATT months, exposed versus unexposed stocks, and specialist versus non-specialist funds. The coefficient on the triple-interaction term captures whether, after ATT, specialists lose more of their edge on exposed stocks relative to unexposed stocks than non-specialists do. Since both specialization and picking ability are inferred rather than directly observed, measurement noise works against finding significant results.

We find that this coefficient is consistently negative and highly significant, indicating that ATT eroded specialists' comparative edge on exposed stocks relative to non-specialists. The effect is economically large: in the 12-month baseline, it is comparable in magnitude to the allocative shift we attribute to ATT's direct impact on firm fundamentals. The estimate remains virtually unchanged with stock-month fixed effects, which absorb any direct impact of ATT on firm performance. Thus, the deterioration of mobile-data precision implies a sizable incremental drop in specialized funds' relative picking ability. Event-time plots show little pre-trend and a persistent decline once post-ATT months enter the forward-looking return window. Estimates remain negative out to an 18-month horizon, suggesting that managers shifted attention across stocks rather than across short- and long-term signals for the same firms. Together, these patterns show that specialists redirected scarce analytical capacity away from exposed mobile-app stocks when their informational edge eroded.

Finally, we examine whether the deterioration of mobile-generated signals translates into lower firm-level price efficiency. If ATT impaired specialists' information production, both by reducing signal precision and by redirecting attention, firms more exposed to these specialists ex ante should experience a relative worsening of their information environment. We test this prediction using bid-ask spreads and post-earnings announcement volatility, which capture ex-ante information asymmetry and post-disclosure residual uncertainty. Consistent with this mechanism, firms followed by specialized analysts or funds exhibit narrower spreads and lower volatility prior to ATT, but experience relative increases in both measures afterward, indicating greater uncertainty and investor disagreement.

More broadly, ATT likely impaired more than app-usage metrics. Other mobile-collected datasets relying on user identifiers, such as geolocation-based foot traffic, may also have deteriorated, so the capital market effects we document likely reflect a broader loss of precision across mobile signals. Such DGP-driven precision shocks can generate systematic mispricing when affected stocks share common traits. In our setting, app-intensive firms have higher market-to-book ratios and attract funds with stronger growth tilts. If information deteriorates disproportionately along the growth-value spectrum, or any exposure-linked dimension, mispricing need not wash out in equilibrium, adding a new layer of fragility to asset prices.

Our results expose a core tension of the digital economy: the same user-generated data that powers marketing and competitive intelligence also shapes investors' expectations. Yet unlike traditional financial disclosures, these data are generated outside the financial system as a byproduct of consumer activity, leaving market participants dependent on informational inputs whose production they cannot easily influence or safeguard. Regulating these data streams is therefore complex. ATT, though designed to protect consumer privacy rather than to affect finance, disrupted the data-generating process and eroded inputs to price discovery. Even if markets were to adapt and any single distortion was transient, the growing reliance of market participants on alternative data implies a persistent vulnerability as these data remain exposed to recurring regulatory and platform shocks. The takeaway is twofold: investors should treat alternative data as fragile, and policymakers should recognize that interventions in the data economy can spill over into price discovery and capital allocation.

Related Literature. This paper contributes to several strands of the literature. A burgeoning line of work examines the economic implications of data privacy regulations. Much of this literature focuses on the European General Data Protection Regulation (GDPR), connecting it to app entry and exit (Janssen et al., 2022), venture capital financing (Jia et al., 2021), shifts in web traffic (Goldberg et al., 2024), and firms' capacity to collect, monetize, and store consumer information (Aridor et al., 2023; Bessen et al., 2020; Demirer et al., 2024; Peukert et al., 2022; Ramadorai et al., 2025).[6] Beyond Europe, research has analyzed the California Consumer Privacy Act, documenting its effects on mortgage lending (Doerr et al., 2023) and firm risk (Wu, 2023). More recently, attention has shifted to Apple's privacy initiatives, such as the introduction of App Store privacy labels (Bian et al., 2021) and the App Tracking Transparency framework (Kesler, 2022; Cheyre et al., 2023; Bian et al., 2023, 2024), with studies examining their direct effects on firms and consumer outcomes.

A gap in this literature is the absence of a financial market perspective from the above discussion. Digital footprints matter not only for firms' production decisions and consumers' behavior, but also for how financial market participants acquire, process, and trade on information. It is well established that data-rich environments improve financial market efficiency and facilitate capital allocation (Bai et al., 2016; Begenau et al., 2018; Farboodi et al., 2022). Our paper is the first to show that privacy-driven restrictions spill over into financial markets by degrading the informational inputs on which market participants rely. We document how such restrictions propagate through forecasting, trading, and price efficiency. While ATT is not a government-led privacy regulation, it shares the core feature of many such policies, making our findings relevant for understanding the informational cost of such policies for financial market.

Our paper also connects with a rich body of research examining the use of alternative data in capital markets. Prior work documented how diverse datasets—e.g., satellite imagery (Zhu, 2019; Katona et al., 2024; Ke, 2023), customer and employer reviews (Huang, 2018; Green et al., 2019), real-time sales data (Froot et al., 2017; Blankespoor et al., 2022; Dichev and Qian, 2022; Jin et al., 2025; Du and Qian, 2024), EDGAR web traffic and Google searches (Chen et al., 2020; Da et al., 2011), robo-journalism (Blankespoor et al., 2018), app traffic (Chen et al., 2024), social media (Antweiler and Frank, 2004; Chen et al., 2014; Dessaint et al., 2024), weather (Hirshleifer and Shumway, 2003; Goetzmann et al., 2015), cloud records (Chang and Da, 2022), crowd-based corporate earnings forecasts (Da and Huang, 2020), and

[6] See Johnson (2022) for a review of the GDPR literature and the papers therein.

foot traffic (Jin et al., 2025)—can be exploited to improve forecasting and trading. This literature shows the broad *supply* of alternative data that can enhance price efficiency.

A complementary strand focuses on the *demand* side of alternative data, providing granular evidence on how market participants use these signals. The proliferation of alternative data vendors and the rise of mobile-based consumer surveillance have created new information channels for analysts and fund managers (Grennan and Michaely, 2021; Green and Zhang, 2024). For analysts, Chi et al. (2024) show that textual references to alternative data improve forecast accuracy, while Dessaint et al. (2024) highlight the trade-off between short- and long-term informativeness. For funds, Bonelli and Foucault (2023) show that a positive satellite imagery shock can displace skill by eroding active managers' stock-picking edge.

Relative to these two lines of work we add three main insights. First, while most studies celebrate the gains from new data, we ask what happens when an established dataset degrades. This "negative shock" perspective highlights a previously overlooked vulnerability of financial markets: as reliance on alternative data grows, adaptation to disruptions may be slow, and even successful adaptation leaves these inputs exposed to future regulatory and platform shocks.[7] Second, the granularity of our data and setting lets us trace the full transmission chain: we link mobile signals to firm fundamentals, track their use by analysts and funds, and show how their deterioration spills over into prices and firms' information environments. Third, Apple's ATT provides a clean, major shock to the DGP, allowing us to document causal evidence on how shocks to information flows governed by actors outside the financial sector propagate through financial markets. In doing so, we join the emerging natural-experiment literature on alternative data (Zhu, 2019; Bonelli and Foucault, 2023; Katona et al., 2024).

[7]This also distinguishes our setting from studies of temporary interruptions to retail trading or information processing, such as brokerage outages (Eaton et al., 2022) or suspension of ChatGPT (Bertomeu et al., 2025), which do not subject markets to the same lasting regulatory and data-provenance risks.

2 Institutional Background

2.1 The Rise of the Mobile App Economy

The mobile app market has expanded rapidly over the past decade, fueled by rising internet access, widespread smartphone adoption, and advances in AI and augmented reality (Statista, 2024b,c; Forbes, 2024). Mobile apps have become an important revenue source for firms. In the U.S., direct revenues reached \$153.04 billion in 2023 and are projected to grow to \$230.44 billion by 2027, 76% of which comes from mobile advertising (Statista, 2024a).[8]

Firms offering mobile apps are not just a niche group of technology companies but represent a broad and economically significant share of the corporate sector. Among Compustat firms, more than 1,000 operate at least one mobile app, together accounting for over 60% of total assets (Bian et al., 2024). These firms span nearly all industries: sectors such as Business Services, Retail, and Telecommunications are overrepresented, while Finance and Oil are notably underrepresented (Figure 1a). As shown in Figure 1b, the coverage of app-owning firms is substantial in many sectors: in 24 of the 48 Fama-French industries, they account for over half of the sector, whether measured by firm count or total assets.

2.2 How Mobile-Generated Signals Are Produced and Used

The rise of mobile apps has made them a central source of granular, real-time user-level data. Each user interaction generates a range of signals. The most direct and widely used are app-traffic data, such as installations, active users, and session frequency or length. Apps also generate location data, collected via embedded tracking code across apps and commercialized by firms such as SafeGraph and X-Mode, capturing users' physical movements in real time. Another example is transaction data, generated through in-app purchases, mobile wallets, or linked financial accounts and aggregated by providers such as Stripe and Plaid to track consumer spending across platforms and merchants.[9] Together, these interactions form a

[8] These exclude indirect benefits such as monetizing user data for optimization and third-party sales.

[9] Other examples include advertising data on ad views and clicks, review data on ratings and customer feedback, social interaction data on sharing and referrals, and retail investing data from platforms like Robinhood that reveal trading activity and investor sentiment.

pyramid of alternative datasets capturing diverse dimensions of consumer activity.

Technology providers and analytics companies clean and aggregate these raw signals. The primary use of such processed data is operational, helping app developers and businesses monitor user engagement and improve products. Firms such as Apptopia and Sensor Tower, though, also compile them across apps and industries, turning them into broader indicators of firm performance and market trends. Financial market participants then rely on these repurposed signals to forecast revenue growth, assess competitive dynamics, and guide investment decisions, even though they originate as a by-product of digital operations.

The Key Role of Tracking and Data Sharing. The quality of mobile-generated signals depends critically on users permitting tracking and cross-app data sharing, a long-standing industry practice. This sharing is enabled by persistent device identifiers, such as Apple's Identifier for Advertisers (IDFA) and Google's Advertising ID (GAID), which allow app developers and data vendors to observe and link user behavior across multiple sources and are distinctive to the mobile ecosystem.[10] With tracking enabled, firms and data vendors can collect activity well beyond any focal app. For example, location-data providers such as Kochava or X-Mode drew GPS signals from many unrelated apps running in the background, generating large streams of raw inputs later repackaged into alternative data products.

What makes these data valuable, however, is not merely collection, but linkage. Device identifiers allow firms to combine scattered signals (e.g., where a user goes, what she buys, and what she browses) into coherent profiles that help identify high-intent buyers, deliver personalized offers, and convert digital activity into revenues. The mapping from observed behavior to real economic outcomes is what gives app-level metrics their value for financial markets, allowing variables such as downloads to serve as leading indicators of firm performance. The same logic extends beyond app traffic. For example, platforms can use cross-app information to identify representative purchasers and selectively solicit reviews, turning dis-

[10] In contrast, website-generated data are much more fragmented, as cookies are typically site-specific and do not allow seamless tracking across domains.

persed opinions into structured indicators of product quality and consumer sentiment.[11]

2.3 ATT as a Privacy-Driven Shock to Mobile-Generated Signals

This data infrastructure changed sharply on April 26, 2021, when Apple introduced the App Tracking Transparency (ATT) framework as part of iOS 14.5. Traditionally, both Android and iOS followed an opt-out policy: sharing identifiers and other data (e.g., geolocation) by default unless explicitly disabled by the user. Under ATT, instead, apps must obtain explicit opt-in consent to share the IDFA or other personally identifiable data with other apps or third parties (Apple, 2024b). On first opening an app after ATT, users see a standardized pop-up asking whether to allow tracking "*across apps and websites owned by other companies,*" and noting that the data will be used for personalized advertising (Apple, 2024a).[12] Users can also disable tracking for all apps globally through system settings.

ATT's rollout had a wide-reaching impact because the IDFA had long been central to the mobile data ecosystem, and no viable substitute existed for linking user data across apps. Neither first-party data nor alternative identifiers could replace the scale, granularity, and real-time cross-app tracking and attribution it enabled, at least over the medium run.[13] Consistent with this, evidence discussed in Section 4 shows that firms did not fully circumvent or substitute for the IDFA after ATT, reinforcing the view that the policy materially impaired cross-app tracking and attribution. Opt-in rates remained low: within one month of release, only 11% of users globally and 4% in the U.S. allowed tracking, among those who saw the prompt. After one year, those rates had risen only modestly, to 25% and 18%, respectively (Balasubramanian, 2022). Crucially, as the first policy of its kind, ATT's disruption of user targeting appears to have been largely unanticipated by financial markets, as reflected in the strong negative market reaction at rollout documented by Bian et al. (2021).

Without tracking and data sharing, firms lose the ability to connect user activity across

[11] Amazon routinely solicits reviews from selected customers through post-purchase prompts and programs like Amazon Vine, which invite selected reviewers to provide detailed feedback.

[12] An example of such prompt is provided in Appendix Figure A.1.

[13] Even by the end of our sample period in June 2023, industry commentary suggested that ad efficiency remained at only about 25% of its pre-ATT level, implying a decline of as much as 75% (Source: Singular).

apps and contexts. As a result, mobile-generated signals may remain observable, but their informational value falls, because the cross-app linkages that once allowed firms to identify and target high-intent, profitable consumers disappear. The remaining subset of users who opt into tracking is also unlikely to be representative of the broader user base. If these "trackable" users are disproportionately low-intent (e.g., more curious but less likely to convert), then strategies optimized on their behavior may mis-target the broader population. This logic applies across a range of alternative datasets. For app traffic, firms can still observe installs and logins, but downloads may no longer predict higher sales and can even move in the opposite direction, as low-intent users inflate volumes without generating corresponding revenue. For foot traffic, ATT can shrink tracking panels and make them less representative. For customer reviews, it can limit platforms' ability to selectively solicit feedback from representative purchasers. In each case, the result is the same: signals remain available, but become less reliable leading indicators of firm performance.

More broadly, the rapid growth of the alternative data market has raised significant regulatory and privacy concerns in financial markets. Deloitte's 2017 report highlighted data provenance and compliance risks as key challenges (Deloitte, 2017); recent surveys confirm that regulatory scrutiny, data quality, and vendor integrity remain top concerns (Moss et al., 2023). A prominent example came in 2021, when the SEC charged App Annie, a leading provider of mobile data, with securities fraud for misusing confidential app information and selling it to trading firms without proper consent (SEC, 2021); according to Eagle Alpha, the case heightened awareness of the legal boundaries of alternative data use (Ryan, 2023).

ATT is one example of this vulnerability. Although it is a unilateral platform policy rather than a government regulation, it restricts tracking and data sharing in ways similar to GDPR or CCPA, illustrating how privacy-driven interventions can alter the quality of data on which financial markets rely. Such vulnerabilities extend beyond mobile data and formal regulation: platform choices such as recent changes at Glassdoor, Twitter/X, or Robinhood,[14] can abruptly alter the quality and availability of widely used datasets, reducing the reliability and breadth of alternative data in financial markets.

Building on the above discussion, our empirical setting combines Apple's ATT policy (an unexpected, large-scale restriction on tracking and data sharing) with widely used mobile signals such as app downloads, which are both high-frequency and firm-specific. This setting enables three tests. First, we validate that app-usage metrics predict firm performance and that ATT weakens their predictive content. Second, we examine how this deterioration affects the forecasting performance of sell-side analysts and fund managers. Third, we study how these disruptions spill over into firms' information environments and price efficiency.

3 Data

Our sample covers publicly listed firms from 2017Q1 to 2023Q2, spanning the pre- and post-ATT periods. We leverage a granular dataset of the mobile apps market, which we merge with firm financials, analyst forecasts, and mutual fund holdings.

Mobile Apps. We obtain mobile app usage data (e.g., downloads, active users) from Apptopia, a leading alternative data provider. Apptopia combines data from partner apps, user panels, and public sources (e.g., App Store rankings), and uses proprietary AI models to estimate usage metrics for over 7 million apps, including more than 3,000 publicly listed firms globally. Since 2019, these data can be purchased through Bloomberg's alternative data suite, making Apptopia one of the main sources of mobile usage data for investors. We aggregate usage across all apps developed by each firm.

Although ATT applies directly to iOS apps, focusing on overall traffic is crucial for two reasons. First, iOS users generate disproportionately higher revenue, making iOS the primary driver of firm performance relative to Android.[15] Moreover, since firms typically do not

[14]In 2023, Glassdoor introduced a requirement that users verify their real names, job titles, and employers, a change that risks deterring candid reviews and skewing submissions. Twitter/X ended free API access in 2023, and Robinhood discontinued publication of Robintrack data in 2020.

[15]E.g., AppsFlyer reports that non-gaming subscription ARPU on iOS is \$8.4 vs. \$1.5 on Android.

separately disclose iOS and Android revenue, platform-specific outcomes are difficult to isolate. Second, ATT's effects can spill over to Android because targeting and advertising strategies are often designed across platforms, so restrictions on iOS data may weaken cross-platform customer profiling.[16]

Analysts. We collect analyst forecasts from I/B/E/S, retaining quarterly (annual) forecasts issued within 90 (365) days of each earnings announcement, and keeping the most recent forecast for each analyst–ticker pair. Additionally, we use the LSEG Workspace (formerly Refinitiv Eikon) to search for mobile-data usage in the full text of analyst reports.

Funds. We obtain mutual fund data from the CRSP Survivorship-Bias-Free Mutual Fund dataset to identify active domestic equity funds. We exclude funds with less than $5 million in Total Net Assets (TNA) (Kacperczyk et al., 2008) and those with fewer than 12 months of observations. To address incubation bias, we drop observations prior to each fund's first offer date (Evans, 2010). We then collect holdings for those funds and further restrict the sample to funds holding at least 10 stocks and, on average, at least 80% of assets in common stock (Kacperczyk et al., 2008). When holdings are missing or only reported quarterly, we forward-fill them to a monthly frequency.[17]

Firms. We obtain quarterly accounting data from Compustat, financial market data from CRSP, and factors from Ken French data library.[18]

[16]For example, a food-delivery app may use data from iOS users to learn that consumers who recently browsed restaurant menus, live in dense urban areas, and respond to late-night promotions are especially likely to place high-value orders. It can then use that cross-platform profile to target similar users on Android. If ATT limits the ability to observe and link iOS behavior, the profile becomes less precise, weakening targeting on Android as well.

[17]In our sample period, the vast majority of funds disclose holdings monthly.

[18]Following Livnat and Mendenhall (2006), we require: earnings announcement dates to be reported; price per share to be available at fiscal quarter end and greater than $1; market (book) equity to be available at fiscal quarter end and larger than $5 million; and, if I/B/E/S forecasts are available, we require earnings announcement dates in Compustat and I/B/E/S not to differ by more than one day.

Firm-level Sample. The key sample of interest includes 809 firms with apps averaging at least 1,000 daily active users (DAU) over our sample period (*"App Sample"*). This excludes a long tail of low-activity apps, unlikely to be relevant for firms. Table 1 shows that firms in the "App Sample" are comparable to the 5,970 unique firms that do not have an app, with the exception that "App sample" firms have higher debt-to-asset and market-to-book ratios.

Variables Definition. A complete list of variables used in our analysis and their definitions is found in Table A.1.

4 Validation: ATT and the Signal Quality of App Downloads

In this section, we validate the key mechanism outlined in Section 2.3: ATT, by disrupting data linkage, impairs firms' ability to act based on detailed user profiles, diminishing the predictive power of mobile-collected data for fundamentals. We focus on app traffic, as it is the most direct outcome of the mechanism and is widely studied in the literature. Two strands of studies are relevant. First, targeted advertising based on rich, linked behavioral data improves marketing effectiveness (Rafieian and Yoganarasimhan, 2021; Wernerfelt et al., 2025), and digital browsing and clickstream data help identify consumers with stronger purchase intent, implying that targeted web and app traffic proxy for latent demand (Requena et al., 2020). Second, privacy restrictions reduce the effectiveness of data-driven advertising (Goldfarb and Tucker, 2011). Specifically, Aridor et al. (2025) show that, post-ATT, exposed firms experience worse advertising performance, weaker customer acquisition, and lower revenues, consistent with a 40% reduction in their targeting abilities.

We verify that those patterns hold in our data by showing that app traffic predicts firm performance, but that relation weakens after ATT. Specifically, we relate quarter-t sales growth or standardized unexpected earnings (SUE), to app downloads in the same quarter. Sales growth is defined as $\frac{Sales_t - Sales_{t-4}}{2(Sales_t + Sales_{t-4})}$, while SUE follows a seasonal random walk as in Livnat and Mendenhall (2006), with $SUE_{it} = \frac{EPS_{i,t} - EPS_{it-4}}{P_{it}}$. As in Bradshaw and Sloan (2002), we adjust EPS to exclude special items. We transform both performance measures into within-quarter decile ranks to reduce the influence of outliers and measurement error. We then estimate:

$$\text{Performance}_{it} = \beta_1 \text{Downloads}_{it} + \beta_2 \text{ATT}_t \times \text{Downloads}_{it} + \Gamma X'_{it-4} + FE + \epsilon_{it}, \qquad (1)$$

where Downloads_{it} is the log of total downloads across firm i's apps in quarter t, ATT_t equals one after ATT, and X'_{it-4} includes lagged size, cash, tangibility, debt-to-assets, and market-to-book. $FE = [\theta_{sic2,t}, \theta_i + \iota_t]$ indicates fixed effects: either two-digit SIC-by-quarter fixed effects $\theta_{sic2,t}$, to compare firms to industry peers facing similar demand shocks and seasonality; or firm and quarter fixed-effects $(\theta_i + \iota_t)$, to study within-firm changes in the download-performance relation. Standard errors are clustered by firm. β_1 captures the pre-ATT relation between downloads and firm performance, β_2 captures the post-ATT change.

Table 2 shows that downloads are positively associated with sales growth and SUE before ATT, consistent with app traffic containing information about firms' near-term fundamentals.[19] After ATT, however, this relation weakens substantially. The interaction coefficient is negative and statistically significant for both outcomes, and its magnitude roughly offsets pre-ATT predictability ($\beta_1 + \beta_2 \simeq 0$). In the specifications with SIC-by-quarter fixed effects and controls (Columns 2 and 4), a one-standard-deviation increase in downloads is associated with a 0.25-decile increase in sales growth (2.52×0.102) and a 0.11-decile increase in SUE (2.52×0.042) pre-ATT; this association is largely eliminated afterward.[20] The same findings hold in the specifications with firm and quarter fixed effects (Columns 3 and 6).

In Appendix C, we examine the temporal pattern in the predictive power of downloads and conduct a broad set of robustness tests. First, we show that results are not driven by secular changes in the performance of firms with specific characteristics that happen to correlate with downloads (e.g., high market-to-book): allowing the effects of controls

[19] This is consistent with Chen et al. (2024), who shows that mobile app downloads predict earnings, consistent with $\hat{\beta}_1 > 0$. That paper focuses on how disclosure in SEC filings shapes investors' understanding of app performance.

[20] For SUE, the post-ATT net effect for affected firms falls slightly below the pre-ATT baseline, although we cannot reject that the sum $\hat{\beta}_1 + \hat{\beta}_2 = 0.042 - 0.065$ is equal to zero at the 10% significance level.

to vary flexibly by quarter leaves the estimates largely unchanged (Table C.1). Second, we document that the decline in predictive power emerges sharply around ATT's implementation (Figure C.1), and it is similar when using alternative mobile traffic measures such as MAU and engagement, or when using lagged downloads (Figure C.2). Finally, we validate that results are quantitatively similar when excluding COVID quarters, alleviating concerns that our findings merely reflect pandemic-era shifts in digital activity (Table C.2 and Figure C.2).

These results should not be interpreted as implying that all mobile-generated signals completely lose predictive power after ATT; we focus on a simple set of measures for expositional clarity. Nonetheless, the evidence points to a meaningful decline in the predictive power of app traffic for firm performance, consistent with a deterioration in targeting precision and a reduced representativeness of the remaining tracked users (see Section 2.3).

Importantly, this validation exercise is not designed to establish a causal effect on financial market outcomes. While we alleviate several concerns, such as secular trends among firms with certain characteristics and pandemic-related shifts in digital activity, we cannot fully rule out the presence of firm-quarter shocks that jointly affect downloads and fundamentals. As such, this section should be viewed as simply establishing a necessary condition for our mechanism: mobile-generated signals become less precise after ATT. In the next section, we turn to settings that allow us to more cleanly isolate how this loss of precision affects the behavior of market participants.

5 ATT Decreases the Forecasting Ability of Market Participants

Building on the evidence above, we examine how the decline in mobile-data precision affects the forecasting and investment decisions of sell-side analysts (Section 5.1) and mutual fund managers (Section 5.2), two key groups whose actions shape information aggregation in capital markets.[21] A central challenge is to disentangle the effect of degraded information from any direct impact of ATT on firm fundamentals.[22] Our empirical design addresses

[21] Our focus on both analysts and funds is similar in spirit to Gao and Huang (2020).

[22] For example, if ATT makes the cash flows of app-owning firms more volatile, forecasting accuracy could decline even for analysts and fund managers who do not use alternative data.

this challenge by exploiting within-firm-quarter variation in pre-ATT reliance on mobile-generated data across users. This approach differences out shocks to a firm at a given point in time, including any direct effects of ATT, while isolating its differential impact across analysts or funds. We can therefore test whether forecasting accuracy and stock-picking ability deteriorate more for users that relied more heavily on mobile data prior to ATT.

5.1 Sell-Side Analysts

We begin by studying sell-side analysts, who offer two advantages. First, their forecasts are directly observable. Second, analysts often disclose the data sources and methodologies underlying their forecasts in accompanying reports. This transparency allows us to identify the variation in the use of mobile data between analysts covering the same firm at a given point in time, helping to isolate differences in information usage from firm-level shocks.

5.1.1 Measures Construction and Samples

Forecast Errors. For each firm-analyst pair and quarter, we compute Proportional Mean Absolute Forecast Errors (PMAFE), defined as Absolute Forecast Errors (AFE, i.e., the absolute difference between the analyst's forecast and actual earnings) scaled by the mean AFE across all analysts making forecasts for the same firm in that quarter. This normalization allows us to benchmark individual performance relative to peers, facilitating cross-firm and cross-time comparisons while controlling for forecast difficulty.

Specialization. We measure analysts' specialization based on their pre-ATT disclosed usage of app-related alternative data in analyst reports. We first identify usage of any mobile-related data by automatically searching in the LSEG workspace all reports of firms in the "App Sample" (see Section 3) for mentions of any mobile-related keyword.[23] Within that list, we manually select the keywords most likely affected by ATT ("App keywords"). Those most

[23] The list is obtained by first combining the alternative data keywords of Chi et al. (2024) with names of data providers from JP Morgan's Alternative Data Handbook (Kolanovic and Krishnamachari, 2019), and additional terms related to app usage (e.g., "in-app purchase," "SensorTower"). Then removing terms unrelated to mobile data (e.g., "satellite image"), and terms frequently used in unrelated contexts (e.g., "downloads"), to minimize false positives. See Appendix Section B for details and illustrative report excerpts.

cited by analysts, with counts in parentheses, are: *mau* (or monthly active users, 62,480), *dau* (or daily active users, 43,619), *user engagement* (15,254), *sensor tower/sensortower* (34,550/14,240), *Apptopia* (9,818), *active users* (8,443), *safegraph* (5,259), and *app usage* (5,138); the full list appears in Appendix Table B.1. We proxy for an analyst's specialization with the share of their reports containing "App keywords"; calculated as the number of such reports divided by the total reports they produced (*Keyword usage*).[24]

Analyst- and Report-level Samples. For analyst-level analyses, we name-match I/B/E/S and LSEG, yielding 1,166 analysts (123 brokerage firms) covering the *App Sample*, and 1,370 analysts (131 brokerage firms) covering a placebo; i.e. firms in Fama-French 48 industries *without* apps with quarterly downloads consistently above the median. The placebo captures firms for which mobile-collected data is unlikely to be informative, either directly (e.g., firms' own app traffic) or indirectly (e.g., foot traffic to stores inferred from other apps' geolocation data). Panels A and B of Table 3 report summary statistics for the "App Sample" and the placebo, respectively. The two samples only differ in average "App keywords" usage (6.4% vs. 2.8%), and length of coverage per firm (13.8 vs. 11.6 quarters). For report-level analyses, we merge the previously searched reports to CRSP, yielding 42,992 reports.[25] Panel C of Table 3 shows that the average report spans 19.9 pages, contains 0.1 "App keywords" per page, and sells at $196.

5.1.2 Analyst Forecast Errors around ATT

We use *Keyword usage* as a proxy for an analyst's attention allocated to mobile-data signals. Because analysts' coverage sets are sticky in the short run,[26] the key margin is how signal deterioration affects forecast accuracy for exposed stocks, across analysts who used mobile

[24] We restrict the sample to lead analysts per report, since I/B/E/S only reports their names and forecasts.

[25] We include LSEG reports available within 4 days after issuance to align them with market reactions.

[26] Analyst coverage may adjust over longer horizons or in response to the addition or removal of data sources, as in Grennan and Michaely (2020). In our sample, however, the average analyst covers 4.50 exposed firms before ATT and 4.45 afterward, indicating little change in coverage. This likely reflects two features of ATT: it affects an alternative dataset originally generated for non-financial purposes, and it degrades rather than eliminates that dataset, making the shock less salient than the entry or disappearance of a purpose-built financial information source such as TipRanks.

data pre-ATT and those who did not. Guided by the multitasking framework of Dessaint et al. (2024), we hypothesize that, before ATT, higher *Keyword usage* should translate into lower forecast errors for exposed stocks, reflecting effective use of an informative signal. After ATT, as signal precision deteriorates, its marginal value should decline, causing high-usage analysts to see their advantage erode and forecast errors rise, particularly at short horizons and when substitutes are scarce. To test this hypothesis, we run the following regression:

$$\text{PMAFE}_{ijt} = \beta_1 \times \text{Keyword usage}_j + \beta_2 \times \text{ATT}_t \times \text{Keyword usage}_j + \Gamma X_{jt} + \theta_{it} + \iota_{ij} + \kappa_d + \epsilon_{ijt} \qquad (2)$$

where j indicates analysts, i firms in the "App Sample", t year-quarters, and d estimation month (i.e., when analyst j provided the forecast for firm i's quarter t earnings). *Keyword usage* is our proxy for analyst specialization. X_{jt} includes analyst and brokerage characteristics: forecast age, analyst experience (since first forecasting and for each firm i), number of firms covered, and brokerage size. We include firm–year–quarter fixed effects (θ_{it}) to absorb time-varying firm-level changes, addressing concerns that reduced consumer data access around ATT could increase firm volatility and affect forecast accuracy. Analyst-firm fixed effects control for persistent match effects, isolating ATT-induced changes in accuracy within each pair. Finally, estimation-month fixed effects (κ_d) absorb aggregate shocks, like market changes or data availability, that could influence forecasting accuracy across analysts. Standard errors are clustered by firm and estimation month. We expect the coefficient of interest $\hat{\beta}_2$ to be positive and statistically significant, indicating that analysts who use app-related data experience an increase in forecast errors for exposed stocks following ATT.

Table 4 reports results consistent with this prediction. Columns 1–2 include firm-year-quarter and estimation-date fixed effects, while Columns 3–4 also include analyst-firm fixed effects. $\hat{\beta}_2$ is positive and statistically significant across all specifications. Economically, in Column 2, the estimate of $\hat{\beta}_2$ is 0.233 (1% significance). This implies that a one-standard-deviation increase in the share of reports referencing "App keywords" corresponds to a 2.5% increase in PMAFE (0.233×0.11), moving an analyst from the median to the 54$^{\text{th}}$

percentile of the post-ATT forecast error distribution.[27] The coefficient $\hat{\beta}_1$ for *Keyword usage*, is -0.095, indicating that higher pre-ATT reliance on mobile data is linked to greater forecast accuracy. The larger post-ATT decline suggests that specialized analysts not only lose their informational edge but also perform worse overall. Other controls behave as expected: greater analyst experience is associated with lower forecast errors, while older forecast vintage is associated with larger errors. Including analyst-firm fixed effects reduces $\hat{\beta}_2$ mildly to 0.176 (Column 4), indicating that selection of analysts into firms is unlikely to explain our findings.[28]

To further validate our interpretation of β_2, we conduct the same analysis using the placebo sample described earlier. $\hat{\beta}_2$ is not statistically significant across specifications, suggesting that the forecast errors of analysts specialized in mobile data decrease only for firms whose future payoffs are likely to be informed by such data (see Appendix Table D.2). Additionally, excluding forecasts between 2020Q2 and 2021Q1 yields quantitatively similar results, ruling out the possibility that results are driven by COVID-19 (see Appendix Table D.3).

Why do specialized analysts perform worse than non-specialized ones post-ATT? We address this by distinguishing between analysts who reduce their keyword usage post-ATT ("switchers")—those with an above-median decline in keyword usage from pre-ATT to post-ATT periods—and "non-switchers". Results, reported in Appendix Table D.4, show that the net performance deterioration is driven by non-switchers, who continue to rely on degraded signals. In contrast, switchers lose their informational edge but do not show a significantly worse net performance. This suggests that the poor post-ATT performance among specialized analysts is due to continued reliance on degraded mobile data or a failure to re-optimize effort allocation in response to the reduced informativeness of these signals.

[27]For context, this effect is about half to two-thirds the size of the 4–5% forecast accuracy gains documented by Green et al. (2014) for conference-hosting brokers, larger than the 1.9% gains documented by Harford et al. (2019) for firms that analysts consider relatively more important, and more than double the accuracy improvement associated with a one standard deviation increase in analyst workday length documented by Ben-Rephael et al. (2022).

[28]Results are robust to using AFE scaled by absolute actual earnings, as the outcome variable, as reported in Appendix Table D.1.

Finally, to examine the timing of the decline in forecast accuracy. We estimate a dynamic version of Equation (2), mirroring the most stringent specification, but replacing the ATT indicator with a set of quarter-specific dummies. Figure 2 visualizes the estimated coefficients using 2021Q2 as the reference quarter. Forecast errors for specialized analysts begin to rise in the second quarter post-ATT and remain elevated through the third and fourth quarters. Importantly, no pre-trend is observed, supporting the parallel trends assumption.

A key contribution of Dessaint et al. (2023) is demonstrating that alternative data is primarily informative about short-term firm outcomes. This has clear implications for analysts' forecast errors in our context. Since analysts cover a sticky set of firms, they cannot easily reallocate attention away from ATT-exposed firms toward less exposed ones. We therefore expect the forecast accuracy deterioration to be most pronounced at short horizons. Results in Figure 3 are consistent with this prediction. Comparing earnings forecasts for months $[t, t+3]$, $[t+3, t+6]$, $[t+6, t+9]$, $[t+9, t+12]$, and $[t+12, t+24]$, the effect of ATT weakens monotonically in both economic magnitude and statistical significance as the forecast horizon extends, becoming statistically insignificant beyond three quarters.

5.1.3 Market Response to Analyst Reports

A natural question is whether these differences in analyst accuracy are reflected in prices, since analysts themselves do not trade. We address this by examining market reactions to analyst reports, where information is incorporated into prices upon issuance. If a signal is informative, reports that rely more heavily on it should elicit stronger price reactions; if its precision deteriorates, that differential response should fade. Studying report-level price impacts complements the forecast error analysis in two ways. First, these reports are discrete, time-stamped information events that often attract media coverage, especially when containing explicit recommendations, allowing us to examine how information is incorporated into prices. Second, they allow us to exploit finer variation in reliance on mobile-generated signals: rather than all reports by a given analyst, we focus on those referencing mobile-related data and measure the intensity of app-related content in each report. As ATT primarily disrupts signals that rely on cross-app linkage, we test whether the market response to reports with greater "App keywords" per page declines following ATT as follows:

$$CAR^{adj,\ 3d}_{ikt} = \beta_1 \times \text{Keyword per page}_{kt} + \boldsymbol{\beta_2} \times \text{ATT}_t \times \text{Keyword per page}_{kt} + \Gamma X_{kt} + \theta_{it} + \iota_{br,i} + \kappa_d + \epsilon_{ikt}, \quad (3)$$

where i indexes firms, k reports, and t the year-quarter corresponding to the earnings period. The outcome variable $CAR^{adj,\ 3d}_{ikt}$ is the direction-adjusted cumulative abnormal return over a 3-day window (returns are multiplied by -1 for negative reports). $Keyword\ per\ page_{kt}$ is the number of "App keywords" in report k divided by the total number of pages in that report. The vector X_{kt} includes report-level characteristics: the natural logarithm of retail price and of number of pages. Consistent with our analyst-level analysis, we include firm-year-quarter (θ_{it}), brokerage-firm ($\iota_{br,i}$), and report-issuance-date (κ_d) fixed effects. These fixed effects allow us to isolate within-firm-quarter variation in market responses attributable to the degree of "App keywords" usage across reports. Standard errors are clustered by firm.

Estimation results appear in Table 5, with Columns 1–2 (3–4) reporting CAR computed using the CAPM (Fama-French 6-factor) model. Across specifications, the market appears to recognize both the pre-ATT value of app-generated signals referenced in these reports and their post-ATT deterioration. Pre-ATT, reports citing more "App keywords" triggered stronger market reactions; post-ATT, this differential response disappears. Quantitatively, pre-ATT, a one standard deviation increase in *keywords per page* (about 0.3 additional) corresponds to a three-day abnormal return of 16 basis points per report (0.3×0.545), an effect that vanishes post-ATT. Controls behave as expected: more expensive and shorter reports (holding price fixed) generate larger market reactions.

5.2 Mutual Funds

We next turn to mutual funds. This setting is less transparent, because we do not observe managers' research inputs or forecasting ability directly and must infer both from holdings data. Its advantages, however, are twofold. First, mutual funds affect stock prices directly through their trades, bringing us closer to the market mechanism of interest. Second, because fund managers can adjust portfolio weights rather than merely revise beliefs about a sticky set of firms, their holdings reveal changes in the allocation of attention and capital across stocks in response to signal precision. This lets us study more directly how a privacy-driven loss of signal precision is transmitted to prices through investors' portfolio decisions.

5.2.1 Measures Construction and Sample

Stock Picking. We proxy for forecast accuracy with stock-level picking ability, measured as the product of a fund's benchmark-adjusted portfolio weight and a stock's future abnormal return. Intuitively, if a fund overweights stocks that subsequently outperform, this indicates informed selection. Formally, for each fund f, stock i, and month t, we define:

$$\text{Picking}_{ift}^{s,h} = 100 \times \Delta w_{ift} \times \text{Return}_{i[t+1,t+h]}^{s}, \tag{4}$$

where $\text{Return}_{i[t+1,t+h]}^{s}$ is stock i's cumulative idiosyncratic return from $t+1$ to $t+h$, computed as the residual return from a factor model $s \in \{\text{CAPM}, \text{FF6}\}$, and Δw_{ift} is the deviation of fund f's portfolio weight in stock i at time t from a benchmark portfolio. Because mutual funds differ substantially in investment style, we assign each fund to the passive benchmark that minimizes the distance between the fund's portfolio weights and the benchmark's weights, following the minimum-active-share approach of Cremers and Petajisto (2009).[29] This measure is related to Kacperczyk et al. (2016) and Bonelli and Foucault (2023), but uses fund-specific benchmarks rather than the market portfolio.[30] As shown by Bonelli and Foucault (2023), a fund's alpha is the sum of its expected stock picking ability in each stock. Hence, our measure can be interpreted as stock i's contribution to a fund's alpha.

Specialization. We infer specialization on mobile-generated signals from revealed trading behavior. A first challenge is that we do not observe mutual fund trades themselves, but only periodic snapshots of portfolio holdings. Changes in portfolio weights constructed from those holdings reflect both active reallocation and passive movements induced by stock returns.

[29] We consider seven passive Vanguard funds as possible benchmarks: broad, large-cap, mid-cap, small-cap, growth, value, and momentum. Details are provided in Appendix Section E.1.

[30] Our results are robust to using the market portfolio as the benchmark; see Appendix Figure E.2.

To isolate the component of portfolio adjustment most plausibly driven by information, we construct *trade-induced* weight changes as the difference between a stock's observed portfolio weight and the counterfactual weight it would have reached absent any trade—i.e., holding prior positions fixed and allowing them to drift only with realized returns.[31]

We then ask whether these trade-induced portfolio changes load on a concrete mobile signal for firms in the "App Sample" and whether that relation changes around ATT. Specifically, we relate trade-induced weight changes to lagged abnormal app downloads by estimating:

$$\begin{aligned} D^h_{ift} = \theta_{ft} + \iota_i + \sum_{k=1}^{h-1} \beta_k \times \text{Abnormal Downloads}_{i,t-k} \\ + \sum_{k=1}^{h-1} \delta_k \times \text{Abnormal Downloads}_{i,t-k} \times \text{ATT}_t + \epsilon_{ift}, \end{aligned} \tag{5}$$

where i indexes firms in the "App Sample"; D^h_{ift} is the trade-induced change in fund f's portfolio weight in stock i over $[t-h,t]$; θ_{ft} are fund-month fixed effects, absorbing time-varying fund-level shocks (e.g., flows); ι_i are stock fixed effects, absorbing persistent firm heterogeneity (e.g., trading intensity); and *Abnormal Downloads*$_{i,t-k}$ is log downloads in month $t-k$ minus their average over the prior 12 months. We set $h=3$, so changes over $[t-3,t]$ are linked to signals observed at $t-2$ and $t-1$, which helps mitigate look-ahead concerns when monthly holdings are missing or reported mid-month. Standard errors are clustered by fund and month. β_k captures pre-ATT trading sensitivity to abnormal downloads k months earlier, and δ_k is the post-ATT change in that sensitivity. We estimate Equation 5 separately for buys and sells, since mutual funds are often subject to long-only mandates, making negative signals harder to implement than positive ones.[32]

Table 7 reports the results. On the buy side, abnormal downloads positively predict active trades before ATT, but this relation disappears afterward. The effect is concentrated in the most recent lag, while adding a second lag provides little additional predictive content. On

[31] See Appendix Section E.2 for details on the construction of trade-induced weight changes.

[32] Positive signals can be acted upon at both the extensive margin (opening new positions) and the intensive margin (increasing existing positions). If negative signals cannot create shorts, they should mostly manifest as trims or non-rebalancing toward the no-trade weight, not as large negative trades.

the sell side, the estimates are smaller and generally imprecise, consistent with mutual funds' limited ability to act on negative signals, due to short-sale constraints. This pattern suggests that abnormal downloads were indeed used by mutual fund managers as an input into stock-level trading decisions before ATT, while the disappearance of this relation after ATT is consistent with managers recognizing that the precision of the signal had deteriorated.

The above evidence guides our measurement of fund-level specialization on mobile-based signals. We aim to identify which funds' trade-induced portfolio changes were more tightly linked to abnormal app downloads in the pre-ATT period, and hence which funds were more exposed to the deterioration of that signal after ATT. Intuitively, a fund should be classified as more specialized if variation in its trades is more systematically aligned with prior abnormal downloads, and if that signal is relevant for a larger share of its portfolio. Accordingly, for each fund f,[33] we estimate the following regression in the pre-ATT period:

$$D^{+}_{ift} = \beta_f \text{ Abnormal Downloads}_{i,t-1} + \mu_t + \varepsilon_{ift}, \quad \forall f \in \mathcal{F},\ \forall t < ATT, \tag{6}$$

where D^{+}_{ift} is the trade-induced buy weight change for stock i over $[t-3, t]$, and *Abnormal Downloads*$_{i,t-1}$ is abnormal downloads observed at the end of month $t-1$. We focus on buy-side changes and only include one lag as Table 7 shows that the trading response to downloads is concentrated there. Month fixed effects μ_t absorb aggregate trading conditions and common shocks to app-related stocks. Our measure of signal reliance is the within-fixed-effects adjusted R^2_f from Equation 6, which captures how much of a fund's trading variation is explained by abnormal downloads. To translate this trading sensitivity into economically meaningful reliance, we scale R^2_f by the fund's average pre-ATT portfolio weight in exposed stocks: $\text{SpecScore}_f = R^2_f \times \overline{W}_f^{\text{pre-ATT}}$. We classify funds above the 75th percentile of SpecScore_f as *specialized* and the remainder as *non-specialized*.

This measure is closest in spirit to Sheng et al. (2024), who infer funds' reliance on

[33] We require funds to (i) hold at least one stock with available abnormal downloads, (ii) do so for at least 2 months of the pre-period, and (iii) to have a panel of more than 10 observations. If funds do not meet these criteria we assume they are not specialized in mobile stocks and set their specialization to zero.

generative-AI from the explanatory power of AI-generated signals for holdings changes. As in their setting, our measure is not designed to capture managerial skill directly, but rather the importance of a specific information source for portfolio decisions. More broadly, the measure builds on Kacperczyk and Seru (2007), who study the sensitivity of portfolio changes to public information. Our implementation differs from these papers in two ways. First, although we focus on a single signal, we do not residualize weight changes against a long list of controls; instead, we use trade-induced weight changes and month fixed effects to isolate signal reliance. This filters out portfolio changes due to flows, public information already impounded in prices, and market-wide news, while keeping the specification parsimonious. Second, since abnormal downloads are informative only for firms in the "App Sample," explanatory power alone would overstate reliance for funds that react strongly to the signal but allocate little capital to that segment. Scaling by pre-ATT average holdings in exposed stocks addresses this issue by combining the intensity of a fund's response to the signal with the share of the portfolio over which that signal can matter.[34]

Fund-level Sample. Panel A of Table 6 shows that specialized and non-specialized funds are broadly similar, with the former being smaller, lower-turnover, and tilted toward more growth-oriented firms, consistent with the composition of the "App Sample". Panel B reports summary statistics on stock-level picking ability, constructed under the Fama-French 6-factor model, split by fund specialization and stocks' ATT-exposure.[35]

5.2.2 Funds' Stock-Picking Ability Around ATT

Mutual funds affect prices through their trades, and fund managers are free to reallocate attention and capital across stocks when signal quality changes. These portfolio considerations are central to the rational inattention framework of Kacperczyk et al. (2016), in which managers allocate scarce attention across assets in response to differences in signal precision. In our setting, the framework predicts that when ATT reduces the precision of

[34] Appendix E.3 provides further discussion of the measure's construction and characteristics.

[35] Stocks in the "App Sample" are classified as exposed to ATT; all others are not.

mobile-generated signals, funds that relied more heavily on those signals should shift attention away from exposed stocks, reducing their relative picking ability in that segment. Our object of interest is therefore the within-fund-month gap in average picking ability between exposed and unexposed stocks (ΔP_{ft}). This gap captures where a manager's scarce forecasting capacity is most effectively deployed. A decline in ΔP_{ft} after ATT indicates a relative loss of advantage on the exposed segment, whether through worse selection on exposed stocks, better selection on unexposed stocks, or both. This comparison is necessary, rather than focusing only on exposed stocks, because selective liquidation can mechanically attenuate the measured decline in exposed-stock picking.[36]

To test whether this decline is stronger for funds specialized in mobile-based signals, we combine our specialization measure with the picking metric in a triple-difference design:

$$\begin{aligned}\text{Picking}_{ift}^{s,h} = {} & \alpha + \beta_1 \times \text{Exposed}_{if} + \beta_2 \times \text{ATT}_t \times \text{Exposed}_{if} + \beta_3 \times \text{Exposed}_{if} \times \text{Specialized}_f \\ & + \beta_4 \times \text{ATT}_t \times \text{Exposed}_{if} \times \text{Specialized}_f + \delta_{ft} + \lambda_{it} + \varepsilon_{ift}, \end{aligned} \tag{7}$$

where ATT_t equals one for months after April 2021. Fund-month fixed effects (δ_{ft}) absorb all time-varying shocks at the fund level, while firm-month fixed effects (λ_{it}) absorb time-varying shocks to each stock, including ATT's direct effects on app-owning firms. Standard errors are triple clustered at the fund, firm, and month levels. We estimate the regression for horizons $h \in \{1, 3, 6, 9, 12, 18\}$ and use $h = 12$ as the baseline, compounding daily abnormal returns over $[t+1, t+12]$. The coefficient of interest, β_4, captures the differential post-ATT change in the exposed–unexposed picking gap for specialist funds relative to non-specialists. A negative $\hat{\beta}_4$ indicates that ATT-induced signal deterioration disproportionately weakened specialists' stock-picking ability on exposed stocks, net of broad fund- and stock-level shocks absorbed by the fixed effects. Since specialization is inferred from revealed trading rather than directly observed, the classification is necessarily noisy; if anything, this measurement

[36] E.g., if only a subset of exposed stocks becomes materially harder to forecast, fund managers may simply liquidate those positions. In that case, observed picking ability among the surviving holdings could remain unchanged, obscuring the true deterioration in the fund's informational advantage on exposed stocks.

error should attenuate the estimate toward zero.

Table 8 reports the baseline results. The dependent variable is stock-level picking ability, based either on the CAPM (Columns 1–2) or the Fama–French 6-factor model (Columns 3–4). Even-numbered columns include firm-month fixed effects. Across all specifications, the triple-interaction coefficient, $\hat{\beta}_4$, is negative and statistically significant, consistent with a larger post-ATT decline in specialists' relative performance on exposed stocks. In Column 3, $\hat{\beta}_2 = -0.030$ captures the decline in exposed-stock picking ability for *non-specialized* funds after ATT, plausibly reflecting ATT's direct effect on the underlying firms. While not central to our mechanism, this sizable negative effect aligns with the pronounced stock market reaction among app-owning firms most dependent on tracking and data-sharing (Bian et al., 2021), suggesting that ATT impaired the ability of all funds to pick exposed stocks, likely due to increased uncertainty. The specialist differential, $\hat{\beta}_4 = -0.033$, is of nearly identical magnitude, implying that the incremental loss associated with reliance on degraded mobile signals is about as large as the baseline deterioration affecting exposed stocks.[37]

We examine how the effect varies with the horizon used to construct the picking measure. Using the most stringent specification in Column 4 of Table 8, Figure 5 plots the estimated $\hat{\beta}_4$ for horizons from 1 to 18 months. The coefficient remains negative at all horizons and declines nearly monotonically, consistent with a reallocation away from exposed names, rather than a shift in attention toward signals at specific horizons.[38]

We study the timing of the decline by estimating a dynamic version of Equation (7) with monthly event-time dummies in place of the ATT_t indicator. Figure 4 plots the resulting β_4 coefficients, using October 2020 as the reference month.[39] Specialists' relative picking ability begins to decline in November 2020, when half of the 12-month return window overlaps with

[37] The effect of β_2 corresponds to moving an unspecialized fund by one quartile of the picking distribution (Row 7 of Table 6 Panel B); the effect of β_4 corresponds to moving a specialized fund by half a quartile (Row 8 of Table 6 Panel B), reflecting the different distribution of the picking measure for the two groups.

[38] Since returns are cumulative, the coefficient at horizon h reflects the sum of effects over shorter horizons. A monotonic decline therefore indicates a roughly constant incremental effect per unit of horizon.

[39] We use an earlier reference month than in other event studies because, with a 12-month forward-looking return window, portfolios formed before April 2021 can already be partially exposed to the post-ATT regime.

post-ATT months, reflecting the forward looking nature of our measure. The decline deepens once the window is fully post-ATT, and remains statistically significant through June 2023.

We subject this result to several robustness checks. First, we re-estimate the baseline using alternative definitions of the picking and specialization measures: we calculate the picking metric using weights bench-marked either against the CRSP universe or each fund's closest passive benchmark; we define specialists using top-quartile, top-quintile, and top-decile cutoffs of the specialization score, which we construct using both raw and benchmark-adjusted weights. All specifications are estimated for horizons from 1 to 18 months. Appendix Figure E.2 shows that the triple-interaction coefficient remains negative throughout, is usually significant at the 5% level, and increases as the specialization threshold tightens (though power falls as the specialized sample shrinks). Second, excluding March 2019 to March 2021 to remove COVID-19-related distortions leaves the estimates virtually unchanged (Appendix Table E.1).[40] Third, replacing actual stock-level ATT exposure with a placebo assignment that preserves the share of exposed stocks yields small, insignificant, and generally opposite-signed coefficients, confirming results are not driven by spurious correlation (Appendix Table E.2). Finally, we aggregate the stock-level picking measure to the exposure-group level by computing, for each fund-month, average picking scores separately for exposed and unexposed stocks, and re-estimate the baseline. Results are quantitatively similar and highly significant, indicating that our findings are not driven by the large number of stock-level observations in the baseline panel (Appendix Section E.5).

Taken together, the evidence indicates that ATT reduced specialist funds' comparative picking ability on exposed stocks, beyond any direct effect on firm fundamentals. This is consistent with ATT degrading a previously informative signal and redirecting informed investors' scarce attention away from that segment of the market. With fewer informed investors monitoring these stocks and less accurate analyst forecasts, price informativeness should decline and mispricing should rise, a hypothesis we examine more directly in Section 6.

[40] We omit both the COVID period itself and the pre-COVID months for which the 12-month forward-return window of the picking measure overlaps with COVID months.

6 ATT Worsens Firms' Information Environment

Section 5 shows that ATT impaired both margins through which mobile-data specialists contribute to price discovery: analysts produced less accurate forecasts, while funds redirected attention and capital away from exposed stocks. We therefore expect price efficiency to deteriorate most for firms that were more heavily covered by these specialists before ATT. We test this prediction using two firm-level measures: quarterly bid-ask spreads, defined as daily closing ask-bid spreads scaled by the mid-price and averaged within quarter; and post-earnings-announcement volatility, computed as the root-mean-squared Fama-French 6-factor abnormal return over trading days 6 to 28 after the earnings announcement (Loughran and McDonald, 2014; Kim et al., 2023). The former captures ex-ante information asymmetry and trading frictions; the latter captures residual uncertainty and disagreement after earnings news is incorporated, proxying for noise in the information environment. We then ask whether firms more exposed to specialized analysts or funds before ATT experienced wider spreads and higher post-earnings volatility relative to otherwise similar peers by estimating:

$$\text{Info Friction}_{it} = \beta_1 \times \text{Exposure}_i + \beta_2 \times \text{ATT}_t \times \text{Exposure}_i + \Gamma_1 X_{it-4} + FE + \epsilon_{it}, \tag{8}$$

where i indexes firms and t year-quarters. *Exposure* is an indicator equal to one if the firm is more exposed to specialized attention, defined as: (i) having analyst reports containing the relevant keywords in more than 50% of pre-ATT quarters, (ii) being above the 75$^{\text{th}}$ percentile of the pre-ATT distribution of market capitalization held by specialized funds. Not surprisingly, relative to firms without an app, firms in the "App Sample" are more likely to be held by specialized funds (42% vs. 25%) and followed by specialized analysts (67% vs. 23%). We control for firm characteristics X_{it-4} (size, cash, tangibles, debt-to-assets ratio, and market-to-book ratio), all measured at $t-4$, and include $FE = [\theta_{sic2,t}, \theta_i + \iota_t]$: either industry-by-quarter fixed effects or firm and quarter-year fixed effects. Standard errors are clustered by firm.

Table 9 shows that, prior to ATT, firms with greater exposure to specialists, especially specialized funds, exhibit lower information asymmetry ($\hat{\beta}_1 < 0$), in line with specialized attention and alternative data improving firms' information environment. This advantage is reduced or reversed after ATT ($\hat{\beta}_1 + \hat{\beta}_2 \geq 0$), suggesting weaker price discovery and a loss of informational advantage. Including firm and year-quarter fixed effects does not materially change $\hat{\beta}_2$. While part of this effect may reflect a direct impact of ATT on firms' earnings volatility as access to cross-app data declines and advertising effectiveness weakens, the overall pattern supports the interpretation that ATT, by degrading mobile-generated data, disproportionately weakens the information environment of firms followed by market participants who rely on these signals.

More broadly, when firms most affected by a precision shock share distinctive characteristics, such as the high market-to-book ratios we observe among app-intensive companies, the resulting mispricing may not average out across the market. If information frictions rise disproportionately for growth-oriented stocks, or for any trait correlated with exposure, these shocks can systematically distort relative valuations and amplify fragility in asset prices.

7 Conclusion

Linked personal information has become a common input to financial forecasting, but its value depends on the integrity of the data chain linking consumer behavior to tradable signals. Using ATT as a laboratory, we study how shocks to mobile-data linkage impact information production and price efficiency. We show that when established alternative-data signals lose precision, the effects propagate beyond the data market: analysts become less accurate, specialized investors reallocate attention, and price discovery deteriorates for firms most exposed to these information producers. These findings point to a new fragility in data-driven financial markets. For investors, they suggest that alternative-data strategies require risk management not only around model performance, but also around the stability of the underlying DGP and the regulatory or platform decisions that can alter it. For policymakers, they highlight an unintended ripple effect of privacy regulation: restrictions on tracking and data sharing can weaken price discovery and capital allocation.

Our findings are only a starting point for understanding this fragility. One natural direction for future research is to study heterogeneity in adaptation among information producers: which analysts and funds adjust most quickly when established signals degrade, what frictions slow this adjustment, and how these differences shape the duration of market disruptions. A second direction is to examine the distributional consequences across investors and firms, especially when affected firms share common traits or attract specialized capital. Third, future work could study whether these asset-pricing effects feed back into real outcomes, for example through firms' cost of capital, investment, or disclosure choices. Such tests would require settings that separate the real effects of a specific data shock from the broader exposure of firms to recurring disruptions in the data ecosystem. Fourth, studying other regulatory and platform shocks would help quantify the aggregate importance and welfare consequences of this new source of market fragility. Finally, future work could examine the long-run dynamics of data production itself: because many alternative datasets are byproducts of digital products monetized through user data, changes in those business models could reshape the ecosystem that generates rich user-level information.

References

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance 59*(3), 1259–1294.

Apple (2024a). If an app asks to track your activity. https://support.apple.com/en-us/102420. Accessed August 02, 2024.

Apple (2024b). User privacy and data use. https://developer.apple.com/app-store/user-privacy-and-data-use/. Accessed August 02, 2024.

Aridor, G., Y.-K. Che, B. Hollenbeck, M. Kaiser, and D. McCarthy (2025). Evaluating the impact of privacy regulation on e-commerce firms: Evidence from apple's app tracking transparency. *Management Science*.

Aridor, G., Y.-K. Che, and T. Salz (2023). The economic consequences of data privacy regulation: Empirical evidence from GDPR. *RAND Journal of Economics 54*(4).

Bai, J., T. Philippon, and A. Savov (2016). Have financial markets become more informative? *Journal of Financial Economics 122*(3), 625–654.

Balasubramanian, M. (2022). App tracking transparency opt-in rate - monthly updates. https://www.flurry.com/blog/att-opt-in-rate-monthly-updates/. Flurry.

Begenau, J., M. Farboodi, and L. Veldkamp (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics 97*, 71–87.

Ben-Rephael, A., B. Carlin, Z. Da, and R. D. Israelsen (2022). All in a day's work: What do we learn from analysts' bloomberg usage? *Available at SSRN 4056506*.

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni (2025). The impact of generative ai on information processing: Evidence from the ban of chatgpt in italy. *Journal of Accounting and Economics 80*(1), 101782.

Bessen, J. E., S. M. Impink, L. Reichensperger, and R. Seamans (2020). GDPR and the importance of data to AI startups. Working paper, New York University, Boston University.

Bian, B., Q. Huang, Y. Li, and H. Tang (2024). Data as a networked asset. *Available at SSRN*.

Bian, B., X. Ma, and H. Tang (2021). The supply and demand for data privacy: Evidence from mobile apps. *Available at SSRN*.

Bian, B., M. Pagel, H. Tang, and D. Raval (2023). Consumer surveillance and financial fraud. Technical report, National Bureau of Economic Research.

Blankespoor, E., E. deHaan, and C. Zhu (2018). Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. *Review of Accounting Studies 23*, 1–36.

Blankespoor, E., B. E. Hendricks, J. Piotroski, and C. Synn (2022). Real-time revenue and firm disclosure. *Review of Accounting Studies 27*(3), 1079–1116.

Bonelli, M. and T. Foucault (2023). Displaced by Big Data: Evidence from Active Fund Managers. *Available at SSRN 4527672*.

Bradshaw, M. T. and R. G. Sloan (2002). GAAP versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research 40*(1), 41–66.

Buffa, A. M., D. Vayanos, and P. Woolley (2022). Asset management contracts and equilibrium prices. *Journal of Political Economy 130*(12), 3146–3201.

Chang, R. and Z. Da (2022). Nowcasting firms' fundamentals: Evidence from the cloud. Technical report, Working Paper.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy (2020). IQ from IP: Simplifying search in portfolio choice. *Journal of Financial Economics 138*(1), 118–137.

Chen, H., P. De, Y. Hu, and B.-H. Hwang (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies 27*(5), 1367–1403.

Chen, S., Y. Liu, and X. Wu (2024). Performance indicators of the digital age: Mobile apps, firm disclosure, and stock returns. *Available at SSRN 4851807*.

Cheyre, C., B. T. Leyden, S. Baviskar, and A. Acquisti (2023). The impact of Apple's App Tracking Transparency framework on the app ecosystem. *Available at SSRN 4453463*.

Chi, F., B.-H. Hwang, and Y. Zheng (2024). The use and usefulness of big data in finance: Evidence from financial analysts. *Management Science*.

comScore (2024). Subscriber share held by smartphone operating systems in the united states from 2012 to 2024. https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/. Statista. Accessed August 02, 2024.

Coval, J. D. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cremers, K. M. and A. Petajisto (2009). How active is your fund manager? A new measure that predicts performance. *The Review of Financial Studies 22*(9), 3329–3365.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Da, Z. and X. Huang (2020). Harnessing the wisdom of crowds. *Management Science 66*(5), 1847–1867.

Deloitte (2017). Alternative data for investment decisions: Today's innovation could be tomorrow's requirement. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-fsi-dcfs-alternative-data-for-investment-decisions.pdf.

Demirer, M., D. J. J. Hernández, D. Li, and S. Peng (2024). Data, privacy laws and firm production: Evidence from the GDPR. Technical report, National Bureau of Economic Research.

Dessaint, O., T. Foucault, and L. Frésard (2023). The horizon of investors' information and corporate investment. *HEC Paris Research Paper No. FIN-2022-1462, Swiss Finance Institute Research Paper* (23-03).

Dessaint, O., T. Foucault, and L. Frésard (2024). Does alternative data improve financial forecasting? The horizon effect. *The Journal of Finance 79*(3), 2237–2287.

Dichev, I. D. and J. Qian (2022). The benefits of transaction-level data: The case of NielsenIQ scanner data. *Journal of Accounting and Economics 74*(1), 101495.

Doerr, S., L. Gambacorta, L. Guiso, and M. Sanchez del Villar (2023). Privacy regulation and fintech lending. *Available at SSRN 4353798*.

Du, M. and J. Qian (2024). The impact of big data on financial information environment: Evidence from big data release schedules. Working paper, National University of Singapore.

Eaton, G. W., T. C. Green, B. S. Roseman, and Y. Wu (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics 146*(2), 502–528.

Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance 65*(4), 1581–1611.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran (2022). Where has all the data gone? *The Review of Financial Studies 35*(7), 3101–3138.

Forbes (2024). 2024 mobile apps: 20 tech experts reveal top design trends. https://www.forbes.com/sites/forbestechcouncil/2024/01/25/2024-mobile-apps-20-tech-experts-reveal-top-design-trends/. Accessed August 02, 2024.

Froot, K., N. Kang, G. Ozik, and R. Sadka (2017). What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics 125*(1), 143–162.

Gao, M. and J. Huang (2020). Informing the market: The effect of modern information technologies on information production. *The Review of Financial Studies 33*(4), 1367–1411.

Goetzmann, W. N., D. Kim, A. Kumar, and Q. Wang (2015). Weather-induced mood, institutional investors, and stock returns. *The Review of Financial Studies 28*(1), 73–111.

Goldberg, S. G., G. A. Johnson, and S. K. Shriver (2024). Regulating privacy online: An economic evaluation of the GDPR. *American Economic Journal: Economic Policy 16*(1), 325–358.

Goldfarb, A. and C. E. Tucker (2011). Privacy regulation and online advertising. *Management Science 57*(1), 57–71.

Green, T. C., R. Huang, Q. Wen, and D. Zhou (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics 134*(1), 236–251.

Green, T. C., R. Jame, S. Markov, and M. Subasi (2014). Access to management and the informativeness of analyst research. *Journal of Financial Economics 114*(2), 239–255.

Green, T. C. and S. Zhang (2024). Alternative data in active asset management. Technical report, Fisher College of Business Working Paper No. 2024-03-012 and Charles A. Dice Center Working Paper No. 2024-12. Available at SSRN: https://ssrn.com/abstract=4932677.

Grennan, J. and R. Michaely (2020). Artificial intelligence and high-skilled work: Evidence from analysts. *Swiss Finance Institute Research Paper* (20-84).

Grennan, J. and R. Michaely (2021). Fintechs and the market for financial analysis. *Journal of Financial and Quantitative Analysis 56*(6), 1877–1907.

Harford, J., F. Jiang, R. Wang, and F. Xie (2019). Analyst career concerns, effort allocation, and firms' information environment. *The Review of Financial Studies 32*(6), 2179–2224.

Hirshleifer, D. and T. Shumway (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance 58*(3), 1009–1032.

Huang, J. (2018). The customer knows best: The investment value of consumer opinions. *Journal of Financial Economics 128*(1), 164–182.

Janssen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022). GDPR and the lost generation of innovative apps. Technical report, National Bureau of Economic Research.

Jia, J., G. Z. Jin, and L. Wagman (2021). The short-run effects of the general data protection regulation on technology venture investment. *Marketing Science 40*(4), 661–684.

Jin, H., S. R. Stubben, and K. Ton (2025). Customer shopping behavior and the persistence of revenues and earnings. *The Accounting Review 100*(3), 307–332.

Johnson, G. (2022). Economic research on privacy regulation: Lessons from the GDPR and beyond. Technical report, National Bureau of Economic Research.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). Time-varying fund manager skill. *The Journal of Finance 69*(4), 1455–1484.

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

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Kacperczyk, M. T. and A. Seru (2007). Fund manager use of public information: New evidence on managerial skills. *The Journal of Finance 62*(2), 485–528.

Katona, Z., M. O. Painter, P. N. Patatoukas, and J. Zeng (2024). On the capital market consequences of big data: Evidence from outer space. *Journal of Financial and Quantitative Analysis*, 1–29.

Ke, S. (2023). The double-edged sword of data mining: Implications on asset pricing and information efficiency. *Available at SSRN*.

Kesler, R. (2022). The impact of Apple's app tracking transparency on app monetization. *Available at SSRN 4090786*.

Kim, A., M. Muhn, and V. Nikolaev (2023). Bloated disclosures: Can ChatGPT help investors process information? *arXiv preprint arXiv:2306.10224*.

Kolanovic, M. and R. T. Krishnamachari (2019). Big data and AI strategies. *JP Morgan*.

Livnat, J. and R. R. Mendenhall (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research 44*(1), 177–205.

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

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *The Journal of Finance 69*(4), 1643–1671.

Moss, S. H., B. Liberman, and G. Danenhauer (2023). Alternative data is now mainstream; AI could be next. Technical report, Lowenstein Sandler.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science 41*(4), 746–768.

Rafieian, O. and H. Yoganarasimhan (2021). Targeting and privacy in mobile advertising. *Marketing Science 40*(2), 193–218.

Ramadorai, T., A. Uettwiller, and A. Walther (2025). Privacy policies and consumer data extraction: Evidence from US firms. *Review of Finance*, rfaf017.

Requena, B., G. Cassani, J. Tagliabue, C. Greco, and L. Lacasa (2020). Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific reports 10*(1), 16983.

Ryan, D. (2023). Alternative data through a compliance lens - Seven key considerations for 2023. https://www.eaglealpha.com/2023/06/22/alternative-data-through-a-compliance-lens-seven-key-considerations-for-2023/. Eagle Alpha.

SEC (2021). SEC charges App Annie and its founder with securities fraud. https://www.sec.gov/newsroom/press-releases/2021-176. Press Release.

Sheng, J., Z. Sun, B. Yang, and A. L. Zhang (2024). Generative ai and asset management. *Review of Financial Studies (Forthcoming)*.

Statista (2024a). App - United States. https://www.statista.com/outlook/amo/app/united-states. Accessed August 02, 2024.

Statista (2024b). Number of available apps in the Apple App Store from 2008 to July 2023. https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/. Accessed August 02, 2024.

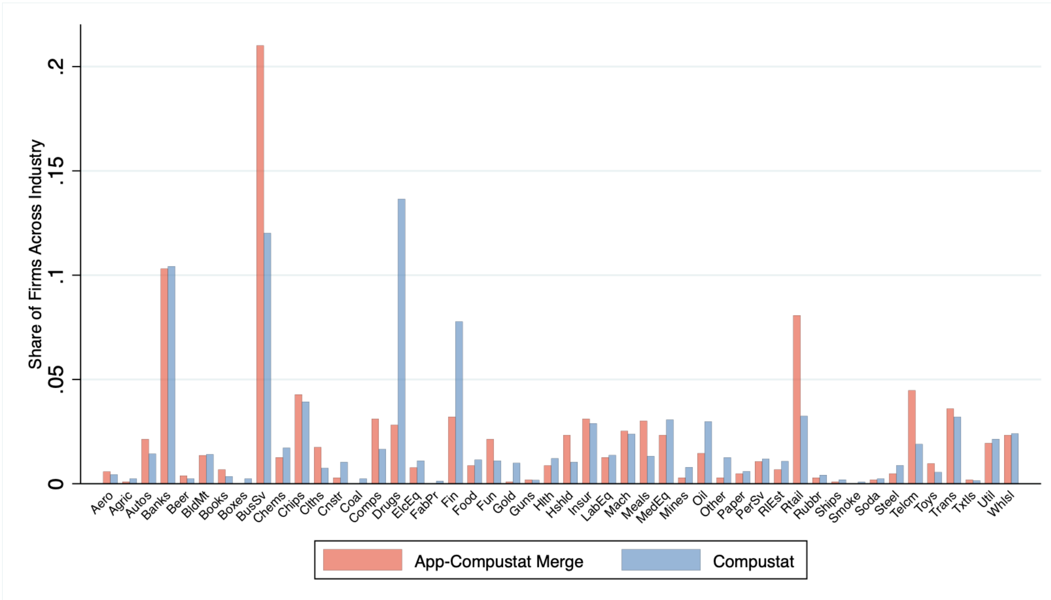
Statista (2024c). Percentage of population using the internet in the United States from 2000 to 2024. https://www.statista.com/statistics/209117/us-internet-penetration/. Accessed August 02, 2024.

Wernerfelt, N., A. Tuchman, B. T. Shapiro, and R. Moakler (2025). Estimating the value of offsite tracking data to advertisers: Evidence from meta. *Marketing Science 44*(2), 268–286.

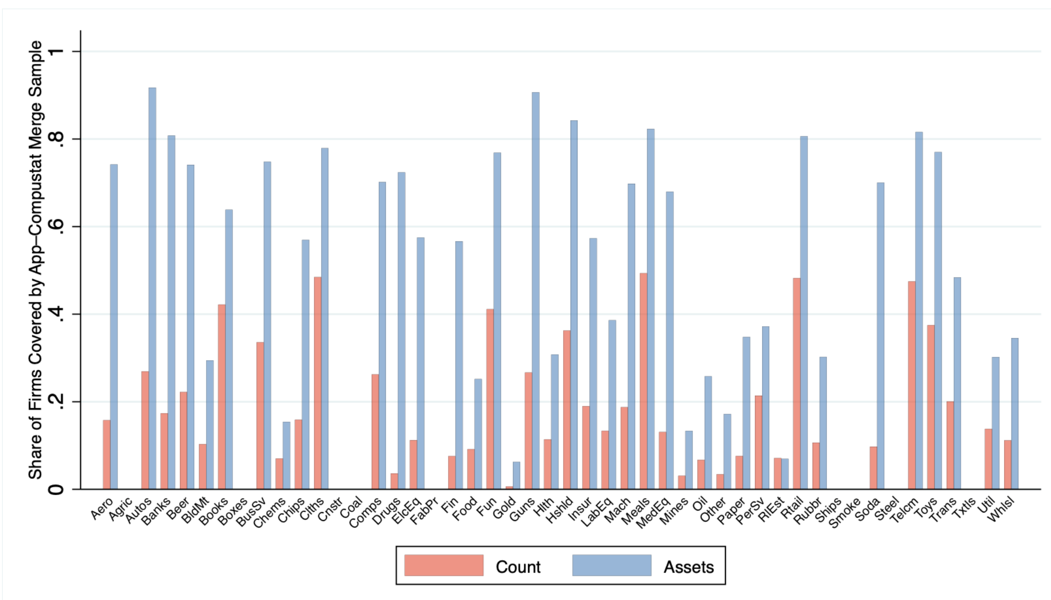
Wu, X. (2023). Mobile app, firm risk, and growth. *Available at SSRN 4519061*.

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

Figure 1: App-Owning Firms vs. Compustat Universe



(a) Industry Distribution



(b) Share of App-Owning Firms Across Industries

Figure 1a compares the industry distribution of app-owning firms ("App-Compustat Merge") with those in the broader Compustat universe ("Compustat"). It shows the distribution over the Fama-French 48 industries. Figure 1b shows the proportion of app-owning firms within each Fama-French 48 industry. Source: Bian et al. (2024).

Figure 2: Analyst Forecast Errors and Usage of Mobile Signals Around ATT

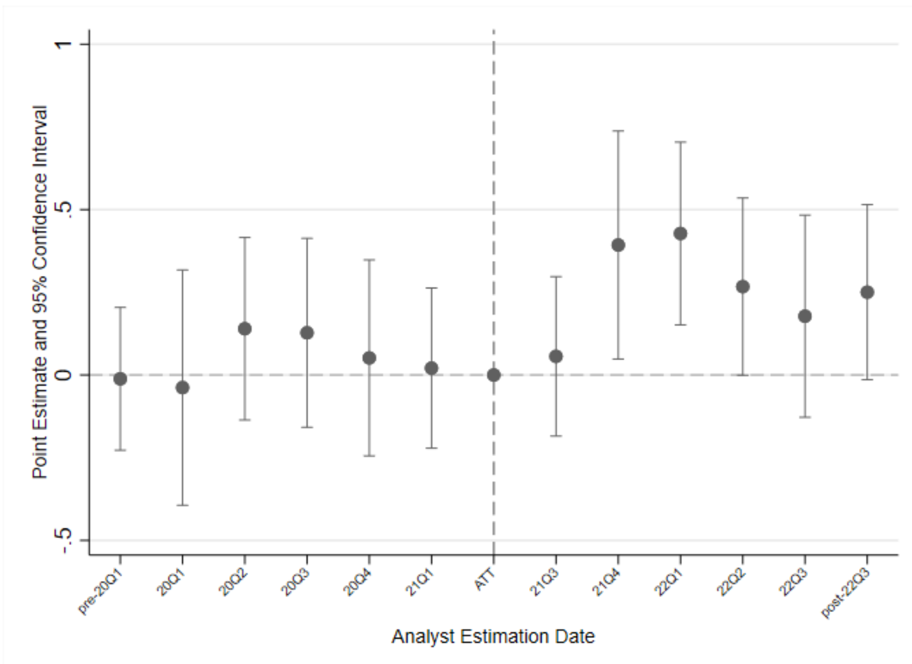


Figure 2 plots the dynamic effect of ATT on the gap in analyst forecast errors between analysts who differ in their reliance on mobile signals. The reference quarter is 2021Q2, which is the quarter of ATT implementation. We estimate coefficients on quarter dummies spanning 2020Q1 through 2022Q3, where all periods prior to 2020Q1 are grouped into one group and all periods after 2022Q3 into another.

Figure 3: Analyst Forecast Errors at Various Horizons Around ATT

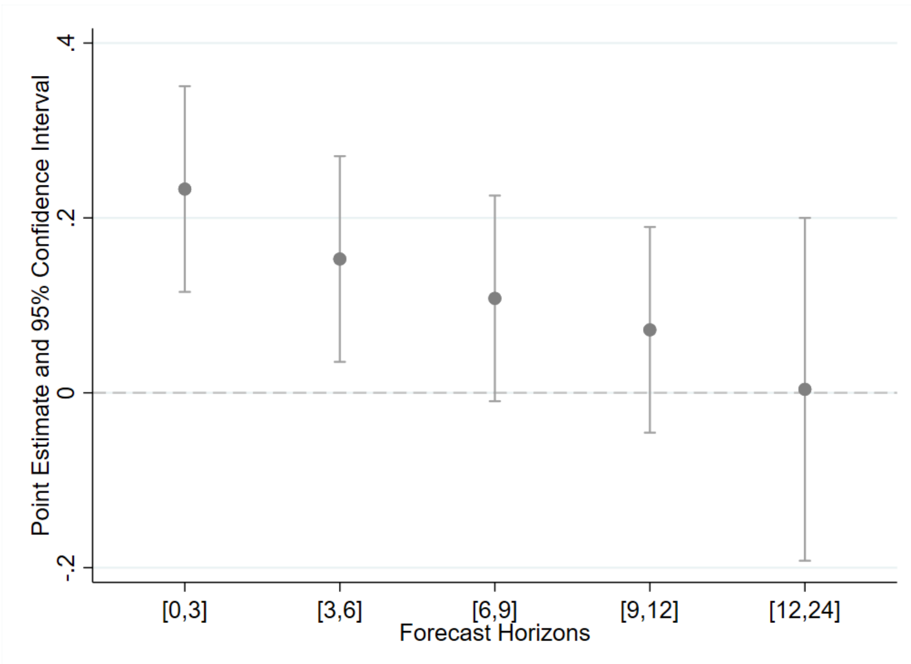


Figure 3 reports estimates of Equation (2) for the app sample. The dependent variable is $PMAFE$ (proportional mean absolute forecast error), computed at the firm-analyst-quarter level for different forecast horizons. From left to right, the estimates consider earnings forecasts for months $[t, t+3]$, $[t+3, t+6]$, $[t+6, t+9]$, $[t+9, t+12]$, and $[t+12, t+24]$. The first estimate corresponds to Column 2 of Table 4. All the other estimates are obtained with the same specification.

Figure 4: Fund Picking Ability at 12-Month Horizon Around ATT

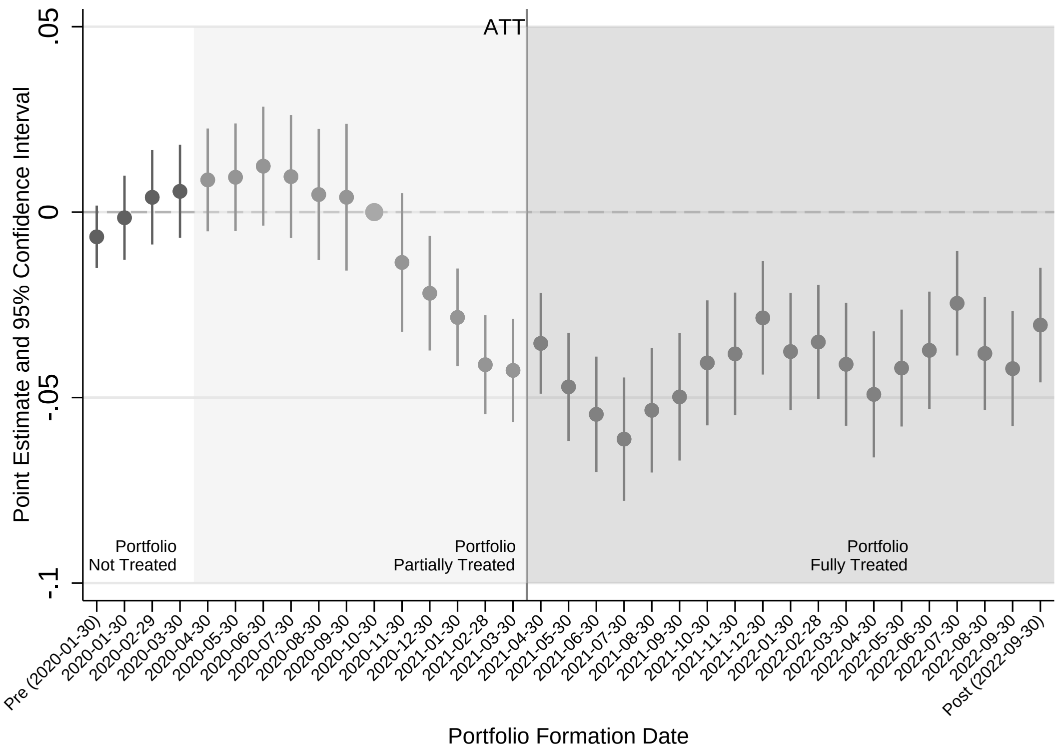


Figure 4 shows the dynamic triple-difference estimates of ATT's impact on specialized funds' picking ability. Specifically, the coefficients trace how the gap in picking ability evolves between specialized and non-specialized funds, for exposed versus non-exposed stocks, over time. The picking measure is a 12-month forward-looking CAR estimated using the Fama–French 6-factor model. The reference portfolio-formation month is October 2020, the midpoint of the partially treated period. Shading highlights how ATT enters the 12-month CAR window: the light-shaded region marks periods where portfolios are partially exposed to ATT (e.g., the 12-month performance window of the portfolio constructed at the end of April 2020 includes the post-ATT month of April 2021); the dark-shaded region marks periods where portfolios are fully exposed (i.e., the entire 12-month window falls after ATT); and the unshaded region marks pre-ATT portfolios that contain no post-ATT months. We plot coefficients on monthly dummy indicators from January 2020 through September 2022, with April 2021 (the ATT month) as the midpoint. We bin all months prior to January 2020 into a single pre-period and all months after September 2022 into a single post-period.

Figure 5: Fund Picking Ability at Various Horizons Around ATT

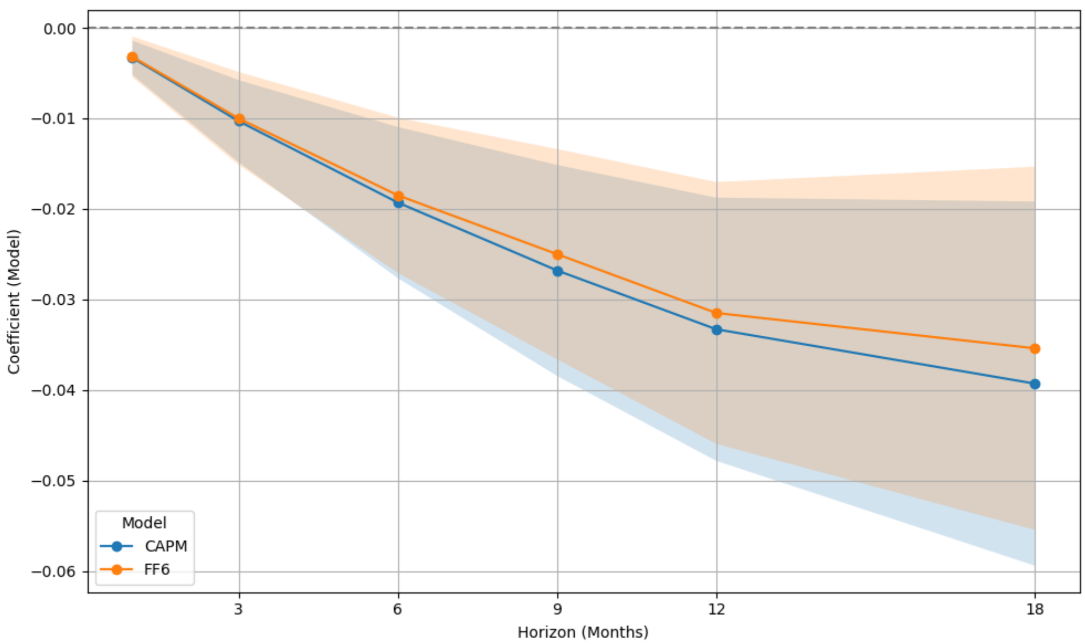


Figure 5 displays the estimated coefficients of $ATT \times Exposed \times Specialized$ as specified in Equation (7). Each dot represents an estimate based on a different *picking* measure, constructed at horizons ranging from 1 to 18 months ahead in increments of 3 to 6 months.

Table 1: Summary Statistics — Firms

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
Sales growth	8.609	30.32	-14.26	-1.62	6.87	18.44	36.00	16,165
SUE	0.003	5.88	-1.95	-0.37	0.10	0.53	1.99	16,445
Size (in log)	22.940	2.11	20.35	21.44	22.90	24.35	25.78	16,404
Cash/Assets	17.897	18.01	1.91	4.59	11.64	24.96	44.94	16,396
Tangibles/Assets	20.667	21.64	1.24	4.38	11.88	30.74	56.96	16,067
Debt/Assets	32.546	24.33	3.60	13.08	30.28	46.01	63.38	16,244
Market-to-Book	1.831	7.53	0.03	0.14	0.35	0.78	1.53	16,387
Downloads (in log)	12.583	2.52	9.45	10.86	12.55	14.21	15.79	16,445
MAU	12.836	2.66	9.60	11.09	12.82	14.49	16.13	16,445
DAU	11.377	2.98	7.64	9.36	11.47	13.35	14.98	16,445
Engagement	0.268	0.12	0.12	0.18	0.26	0.34	0.40	16,445
Bid-ask spread	3.427	1.84	1.66	2.15	2.93	4.19	5.87	14,443
Post-EA vol. (CAPM)	2.003	1.44	0.83	1.11	1.59	2.41	3.56	14,376
Post EA vol. (FF6)	1.827	1.33	0.77	1.02	1.45	2.18	3.21	14,376
Exposure to specialized analysts	0.667	0.47	0.00	0.00	1.00	1.00	1.00	14,443
Exposure to specialized funds	0.424	0.49	0.00	0.00	0.00	1.00	1.00	14,443

Panel B. No-App Sample

	mean	sd	p10	p25	p50	p75	p90	count
Sales growth	9.200	50.59	-27.16	-4.61	7.39	22.24	50.00	76,055
SUE	0.918	11.97	-3.83	-0.79	0.11	1.19	5.25	84,697
Size (in log)	20.560	2.06	17.80	19.13	20.66	21.96	23.09	91,067
Cash/Assets	24.585	28.80	1.20	3.54	10.85	36.76	77.90	90,974
Tangibles/Assets	21.132	25.60	0.49	2.00	9.88	30.62	66.69	86,689
Debt/Assets	26.716	25.47	0.28	4.81	21.17	41.98	60.18	90,282
Market-to-Book	1.204	4.51	0.09	0.26	0.55	0.96	1.61	90,128
Bid-ask spread	4.478	2.50	1.89	2.55	3.81	5.89	8.09	75,408
Post EA vol. (FF6)	2.582	1.92	0.92	1.27	1.99	3.24	4.93	74,944
Exposure to specialized analysts	0.228	0.42	0.00	0.00	0.00	0.00	1.00	91,125
Exposure to specialized funds	0.249	0.43	0.00	0.00	0.00	0.00	1.00	91,125

Table 1 Panel A includes all firms that have an app. Panel B includes all firms without an app. CAR and SUE are expressed in percentage points.

Table 2: App Downloads and Firm Performance Around ATT

Decile	Sales Growth			SUE (random walk)		
	(1)	(2)	(3)	(4)	(5)	(6)
Downloads	0.085***	0.102***	0.228***	0.031***	0.042***	0.084***
	(0.03)	(0.03)	(0.04)	(0.01)	(0.01)	(0.03)
ATT $\times$ Downloads	-0.101***	-0.111***	-0.061**	-0.064***	-0.065***	-0.038*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Industry-Quarter FE	Y	Y	N	Y	Y	N
Firm FE	N	N	Y	N	N	Y
Year-Quarter FE	N	N	Y	N	N	Y
Controls	N	Y	Y	N	Y	Y
Observations	16,165	15,645	15,647	16,445	15,619	15,621
R-sq	0.248	0.292	0.341	0.178	0.190	0.083

Table 2 presents estimation results from Equation (1) using the app sample. The dependent variable is *Sales Growth* in Columns 1–3 and *SUE* in Columns 4–6 (i.e., earnings surprises based on a random-walk model, calculated as $SUE_{it} = \frac{EPS_{i,t} - EPS_{it-4}}{P_{it}}$. Both variables are expressed in deciles. The variable *Downloads* denotes the natural logarithm of total downloads for a firm's apps in a given quarter. The indicator variable *ATT* equals to 1 for all quarters following the ATT announcement. Control variables include firm size, cash holdings, tangible assets, debt-to-asset ratio, and market-to-book ratio, all measured in quarter $t-4$. Columns 3 and 6 include firm and quarter fixed effects; the remaining columns include industry-quarter fixed effects. Standard errors, clustered at the firm level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Summary Statistics — Analysts and Reports

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
Keyword usage intensity	0.064	0.11	0.00	0.00	0.02	0.08	0.20	90,606
PMAFE	-0.025	0.53	-0.70	-0.34	-0.03	0.20	0.61	90,606
Forecast age	48.196	33.14	7.00	15.00	44.00	84.00	90.00	90,606
Analyst-firm experience	13.842	9.20	1.72	6.39	12.77	20.97	27.46	90,606
#Firms covered	14.754	7.39	6.00	10.00	14.00	19.00	24.00	90,606
Brokerage size	52.613	30.02	14.00	26.00	51.00	74.00	100.00	90,606

Panel B. Placebo Sample

	mean	sd	p10	p25	p50	p75	p90	count
Keyword usage intensity	0.028	0.08	0.00	0.00	0.00	0.02	0.08	164,275
PMAFE	-0.024	0.55	-0.73	-0.36	-0.03	0.23	0.67	164,275
Forecast age	51.214	30.65	11.00	21.00	55.00	83.00	90.00	164,275
Analyst-firm experience	11.575	8.80	0.78	4.11	10.09	17.41	24.63	164,275
#Firms covered	15.890	8.06	6.00	10.00	15.00	21.00	26.00	164,275
Brokerage size	50.740	30.45	13.00	23.00	50.00	71.00	99.00	164,275

Panel C. Analyst Report Sample

	mean	sd	p10	p25	p50	p75	p90	count
#Keywords per page	0.108	0.30	0.00	0.00	0.00	0.06	0.33	42,992
Direction-adjusted CAR - CAPM (3d)	0.498	6.49	-5.64	-2.11	0.04	2.59	6.97	42,992
Direction-adjusted CAR - FF6 (3d)	0.506	6.24	-5.34	-1.96	0.04	2.48	6.61	42,992
Price of the report	196.114	344.35	34.50	69.00	115.00	200.00	400.00	42,992
#Pages	19.907	27.83	7.00	9.00	14.00	21.00	37.00	42,992

Table 3 Panel A includes all firms that have an app averaging at least 1,000 daily active users over the sample period 2017-2023. Panel B includes firms within any of the fine-grained Fama-French 48 industries with no firms whose apps have quarterly downloads consistently above the sample median. Panel C reports summary statistics for analyst reports that reference any of the keywords listed in Appendix Table B.1, cover app-owning firms, and are available in LSEG within 4 days of the report issuance date.

Table 4: Analyst Forecast Errors Around ATT

	PMAFE			
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.236***	0.233***	0.174***	0.176***
	(0.06)	(0.06)	(0.06)	(0.06)
Keyword usage intensity	-0.089*	-0.095**		
	(0.05)	(0.05)		
ln(Forecast age)		0.012*		0.004
		(0.01)		(0.01)
Analyst-firm experience		-0.001**		-0.064*
		(0.00)		(0.04)
ln(#Firms covered)		0.001		-0.001
		(0.01)		(0.01)
ln(Brokerage size)		0.008**		-0.004
		(0.00)		(0.01)
Firm-Quarter FE	Y	Y	Y	Y
Estimation-Date FE	Y	Y	Y	Y
Analyst-Firm FE	N	N	Y	Y
Observations	90,606	90,606	89,938	89,938
R-sq	0.093	0.093	0.240	0.240

Table 4 reports the estimation results of Equation (2) based on the app sample. The dependent variable is $PMAFE$ (Proportional Mean Absolute Forecast Error), defined as the absolute forecast error (i.e., the absolute difference between an analyst's forecast and actual earnings), scaled by the mean absolute forecast error across all analysts covering the same firm in the same quarter. $PMAFE$ is computed for each firm-analyst pair in each quarter. The variable ATT is an indicator equal to one for all quarters following the ATT implementation in April 2021. The variable *Keyword usage intensity* denotes the share of reports issued by an analyst before ATT that contain mobile-related keywords, as listed in Table B.1. To construct *Keyword usage intensity*, we restrict the sample to lead analysts issuing reports on app-owning firms (available at LSEG) during 2017–2020 and whose names can be matched to I/B/E/S. *Keyword usage intensity* is computed as the number of reports referencing any mobile-related keyword divided by the total number of reports written on app-owning firms available in LSEG during 2017-20. Control variables include forecast age, analyst experience (both overall and firm-specific), the number of firms covered, brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimation-date fixed effects. Columns 3–4 additionally include analyst-firm fixed effects. Standard errors, double-clustered at the firm and estimation month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Market Reactions to Analyst Reports Around ATT

	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × #Keywords per page	-0.688**	-0.674**	-0.629**	-0.616**
	(0.28)	(0.28)	(0.28)	(0.28)
#Keywords per page	0.576**	0.545**	0.571**	0.543**
	(0.28)	(0.27)	(0.26)	(0.26)
ln(Price of the report)		0.416**		0.376**
		(0.18)		(0.17)
ln(#Pages)		-0.630***		-0.580***
		(0.24)		(0.22)
Firm-Quarter FE	Y	Y	Y	Y
Report-Date FE	Y	Y	Y	Y
Brokerage-Firm FE	Y	Y	Y	Y
Observations	42,992	42,992	42,992	42,992
R-sq	0.171	0.171	0.170	0.170

Table 5 reports the estimation results of Equation (3) based on the app sample. The outcome variable is direction-adjusted cumulative abnormal returns, where returns are multiplied by -1 when the recommendation in the analyst report is "sell". We compute direction-adjusted CAR during the three-day window following the report issuance date. In Columns 1–2 (3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. The variable *ATT* is an indicator equal to one for all quarters following the ATT announcement. The variable *#Keyword per page* is the number of mobile-related keywords divided by the number of pages in the report. The sample includes all reports on the app-owning firms over 2017-2023 that are available on LSEG within 4 days of the report issuance date. In all columns, we include firm-quarter, report-issuance-date, and brokerage-firm fixed effects. In Columns 2 and 4, we additionally control for report characteristics, including the retail price of the report as listed in LSEG and the number of pages. Standard errors, clustered at the firm level, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Summary Statistics — Funds

Panel A: Fund characteristics

Variable	Specialized	mean	std	10%	25%	50%	75%	90%	count
Fund age	N	5.218	0.89	4.01	4.81	5.43	5.77	6.06	93,545
Fund age	Y	5.153	0.86	3.99	4.68	5.38	5.71	6.01	29,346
Fund size	N	6.073	1.98	3.42	4.62	6.09	7.45	8.66	93,674
Fund size	Y	5.806	2.01	3.21	4.25	5.73	7.31	8.61	29,364
Expense ratio	N	0.009	0.00	0.01	0.01	0.01	0.01	0.01	93,353
Expense ratio	Y	0.010	0.00	0.01	0.01	0.01	0.01	0.01	29,295
Turnover	N	0.581	0.50	0.14	0.25	0.44	0.75	1.15	93,394
Turnover	Y	0.506	0.45	0.12	0.22	0.38	0.63	1.01	29,298
Flow growth	N	-0.004	0.04	-0.03	-0.01	-0.01	0.00	0.02	93,505
Flow growth	Y	-0.004	0.04	-0.03	-0.01	-0.01	0.00	0.02	29,302
Flow vol	N	0.034	0.06	0.01	0.01	0.02	0.03	0.07	90,836
Flow vol	Y	0.035	0.06	0.01	0.01	0.02	0.04	0.07	28,486
Style-MKT	N	0.981	0.12	0.84	0.92	0.98	1.05	1.12	93,674
Style-MKT	Y	0.983	0.13	0.83	0.91	0.99	1.06	1.13	29,364
Style-SMB	N	0.222	0.38	-0.17	-0.08	0.10	0.53	0.79	93,674
Style-SMB	Y	0.195	0.36	-0.19	-0.08	0.09	0.44	0.76	29,364
Style-HML	N	0.068	0.25	-0.25	-0.09	0.06	0.24	0.40	93,674
Style-HML	Y	0.041	0.26	-0.28	-0.14	0.04	0.22	0.37	29,364
Style-MOM	N	-0.037	0.15	-0.22	-0.13	-0.04	0.05	0.14	93,674
Style-MOM	Y	-0.042	0.16	-0.24	-0.14	-0.04	0.05	0.15	29,364
Style-RMW	N	-0.027	0.25	-0.34	-0.16	-0.00	0.13	0.25	93,674
Style-RMW	Y	-0.029	0.28	-0.38	-0.18	-0.01	0.14	0.29	29,364
Style-CMA	N	-0.135	0.30	-0.53	-0.32	-0.12	0.06	0.23	93,674
Style-CMA	Y	-0.161	0.32	-0.58	-0.37	-0.15	0.05	0.23	29,364

Panel B: Stock-picking ability

Variable	Specialized	Exposed	Mean	Std	10%	25%	50%	75%	90%	Count
Picking (CAPM)	N	N	-0.014	0.224	-0.148	-0.026	-0.001	0.007	0.079	8,997,884
Picking (CAPM)	Y	N	-0.035	0.381	-0.381	-0.146	-0.012	0.054	0.271	1,116,015
Picking (CAPM)	N	Y	-0.008	0.268	-0.184	-0.040	-0.001	0.022	0.142	6,085,105
Picking (CAPM)	Y	Y	-0.013	0.410	-0.371	-0.129	-0.004	0.080	0.321	1,078,312
Picking (FF6)	N	N	0.000	0.223	-0.121	-0.019	-0.000	0.011	0.101	8,997,884
Picking (FF6)	Y	N	-0.001	0.381	-0.331	-0.115	-0.003	0.081	0.321	1,116,015
Picking (FF6)	N	Y	0.001	0.248	-0.159	-0.033	-0.000	0.025	0.147	6,085,105
Picking (FF6)	Y	Y	0.003	0.379	-0.326	-0.108	-0.001	0.089	0.321	1,078,312

Table 6 Panel A reports summary statistics on time-varying fund characteristics by fund specialization as measured in Section 5.2.1. Fund-level characteristics are obtained by aggregating observations across all share classes of each fund. Fund age is based on the first offer date of the oldest share class. Fund size is calculated as the sum of the TNA of all share classes, with the natural logarithm taken of both fund age and fund size. Other variables (Expense ratio, Turnover) are weighted by lagged TNA. We measure flow growth as the percentage change in TNA not due to returns, and we compute flow volatility using 12-month rolling windows. Fund style is then constructed as the average exposure of stocks held to Fama-French 6 factors. Panel B reports summary statistics on fund stock-picking ability at the fund-stock-month level. Funds are partitioned by their specialization in mobile-generated alternative data, as defined in Section 5.2.1. Exposed stocks are those in the "App Sample", as defined in Section 3. The stock-picking measure is constructed from 12-month future idiosyncratic returns, estimated using both the CAPM and the Fama–French 6-factor (FF6) models.

Table 7: App Downloads and Changes in Portfolio Weights Around ATT

Direction of Trades	Buy	Sell	Buy	Sell
	(1)	(2)	(3)	(4)
L1.Abnormal Downloads	0.336**	-0.224*	0.288*	-0.099
	(0.136)	(0.124)	(0.157)	(0.135)
ATT × L1.Abnormal Downloads	-0.560***	0.218	-0.455**	0.352
	(0.191)	(0.190)	(0.223)	(0.228)
L2.Abnormal Downloads			0.072	-0.186
			(0.145)	(0.141)
ATT × L2.Abnormal Downloads			-0.181	-0.183
			(0.217)	(0.199)
Fund-Month FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Observations	2,602,610	2,719,110	2,594,171	2,709,828
R-sq	0.48	0.48	0.48	0.48

Table 7 presents the estimation results of Equation (5), where the outcome variable is the trade-induced weight change for stock i, fund f, and month t, calculated as the difference between the actual and no-trade weights over the three-month window from the end of month $t-3$ to the end of month t. For each month, *abnormal downloads* are defined as the log of downloads in month t minus the average log of downloads over the previous 12 months. We include up to 2 lags of abnormal downloads as explanatory variables, measured at the end of months $t-1$ and $t-2$, respectively. The variable *ATT* is an indicator set to 1 for all months following the implementation of ATT in April 2021. The results are reported separately for positive and negative trades. For each sample of fund-stock-month-level trades, we estimate two specifications. Columns 1–2 include only the abnormal downloads from month $t-1$, while Columns 3–4 include both lags. In all columns, we include fund-month and stock fixed effects. Standard errors, clustered at the fund and month levels, are reported in parentheses. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% levels, respectively.

Table 8: Fund Picking Ability at 12-Month Horizon Around ATT

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	-0.038***	-0.033***	-0.033***	-0.032***
	(0.009)	(0.007)	(0.009)	(0.007)
Exposed × Specialized	0.019***	0.017***	0.013**	0.012***
	(0.005)	(0.004)	(0.005)	(0.004)
ATT × Exposed	-0.031***		-0.030***	
	(0.006)		(0.006)	
Exposed	0.014***		0.011***	
	(0.004)		(0.003)	
Fund-Month FE	Y	Y	Y	Y
Firm-Month FE	N	Y	N	Y
Observations	17,296,237	17,277,316	17,296,237	17,277,316
R-sq	0.09	0.38	0.04	0.36

Table 8 reports the estimation results of Equation (7). The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the twelve-month horizon $[t+1, t+12]$. In Columns 1–2 (3–4), abnormal returns are estimated using the CAPM (Fama–French 6-factor) model. For each fund, stocks are classified into two groups according to their exposure to ATT. A stock is defined as *Exposed* if the issuing firm owns at least one mobile app preceding the introduction of ATT (i.e. belongs to the "App Sample"). Funds are classified as *Specialized* if their continuous specialization measure exceeds the 75^{th} percentile of the distribution, as defined in Section 5.2.1. The variable *ATT* is an indicator equal to one for all months following the ATT implementation in April 2021. For each picking measure, we report two specifications from left to right: the first includes fund×month fixed effects, and the second additionally includes firm×month fixed effects. Standard errors, triple-clustered at the fund, firm, and month levels, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Information Friction Around ATT
– by Firm Exposure to Specialized Funds and Analysts

Panel A. Exposure to Specialized Analysts

	Bid-ask Spread		Post EA Vol. (FF6)	
	(1)	(2)	(3)	(4)
ATT × Exposure	0.430***	0.351***	0.310***	0.265***
	(0.09)	(0.06)	(0.07)	(0.06)
Exposure	−0.149		−0.122*	
	(0.10)		(0.07)	
Industry-Quarter FE	Y	N	Y	N
Firm FE	N	Y	N	Y
Year-Quarter FE	N	Y	N	Y
Controls	Y	Y	Y	Y
Observations	14,443	14,445	14,376	14,378
R-sq	0.617	0.811	0.426	0.494

Panel B. Exposure to Specialized Funds

	Bid-ask Spread		Post EA Vol. (FF6)	
	(1)	(2)	(3)	(4)
ATT × Exposure	0.261***	0.166***	0.139**	0.074*
	(0.09)	(0.05)	(0.06)	(0.04)
Exposure	−0.268***		−0.210***	
	(0.08)		(0.06)	
Industry-Quarter FE	Y	N	Y	N
Firm FE	N	Y	N	Y
Year-Quarter FE	N	Y	N	Y
Controls	Y	Y	Y	Y
Observations	14,443	14,445	14,376	14,378
R-sq	0.618	0.810	0.428	0.492

Table 9 reports the estimation results of Equation (8) based on the app sample. We examine two measures of information friction: 1. average bid-ask spread during the quarter; 2. volatility of daily stock returns following earnings announcements, where volatility is defined as the root-mean-square error of the Fama-French 6-factor model. The variable *ATT* is an indicator equal to one for all quarters following the ATT implementation in 2021Q2. The variable *Exposure* equals to one if, prior to ATT, the firm (i) has analyst reports containing the relevant keywords in more than 50% of quarters (Panel A) (ii) is above the 75th percentile in terms of the share of market capitalization held by specialized funds (Panel B). Keyword-referencing reports are defined as those mentioning any app-related keywords listed in Table B.1. Specialized funds are defined as those that incorporate mobile signals into their trading strategies (Section 5.2.1). In Columns 1–2, we include industry-quarter fixed effects and in Columns 3–4, firm and year-quarter fixed effects. All regressions include control for firm-level characteristics—firm size, cash holdings, tangible assets, debt-to-asset ratio, and market-to-book ratio (all measured in quarter $t-4$)—as well as their interactions with *ATT*. Standard errors, clustered at the firm level, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

A Variables Definition

Table A.1: Variable Definition

Variable Name	Definition
Firm-Level Variables	
SUE	Standardized Unexpected Earnings. The difference between actual earnings per share (EPS) in calendar quarter t and that in calendar quarter $t-4$, scaled by stock price. Values are expressed in percentage points and winsorized at the 1% level in both tails.
CAR – CAPM	Cumulative abnormal return in a 10-day window following earnings announcements, where abnormal returns are estimated using the CAPM. Values are expressed in percentage points and winsorized at the 1% level in both tails.
CAR – FF6	Cumulative abnormal return in a 10-day window following earnings announcements, where abnormal returns are estimated using the Fama-French 6-factor model. Values are expressed in percentage points and winsorized at the 1% level in both tails.
Sales growth	Sales growth over a four-quarter period, defined as $Sales\ Growth = \frac{Sales_t - Sales_{t-4}}{2(Sales_t + Sales_{t-4})}$. Values are expressed in percentage points and winsorized at the 1% level in both tails.
Size	Natural logarithm of total assets (in dollars), winsorized at the 1% level in both tails.
Cash/Assets	Cash scaled by total assets. Expressed in percentage points and winsorized at the 1% level in both tails.
Tangibles/Assets	Tangible assets scaled by total assets. Expressed in percentage points and winsorized at the 1% level in both tails.
Debt/Assets	Total debt scaled by total assets. Expressed in percentage points and winsorized at the 1% level in both tails.
Market-to-Book	Market capitalization divided by book equity. Expressed in percentage points and winsorized at the 1% level in both tails.
Bid-ask spread	Relative bid-ask spread, defined as $\frac{Ask_{it} - Bid_{it}}{(Ask_{it} + Bid_{it})/2}$. Daily spreads are averaged at the quarterly level. Expressed in percentage points and winsorized at the 1% level in both tails.
Post-EA Vol. (FF6)	Root-mean-squared error of abnormal returns over trading days 6 to 28 following the earnings announcement, where abnormal returns are estimated using the Fama-French 6-factor model. Winsorized at the 1% level in both tails.
Exposure to specialized analysts	Indicator equal to one if the firm is covered by keyword-referencing reports in more than 50% of pre-ATT quarters.
Exposure to specialized funds	Indicator equal to one if, prior to ATT, the firm is above the 75^{th} percentile in terms of the share of market capitalization held by specialized funds.
Downloads	Natural logarithm of total quarterly app downloads for all apps owned by the firm. Winsorized at the 1% level in both tails.
MAU	Natural logarithm of total monthly active users across all apps owned by the firm, aggregated at the quarterly level. Winsorized at the 1% level in both tails.
Engagement	Ratio of daily active users to monthly active users. Winsorized at the 1% level in both tails.
Analyst-Level Variables	
Keyword usage intensity	Fraction of pre-ATT reports written by the analyst that include any app keywords.
AFE	Absolute forecast error: the absolute difference between the analyst's forecast and actual earnings.
PMAFE	Proportional mean absolute forecast error: AFE scaled by the mean AFE across all analysts forecasting the same firm in the same quarter.
Forecast age	Number of calendar days between the forecast date and the corresponding IBES report date.
Analyst-firm experience	Number of quarters since the analyst first issued a forecast for any firm. Serves as a proxy for the analyst's forecasting tenure.
#Firms covered	Number of firms covered by the analyst in a given quarter.
Brokerage size	Number of analysts employed by the brokerage firm in the quarter in which the forecast is issued.
Report-Level Variables	
#Keywords per page	Number of app keywords in the report divided by the total number of report pages.
Direction-adjusted CAR – CAPM (3d)	Direction-adjusted cumulative abnormal return over the 3-day window following report issuance, where returns are multiplied by -1 if the recommendation is "sell." Abnormal returns are estimated using CAPM. Expressed in percentage points and winsorized at the 1% level in both tails.
Direction-adjusted CAR – FF6 (3d)	Direction-adjusted cumulative abnormal return over the 3-day window following report issuance, where returns are multiplied by -1 if the recommendation is "sell." Abnormal returns are estimated using the Fama-French 6-factor model. Expressed in percentage points and winsorized at the 1% level in both tails.
Price of the report	Retail price of the report in U.S. dollars.
#Pages	Total number of pages in the analyst report.
Fund-Level Variables	
Picking	The stock-level picking ability is defined as the product of 12-month forward-looking cumulative abnormal returns (estimated using the Fama–French 6-factor model) and the portfolio's weight deviations from the benchmark. The variable is scaled by 100.
Exposed (stocks)	Indicator equal to one for stocks issued by firms that own an app pre-ATT.
Specialized (funds)	Indicator equal to one if a fund's specialization measure is above the 75^{th} percentile of the distribution. The specialization measure is defined as the product of (i) the within fixed-effect adjusted R^2 from Equation (6) and (ii) the fund's average portfolio weights in exposed stocks prior to ATT.
Fund age	Natural logarithm of the number of months since fund inception.
Fund size	Natural logarithm of total net assets (TNA) in millions of dollars.
Expense ratio	Annual ratio of total investment costs paid by shareholders to operate the fund. Winsorized at the 1% level and multiplied by 100.
Turnover	Minimum of aggregate sales or aggregate purchases of securities, divided by the average 12-month TNA. Winsorized at the 1% level and multiplied by 100.
Flow growth	Monthly percentage change in TNA not attributable to returns. Winsorized at the 1% level and multiplied by 100.
Flow volatility	12-month rolling standard deviation of monthly flow growth. Winsorized at the 1% level.
Style	TNA-weighted sensitivities (betas) of stocks held to the market (MKT), size (SMB), value (HML), momentum (MOM), profitability (RMW), and investment (CMA) factors. Betas are estimated using 24-month rolling regressions of excess stock returns on the factors. Winsorized at the 1% level.

For Online Publication:

Internet Appendix to "Breaking the Data Chain: The Ripple Effect of Data Sharing Restrictions on Financial Markets"

Table of Contents

List of Appendix Figures

List of Appendix Tables

A Pop-Up Notification Mandated by ATT

Figure A.1: An Example Prompt Mandated by Apple's App Tracking Transparency Policy

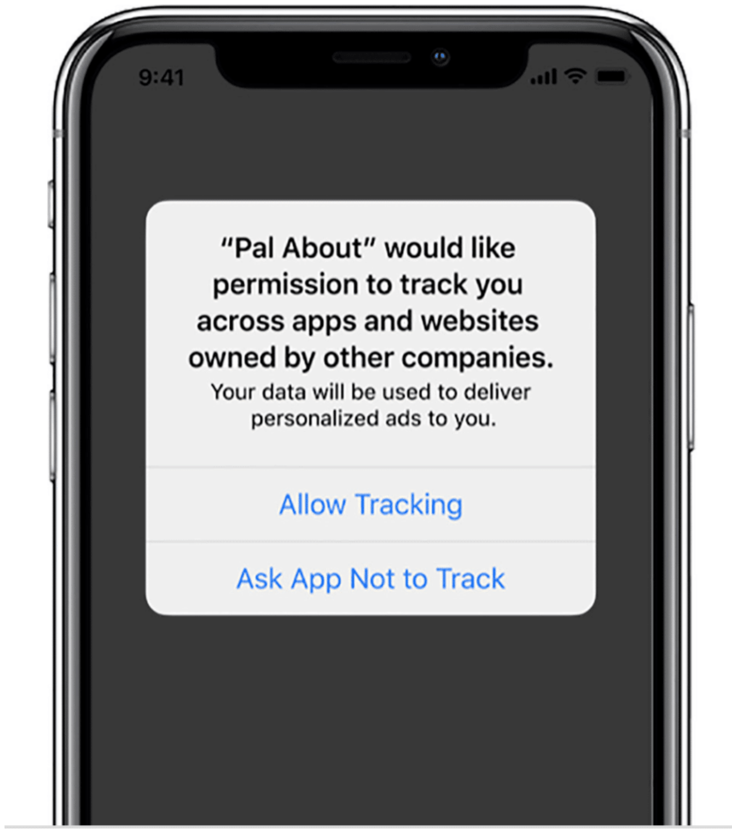


Figure A.1 presents an example of a prompt required by the ATT policy. By default, users are opted out of data sharing, and firms may share data only if users explicitly consent by clicking "*Allow Tracking*". Source: https://www.branch.io/resources/blog/what-happens-to-idfas-if-you-stop-showing-the-att-prompt/.

B Analyst Reports from LSEG Workspace

B.1 Scraping, Cleaning, and Merging

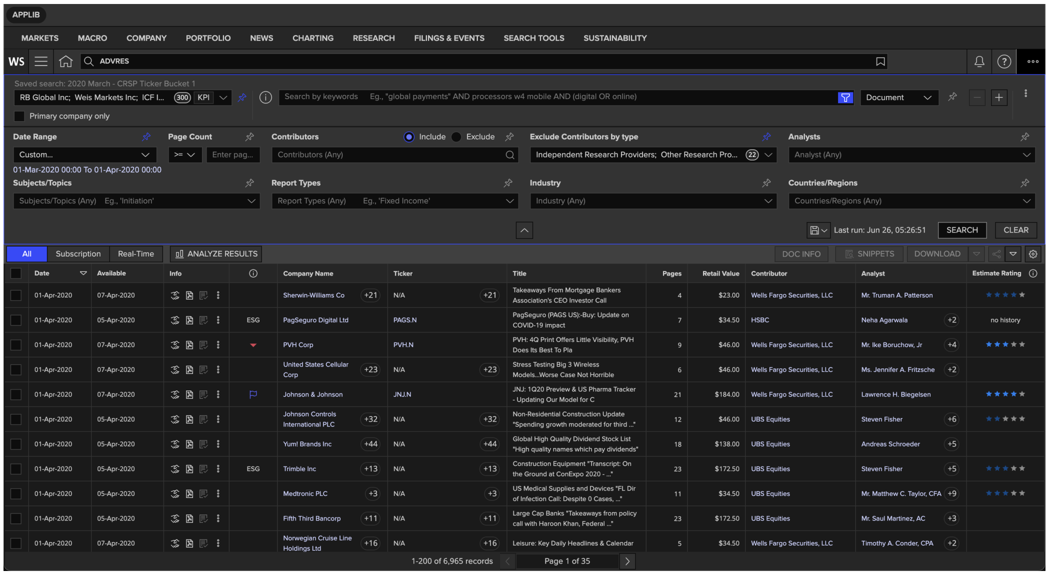
This section discusses how we obtain analyst reports from LSEG Workspace and link them to CRSP and I/B/E/S.

Scraping Keywords Data We developed an automated pipeline to extract keyword-associated metadata from the LSEG Workspace "Advanced Research" (ADVRES) portal (see Figure B.1). The aim was to identify the presence of specific keywords, covering themes such as app usage, sentiment, employment, and web traffic, within broker-authored reports. LSEG Workspace allows a maximum of 300 companies per search query, so we partitioned our full list of tickers into non-overlapping ticker buckets. For each bucket, we used LSEG's portal feature to save a pre-configured query that included the selected tickers and contributor-type filters (e.g., excluding non-broker contributors). These saved searches were recalled during each run, avoiding the need to reconfigure filters manually. The scraping was executed using a Selenium-based script on a headless Firefox browser, navigating the portal's JavaScript-heavy structure, including nested iframes and shadow DOM elements. A curated list of keywords (specifically set to comply with LSEG's keyword input constraints) was used for all queries (see Table B.1). The script looped over each keyword batch across monthly date windows, logging in, applying the filters, iterating over each report, and scraping fields such as report title, contributor, analysts, pages, and keyword match summaries (see Figure B.2 and Figure B.3). The output file shows, for each report, the frequency of every keyword in Table B.1 and the corresponding text snippets. In total, 148,005 analyst reports reference at least one keyword from Table B.1. On average, each report contains 5.2 keyword mentions.

Scraping Metadata To build a complete index of analyst reports from LSEG Workspace, we then developed a dedicated scraping pipeline focused on extracting report-level metadata. While the keyword scraping pipeline targeted documents mentioning pre-specified terms, this aimed to capture the full population of available reports, including those with no keyword matches. The setup mirrored the same browser automation stack as the keyword scraper. Each execution began by loading a saved search query preconfigured with filters on monthly date ranges, contributor types (excluding non-broker contributors), and one of the ticker buckets. In addition, within each saved query, we manually set the display option to show 50 reports per page, which helped reduce scroll-based rendering issues and ensured consistent pagination behavior during scraping. Once the results were loaded, the script paginated through them using JavaScript selectors. Because the HTML table was structured column-wise using div tags, we scraped each metadata field (e.g., contributor, analyst, date, title) column-by-column, appending entries in sequence across pages. This process produced structured metadata for every report in our firm sample, allowing us to compute the proportion that mentions specific keywords. The metadata identified 1,481,039 distinct LSEG reports issued from 2017 to 2023.

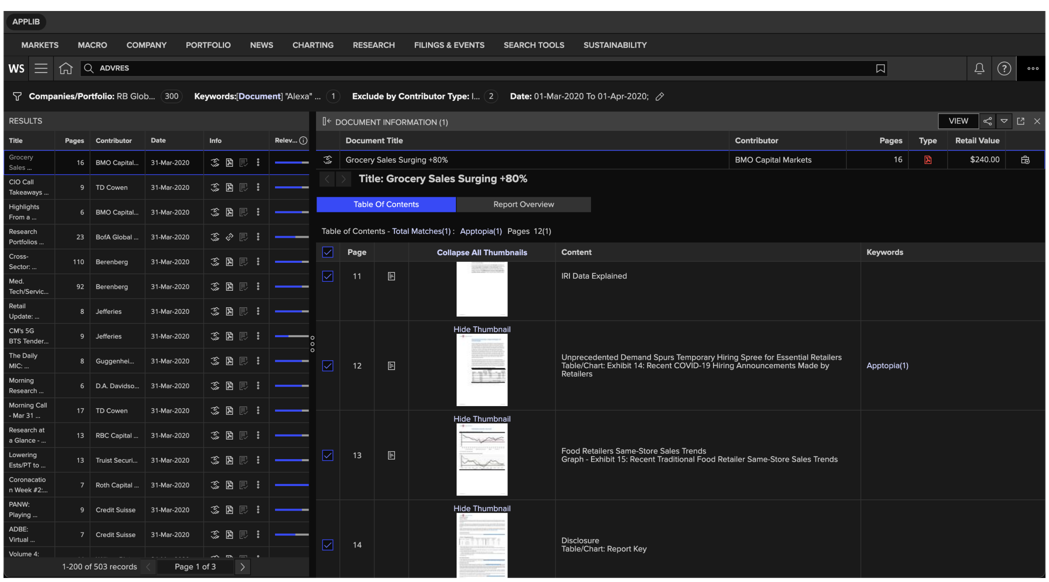
Mapping with I/B/E/S To integrate the LSEG Workspace scraped keywords data and reports metadata with standardized institutional identifiers, we developed a matching framework that linked broker and analyst names to I/B/E/S. In the first stage, we matched LSEG contributor names to I/B/E/S brokerage names. This involved cleaning the raw names by standardizing casing, removing punctuation and corporate suffixes, and standardizing whitespace. We then applied fuzzy string matching with a confidence threshold of 75% to identify likely matches. Furthermore, additional filters were used based on first-character overlap and partial substring alignment. Once the brokers were reliably matched, we resolved analyst names by cleaning and reformatting them to a "Last Name + First Initial" convention, then performing exact and fuzzy matching within each broker's analyst subset. After an initial run, we undertook several rounds of manual refinement. This involved cross-verifying analyst-broker associations online and updating the I/B/E/S file by inserting duplicate analyst entries for alternative broker-ESTIMID combinations. These adjustments were necessary because some analysts in LSEG were tied to outdated or unmatched broker IDs in I/B/E/S. By inspecting clusters of analysts repeatedly linked to the same mismatched ID, we reassigned them to more accurate identifiers based on scraping evidence and online records (e.g., LinkedIn). This manual curation process was repeated multiple times to improve both accuracy and coverage. We successfully linked more than 98% of brokerages and around 67% of analysts in LSEG to their corresponding I/B/E/S identifiers. The analyst match rate was lower because reports often list the lead analyst as “nan” or simply use the brokerage’s name. The final merged dataset incorporated both broker (ESTIMID) and analyst (AMASKCD) identifiers from I/B/E/S.

Figure B.1: Search Query for a Ticker Bucket in LSEG Workspace



[Figure B.1](#) shows the use of a saved search query configuration in the Advanced Research (ADVRES) portal, where a predefined set of 300 tickers is selected for scraping along with contributor-type and date range filters.

Figure B.2: Keyword Match Panel and Snippets in LSEG Workspace



[Figure B.2](#) displays the search results for a keyword-based query. The report-level match (e.g., "Apptopia(1)") is highlighted along with its corresponding location in the document content via a page-by-page preview of matched terms.

Figure B.3: Report Overview Panel with Attributes in LSEG Workspace

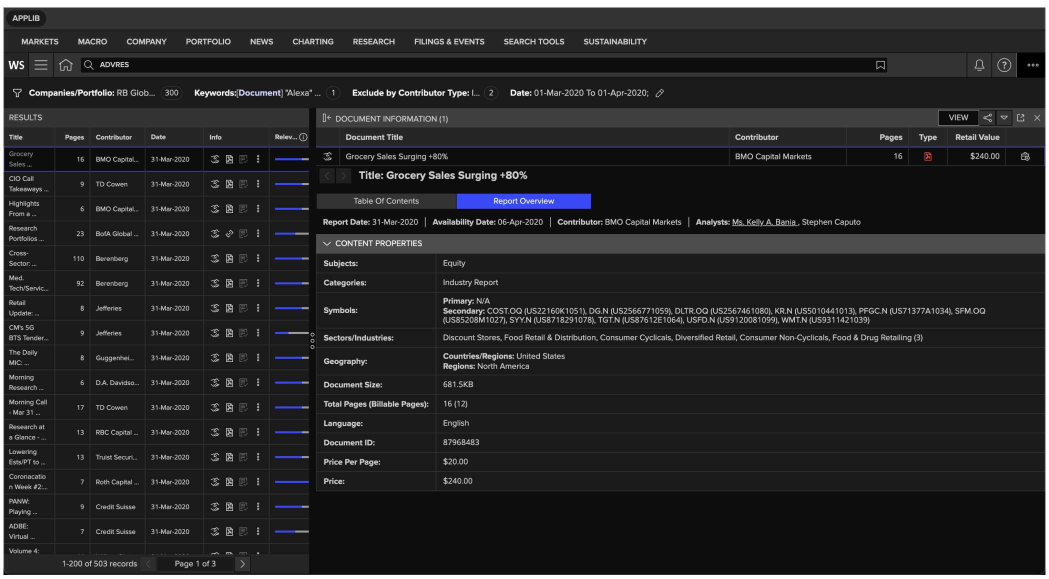


Figure B.3 presents detailed metadata for a selected broker report, including contributor, analyst(s), publication dates, page count, document size, research category, and geographic focus.

B.2 Keyword List

Table B.1: List of Mobile Data Keywords

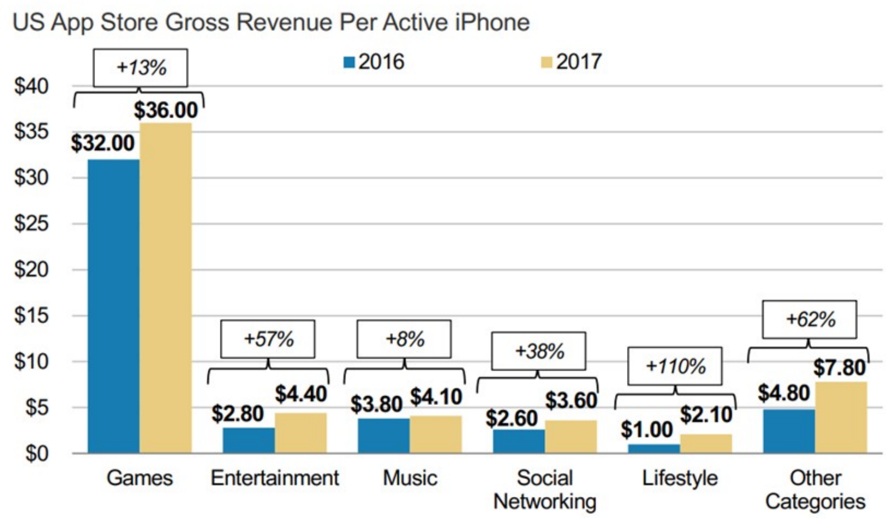
active user (8443)	alexa (10782)	app activity (340)	app annie (3578)
app data (2740)	app engagement (693)	app payment (391)	app performance (283)
app purchase (1091)	app tracking (970)	app usage (5138)	app use (184)
appannie (2009)	apptopia (9818)	brand sentiment (495)	churn rate (7925)
comscore (19362)	consumer sentiment (25703)	consumer tracking (1266)	contactless payment (922)
contactless transaction (66)	daily sessions (134)	dau (43619)	digital purchase (81)
digital transaction (695)	digital wallet (5731)	downloads (91356)	envestnet (6672)
envestnet yodlee (67)	facebook data (262)	facebook post (165)	first data (20366)
foursquare (205)	geospatial (6129)	google trend (894)	in app purchase (608)
in-app purchase (235)	instagram engagement (513)	instagram follower (173)	iresearch (5308)
mau (62480)	mobile commerce (2059)	mobile financial (507)	mobile payment (5338)
mobile transaction (332)	monthly sessions (276)	opensignal (6548)	otas (18121)
point of sale (14243)	point-of-sale (2995)	price intelligence (427)	proprietary data (14571)
quest mobile (622)	questmobile (10668)	retention rate (23066)	safegraph (5259)
search interest (10951)	search volume (4062)	sensor tower (34550)	sensortower (14240)
sentiment analysis (1068)	sentiment data (492)	session length (116)	share this (12615)
similar web (2206)	similarweb (35813)	social media analysis (170)	social media engagement (777)
spend tracker (4177)	superdata (1308)	superfly (156)	suzy (2047)
targeted advertising (1933)	thinknum (2827)	traffic analysis (645)	transaction activity (2918)
trust data (11056)	trustdata (258)	unionpay (2198)	user behavior (2196)
user data (5719)	user engagement (15254)	user interaction (315)	user spending (761)
web analytics (371)	yelp (68314)	yodlee (628)	zillow (72296)

This table lists the keywords used to identify analyst reports that reference potentially mobile-related alternative data ("Mobile keywords"). Keywords highlighted in gray correspond to all app-related terms ("App keywords"), including mobile data providers and commonly used app performance metrics. The number in brackets indicates how many times the corresponding term appears in reports issued between 2017 and 2023. We exclude certain ambiguous keywords (e.g., downloads), even though they are commonly used, as they are prone to false positives. For example, analyst reports often include phrases such as "the report is available for downloads...", which are unrelated to app performance. More generally, some terms (e.g., user interaction, mobile transaction) are not uniquely tied to app-related activity and may not be exposed to ATT, and are therefore omitted.

B.3 Examples of Mobile Data Usage

Figure B.4: Usage of Mobile Data (Sensor Tower) in Analyst Reports – An Example

Exhibit 23: US App Store in-app purchases grew 23% Y/Y in 2017, with gaming accounting for the majority of that growth.



Source: Sensor Tower Store Intelligence, Tech Crunch, Morgan Stanley Research

Japan: In late 2015 and early 2016 Japan was the top grossing App Store country in the world, despite the fact that it ranked 3rd in total downloads, after the US and China. This is because Japan is the highest grossing country per app download. More recently, the US and China overtook Japan, however Japan app revenue still accelerated in recent quarters. Looking at Sensor Tower data from C1Q18, Apple collected $1.89 (in net revenue) per download from Japanese consumers, 67% higher than the next closest country/region, Macau, which generated $1.13 per download, and over 280% greater than net revenue per download in the US ($0.49) and China ($0.43) (Exhibit 24). While Japan faces a headwind from slowing downloads (Exhibit 25), their app usage and purchasing intensity provide offsets. According to App Annie, Japan is the top market for mobile game revenue, which is the primary driver of Japan's strong App Store (and Google Play) spend. Additionally, Japanese smartphone owners use gaming Apps twice as frequently than in the US and Japan is #1 in terms of sessions in games per app user. In C1Q18, App Store net revenue in Japan grew 19% Y/Y, a 6 percentage point acceleration from C4Q18 and the second consecutive quarter of accelerating revenue growth.

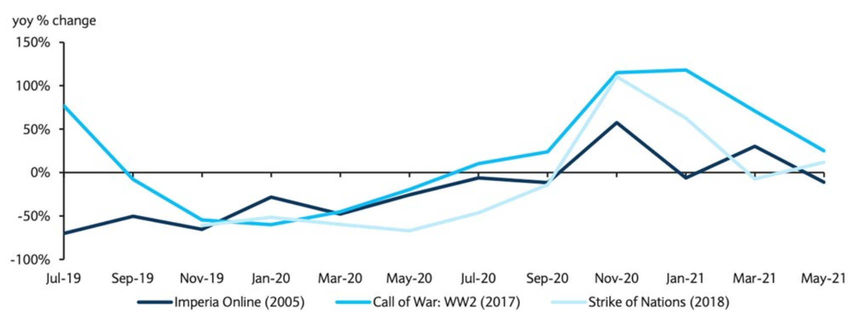
Figure B.4 comes from a Morgan Stanley report on Apple Inc., dated May 24, 2018, titled "The Emerging Power of Apple Services, Part 2: The App Store." The analysts draw on Sensor Tower's Store Intelligence data to examine App Store revenue and download trends in Japan.

Figure B.5: Usage of Mobile Data (Apptopia) in Analyst Reports – An Example

Below we look at MAUs in more detail, showing yoy % change in MAUs for key games in each product area.

As this is a portfolio of games with some improving and some declining all the time, it is difficult to read too much into the group's performance, but we show below that some Casual/Mash-up titles were really boosted in the first lockdown last year but others had interesting spikes at other times – likely as new content was added or there was a new marketing push.

FIGURE 34 **Strategy key games, yoy % change in MAUs**



Source: Apptopia

Figure B.5 is taken from a Barclays report on Stillfront titled "Recent Doubts Create Opportunity; Initiate at Overweight." The analysts use Apptopia data to illustrate year-over-year percentage changes in monthly active users (MAUs) for the company's key games.

C Predictive Power of Mobile Metrics on Firm Performance: Time-Series Pattern and Robustness

In this Appendix section, we examine the time-series pattern in the predictive power of mobile metrics for firm performance and assess the robustness of the results to alternative mobile metrics, regression specifications, and the exclusion of the COVID-19 period.

Temporal variation. Does the reduction in predictive power immediately follow the ATT implementation? We visually examine the evolution of the predictive power of downloads over time by estimating the following equation:

$$\begin{aligned}\text{Performance}_{it} = \alpha + \delta_{-6} \times \mathbb{1}\{t < -5\} \times \text{Downloads}_{it} + \sum_{k=-5, k\neq 0}^{5} \delta_k \times \mathbb{1}\{t = k\} \times \text{Downloads}_{it} \\ + \delta_6 \times \mathbb{1}\{t > 5\} \times \text{Downloads}_{it} + \Gamma_1 X'_{it-4} + \theta_{sic2,t} + \varepsilon_{it}, \quad \text{(C.1)}\end{aligned}$$

where we replace ATT_t with a set of quarter dummies, denoted as $\mathbb{1}\{t = k\}$ for $k \in [-5, 5]$. In addition, $\mathbb{1}\{t < -5\}$ and $\mathbb{1}\{t > 5\}$ correspond to all quarters prior to 2020Q1 and after 2022Q3, respectively. The indicator $\mathbb{1}\{t = 0\}$, corresponding to 2021Q2, is omitted and serves as the reference group.

Figure C.1 visualizes the estimated δ_k coefficients for sales growth (Panel (a)) and SUE (Panel (b)). For both outcomes, we observe a sharp and persistent decline in the predictive power of downloads around 2021Q2. Prior to ATT, the predictive power of downloads for sales growth is elevated in certain quarters (e.g., 2020Q3–2021Q1) relative to 2021Q2, while appearing relatively stable for SUE. One possible explanation is that app traffic contributes more directly to sales than to earnings, as increases in app traffic may be driven by temporary spikes in marketing expenditures rather than sustained improvements in firm profitability.[1]

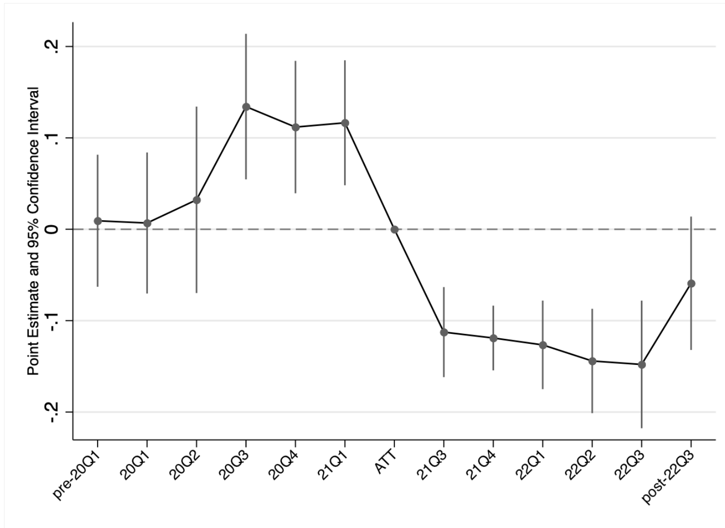
[1]Note that this is not a standard DiD dynamics figure, and parallel trends are not expected. The upward trend in Panel (a) indicates that, prior to ATT, growing app usage likely made downloads an increasingly important predictor of firm performance.

Additional mobile performance metrics. We report the predictive power of mobile data based on downloads as it is the most widely used metric, and it is highly correlated with other app traffic metrics that are frequently referenced in analyst reports. In Appendix Figure C.2, we verify that results are robust to alternative measures of traffic, such as monthly active user *MAU* or *Engagement*, the ratio between DAU and MAU, capturing the share of users who interact with the app on a frequent basis.[2] These two alternative measures were mentioned by analysts for 62,480 and 15,254 times based on Appendix Table B.1. Moreover, we show in Appendix Figure C.2 that the results are robust to using lagged downloads as the app traffic metric, alleviating concerns that our findings may be driven by reverse causality. We focus on contemporaneous app downloads in the main analysis because app traffic data is available to investors at high frequency, and downloads from the previous quarter may constitute stale information.

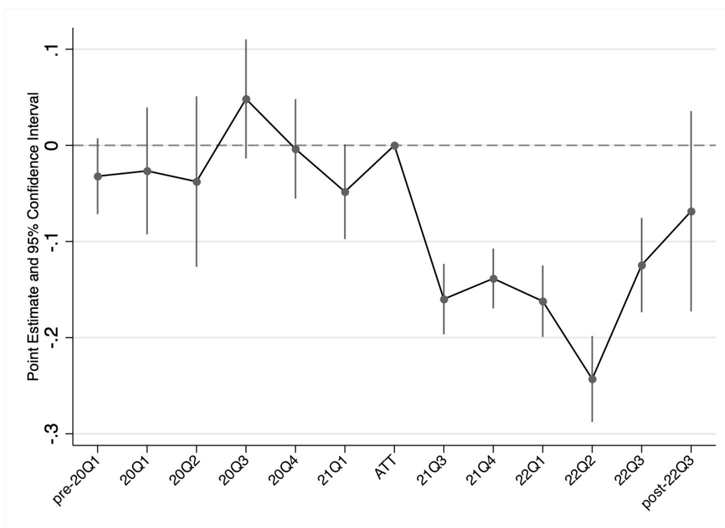
The impact of COVID-19. A potential concern is that our results may be confounded by the effects of COVID-19. Bian et al. (2021) document a surge in app traffic following the lockdown restrictions implemented in March 2020. This structural shift in user behavior from offline to online could mechanically increase the predictive power of app traffic for firm performance. To address this concern, we exclude one year of observations spanning 2020Q2–2021Q1 and compare the predictive power of app traffic in the pre-2020Q2 period and after the introduction of the ATT policy. The results, reported in Appendix Table C.2, are quantitatively similar to those from the full sample. Moreover, the evolution of predictive power exhibits the same pattern: stable prior to 2020Q2 and sharply declining in the ATT quarter, as shown in the bottom panel of Appendix Figure C.2.

[2]The median firm in our sample has 379,611 monthly active users per quarter and an engagement ratio of 26.9%, as reported in Table 1.

Figure C.1: App Downloads and Firm Performance Around ATT



(a) Sales Growth

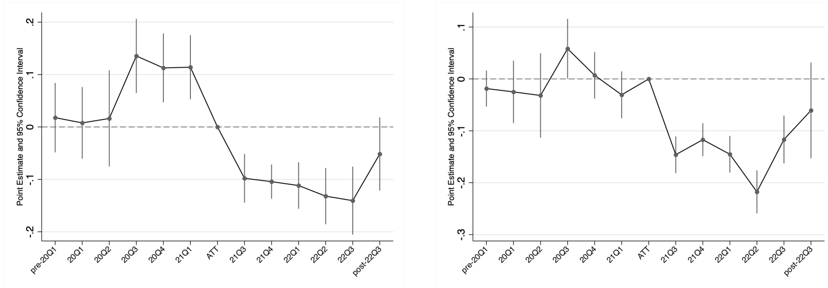


(b) SUE

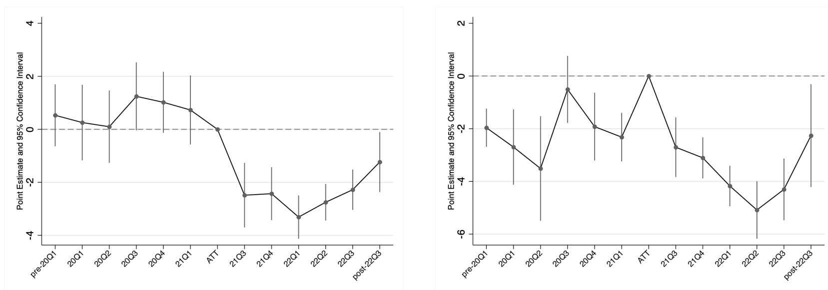
Figure C.1 plots the estimated coefficients for the quarterly time dummies from Equation C.1. Dummy variables are included for each quarter from 2020Q1 through 2022Q3, while quarters before 2020Q1 and after 2022Q3 are grouped into respective indicators. The reference category is 2021Q2, corresponding to the quarter of ATT implementation. Panel (a) shows the yearly sales growth (sales in the current quarter relative to four quarters ago), and Panel (b) displays the earnings surprise based on a random-walk model. Both variables are in deciles.

Figure C.2: Mobile Metrics and Firm Performance Around ATT – Robustness

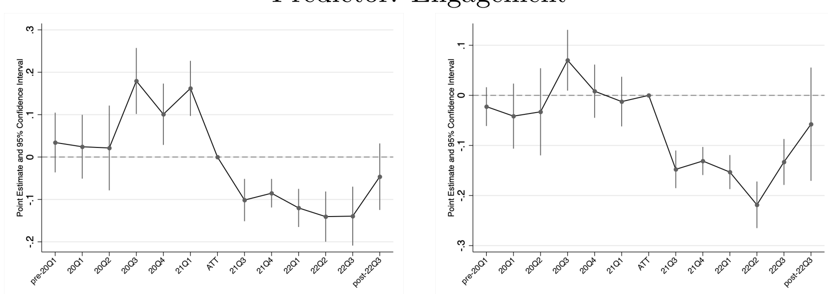
Outcome: Sales Growth **Outcome: SUE**



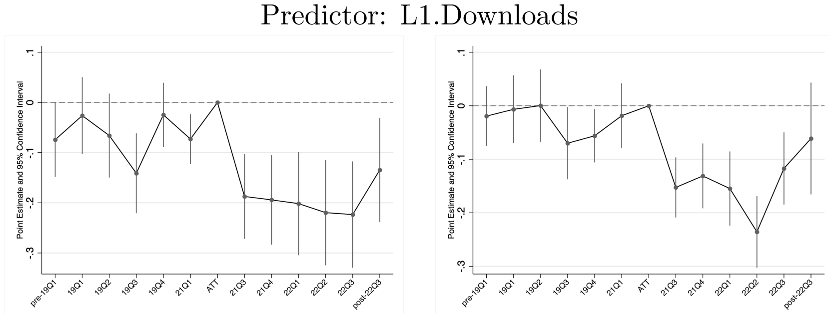
Predictor: MAU



Predictor: Engagement



Predictor: L1.Downloads



Excl. COVID period

Figure C.2 plots the estimated coefficients for the time dummy variables from Equation (C.1), where the mobile metrics are MAU, user engagement ratio, and L1.Downloads, in the top three panels. In the bottom panel, we exclude observations during the COVID-19 period (2020Q2–2021Q1). *MAU* denotes the natural logarithm of total monthly active users for a firm's apps in a given quarter. *Engagement* is defined as the ratio of DAU to MAU. *L1.Downloads* is the natural logarithm of downloads from previous quarter. In the left panel, the outcome variable is the yearly sales growth (sales in the current quarter relative to the same quarter in the prior year), and in the right panel, the outcome variable is the earnings surprise based on a random-walk model. Both variables are in deciles.

Table C.1: Mobile Metrics and Firm Performance Around ATT
– Allowing for Time-Varying Effects of Firm Characteristics

	Sales Growth		SUE	
	(1)	(2)	(3)	(4)
Downloads	0.103***	0.101***	0.050***	0.048***
	(0.03)	(0.03)	(0.01)	(0.01)
ATT × Downloads	-0.120***	-0.139***	-0.090***	-0.083***
	(0.04)	(0.03)	(0.02)	(0.02)
Industry-Quarter FE	Y	Y	Y	Y
Controls	Continuous	Decile	Continuous	Decile
Controls × Quarter	Continuous	Decile	Continuous	Decile
Observations	15,534	15,534	15,508	15,508
R-sq	0.300	0.370	0.196	0.265

Table C.1 reports estimates of Equation (1) for the app sample. To allow firm-level controls to vary over time, each control is interacted with quarter dummies. Odd-numbered columns use the continuous controls; even-numbered columns use their within-quarter deciles. The dependent variable is *Sales Growth* in Columns 1–32 and *SUE* in Columns 3–4 (i.e., earnings surprises based on a random-walk model, described in Section 4. Both variables are expressed in deciles. *ATT* equals to one for all quarters following the ATT announcement. Control variables include firm size, cash holdings, tangible assets, debt-to-asset ratio, and market-to-book ratio, all measured in quarter $t-4$. All specifications include firm and year-quarter fixed effects. Standard errors, clustered at the firm level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.2: App Downloads and Firm Performance Around ATT – Excluding COVID-19 Period

Deciles	Sales Growth		SUE (random walk)	
	(1)	(2)	(3)	(4)
Downloads	0.060*	0.078**	0.025**	0.031*
	(0.04)	(0.04)	(0.01)	(0.02)
ATT × Downloads	-0.076**	-0.090***	-0.058***	-0.059***
	(0.03)	(0.03)	(0.02)	(0.01)
Industry-Quarter FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Observations	13,497	13,061	13,722	13,037
R-sq	0.227	0.273	0.158	0.168

Table C.2 presents estimation results from Equation (1) using the app sample, excluding observations during the COVID-19 period (2020Q2–2021Q1). The dependent variable is *Sales Growth* in Columns 1–2 and SUE in Columns 3–4 (i.e., earnings surprises based on a random-walk model). Both variables are expressed in deciles. The variable *Downloads* denotes the natural logarithm of total downloads for a firm's apps in a given quarter. The indicator variable ATT equals to one for all quarters following the ATT announcement. Control variables include firm size, cash holdings, tangible assets, debt-to-asset ratio, and market-to-book ratio, all measured in quarter $t-4$. In even columns, we include control variables and their interactions with ATT. All specifications include firm and year-quarter fixed effects. Standard errors, clustered at the firm level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

D Analyst-Level Analysis: Robustness

Table D.1: Analyst Forecast Errors Around ATT – Absolute Forecast Error

	AFE/Absolute Actual Earnings			
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.097**	0.096**	0.053*	0.055*
	(0.04)	(0.04)	(0.03)	(0.03)
Keyword usage intensity	-0.038*	-0.039*		
	(0.02)	(0.02)		
ln(Forecast age)		0.004		0.002
		(0.00)		(0.00)
Analyst-firm experience		-0.000*		-0.022
		(0.00)		(0.02)
ln(#Firms covered)		0.000		-0.004
		(0.00)		(0.00)
ln(Brokerage size)		0.001		-0.003
		(0.00)		(0.00)
Firm-Quarter FE	Y	Y	Y	Y
Estimation-Date FE	Y	Y	Y	Y
Analyst-Firm FE	N	N	Y	Y
Observations	90,378	90,378	89,702	89,702
R-sq	0.851	0.851	0.882	0.882

Table D.1 reports the estimation results of Equation (2) based on the app sample. The dependent variable is AFE (Absolute Forecast Error), defined as the absolute difference between an analyst's forecast and actual earnings, scaled by the absolute actual earnings. AFE is computed for each firm-analyst pair in each quarter. The variable ATT is an indicator equal to one for all quarters following the ATT implementation in April 2021. The variable *Keyword usage intensity* denotes the share of reports issued by an analyst before the ATT that contain mobile-related keywords, as listed in Table B.1. To construct *Keyword usage intensity*, we restrict the sample to lead analysts issuing reports on app-owning firms (available at LSEG) during 2017–2020 and whose names can be matched to I/B/E/S. *Keyword usage intensity* is computed as the number of reports referencing any mobile-related keyword divided by the total number of reports written on app-owning firms available in LSEG during 2017-20. Control variables include forecast age, analyst experience (both overall and firm-specific), the number of firms covered, brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimation-date fixed effects. Columns 3–4 additionally include analyst-firm fixed effects. Standard errors, double-clustered at the firm and estimation month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D.2: Analyst Forecast Errors Around ATT – Placebo

	PMAFE			
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.044	0.046	0.027	0.032
	(0.08)	(0.08)	(0.09)	(0.09)
Keyword usage intensity	-0.001	-0.003		
	(0.05)	(0.05)		
ln(Forecast age)		0.029***		0.024***
		(0.01)		(0.01)
Analyst-firm experience		-0.000		-0.072*
		(0.00)		(0.04)
ln(#Firms covered)		0.002		-0.008
		(0.00)		(0.01)
ln(Brokerage size)		-0.005		0.000
		(0.00)		(0.01)
Firm-Quarter FE	Y	Y	Y	Y
Estimation-Date FE	Y	Y	Y	Y
Analyst-Firm FE	N	N	Y	Y
Observations	164,275	164,275	161,051	161,051
R-sq	0.130	0.131	0.301	0.302

Table D.2 reports the estimation results of Equation (2) based on the placebo sample, defined as firms within the Fama-French 48 industries where no composite firm has a popular app (e.g., download above sample median). The dependent variable is *PMAFE* (Proportional Mean Absolute Forecast Error), defined as the absolute difference between an analyst's forecast and actual earnings, scaled by the mean absolute forecast error across all analysts covering the same firm in the same quarter. *PMAFE* is computed for each firm-analyst pair in each quarter. The variable *ATT* is an indicator equal to one for all quarters following the ATT announcement. The variable *Keyword usage intensity* denotes the share of reports issued by an analyst before the ATT that contain mobile-related keywords, as listed in Table B.1. To construct *Keyword usage intensity*, we restrict the sample to lead analysts issuing reports on app-owning firms (available at LSEG) during 2017–2020 and whose names can be matched to I/B/E/S. *Keyword usage intensity* is computed as the number of reports referencing any mobile-related keyword divided by the total number of reports written on app-owning firms available in LSEG. Control variables include forecast age, analyst experience (both overall and firm-specific), the number of firms covered, brokerage size, along with their interactions with *ATT*. Columns 1–2 include firm-quarter and estimation-date fixed effects. Columns 3–4 additionally include analyst-firm fixed effects. Standard errors, double-clustered at the firm and estimation month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D.3: Analyst Forecast Errors Around ATT – Excluding COVID-19 Period

	PMAFE			
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.291***	0.287***	0.176**	0.178**
	(0.07)	(0.07)	(0.08)	(0.08)
Keyword usage intensity	-0.134**	-0.138**		
	(0.06)	(0.06)		
ln(Forecast age)		0.015**		0.010
		(0.01)		(0.01)
Analyst-firm experience		-0.001*		-0.026
		(0.00)		(0.04)
ln(#Firms covered)		-0.001		-0.007
		(0.01)		(0.01)
ln(Brokerage size)		0.006		-0.004
		(0.00)		(0.01)
Firm-Quarter FE	Y	Y	Y	Y
Estimation-Date FE	Y	Y	Y	Y
Analyst-Firm FE	N	N	Y	Y
Observations	73,841	73,841	73,103	73,103
R-sq	0.096	0.096	0.262	0.262

Table D.3 reports the estimation results of Equation (2), excluding observations during the COVID-19 period (2020Q2–2021Q1). The dependent variable is *PMAFE* (Proportional Mean Absolute Forecast Error), defined as the absolute difference between an analyst's forecast and actual earnings, scaled by the mean absolute forecast error across all analysts covering the same firm in the same quarter. *PMAFE* is computed for each firm-analyst pair in each quarter. The variable *ATT* is an indicator equal to one for all quarters following the ATT announcement. The variable *Keyword usage intensity* denotes the share of reports issued by an analyst before the ATT that contain mobile-related keywords, as listed in Table B.1. To construct *Keyword usage intensity*, we restrict the sample to lead analysts issuing reports on app-owning firms (available at LSEG) during 2017–2020 and whose names can be matched to I/B/E/S. *Keyword usage intensity* is computed as the number of reports referencing any mobile-related keyword divided by the total number of reports written on app-owning firms available in LSEG. Control variables include forecast age, analyst experience (both overall and firm-specific), the number of firms covered, brokerage size, along with their interactions with *ATT*. Columns 1–2 include firm-quarter and estimation-date fixed effects. Columns 3–4 additionally include analyst-firm fixed effects. Standard errors, double-clustered at the firm and estimation month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D.4: Analyst Forecast Errors – Switcher vs. Non-Switcher

	App sample		Placebo sample	
	non-switcher	switcher	non-switcher	switcher
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.269**	0.073	-0.213	0.024
	(0.12)	(0.12)	(0.22)	(0.17)
ln(Forecast age)	0.009	0.016	0.016*	0.059**
	(0.01)	(0.02)	(0.01)	(0.02)
Analyst-firm experience	-0.081	0.075	-0.119***	0.044
	(0.05)	(0.08)	(0.05)	(0.11)
ln(#Firms covered)	-0.009	0.007	0.003	-0.015
	(0.01)	(0.02)	(0.01)	(0.03)
ln(Brokerage size)	0.010	-0.011	0.001	-0.007
	(0.01)	(0.02)	(0.01)	(0.02)
Firm-Quarter FE	Y	Y	Y	Y
Estimation-Date FE	Y	Y	Y	Y
Analyst-Firm FE	Y	Y	Y	Y
Observations	53,157	24,367	102,830	18,713
R-sq	0.292	0.408	0.348	0.488

Table D.4 presents the estimation results of Equation (2) for both the app sample and the placebo sample, distinguishing between analysts who reduce keyword usage and those who do not. We define "non-switcher" as analysts with a below-median change in keyword usage following ATT, and "switchers" as those with an above-median change. The variable *Keyword usage intensity* denotes the share of reports issued by an analyst before the ATT that contain mobile-related keywords, as listed in Table B.1. To construct *Keyword usage intensity*, we restrict the sample to lead analysts issuing reports on app-owning firms (available at LSEG) during 2017–2020 and whose names can be matched to I/B/E/S. *Keyword usage intensity* is computed as the number of reports referencing any mobile-related keyword divided by the total number of reports written on app-owning firms available in LSEG during 2017-20. The dependent variable is $PMAFE$ (Proportional Mean Absolute Forecast Error), defined as the absolute difference between an analyst's forecast and actual earnings, scaled by the mean absolute forecast error across all analysts covering the same firm in the same quarter. $PMAFE$ is computed for each firm-analyst pair in each quarter. The variable ATT is an indicator equal to one for all quarters following the ATT implementation in April 2021. Control variables include forecast age, analyst experience (both overall and firm-specific), the number of firms covered, brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimation-date fixed effects. Columns 3–4 additionally include analyst-firm fixed effects. Standard errors, double-clustered at the firm and estimation month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

E Fund-Level Analysis

E.1 Determining the Benchmark Portfolio

To assess how actively a fund deviates from standard benchmarks, we compute weight-based comparisons between each fund's holdings and two benchmarks: (1) the CRSP market portfolio, and (2) a set of passive benchmark funds (e.g., Vanguard funds). These deviations are used to construct Active Share metrics and weight differentials for picking analysis.

CRSP Market Benchmark We use the CRSP monthly stock file to compute the market-cap weight of each stock i at each calendar month-end t:

$$w_{it}^{\text{CRSP}} = \frac{P_{it} \cdot \text{Shares}_{it}}{\sum_j P_{jt} \cdot \text{Shares}_{jt}} \quad \text{where } P_{it} = |\text{price}_{it}| \; and \; \text{Shares}_{jt} = \text{Shares outstanding}_{jt}$$

We define the market deviation for each stock held in a fund as:

$$\Delta w_{ift}^{\text{CRSP}} = w_{ift} - w_{it}^{\text{CRSP}}$$

This provides a baseline measure of how a fund deviates from the passive market portfolio.

Benchmark Funds In addition to the CRSP benchmark, we compare each fund's portfolio to a set of $\mathcal{B} = 7$ passive benchmark funds from Vanguard, including broad, large-cap, mid-cap, small-cap, growth, value, and momentum. For each benchmark $b \in \mathcal{B}$, we compute the Active Share for fund f in month t as:

$$\Delta W_{ft}^{\text{b}} = 0.5 \sum_i \left| w_{ift} - w_{it}^b \right|$$

where:

- w_{ift} is the weight of stock i in fund f in month t

- w_{it}^{b} is the weight of stock i in benchmark b in month t

Selecting the Minimum Active Share Benchmark For each fund-month (f, t), we identify the benchmark b_{ft}^{*} that minimizes the Active Share:

$$b_{ft}^{*} = \arg\min_{b \in \mathcal{B}} \Delta W_{ft}^{\mathrm{b}}$$

This approach allows us to assign each fund a dynamic benchmark that it most closely tracks, enabling more meaningful analysis of performance deviations. Our main analysis utilizes this fund-specific benchmark portfolio.

E.2 Trade-Induced Weight Changes

We use the holdings of active equity mutual funds to construct *trade-induced* changes in portfolio weights; i.e., the portion of a weight change not explained by passive returns. This measure captures *active* trading choices rather than mechanical effects from flows or price drift.[3] More specifically, we decompose each change in portfolio weight into a passive return component and an active trading component. The passive component is computed as a hypothetical *no-trade* weight that merely compounds prior holdings by realized returns:

$$w_{f,i,t}^{\mathrm{noT},h} = \frac{w_{f,i,t-h}\left(1 + \mathrm{cumret}_{i,t}^{h}\right)}{\sum_j w_{f,j,t-h}\left(1 + \mathrm{cumret}_{j,t}^{h}\right)}$$

$$for \quad \mathrm{cumret}_{i,t}^{h} = \exp\left(\sum_{\tau=t-h+1}^{t} \log(1 + r_{i,\tau})\right) - 1,$$

where $w_{f,i,t-h}$ is the weight of stock i in the portfolio of fund f at the end of month $t-h$, $\mathrm{cumret}_{i,t}^{h}$ is the cumulative return of stock i from $t-h$ to t, and j indexes all stocks in the portfolio of fund f at time $t-h$. The *trade-induced* change in the weight of stock i in the portfolio of fund f for the period $[t-h, t]$ is then defined as: $D_{ift}^{h} = 100\left(w_{i,t} - w_{i,t}^{\mathrm{noT},h}\right)$.

[3] Flow effects are absorbed when one assumes proportional allocation of flows across existing holdings. See Coval and Stafford (2007), Lou (2012), and Buffa et al. (2022) for related decompositions and discussions of flow- or return-induced changes in fund holdings.

E.3 Additional Discussion of the Specialization Measure

This section elaborates on our measurement choices that depart from comparable measures (e.g., Sheng et al. (2024) and Kacperczyk et al. (2008)), and describes the final measure.

Public–news filtration and conservative bias: our trade-induced-weight construction nets out price drift over $[t-3, t]$, while month fixed effects absorb macro or sector shocks and market-wide liquidity or flow constraints that could induce correlated trading across a fund's positions. Any residual stock-level public information not impounded in prices but correlated with abnormal downloads merely adds noise to the ranking and, once the score is applied *conditional* on a mobile-data shock, can only attenuate the estimated effect toward zero. Purging this residual with a long, ad hoc list of controls, for which no consensus exists, would substantially increase model complexity with limited additional purification.

Statistical power: month fixed effects consume one parameter per month, yet for the vast majority of funds the residual degrees of freedom remain ample. At the median, each fund-level regression has $T_{50} = 51$ months and $N_{50} = 556$ stock–month observations, leaving roughly $N_{50} - T_{50} \approx 505$ residual degrees of freedom. Fewer than 10% of funds fall below $T = 19$ and $N = 120$, where the observation-to-parameter ratio drops below 6:1. These thin panels naturally receive less weight because the within-FE-adjusted R^2 penalizes low-T, and because the specialization score is further scaled by each fund's pre-ATT average portfolio weight in signal-covered stocks. By contrast, replacing fixed effects and trade-induced weight changes with an extensive, ad hoc set of controls would use as many or more parameters while reintroducing specification risk. Our parsimonious approach therefore preserves precision and offers a transparent template for other asset-level alternative-data settings, where conditioning on a single signal and pre-shock periods makes power considerations paramount.

Relevance: abnormal downloads are informative only for a subset of firms, namely those in the "App Sample." For that reason, explanatory power alone is not sufficient to characterize economically meaningful reliance. A fund may react strongly to abnormal downloads when trading app-related names, yet devote little capital to that segment overall. Scaling the fund-level adjusted R^2 by the fund's average pre-ATT portfolio weight in exposed stocks addresses this issue by combining two margins: the intensity with which the fund's active trades respond to the signal, and the share of the portfolio for which the signal is relevant.

Measure Characteristics: Figure E.1 plots the joint distributions of the composite specialization score, the within-fixed-effects adjusted R^2, average pre-ATT exposed-stock portfolio weights, and the p-value on the fund-level abnormal-downloads coefficient β_f, separately for funds above and below the 75$^{\text{th}}$ percentile cutoff. Three patterns are worth noting. First, specialization is highly concentrated: most funds exhibit little or no trading sensitivity to the app signal and/or low exposure to app-linked stocks. Second, specialized funds cluster in the right tails of both the composite score and adjusted-R^2 distributions and have systematically lower p-values. Third, exposed-stock weights remain broadly distributed in both groups. This indicates that the classification is driven primarily by active trading variation aligned with the signal rather than by passive exposure alone. The sharp left-skew and mass near zero further support the use of a percentile-based split.

Figure E.1: Distributions of Regression Statistics by Fund Specialization

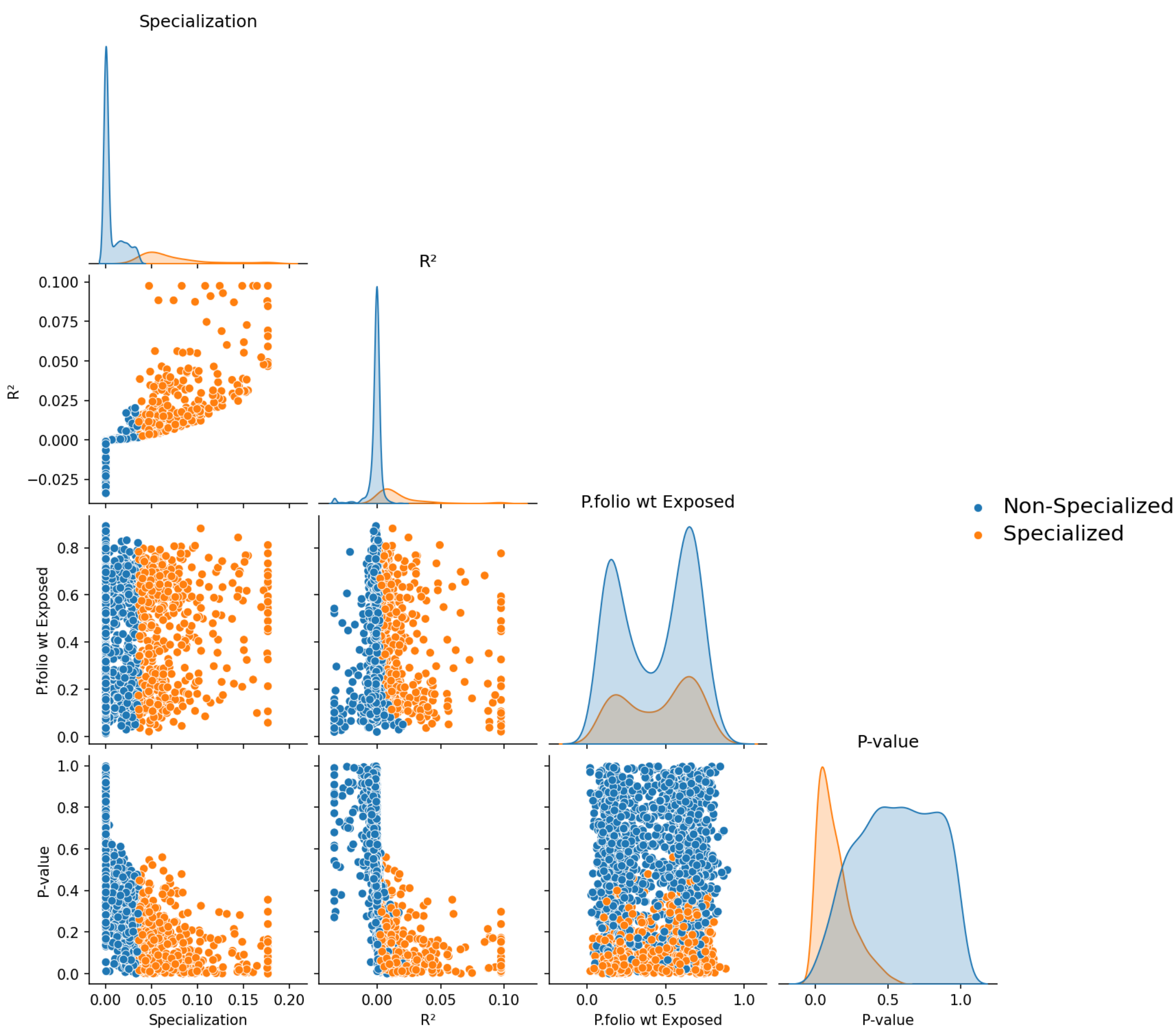


Figure E.1 displays pairwise distributions between regression statistics and components of the fund specialization measures: (i) the composite measure (portfolio weight in exposed stocks $\times$ within–fixed effect adjusted R^2, both measured prior to ATT), (ii) within–fixed effect adjusted R^2, (iii) portfolio weight in exposed stocks, and (iv) p-value of the fund-level coefficient of abnormal downloads from Equation (6), ordered from left to right and top to bottom. For visual clarity, we substitute negative within–fixed effect adjusted R^2 with zero in measure construction and take the measure's squared root. We then winsorize both the specialization measure and within–fixed effect adjusted R^2. Those transformations do not affect stocks ranking or the categorical measure.

Figure E.2: Fund Picking Ability Around ATT
– Robustness: Benchmark portfolio, Horizons, Specialized Funds

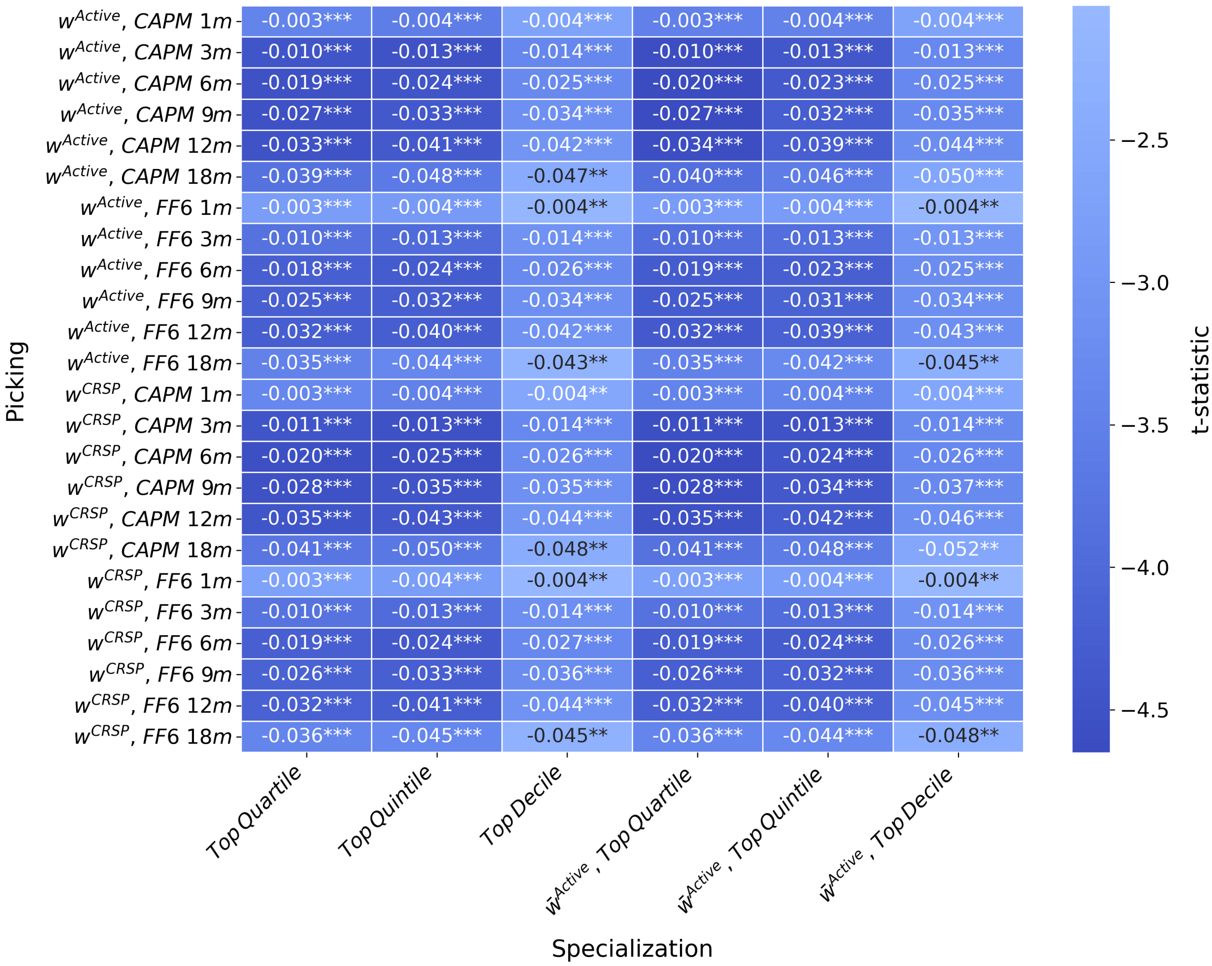


Figure E.2 reports estimates of the coefficient on the triple-interaction term ATT × Exposed × Specialized from Equation (7). The vertical axis lists alternative stock-picking measures, defined using different benchmark portfolios (CRSP market vs. the closest passive benchmark funds) and horizons for future idiosyncratic returns (1-, 3-, 6-, 9-, 12-, and 18-month). The horizontal axis shows different classifications of specialized funds, based on alternative cutoffs of the continuous specialization measure (75^{th}, 80^{th}, and 90^{th} percentiles) and on whether portfolio weights of exposed stocks are measured in raw terms or relative to the closest passive benchmark funds.

E.4 Additional Robustness

Table E.1: Fund Picking Ability at 12-Month Horizon Around ATT – Excluding COVID-19 Period

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	-0.045***	-0.037***	-0.039***	-0.036***
	(0.010)	(0.008)	(0.009)	(0.008)
Exposed × Specialized	0.026***	0.021***	0.019***	0.016***
	(0.007)	(0.005)	(0.006)	(0.005)
ATT × Exposed	-0.036***		-0.034***	
	(0.006)		(0.006)	
Exposed	0.019***		0.016***	
	(0.005)		(0.004)	
Fund-Month FE	Y	Y	Y	Y
Firm-Month FE	N	Y	N	Y
Observations	11,921,247	11,907,615	11,921,247	11,907,615
R-sq	0.07	0.37	0.04	0.36

Table E.1 reports the estimation results of Equation (7), excluding 18 COVID-relevant months from 2019m4 to 2021m3. The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the twelve-month horizon $[t+1, t+12]$, using the CRSP market portfolio as the benchmark portfolio. In Columns 1–2 (3–4), abnormal returns are estimated using the CAPM (Fama–French 6-factor) model. For each fund, stocks are classified into two groups according to their exposure to ATT. A stock is defined as *Exposed* if the issuing firm owns at least one mobile app preceding the introduction of ATT. For each exposure group. Funds are classified as *Specialized* if their continuous specialization measure exceeds the 75th percentile of the distribution, as defined in Section 5.2.1. The variable *ATT* is an indicator equal to one for all months following the ATT implementation in April 2021. For each picking measure, we report two specifications from left to right: the first includes fund×month fixed effects, and the second additionally includes firm×month fixed-effects. Standard errors, triple-clustered at the fund, firm, and month levels, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table E.2: Fund Picking Ability at 12-Month Horizon Around ATT – Placebo Stock-Level ATT Exposure

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	0.007	0.004	0.014	0.009
	(0.012)	(0.009)	(0.010)	(0.008)
Exposed × Specialized	0.001	0.000	-0.004	-0.003
	(0.006)	(0.005)	(0.005)	(0.004)
ATT × Exposed	-0.002		0.001	
	(0.007)		(0.006)	
Exposed	-0.000		-0.001	
	(0.003)		(0.003)	
Fund×Month FE	Y	Y	Y	Y
Firm × Month FE	N	Y	N	Y
Observations	17,296,237	17,277,316	17,296,237	17,277,316
R-sq	0.09	0.38	0.04	0.36

Table E.2 reports the estimation results of Equation (7), where we replace the actual stock-level ATT exposure measure with its placebo counterpart. The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the twelve-month horizon $[t+1, t+12]$. In Columns 1–2 (3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. For each fund, we classify stocks into two groups according to their exposure to ATT. We randomly assign a stock to the *Exposed* group to preserve the same fraction of *Exposed* stocks as the main sample. The variable *ATT* is an indicator equal to one for all months following the ATT implementation in April 2021. For each picking measure, we report two specifications from left to right: the first includes fund×month fixed effects, and the second additionally includes firm×month fixed-effects. Standard errors, triple-clustered at the firm, fund, and month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

E.5 Aggregating Stock-Picking Ability by Stock Exposure

In this subsection, we show that our results and their statistical significance on mutual funds' stock-picking ability are robust to an alternative aggregation at the fund-month by stock-exposure-group level. As in our baseline analysis, we classify *exposed* stocks as those in the "App Sample" ($e = 1$) and all other stocks as the comparison group ($e = 0$). Different from our baseline analysis, for each fund-month, we compute the average stock-level picking measure separately within the exposed and unexposed groups. This corresponds to the within-fund, within-group average covariance between active weights and subsequent idiosyncratic returns. Let $\mathcal{I}_{fte}$ denote the set of stocks held by fund f in month t with exposure status e. The group-level picking measure for horizon h and return definition s is:

$$\text{Picking}^{s,h}_{fte} = \frac{1}{|\mathcal{I}_{fte}|} \sum_{i \in \mathcal{I}_{fte}} \text{Picking}^{s,h}_{ift}.$$

We then estimate:

$$\begin{aligned} \text{Picking}^{s,h}_{fte} = \alpha &+ \beta_1 \, \text{Exposed}_e + \beta_2 \, \text{ATT}_t \times \text{Exposed}_e + \beta_3 \, \text{Exposed}_e \times \text{Specialized}_f \\ &+ \beta_4 \, \text{ATT}_t \times \text{Exposed}_e \times \text{Specialized}_f + \delta_{ft} + \lambda_{te} + \varepsilon_{fte}, \end{aligned} \tag{E.2}$$

where ATT_t equals one for months after April 2021. Fund-month fixed effects δ_{ft} absorb all time-varying shocks at the fund level. Exposure-by-month fixed effects λ_{te} absorb shocks common to exposed stocks in a given month, effectively controlling for sector-wide changes around ATT. We estimate Equation (E.2) for our baseline horizon $h = 12$. The magnitudes of $\hat{\beta}_4$ in Table E.3 are similar in magnitude to those in Table 8 and remain highly statistically significant, suggesting that the fund-month-stock-level results are not mechanically driven by the granularity of the data.

Table E.3: Fund Picking Ability at 12-Month Horizon Around ATT – Picking Measure Aggregated by Exposure

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	-0.043***	-0.042***	-0.034***	-0.034***
	(0.012)	(0.012)	(0.011)	(0.011)
Exposed × Specialized	0.020***	0.019***	0.014**	0.014**
	(0.007)	(0.007)	(0.006)	(0.006)
ATT × Exposed	-0.074***		-0.070***	
	(0.010)		(0.009)	
Exposed	0.035***		0.026***	
	(0.006)		(0.006)	
Fund-Month FE	Y	Y	Y	Y
Exposed-Month FE	N	Y	N	Y
Observations	238,310	238,144	238,310	238,144
R-sq	0.24	0.71	0.16	0.58

Table E.3 reports the estimation results of Equation (E.2). The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the twelve-month horizon $[t+1, t+12]$. In Columns 1–2 (3–4), abnormal returns are estimated using the CAPM (Fama–French 6-factor) model. For each fund, stocks are classified into two groups according to their exposure to ATT. A stock is defined as *Exposed* if the issuing firm owns at least one mobile app preceding the introduction of ATT. For each exposure group, stock-level picking measures are aggregated to the fund level by simple averaging. Funds are classified as *Specialized* if their continuous specialization measure exceeds the 75^{th} percentile of the distribution, as defined in Section 5.2.1. The variable *ATT* is an indicator equal to one for all months following the ATT implementation in April 2021. For each picking measure, we report two specifications from left to right: the first includes fund×month fixed effects, the second additionally includes Exposed×Month fixed-effects. Standard errors, double-clustered at the fund and month levels, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.