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

Simona Abis[†] Bo Bian[‡] Huan Tang[§]

August 31, 2025

Abstract

Alternative data has reshaped financial markets, yet it arises outside traditional disclosure channels, beyond market participants' control. We exploit Apple's App Tracking Transparency as a privacy-driven shock to alternative data generation, revealing a previously overlooked vulnerability in financial markets. The policy weakens the predictive power of mobile traffic for firm performance and mutual fund trading. Funds reliant on such data lose their edge in exposed stocks, while analyst forecasts citing mobile traffic become less accurate and elicit weaker reactions. Firms more dependent on these market participants face greater information frictions.

Keywords: Alternative Data, Mobile Apps, Data Sharing, Privacy regulation, Information friction, Analyst Forecast, Mutual Fund

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

*We thank Yakov Amihud, Jennie Bai, Maxime Bonelli, Andrea Buffa, Sean Cao, Tony Cookson, Zhi Da, Olivier Dessaint (discussant), Mengqiao Du, Vivian Fang, Thierry Foucault, Itay Goldstein, Allen Hu, Jiekun Huang, Byoung-Hyoun Hwang, Ryan Israelsen, Wei Jiang, Kai Li, Connie Mao, Ye Mao, Gordon Phillips, Laura Veldkamp, Lulu Wang, Ting Xu, Xiaoyun Yu, Xin Zheng, and seminar/conference participants at Indiana University, University of Colorado Boulder, NBER Summer Institute (Big Data and High-Performance Computing for Financial Economics), Penn/NYU Conference on Law and Finance, 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 financial market has traditionally relied on firm disclosures, regulated by the Securities and Exchange Commission (SEC) and shaped in part by market participants, who influence how such information is generated through their interactions with firms and feedback to regulators. This regime produces standardized, investment-oriented data with a stable and well-understood data generating process (DGP). However, with the rise of the digital economy and pervasive consumer surveillance, markets increasingly draw on alternative data—such as geolocation records, mobile app usage, web traffic and reviews, and credit card transactions—originating outside the traditional disclosure system and produced for purposes unrelated to investment.[1] This shift creates a new vulnerability: the DGP for alternative data depends on upstream platforms, technologies, and regulations, all of which can change unexpectedly and reshape the flow of information to markets.[2]

We trace how this vulnerability propagates in the financial market through the lens of mobile-generated data, an important category of alternative data widely used by market participants.[3] Mobile-generated data is also uniquely suited for causal analysis: Apple's 2021 App Tracking Transparency (ATT) policy introduced an unexpected shock to data generation at the source, which did not eliminate mobile-generated data but disrupted the cross-app linkages that underpinned its predictive power. We use this setting to show that adverse shocks to the DGP of alternative data reduce their predictive content, increase forecast errors, and weaken price efficiency; effects that could systematically tilt valuations of exposed firms that tend to share similar traits; e.g., a growth orientation.

[1] Nasdaq Data Link was launched in 2021 as a platform for alternative data distribution, while Bloomberg began offering alternative data products in 2019 and added its dedicated Alternative Data function (`ALTD <GO>`) in 2023, featuring products such as *Second Measure* for consumer transactions, *Similarweb* for web traffic, and *Placer.ai* for foot traffic.

[2] For example, Eagle Alpha notes that changes by upstream data providers, such as updates to terms of service or platform policies, "*can have ripple effects on data quality, security, and governance,*" highlighting how operational shifts unrelated to financial use cases can disrupt the integrity and usability of alternative datasets. See https://www.eaglealpha.com/2023/06/22/alternative-data-through-a-compliance-lens-seven-key-considerations-for-2023/.

[3] 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.

Introduced by Apple in 2021 as a privacy and consumer protection measure, ATT requires mobile users to opt in to cross-app tracking on iOS devices, effectively removing device-specific identifiers that previously facilitated linking user behavior across apps, websites, and firms. Such linkage underpins the predictive power of many alternative datasets: it enables firms to construct behavioral profiles, detect patterns, and generate precise signals that reflect firm fundamentals—signals that are accessible to and used by market participants through data vendors. For example, before ATT, McDonald's app download was a popular and meaningful signal of sales, as the company could identify potential app users who had recently checked competitors' menus on DoorDash or Uber Eats, passed near a McDonald's location as detected through Google Maps, or browsed late-night snack recipes on Instagram, and then target them with time- and location-specific promotions or meal offers.[4] After ATT, without cross-app tracking, such marketing campaigns lose precision, meaning that app downloads no longer effectively convert into sales or profit, and mobile signals lose their relevance for predicting firm fundamentals.

The scale of ATT was substantial: iOS devices account for over 50% of the U.S. smartphone market (comScore, 2024), yet only 4–18% of users opted in to tracking (Balasubramanian, 2022), abruptly disabling cross-app tracking and data sharing for a large share of mobile users. By severing the linkages that allowed targeting and signal construction, ATT sharply altered the DGP and reduced the predictive content of mobile-generated data. Importantly, ATT did not eliminate these mobile data—existing signals remained observable, yet their informational value deteriorated. In financial markets, where integrating new data sources can be costly and slow, the impact ultimately depends on how market participants adapt to such upstream vulnerabilities in alternative data generation.

Our empirical analysis combines detailed app-level activity data with firm fundamentals, mutual fund portfolio holdings, and sell-side analyst forecasts. We begin by hypothesizing that ATT's restrictions on user-level data sharing should reduce the predictive power of

[4]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`.

mobile-generated signals for firm fundamentals. Consistent with this hypothesis, we show that app traffic (e.g., downloads) positively predicts sales growth and unexpected earnings prior to ATT. Post-ATT, however, this correlation becomes statistically indistinguishable from zero, dropping sharply around the time of ATT's introduction.

Whether the decline in predictability matters for market participants depends on the incremental value of mobile data and the extent to which investors use them. Since information sets of investors are unobserved, we infer this from mutual funds' active rebalancing. Using monthly holdings, we compute trade-induced changes in portfolio weights via a no-trade counterfactual. These represent changes in a stock's portfolio share attributable to trading rather than price drift or flows, pointing to funds' active decisions. We test whether lagged abnormal app downloads explain these trades. Prior to ATT, mutual fund buying is positively related to past abnormal downloads; after ATT, this relation becomes statistically indistinguishable from zero. The decoupling of trades from app traffic is consistent with a post-ATT deterioration in the precision of mobile-generated signals.

When a key information signal becomes less precise, agents with limited analytical capacity–whether constrained by cognition or research budgets–must redirect scarce effort. In the rational-inattention model of Kacperczyk et al. (2016), portfolio managers reallocate attention, and thus capital, toward stocks whose signals remain relatively more precise. Sell-side analysts, bound by a fixed coverage list, face the multitasking trade-off of Dessaint et al. (2024): they can only shift effort between forecasting horizons, so noisier data at one horizon reduces accuracy there unless switching costs are negligible. ATT, which injects noise into mobile-app usage data, therefore predicts weaker information production for the most exposed firms: managers should reallocate away from these stocks, and analysts should post larger forecast errors. We test these predictions by tracking changes in mutual-fund stock picking and sell-side forecast accuracy around ATT.

We begin with mutual funds. We construct a fund-level measure of reliance on mobile signals (*specialization*) that captures both how systematically a fund's active trades load on abnormal downloads and how relevant those signals are for its portfolio.[5] Specifically, for each fund we regress trade-induced weight changes on lagged abnormal downloads with month fixed effects and take the resulting *within-fixed-effects adjusted* R^2 as reliance. We then define our *specialization* measure as this reliance scaled by the fund's pre-ATT average portfolio share invested in app-exposed stocks. This *specialization* measure delivers an *ex-ante* ranking of funds most reliant on mobile-generated signals prior to ATT.

We test whether funds that relied more heavily on mobile signals before ATT suffered a larger loss of comparative picking skill on the affected stocks. For each fund–month, we multiply each stock's active weight (excess over benchmark) by its idiosyncratic return over the next 1–18 months. Averaging these products separately for stocks associated with an app–thus affected by ATT–(*exposed*) and those without (*unexposed*) yields two stock-picking scores per fund-month. Formally, each score is the within-fund-month covariance between active weights and future residual returns, with higher values indicating better selection.[6] Since managers allocate limited capacity across the two sets of firms, our object of interest is the gap between exposed and unexposed picking abilities. This gap captures where forecasting effort is most effective. We then estimate a triple-difference specification that contrasts (a) pre- versus post-ATT months, (b) exposed versus unexposed stocks, and (c) specialist versus non-specialist funds. The coefficient on the triple interaction term (*Post* $\times$ *Exposed* $\times$ *Specialized*) captures whether specialists' edge on exposed stocks shrank relative to unexposed stocks, compared with the corresponding gap for non-specialists around ATT.

Estimating this specification, we find that the coefficient on *Post* $\times$ *Exposed* $\times$ *Specialized* is consistently negative and highly significant, indicating that ATT eroded specialists' comparative edge over exposed stocks relative to non-specialists. This effect through ATT's informational channel is economically large: in the 9-month baseline specification, it amounts to about half of the allocative shift we attribute to ATT's direct impact on firm fundamentals.

[5] Our measure adapts the *Reliance on Public Information* (RPI) index of Kacperczyk and Seru (2007) to an alternative–data setting (see Section 5.1.1 for details on measure construction and motivation).

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

Thus, the deterioration of mobile-data precision adds a sizable incremental effect. Event-time plots show little pre-trends and a swift, persistent widening of the estimated coefficient when post-ATT returns enter the window. The estimates remain negative out to an 18-month return horizon, suggesting that managers shifted attention across stocks, rather than simply reallocating between short- and long-term signals within the same firms. Together, these patterns show that specialists redirected scarce analytical capacity away from exposed mobile-app stocks when their informational edge eroded.

To further test our hypothesis across different types of market participants, we complement the mutual fund analysis with an investigation of sell-side analysts, whose forecast histories and information sources are directly observable. Analysts experience the same precision shock as funds, but, because their coverage lists are largely fixed, they cannot reallocate attention across stocks. Instead, the key margin is how they reallocate effort across signals about the same firm (Dessaint et al., 2024). Our empirical tests thus centers around the *within-stock* forecast gap between analysts who differ in their reliance on mobile signals.

We proxy for analyst specialization on mobile data in two ways. First, for a subset of analysts we directly identify mobile-data users by searching their reports for app-related terms or providers (e.g., daily active users, App Annie) and define "*keyword intensity*" as the share of an analyst's reports containing such references. Second, each quarter we label an analyst specialized if her coverage is concentrated in retail and services firms whose fundamentals depend heavily on app traffic (Bian et al., 2021). For every earnings forecast we calculate the analyst's error relative to same-quarter consensus and include firm-quarter fixed effects, stripping out any direct effect of ATT on firm fundamentals. Hence, a post-ATT widening of the forecast error gap between high- and low-specialization analysts pinpoints 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. The break occurs with virtually no pre-trend—consistent with a causal interpretation.

The shift is absent when the same specialists cover firms unlikely to rely on app traffic, ruling out analyst-level confounds. Interestingly, because the post-ATT penalty is larger than the prior edge, these specialized analysts end up, on average, less accurate than non-specialists.

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 now-uninformative signal and become less accurate than analysts who never used mobile data.

To see how forecast errors translate into prices, 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 the market processed and responded to this content, and that investors regarded the underlying data as especially reliable. After ATT, the premium disappears, indicating that the market no longer values the now-noisy signal. The loss of this "app-keyword premium" mirrors the post-ATT decline in specialists' forecast accuracy and underscores how ATT eroded the informational content of mobile-data-based research.

Finally, we study whether signal deterioration and the reallocation of specialist attention had measurable consequences for the firms most dependent on mobile data. Our hypothesis is that ATT weakened firms' information environments, particularly for firms whose market participants had previously relied heavily on mobile-generated alternative data. This hypothesis links the market participant-level frictions to firm-level distortions in price efficiency. Specifically, we analyze changes in bid-ask spreads and post-earnings announcement volatility, two complementary measures that capture ex-ante information asymmetry and post-disclosure residual uncertainty. Consistent with our hypothesis, we find that firms followed by specialized funds or analysts enjoyed a better information environment before ATT, characterized by narrower spreads and lower volatility. Following ATT, however, these firms experience a relative increase in both measures, suggesting more uncertainty and investor disagreements.

More broadly, DGP-driven precision shocks can create systematic mispricing whenever the affected firms share common traits. In our setting, app-intensive companies have markedly higher market-to-book ratios and, in turn, draw funds with larger growth tilts. If information slips disproportionately on one side of the growth–value spectrum, or on any other dimension that happens to coincide with exposure, the resulting mispricing need not wash out in equilibrium, adding a new layer of fragility to asset prices.

Admittedly, app-usage metrics are just one class of alternative data impaired by ATT. Any mobile-collected dataset relying on user identifiers—such as geolocation feeds tracking foot traffic—likely deteriorates post-ATT. Thus, the capital-market effects we document should be viewed as the compounded impact of lower precision across a broad range of mobile-collected datasets used by market participants.

Our results expose a core tension of the digital economy: the same user-generated data that powers marketing, competitive intelligence, and policymaking also shapes investors' expectations. Regulating these data streams is therefore inherently complex. Apple's ATT—designed for consumer privacy, not finance—disrupted the data-generating process and eroded the information feeding price discovery. The takeaway is twofold: investors should treat alternative data as fragile, and policymakers should recognize that interventions in the data economy can ripple into financial information production 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).[7] 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., 2024), with studies examining their direct effects on firms' real 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 the 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 document a rising counterforce, namely, that the very data intensity that improves market efficiency invites regulatory interventions, motivated by growing concerns over intrusive and opaque data use. We fill the gap and document that such interventions create ripple effects on financial markets. While ATT is not a government-led privacy regulation, it embodies the central feature of many such policies—the restriction of tracking and data sharing—making our findings broadly relevant for understanding the informational cost of such policies to the financial market.

Our paper also connects with a rich body of research examining the rising use of alternative data in capital markets. Prior work has documented how diverse datasets—ranging from 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

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

Google searches (Chen et al., 2020; Da et al., 2011), robo-journalism (Blankespoor et al., 2018), 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), and foot traffic (Jin et al., 2025)—can be exploited to improve forecasting and trading. This literature demonstrates the broad *supply* of alternative data that can enhance price efficiency.

A complementary strand focuses more on the *demand* side of alternative data, providing granular evidence on how specific market participants process and use such signals. The proliferation of alternative data vendors (Grennan and Michaely, 2021; Green and Zhang, 2024) and the rise of mobile-based consumer surveillance have created new information channels for analysts and fund managers. For analysts, Chi et al. (2024) show that textual references to alternative data improve forecast accuracy, while Dessaint et al. (2024) highlight the tradeoff between short- and long-term informativeness. For funds, Bonelli and Foucault (2023) exploit a positive shock from satellite imagery to show how big data 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 reveals both the limits of investors' adaptability and the fragility of information production in a market increasingly dependent on alternative data. Second, the micro detail of our setting lets us follow the entire transmission chain: we link mobile signals to firm fundamentals and track their use by both funds and analysts, and the market reactions induced by analyst reports. For funds, we extend the Reliance-on-Public-Information approach of Kacperczyk and Seru (2007) to an alternative-data feed, creating a fund-level dependence metric that can be ported to any setting where non-traditional data drive investment decisions—an independent methodological contribution that, to the best of our knowledge, is new to the literature. Third, Apple's ATT provides a clean shock in the data-generating process, giving causal evidence on how platform-controlled data streams 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) and highlight a new systemic vulnerability: capital allocation now hinges on information flows governed by actors outside the financial sector.

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 technological 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 from the mobile app market reached $153.04 billion in 2023 and are projected to grow to $230.44 billion by 2027, with mobile advertising accounting for 76% of revenues (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 DGP of Mobile-Generated Signals

The rise of mobile apps has made them a central channel for generating granular and real-time user-level data. Every time a user interacts with an app, a wide range of signals can be produced. The most direct and widely used of these are usage data, such as app installations, active users, and the frequency or length of sessions. Apps also generate loca-

[8] These figures exclude indirect benefits, such as the monetization of user data for business optimization and third-party sales.

tion data, which is harvested and commercialized by firms such as SafeGraph and X-Mode. Their business models rely on embedding tracking code across thousands of mobile apps to capture users' physical movements in real time. Transaction data is another example, generated when users make in-app purchases, pay through mobile wallets, or connect their accounts through Fintech services. Aggregators such as Stripe or Plaid consolidate these records to provide visibility into consumer spending across platforms, merchants, and financial products.[9] Together, various user interactions, whether deliberate or occurring passively in the background, form a pyramid of alternative datasets that capture diverse dimensions of consumer activity.

These raw signals are cleaned and aggregated by technology providers and analytics companies. The primary use of such processed data is operational for app developers and businesses, such as tracking user engagement and improving products. At the same time, some analytics firms (such as Apptopia and Sensor Tower) also compile these datasets across many apps and industries, turning them into broader leading 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 their use is ultimately 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 allowing their information to be shared across apps and contexts, a practice that has long been standard in the industry. 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 to observe and link user behavior across many sources and are unique to the mobile ecosystem.[10]

[9] Other examples include advertising data reflecting how users view or click on promotions, review data capturing ratings and customer feedback, social interaction data showing patterns of sharing, referrals, or community engagement, and retail investing data from platforms like Robinhood that reveal trading activity and investor sentiment.

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

With tracking enabled, firms and data vendors can start with capturing 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—producing vast streams of raw inputs that can later be assembled into alternative data products.

Besides collecting signals from many apps, a far more important step is turning those scattered pieces of information into a full picture of consumer behavior. What makes these data streams powerful is the ability to connect them across domains via the device identifiers mentioned above. This linkage allows firms to combine where a user goes, what they buy, and what they browse into a coherent profile that helps identify high-intent buyers, deliver personalized offers, and convert app usage into real sales and revenue. Those conversions are what give app-level metrics their value for financial markets, allowing measures such as downloads to serve as useful indicators of a company's performance. The same logic goes beyond app usage to other signals, such as customer reviews, where platforms draw on cross-app information to identify representative purchasers and selectively prompt them for feedback, turning scattered opinions into structured indicators of product quality and consumer sentiment.[11]

2.3 ATT as a Major Shock to DGP of Mobile-Generated Signals

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. This ecosystem created rich, high-frequency data flows across apps, and with third parties as well as data aggregators, making the mobile environment a central pillar of the alternative data economy.

This data infrastructure changed dramatically on April 26, 2021, when Apple introduced the App Tracking Transparency (ATT) framework as part of iOS 14.5. Under ATT, apps must now obtain explicit, opt-in user consent to share the IDFA or other personally identifiable data with other apps or third parties (Apple, 2024b). Upon opening an app for

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

the first time post-ATT, users are prompted with a standardized pop-up message asking whether they allow tracking "*across apps and websites owned by other companies,*" with a note that the data will be used for personalized advertising (Apple, 2024a). An example of such prompt is provided in Appendix Figure A.1. Users can also disable tracking for all apps globally via system settings.

ATT's rollout had a wide-reaching impact due to the ubiquity of the IDFA, which had previously powered the data ecosystem, and the absence of viable substitutes for linking user data across firms. Opt-in rates remained low: within one month of release, only 11% of global users and just 4% in the U.S. allowed tracking. After one year, these rates increased only modestly to 25% and 18%, respectively (Balasubramanian, 2022). Crucially, as the first policy of its kind, its impact was unanticipated by financial markets, as documented in the event study by Bian et al. (2021).

Without tracking and data sharing, firms lose the ability to connect user activity across apps and contexts. They can still record in-app events such as installs and logins, but these isolated signals no longer map reliably into revenues or customer value because the cross-app linkages that once allowed firms to target high-intent, profitable consumers disappear. Moreover, the small subset of users who opt into tracking is unlikely to be representative of the broader user base. If these "trackable" users are disproportionately low-intent—more curious but less likely to convert—then strategies optimized on their behavior will mis-target the larger population. In this environment, downloads may no longer predict higher sales and can even move in the opposite direction, as strategies based on low-intent users inflate volumes with unprofitable or non-paying customers.

Beyond app usage, ATT can also alter the DGP and reduce the quality of other alternative data sources by breaking the cross-app linkages that once underpinned their representativeness and predictive power. For instance, before ATT, location data providers could build large and representative tracking panels, since most consumers by default allowed their movements to be recorded across many apps. After ATT, as the majority of iOS users declined tracking, these panels shrank and became less representative, reducing the reliability of foot traffic measures for forecasting firm performance. A similar dynamic likely holds for customer reviews. Without access to cross-app signals, platforms like Amazon may have lost the ability to selectively solicit reviews from representative purchasers, resulting in a pool more skewed toward extreme buyers and less reflective of overall demand, thereby limiting their value in forecasting sales.

More generally, the rapid growth of the alternative data market has raised significant regulatory and data privacy concerns. Deloitte's 2017 report highlighted data provenance and compliance risks as key challenges (Deloitte, 2017), and recent surveys confirm that "increasing regulatory scrutiny" and concerns over data quality and vendor integrity remain top industry 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 disclosure or consent (SEC, 2021). According to Eagle Alpha, this case heightened industry awareness of the legal boundaries of alternative data use (Ryan, 2023). ATT is one example of a privacy-driven intervention, alongside broader regulations such as the EU's GDPR and California's CCPA, which similarly reshape data collection and sharing practices and underscore the regulatory risks that alternative data pose for financial markets.

These vulnerabilities also extend beyond mobile data and regulatory interventions, as idiosyncratic platform choices—such as recent changes at Glassdoor, Twitter/X, or Robinhood—can abruptly alter the quality and availability of widely used datasets, reducing the reliability and breath of alternative data in financial markets.[12]

Building on the above discussion, our empirical setting combines Apple's ATT policy as an unexpected sweeping shock to the GDP with widely used mobile usage signals, such as app downloads, which are both high-frequency and firm-specific. This setting enables

[12]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.

three sets of tests. First, we assess whether app usage metrics contain predictive value and are actively exploited by market participants, and whether ATT weakens this predictive power. Second, we examine how the deterioration in signal quality affects the forecasting performance of fund managers and sell-side analysts. Third, we study how these disruptions propagate to firms' information environments, with implications for price efficiency.

3 Data and Measurement

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

3.1 Mobile Apps

We obtain mobile app usage data (e.g., downloads, active users) from Apptopia, a leading alternative data provider. Apptopia combines information from partner apps, user panels, and public sources (e.g., App Store rankings), and applies proprietary AI models to estimate usage metrics for over 7 million apps, including more than 3,000 publicly listed firms worldwide. These data have been distributed via Bloomberg Terminal since 2019, making them widely accessible to financial market participants. For our analysis, we aggregate usage across all apps associated with each firm. While ATT applies directly to iOS apps, focusing on overall traffic is crucial. Because iOS users generate disproportionate revenue, firms typically design targeting and advertising strategies across platforms.[13] In practice, a restriction on iOS data could also weaken the effectiveness of Android targeting, since many advertising campaigns use cross-platform data to identify and reach high-value customers.

3.2 Fund Holdings

We obtain mutual fund data from the CRSP Survivorship-Bias-Free Mutual Fund dataset, which we use to identify active equity domestic funds. Additionally, we exclude funds with less than \$5 million in Total Net Assets (TNA) (Kacperczyk et al., 2008), and those with

[13] For example, AppsFlyer reports that non-gaming subscription ARPU on iOS is about \$8.39 versus \$1.54 on Android—over five times higher (https://www.appsflyer.com/resources/reports/app-marketing-monetization/).

less than 12 months of observations. To account for incubation bias, we exclude observations dated before each fund's first offer date (Evans, 2010). Next, we gather holdings for those funds and obtain our final sample by further requiring funds to hold at least 10 stocks, and to hold, on average, at least 80% of their assets (excluding cash) in common stock (Kacperczyk et al., 2008). When observations are missing or only available quarterly, we forward-fill them to a monthly frequency.[14]

3.3 Analyst Forecasts and Alternative Data Usage

We collect analyst forecast information from I/B/E/S. We keep forecasts issued within 90 days of each earnings announcement, taking the most recent forecast from each analyst–ticker pair. Additionally, we use LSEG Workspace (formerly Refinitiv Eikon) to extract information on the use of mobile-generated alternative data in analysts' forecasts by searching the full text of each report for keywords, based on the lists provided in Appendix Table B.1. We build our keyword list on Chi et al. (2024), augment it with app usage terms and data provider names from Dessaint et al. (2024) and Kolanovic and Krishnamachari (2019), and refine it by removing categories unlikely to be affected by ATT. Details of the keyword curation procedure appear in Appendix Section B.[15]

Among this list of keyword, we further highlight those that are particularly susceptible to ATT and frequently cited by analysts, with frequency counts reported in parentheses: *mau* (or monthly active users, 62,480), *dau* (or daily active users, 43,619), *user engagement* (15,254), *sensor tower* (34,550), *sensortower* (14,240), *Apptopia* (9,818), *active users* (8,443), *safegraph* (5,259), *app usage* (5,138), *app annie* (3,578), and *app data* (2,470). While other mobile-generated alternative data sources, such as *Facebook data* or *mobile commerce*, are

[14] In our sample period, the vast majority of funds disclose holdings at a monthly frequency.

[15] To curate this list, we begin with the set of alternative data keywords from Chi et al. (2024). We then augment this baseline with two additional keyword lists. The first focuses on app usage; for example, we include terms such as "in-app purchase" and "SensorTower," which are not included in the "app usage" category in Chi et al. (2024). Second, following Dessaint et al. (2024), we incorporate the names of data providers listed in JP Morgan's Alternative Data Handbook (Kolanovic and Krishnamachari, 2019). After merging all three sources, we manually remove data products unlikely to be affected by ATT, such as "satellite image" and "dun & bradstreet." For each firm in our app sample introduced in Section 4.1, we automate the search of each report's text in the LSEG Workspace for mentions of these terms, thereby identifying references to mobile-based alternative data.

more likely to reflect general alternative data expertise, keywords like *mau* and *user engagement* are more likely to capture ATT-impacted inputs for analysts covering firms with large mobile user bases. Illustrative report excerpts are provided in Appendix Section B.3. We refer to this set of most relevant terms as "app keywords."

We construct measures of analyst specialization in mobile-generated data using both the LSEG-scraped keyword dataset and I/B/E/S. Our first proxy is based on the app keywords mentioned in analyst reports. We measure *Keywords usage intensity* as the share of an analyst's reports that contain app-related keywords, calculated by dividing the number of reports with such keywords by the analyst's total number of reports.[16] Our second proxy is based on I/B/E/S data. We classify an analyst as specialized if their coverage is concentrated in the retail and services industries (SIC 5200–5999 or 7000–8999), where performance is closely tied to app traffic (Bian et al., 2021). Specifically, an analyst is deemed specialized when the share of forecasts on these firms in their portfolio exceeds the 90$^{\text{th}}$ percentile across all analysts in the pre-ATT period. In subsequent analysis, we measure firm-level exposure to specialized analysts using an indicator variable that equals one if the firm is followed by at least one specialized analyst in at least half of the quarters prior to ATT.

3.4 Firm Financials

We obtain financial data from various standard sources: quarterly accounting information is from Compustat, analyst forecasts are from I/B/E/S, financial market data (price, market capitalization, returns) is from CRSP, and factors are from Ken French data library.[17] A complete list of variables and their definitions used in our analysis is provided in Table A.1.

[16] We restrict our sample to lead analysts on each report because I/B/E/S only reports names and forecasts of the lead analysts.

[17] 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.

4 Decreased Quality of Mobile-Generated Alternative Data

We start by testing whether ATT lowers the signal-to-noise ratio in mobile-app-collected alternative data available to market participants. In particular, we show that after ATT, app traffic becomes significantly less predictive of firm performance, and mutual fund trading behavior decouples from these signals.

4.1 App-Relevant Firm Sample

Our main sample includes all firms with an app that averaged at least 1,000 daily active users during 2017–2023, thereby excluding the long tail of little-used apps unlikely to be relevant for firms. This yields 809 unique firms in the *"App sample"*, summarized in Panel A of Table 1. For comparison, Panel B reports statistics for 5,970 public firms without an app. Firms in the app sample are larger, more levered, and hold stronger liquidity positions than firms without an app. They also trade at higher market-to-book ratios (1.65 vs. 1.20) and exhibit substantially less volatile sales growth (standard deviation 30.3% vs. 50.6%). The median firm in the app sample has 290,140 downloads.

4.2 Predicting Firm Performance

We first examine changes in the predictability of app traffic for firm performance. App traffic data is available to market participants at a weekly or even daily frequency, allowing them to forecast firm cash flows well in advance of quarterly earnings releases. We verify this by relating firm performance measured as either yearly sales growth (i.e., $Sales\ Growth = \frac{Sales_t - Sales_{t-4}}{2(Sales_t + Sales_{t-4})}$) or unexpected earnings (SUE), as in Equation (1), in quarter t to downloads in the same quarter. In constructing SUE_{it}, we assume that earnings per share (EPS) follows a seasonal random walk (e.g., Livnat and Mendenhall (2006)). So, the best estimate of EPS for stock i in quarter t is EPS from the same quarter of the previous year ($t-4$).[18] We transform both variables into decile ranks based on the within-quarter distribution. This approach reduces the influence of outliers and measurement noise, particularly in cases of extreme volatility.

[18] As in Bradshaw and Sloan (2002), we adjust EPS to exclude special items.

$$SUE_{it}^{random\ walk} = \frac{EPS_{i,t} - EPS_{it-4}}{P_{it}} \tag{1}$$

We estimate the following regression:

$$\text{Performance}_{it} = \alpha + \theta_{sic2,t} + \beta_1 \times \text{Downloads}_{it} + \beta_2 \times \text{ATT}_t \times \text{Downloads}_{it} + \Gamma_1 X'_{it-4} + \epsilon_{it}, \tag{2}$$

where i indicates firms, t year-quarters; $Downloads_{it}$ is the natural logarithm of total downloads for apps of firm i in quarter t; ATT_t is an indicator variable equal to 1 for all quarters after ATT announcement; X'_{it-4} is a vector of control variables measured in quarter $t-4$, namely size, cash, tangibles, debt-to-assets ratio, and market-to-book ratio. To account for industry-specific factors that may influence app traffic and firm cash flows, such as demand shocks or seasonality within an industry, we include industry-by-year-quarter fixed effects ($\theta_{sic2,t}$), where industries are defined using two-digit SIC codes. This specification also enables benchmarking firms against their industry peers within the same quarter. In other words, we test whether an increase in app downloads predicts outperformance relative to peer firms in the same sector.[19]

The coefficients of interest are β_1, representing the average correlation between downloads and firm performance, and β_2, capturing the changes in the correlation post-ATT. We expect $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$.

Table 5 reports the regression results using Sales Growth (Columns 1 and 2) or SUE (Columns 3 and 4) as dependent variables. For both dependent variables, $\hat{\beta}_1$ is positive and statistically significant, while $\hat{\beta}_2$ is negative and statistically significant, indicating a decrease in the predictive power of downloads on performance post-ATT.[20] To evaluate the economic magnitude of these results, we consider Columns 2 and 4, which include firm-level control variables. An increase in downloads by 1 standard deviation moves sales growth upward by

[19]Results are similar when we include only quarter fixed effects, focusing on the cross-sectional predictive power of downloads on firm performance in the entire app sample.

[20]A contemporaneous paper, Chen et al. (2024), also shows that mobile app downloads predict earnings, consistent with $\hat{\beta}_1 > 0$. Their analysis focuses on how firm disclosure through SEC filings shapes investors' understanding of app performance.

0.25 deciles (2.52×0.104) and SUE by 0.126 deciles (2.52×0.050). Post-ATT, we observe a significant reduction in predictive power, with the magnitude of $\hat{\beta}_2$ exceeding the baseline effect. For sales growth, this implies a loss of all the additional predictive power observed pre-ATT.[21]

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:

$$\begin{aligned}\text{Performance}_{it} = \alpha + \delta_{-5} \times \mathbb{1}\{t < -4\} \times \text{Downloads}_{it} + \sum_{k=-4, k\neq 0}^{4} \delta_k \times \mathbb{1}\{t = k\} \times \text{Downloads}_{it} \\ + \delta_5 \times \mathbb{1}\{t > 4\} \times \text{Downloads}_{it} + \Gamma_1 X'_{it-4} + \theta_{sic2,t} + \varepsilon_{it}, \quad (3)\end{aligned}$$

where we replace ATT_t with a set of quarter dummies, denoted as $\mathbb{1}\{t = k\}$ for $k \in [-4, 4]$. In addition, $\mathbb{1}\{t < -4\}$ and $\mathbb{1}\{t > 4\}$ correspond to all quarters prior to 2019Q2 and after 2022Q2, respectively. The indicator $\mathbb{1}\{t = 0\}$, corresponding to 2021Q2, is omitted and serves as the reference group.

Figure 2 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., 2019Q3–2021Q1) relative to 2021Q2, while remaining 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.[22]

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

[21] In the case of SUE, the post-ATT net effect for affected firms falls slightly below the pre-ATT baseline, although we cannot reject that the sum $\hat{\beta}_1 + \hat{\beta}_2 = 0.05 - 0.09$ is equal to zero at the 5% significance level.

[22] 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.

referenced in analyst reports. In Appendix Table C.1 and Appendix Figure C.1, 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.[23] 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 Table C.1 and Appendix Figure C.1 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.

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.1.

While one might expect some residual predictive power from mobile signals like downloads following a data degradation shock, our findings reveal a near-complete breakdown. As discussed in Section 2.3, two factors contribute to this decline. First, ATT severely limits firms' ability to track and target users across apps and sessions, weakening the link between downloads and monetization. Without visibility into user behavior, downloads become a much noisier proxy for revenue. Second, ATT introduces sample selection: the small group

[23]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.

of users who opt into tracking are unlikely to represent the broader population. Their behavior provides weak or even misleading signals of purchase intent, and strategies built on them tend to attract low-intent customers. As a result, download spikes may increasingly reflect low-value acquisitions, with little or even negative correlation to actual sales or profits.

4.3 Predicting Trading Activities

The preceding evidence shows that app downloads lose their predictive power for firm fundamentals after ATT. Whether this matters for asset prices hinges on how intensively investors used that signal before—and continue to use it after—the shock. Although information sets are unobservable, portfolio allocations often are. We therefore use the holdings of active equity mutual funds to construct *trade-induced* changes in portfolio weights, 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.[24] We then test whether these active buy- and sell decisions are predicted by lagged abnormal downloads, and whether that predictive link weakens after ATT, consistent with a loss of signal precision.

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:

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

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 $t-h$.[25] 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: $D_{ift}^{h} = 100\Big(w_{i,t} - w_{i,t}^{\mathrm{noT},h}\Big)$.

[24] 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.

[25] We compute $\mathrm{cumret}_{i,t}^{h} = \exp\Big(\sum_{\tau=t-h+1}^{t} \log(1 + r_{i,\tau})\Big) - 1$.

We then relate trade-induced weight changes to abnormal downloads in preceding months, by estimating:

$$D_{ift}^{h} = \alpha + \theta_{ft} + \iota_i + \sum_{k=1}^{h-1} \beta_k \times \text{Abnormal Downloads}_{i,t-k} + \sum_{k=1}^{h-1} \delta_k \times \text{Abnormal Downloads}_{i,t-k} \times \text{ATT}_t + \epsilon_{ift}, \quad (4)$$

where D_{ift}^{h} is the trade-induced change in the portfolio weight of stock i held by fund f over $[t-h, t]$; θ_{ft} are fund-by-month fixed effects; ι_i are stock fixed effects; Abnormal $\text{Downloads}_{i,t-k}$ denotes abnormal downloads for firm i measured at month end $t-k$, defined as log downloads in month $t-k$ minus the average log downloads over the prior 12 months.[26] We set $h = 3$, so changes over the three-month window $[t-3, t]$ are linked to signals observed at $t-2$ and $t-1$, which mitigates lookahead concerns when monthly holdings are missing or reported mid-month. We report standard errors two-way clustered by fund and month.

Fund–month fixed effects absorb time-varying fund shocks (e.g., flows), and stock fixed effects capture persistent firm heterogeneity (e.g., trading intensity), so identification comes from within–fund–month differences in trading across stocks as a function of abnormal downloads. In this setting, 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 4 separately for buys and sells because mutual funds are often subject to long-only mandates, making negative signals harder to implement than positive ones.[27] Accordingly, we expect $\hat{\beta}_k > 0$ and $\hat{\delta}_k < 0$ for buys, with smaller effects for sells.

Table 6 presents the results. On the buy side (Column 1), an increase in abnormal downloads in month $t-1$ predicts positive trades in month t before ATT. This predictive

[26] As in the firm-level analysis we restrict the sample to firms that have an app with on average at least 1,000 daily active users throughout our sample period.

[27] 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.

power disappears entirely afterward. Adding a second lag (Columns 3) provides no additional predictive content. On the sell side (Column 2 and 4), the effects are noisy and mostly insignificant, likely due to sparse sell-side activity due to mutual funds' long-only mandates.

5 ATT Decreases the Forecasting Ability of Market Participants

So far we have shown that ATT reduced the usefulness of mobile-generated data and that investors began relying less on them. Next we trace how this decline in information precision affected mutual-fund managers (Section 5.1) and sell-side analysts (Section 5.2).[28] We rank each group by its pre-ATT reliance on mobile signals and then test how their stock-picking ability or forecasting accuracy changed post-ATT, relative to their lower-reliance peers.

5.1 Mutual Funds

5.1.1 Funds' Specialization in Mobile-Generated Alternative Data

Although we cannot observe a fund's full information set, we can infer its reliance on mobile-app data from how strongly its trades covary with abnormal downloads. We label this reliance *specialization*. The concept builds on Kacperczyk and Seru (2007), which uses the R^2 of regressing percentage holding changes on analyst-forecast revisions as a measure of "Reliance on Public Information" (RPI). Adapting RPI to an alternative-data setting requires a few modifications. First, we regress on *trade-induced* weight changes, i.e., weights net of mechanical price- and flow-driven drift (Section 4.3), so any public news already impounded in prices is stripped out, and only intentional, signal-driven trades remain. Second, our regressor is abnormal app downloads, an alternative-data feed whose precision is plausibly altered by ATT. Third, we absorb month fixed effects and take the *within-fixed-effects (FE) adjusted* R^2 as metric of signal reliance, which both adjusts for market-wide news, and corrects for degrees of freedom. Fourth, because such signals exist only for a subset of stocks, we scale each fund's signal reliance metric by its pre-ATT average portfolio weight in exposed stocks, so the score reflects both *how strongly* a manager reacts to the signal and *how much capital* that signal can move. These changes allow the specialization score to measure a fund's

[28] The dual focus on trades and analyst forecasts is similar to Gao and Huang (2020).

effective dependence on a stock-specific alternative-data source, something the original RPI framework, geared toward market-wide public news, does not capture.

More in detail, for every fund f,[29] we run a separate time–series regression:

$$D^{+}_{ift} = \beta_f \text{ Abnormal Downloads}_{i,t-1} + \mu_t + \varepsilon_{ift}, \quad \forall f \in \mathcal{F} \quad \& \quad \forall t < ATT, \tag{5}$$

where D^{+}_{ift} is the trade-induced buy weight change for stock i over $[t-3, t]$ (Section 4.3) and Abnormal $\text{Downloads}_{i,t-1}$ is abnormal downloads observed at the end of month $t-1$.[30] Month fixed effects μ_t purge aggregate shocks, so β_f and the within–FE adjusted R^2_f capture the extent to which fund-specific trading reacts to the mobile-data signal. Our measure of signal reliance is the within-FE adjusted R^2_f from (5), which quantifies the portion of each fund's active trading variance that is explained by the variance of abnormal downloads.

We add no further controls for two main reasons. (i) *Public–news filtration and conservative bias:* our trade-induced-weight construction nets out the full price drift over $[t-3, t]$, while the month fixed effects (a) absorb month-specific macro or sector shocks that could move all app-related trades, (b) sweep out market-wide liquidity or flow constraints that induce correlated trading across a fund's positions. Any residual stock-level public information not impounded in price changes but still covarying 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, biasing it toward zero.[31] (ii) *Statistical power:* month fixed effects consume one parameter per month, yet for the vast majority of funds

[29]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.

[30]Motivated by the results in Table 6, we include only one lag of abnormal downloads, corresponding to portfolio changes in month $t-1$. We use buy-side changes because long-only constraints make negative shocks harder to translate into active sales.

[31]Attempting to purge this residual with a long, ad-hoc list of controls, on which no consensus exists, would impose a disproportionate increase in model complexity relative to the marginal gains in purification.

the residual degrees of freedom remain ample.[32] By contrast, replacing fixed effects and no-trade-adjusted weight changes with an extensive, ad-hoc set of macro, fundamental, and flow controls would consume a similar (or larger) number of parameters while re-introducing specification risk. Our parsimonious approach therefore preserves precision, honors the one-regressor spirit of the original RPI, and offers a transparent template for other asset-level alternative-data settings, in which conditioning on a single signal and a pre-shock window makes power considerations paramount.

To translate signal sensitivity into economic reliance, we multiply sensitivity by the fund's average pre-ATT portfolio weight in exposed stocks: $\text{SpecScore}_f = R^2_f \times \overline{W}_f^{\text{pre-ATT}}$ where W_{ft} is the total portfolio weight in stocks for which the signal is available (the same stocks as in Equation 5). Since SpecScore_f is highly left-skewed, we classify funds above the 75$^{\text{th}}$ percentile as *specialized* and all others as *non-specialized*. This binary split forms the main stratification in our subsequent analysis of ATT's impact on stock-picking ability.

To illustrate our specialization classification, Figure D.1 displays the joint distributions of the composite measure, within–fixed effect adjusted R^2, average pre-ATT portfolio weights in exposed stocks, and the p-value of the fund-level abnormal downloads coefficient β_f, separately for funds above and below the 75$^{\text{th}}$ percentile cutoff. Specialization is highly concentrated: most funds exhibit little or no trading sensitivity to the app signal and/or low exposure to app-linked stocks, while specialists combine high regression fit and meaningful portfolio weights. Notably, our classification is driven mainly by differences in R^2, not passive exposure: while specialized funds cluster in the right tails of both the composite measure and R^2 distributions (with systematically lower p-values), exposed-stock weights are broadly distributed in both groups. This demonstrates that the measure captures active trading associated with the signal rather than passive holdings. The sharp left-skew and mass at or

[32] 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 in the cross-fund analysis because the within-FE-adjusted R^2 penalizes low-T, and because our specialization score is further scaled by each fund's pre-ATT average portfolio weight in signal-covered stocks.

below zero further justify our percentile-based split.

Table 2 provides summary statistics for the funds included in our sample, split by the specialization measure described above. The two groups are similar, with *specialized* funds displaying marginally smaller size and turnover. The main notable difference is that *specialized* funds hold stocks that are relatively more "growth" (with smaller HML exposure). This is in line with firm-level summary statistics in Table 1 which show that firms in the app sample have greater market-to-book ratios.

5.1.2 Funds' Stock-Picking Ability

Having classified specialized funds, we next assess their stock-picking ability, defined as the product of benchmark-adjusted weights and future abnormal returns. Intuitively, if a fund overweights stocks that subsequently outperform, this signals informed selection skill. Formally, for each fund f, stock i, and month t, we define:

$$\text{Picking}^{s,h}_{ift} = 100 \times \Delta w_{ift} \times \text{Return}^{s}_{i[t+1,t+h]}, \tag{6}$$

where $\text{Return}^{s}_{i[t+1,t+h]}$ is the stock-level cumulative idiosyncratic return between $(t+1)$ and $(t+h)$, computed as the residual return from a factor regression using models $s = [CAPM, FF6]$; and Δw_{ift} is the portfolio weight deviation for stock i relative to a benchmark portfolio (*active weight*). Since mutual funds vary substantially in their investment styles, we determine the relevant benchmark for each fund from a set of passive benchmarks by selecting the one that minimizes the sum of absolute difference in portfolio weights across all stocks, i.e., their minimum active share benchmark (Cremers and Petajisto, 2009).[33]

To examine changes in picking ability for stocks whose performance was predictable using mobile data, we define *exposed* stocks as those in the *"app sample"* ($e = 1$), and all others as the comparison group ($e = 0$). For each fund-month, we compute the mean stock-level picking ability within the exposed and unexposed groups, corresponding to the within-fund,

[33] We consider 7 passive funds from the Vanguard family as possible benchmarks: broad, large-cap, mid-cap, small-cap, growth, value, and momentum. Measure construction is described in Appendix Section D.1.

within-group covariance between active weights and future idiosyncratic returns. Let $\mathcal{I}_{fte}$ denote the set of stocks held by fund f in month t with exposure e. The average picking measure for horizon-h return s is then:

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

Our measure builds on Kacperczyk et al. (2016), who average stock-level picking ability across the entire portfolio, and on Bonelli and Foucault (2023), who consider stock-level ability without any aggregation. Our approach lies in between these two.[34]

Table 3 reports summary statistics on average picking ability, constructed under the Fama-French 6-factor model, split by fund specialization and stock exposure.

5.1.3 Stock-Picking Ability Around ATT by Fund Specialization

Our object of interest is the within-fund-month gap in picking ability between exposed and unexposed stocks (ΔP_{ft}). This gap reflects how managers allocate attention, as rational inattention theory predicts they will adjust focus when signal precision changes (Kacperczyk et al., 2016)—for example, after ATT reduces the value of mobile-app data. A decline in ΔP_{ft} after ATT indicates a relative loss of forecasting advantage on the exposed segment, whether through worse selection on exposed stocks, better selection on unexposed, or both. In contrast, picking ability measured on exposed stocks alone may be confounded by managers liquidating the most information-sensitive positions.[35]

To test whether this effect is concentrated among funds most reliant on mobile-based signals, we combine our specialization score with the picking metric in a triple-difference design. This approach simultaneously contrasts (i) exposed vs. unexposed stocks, (ii) pre- vs. post-ATT periods, and (iii) specialist vs. non-specialist funds. The triple-difference

[34] Both Kacperczyk et al. (2016) and Bonelli and Foucault (2023) use the market portfolio as the benchmark for all funds. Our results are robust to this alternative benchmark, as reported in Appendix Figure D.2.

[35] For example, if only a subset of exposed stocks becomes materially harder to forecast, fund managers may simply drop those positions. In that case, observed picking ability among the surviving holdings would remain unchanged, thereby obscuring the true deterioration in overall picking ability.

framework isolates the effect of mobile signal deterioration from contemporaneous shocks affecting all funds or all stocks, providing a clean test of our prediction that the within-fund gap ΔP_{ft} should decline more for specialists following ATT. Concretely, we estimate:

$$\begin{aligned}\text{Picking}^{s,h}_{fte} &= \alpha + \beta_1 \times \text{Exposed}_{fe} + \beta_2 \times \text{ATT}_t \times \text{Exposed}_{fe} + \beta_3 \times \text{Exposed}_{fe} \times \text{Specialized}_f \\ &\quad + \beta_4 \times \text{ATT}_t \times \text{Exposed}_{fe} \times \text{Specialized}_f + \delta_{ft} + (\lambda_{te}) + \varepsilon_{fte}, \end{aligned} \tag{7}$$

where $\text{ATT}_t = 1$ for months after April 2021. Fund-month fixed effects (δ_{ft}) absorb all time-varying shocks at the fund level. Exposed-month fixed effects (λ_{te}) absorb sector-wide shocks to app-owning firms around ATT, effectively controlling for the real effects of ATT on this sector. We run the regression for horizons $h \in \{1, 3, 6, 9, 12, 18\}$ and adopt $h = 9$ as the baseline, compounding daily abnormal returns over $[t{+}1, t{+}9]$.

The coefficient of interest is β_4 on *Post* $\times$ *Exposed* $\times$ *Specialized*. β_4 measures how the exposed–unexposed picking gap changes for specialist funds after ATT, relative to non-specialists. A negative $\hat{\beta}4$ indicates that specialists experienced a larger decline in their comparative advantage on exposed stocks, consistent with the loss of precision in a key signal. Because our fixed effects absorb broad fund- and sector-level shocks, a significant negative $\hat{\beta}4$ can be interpreted as evidence that ATT-induced signal deterioration disproportionately weakened specialists' stock-picking ability.

Table 7 presents the results, with the dependent variable defined as picking ability measured by cumulative abnormal returns (CAR), based on either the CAPM (Columns 1–2) or Fama–French 6-factor model (Columns 3–4). Even-numbered columns include Exposed-Month fixed effects. In all specifications, the estimated coefficient on the triple interaction, $\hat{\beta}_4$, is negative and statistically significant. This indicates that, following ATT, specialists' stock-picking ability on mobile-app stocks declined relative to their own performance on unexposed stocks, and by more than the corresponding change for non-specialists. In other words, specialists experienced a greater loss of comparative advantage on exposed stocks.

To assess magnitudes, we focus on Column 3 with Fund-Month fixed effects. The coefficient $\hat{\beta}_2 = -0.055$ reflects the impact of ATT on *unspecialized* funds' ability among exposed stocks, likely through ATT's direct effect on firm performance. This would move an unspecialized fund from the 75^{th} to about the 25^{th} percentile of the distribution (Table 3). While not central to our mechanism, this sizable negative effect on unspecialized funds' picking ability aligns with the pronounced stock market reaction among app-owning firms most dependent on tracking and data-sharing (Bian et al., 2021). When uncertainty rises unexpectedly, the picking ability of all funds tends to deteriorate. In comparison, $\hat{\beta}_4 = -0.025$, capturing the differential effect on specialized funds, is about half as large and represents roughly a half-quartile shift in picking ability, which shows broader cross-sectional dispersion. This back-of-the-envelope comparison suggests that the incremental informational channel we document is economically meaningful, amounting to about half of ATT's direct effect on funds through fundamentals.

We then examine $\hat{\beta}_4$ across horizons from 3 to 18 months using the most stringent specification in Column 4 of Table 7. Figure 3 shows that the loss of advantage for specialists extends beyond short-term predictions, consistent with funds shifting attention away from exposed firms rather than merely across signal maturities. Moreover, the estimates of $\hat{\beta}_4$ decline monotonically with horizon. Because longer-horizon returns (e.g., $[t+1, t+18]$) embed shorter-horizon ones (e.g., $[t+1, t+9]$), this pattern suggests that managers do not systematically shift attention toward long-horizon signals; otherwise, if active weights covaried more with long-horizon performance, the slope would flatten at medium-to-long horizons.

Next, we examine the timing of the decline in stock-picking ability by estimating a dynamic version of Equation (7), using our most stringent specification with Exposed-Month fixed effects. Instead of a single post-ATT indicator, we use quarterly event-time dummies to trace dynamics around ATT.[36] Figure 4 visualizes the coefficients, using 2021Q1 as the

[36]Quarterly dummies in monthly data estimate the average monthly effect within each quarter.

reference quarter.[37] A clear decline in specialists' relative picking ability begins in 2020Q4, when the return window first overlaps post-ATT months. This "partially treated" period reflects the forward-looking nature of our measure. The decline sharpens from 2021Q2, when the return window fully overlaps with the post-ATT regime, marking a persistent, statistically significant drop in specialists' advantage on exposed stocks.

For robustness, we re-estimate our baseline using several alternative definitions of the picking and specialization measures. To start with, the picking metric is recalculated using active weights benchmarked against either the CRSP universe or each fund's closest passive benchmark. We then redefine specialists as funds in the top quintile or decile of the score distribution, using both raw and benchmark-relative exposed-stock weights. The estimated results from these robustness checks are summarized in Appendix Figure D.2. In every case, the triple-interaction coefficient $\hat{\beta}_4$ stays negative and is usually significant at the 5% level, with its magnitude rising as the specialization threshold tightens (though power falls as the specialized sample shrinks). Additionally, to rule out pandemic-related distortions, we omit July 2019 to March 2021: March 2020 to March 2021 covers COVID, and July 2019 to February 2020 avoids overlap with our nine-month forward-return window. Appendix Table D.1 shows results are virtually unchanged.

As a falsification test, we replace actual stock-level ATT exposure with a placebo assignment preserving the share of exposed stocks. Appendix Table D.2 shows that placebo coefficients are small, opposite in sign, and generally insignificant, confirming our main results are not driven by spurious correlation.

Taken together, the evidence indicates that ATT redirected investor attention away from exposed firms, over and above any direct impact on those firms' fundamentals. With fewer informed investors monitoring these stocks, we expect price informativeness to decline and mispricing to rise, a hypothesis we examine more directly in Section 6.

[37] We set the reference quarter earlier than in other plots because the nine-month forward-looking window means effective treatment begins before 2021Q2.

5.2 Sell-Side Analysts

5.2.1 Earnings Forecast Errors

We now turn to the forecast errors of financial analysts. Compared to the analysis of fund managers, examining analyst forecasts offers two advantages. First, unlike fund trading, analysts' actions, such as earnings forecasts and stock recommendations, are directly observable. Second, analyst forecasts are typically accompanied by textual reports that often disclose the data sources and methodologies used. By searching for references to alternative data providers or specific metrics in these reports, we can identify which analysts utilize alternative data and exploit variation across analysts within the same firm-quarter. This feature allows us to more effectively control for firm-level real effects of ATT and isolate variation attributable to differential usage of alternative data.

We identify 2,981 unique analysts from 274 brokerage firms covering the app sample. Panel A of Table 4 reports forecast-level summary statistics. On average, analysts covering the *"App sample"* use app-related keywords in 6.4% of reports and maintain coverage of the same firm for 13.8 quarters. By comparison, analysts covering a placebo set of firms without popular apps use app keywords in only 2.8% of reports and maintain coverage for 11.6 quarters.[38] The two samples are otherwise similar in forecast age (days between forecast and announcement), firm coverage breadth, and brokerage size.

Guided by the multitasking framework of Dessaint et al. (2024), we use an analyst's keyword intensity as a proxy for the attention allocated to mobile-data signals within a fixed coverage set. Since analysts cannot reallocate attention across stocks, the relevant margin is how signal deterioration affects forecast accuracy for exposed stocks among analysts who previously relied most on mobile data. Before ATT, higher keyword intensity should translate into lower forecast errors on exposed stocks, reflecting effective use of an informative signal. After ATT, as signal precision falls, the marginal value of that effort declines, so

[38] The placebo sample consists of forecasts for firms in Fama-French 48 industries without major apps, defined as those with quarterly downloads consistently above the sample median. Firms in the placebo sample may also be followed by specialized analysts.

high-intensity analysts should see their advantage erode and forecast errors rise relative to peers—especially at short horizons and when substitutes are not readily available. To test that prediction, for each quarter and each firm-analyst pair, 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 analyst performance relative to their peers, facilitating cross-firm and cross-time comparisons while controlling for variation in forecast difficulty.[39] We then run the following regression:

$$\text{PMAFE}_{ijt} = \alpha + \beta_1 \times \text{Keyword usage}_j + \beta_2 \times \text{ATT}_t \times \text{Keyword usage}_j + \Gamma X_{jt} + \theta_{it} + \iota_{ij} + \kappa_d + \epsilon_{ijt} \quad (8)$$

where j indicates analysts, i firms, t year-quarters, d estimation month (i.e., the month in which analyst j provided the forecast for firm i's quarter t earnings), and b brokerage firms. *Keyword usage* is the number of app keywords divided by the total number of reports written by the analysts during the pre-ATT period, as defined in Section 3.3. X_{jt} is a list of analyst and brokerage firm characteristics, namely: the age of the forecast, analyst experience (since first issuing forecasts, and in relation to each firm i), number of firms covered, and size of brokerage firm. We include firm–year–quarter fixed effects (θ_{it}) to absorb time-varying changes at the firm level. This helps address concerns that firms' business operations and earnings may become more volatile around ATT due to reduced access to consumer data, which in turn could affect forecast accuracy. Additionally, we include analyst-firm fixed effects to account for the average forecast accuracy of a given analyst for a given firm, thereby controlling for any persistent analyst-firm match effects and isolating deviations in accuracy specifically around ATT. Finally, estimation-month fixed effects (κ_d) absorb time trends or aggregate shocks, such as changes in market conditions or widespread data availability, which

[39] Our results are robust to using AFE scaled by absolute actual earnings as the outcome variable, as reported in Appendix Table E.1.

could affect forecasting behavior or accuracy across all analysts in a given month. Standard errors are clustered at the firm and estimation month levels.

The coefficient of interest is $\hat{\beta}_2$. We expect $\hat{\beta}_2$ to be positive and statistically significant for the *App Sample*, indicating that analysts who use alternative data experience an increase in forecast errors following ATT. Table 8 reports regression results consistent with this prediction. Columns 1–2 include firm-year-quarter and estimation-date fixed effects, while Columns 3–4 additionally include analyst-firm fixed effects. The coefficient $\hat{\beta}_2$ is positive and statistically significant across all specifications.

We interpret the magnitude of $\hat{\beta}_2$ using the specification in Column 2, where it is estimated at 0.233 and is significant at the 1% level. Economically, a one standard deviation increase in the share of analyst reports referencing any of app keywords is associated with a 2.5% increase in PMAFE (0.233×0.11), which would move an analyst at the median of the forecast error distribution to approximately the 54$^{\text{th}}$ percentile post-ATT.[40] The coefficient $\hat{\beta}_1$, corresponding to *Keyword usage intensity*, is estimated at -0.095, consistent with higher forecast accuracy being associated with greater reliance on alternative data in the pre-ATT period. The larger reduction of the post-ATT effect implies that specialized analysts not only lose their informational edge but also perform worse overall. Other control variables behave as expected: longer analyst experience, older forecast vintage, and affiliation with larger brokerage firms are all associated with lower forecast errors. Including analyst-firm fixed effects reduces $\hat{\beta}_2$ mildly to 0.176 (Column 4), suggesting that our findings are not driven by selection effects related to which analysts cover app-owning firms. Our results are robust to using AFE as the outcome variable, as reported in Appendix Table E.1.

To further validate our interpretation of β_2, we repeat the above analysis using a placebo set of firms. The placebo is identified as the analyst forecasts for firms in any of the fine-grained Fama-French 48 industries that does not contain firms with major apps. This choice

[40] For context, this effect is about one-half to two-thirds the size of the forecast accuracy gains documented for conference-hosting brokers: Green et al. (2014) report forecast improvements of about 4–5% when analysts host conferences. As another benchmark, Harford et al. (2019) find that analysts are about 1.9% more accurate for firms they consider relatively more important, and our effect is larger.

aims at isolating firms for which data collected through mobile apps would not be useful, either directly (e.g., app usage for the app of the firm being analyzed) or indirectly (e.g., foot traffic to the firm's stores, collected through any mobile app). Results of those regressions are reported in Appendix Table E.2. $\hat{\beta}_2$ is not statistically different from zero across specifications. This indicates that the forecast errors of analysts specialized in alternative data only decrease for their forecasts of firms for which mobile-collected data should be relevant in forecasting future payoffs. Moreover, excluding an entire year of forecasts for earnings periods between 2020Q2 and 2021Q1 yields similar results, as reported in Appendix Table E.3, thereby ruling out the possibility that the results are driven by COVID-19.

Why do specialized analysts perform worse overall relative to non-specialized analysts post-ATT? We shed light on this question by distinguishing between analysts who reduce their keyword usage following ATT ("switchers"), defined as those with an above-median decline in keyword usage intensity from the pre-ATT to post-ATT period, and the "non-switchers". Regression results reported in Appendix Table E.4 show that the overall deterioration in performance is entirely driven by non-switchers, those who continue to rely on now-uninformative signals to form earnings forecasts. In contrast, switchers merely lose their prior informational advantage but do not exhibit a significantly worse performance. This suggests that the poor post-ATT performance among specialized analysts is driven by continued reliance on degraded alternative data or a failure to re-optimize effort allocation in response to the reduced informativeness of such signals.

Last, to examine the timing of the decline in analyst forecast accuracy, we estimate a dynamic version of Equation (8), mirroring the most stringent regression specification and replacing the *ATT* indicator with a set of quarter-specific dummies. Figure 5 visualizes the estimated coefficients with 2021Q2 as the reference quarter. Forecast errors of specialized analysts begin to rise in the second quarter following ATT and remain elevated through the third and fourth quarters. Importantly, no evidence of a pre-trend is observed, supporting the parallel trends assumption.

5.2.2 Market Responses to Analyst Reports

The information content of analyst reports is typically incorporated into prices through market reactions upon report issuance. When individual reports become less informative, the corresponding market response should weaken. Studying the price impact of analyst reports complements the forecast error analysis in several ways. First, these reports often attract media coverage—particularly those containing explicit buy or sell recommendations—leading to faster incorporation of information into prices. Second, and more importantly, we can measure each report's reliance on mobile-generated signals, enabling more granular, report-level variation than the analyst-level variation in the pre-ATT period. In this way, we provide direct evidence that the market processes and responds to the specific content of each report.

We examine the impact of ATT on report-level market reactions by estimating the following regression:

$$CAR_{ikt}^{adj,\ 3d} = \alpha + \beta_1 \times \text{Keyword per page}_{kt} + \boldsymbol{\beta_2} \times \text{ATT}_t \times \text{Keyword per page}_{kt} + \Gamma X_{kt} + \theta_{it} + \iota_{br,i} + \kappa_d + \epsilon_{ikt}, \quad (9)$$

where i indexes firms, k indexes reports, and t indexes the year-quarter corresponding to the earnings period. The outcome variable $CAR_{ikt}^{adj,\ 3d}$ is the direction-adjusted cumulative abnormal return over a 3-day window, where returns are multiplied by -1 if the analyst issues a negative report. The variable Keyword per page$_{kt}$ is defined as the number of app keywords in report k divided by the number of pages of that report. The control vector X_{kt} includes report-level characteristics such as the retail price and total number of pages (both in natural logarithms). The sample includes all reports on app-owning firms from 2017 to 2023 that appear on LSEG in less than five days from issuance, ensuring that measured market responses reflect reports that were broadly accessible at the time of release. 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 keyword usage across reports. Standard errors are clustered at the firm level.

Our regression sample includes 42,992 relevant reports. As reported in Panel C of Table 4, the average report contains 0.11 app keyword per page, has 19.9 pages, and sells at $196.

The estimation results are presented in Table 9, with Columns 1–2 (respectively, 3–4) reporting CAR computed using the CAPM (Fama-French 6-factor) model. Across all specifications, the results consistently indicate that the market recognized the deterioration in the quality of mobile-generated signals referenced in those reports. Prior to ATT, analyst reports citing more app keywords triggered stronger market reactions. In contrast, following ATT, there is no longer a differential market response based on keyword intensity. Quantitatively, before ATT, a one standard deviation increase in keyword intensity (about 0.3 additional keyword per page) corresponds to a three-day abnormal return of about 20 basis points at the report level, an effect that disappears entirely after ATT. Control variables behave as expected: more expensive reports and shorter reports (conditional on price) are associated with larger market reactions.

6 ATT Worsens Firms' Information Environment

Since ATT diminished the informational value of mobile data, funds reallocated attention and capital away from the most exposed firms, while analysts, unable to switch coverage, experienced a decline in forecast accuracy. We therefore expect price efficiency to deteriorate most for firms that were heavily covered by mobile-data specialists prior to ATT. To quantify this, we construct two firm-level measures. The first is the average quarterly bid-ask spread. For each trading day, the spread is defined as the difference between the closing ask and bid prices, divided by the mid-price, and then averaged across the quarter. This serves as a proxy for information asymmetry and trading frictions. The second measure is post-earnings announcement volatility. Following Loughran and McDonald (2014) and Kim et al. (2023), we compute the root-mean-squared error of abnormal returns over trading days 6 to 28 after the earnings announcement. Abnormal returns are estimated using the Fama-

French 6-factor model. This captures residual uncertainty and disagreement after earnings are initially incorporated and is interpreted as a proxy for noise in the information environment. These measures allow us to test whether ATT-induced signal deterioration and the ensuing reallocation of specialist fund attention translated into wider spreads and greater post-earnings volatility for the most exposed firms, compared to otherwise similar peers. To do so, we estimate the following specification:

$$\text{Info Friction}_{it} = \alpha + \beta_1 \times \text{Exposure}_i + \beta_2 \times \text{ATT}_t \times \text{Exposure}_i + \Gamma_1 X_{it-4} + \theta_{sic2,t} + \epsilon_{it}, \quad (10)$$

where i indexes firms and t indexes year-quarters. *Exposure* is an indicator equal to one if the firm is more exposed to specialized attention, defined in three ways: (i) above the 75^{th} percentile of market capitalization held by specialized funds, (ii) referenced in keyword-based analyst reports in more than half of quarters, or (iii) covered by specialized analysts, based on industry focus, in more than half of quarters. As described in Section 3.3, the keyword-based exposure measure captures only the subset of analysts who issue written reports alongside their forecasts. As an alternative, we construct an analyst specialization measure based on industry focus, which considers the full set of analysts covering the firm. Not surprisingly, relative to firms without an app, firms in the app sample are more likely to be held by specialized funds (25% vs. 18%) and followed by specialized analysts (63% 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 industry-by-quarter fixed effects. Standard errors are clustered by industry and quarter.

The regression results show that firms with exposure to specialized funds and analysts exhibit lower information asymmetry prior to ATT ($\hat{\beta}_1 < 0$). However, this advantage is reversed following ATT ($\hat{\beta}_1 + \hat{\beta}_2 \geq 0$, significant in Columns 3–6), consistent with increased uncertainty, investor disagreement, and greater opacity in the firm's information environment. A caveat is that part of the observed effect may reflect a direct impact of ATT on firms' earnings volatility as they lose access to cross-app data. Nevertheless, the overall pattern of results supports the interpretation that ATT, by reducing the quality of alternative data, disproportionately weakens the information environment of firms followed by market participants who rely on such data.

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 simply 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

A key feature of modern financial markets is their growing reliance on the booming alternative data industry. Until recently, the trajectory of this market seemed one of ever-increasing data availability, with new sources continually expanding the informational frontier. Yet rising concerns over privacy, compliance, and platform control have introduced a very different trend: disruptions to the DGP itself. Our study examines one such disruption through the lends of Apple's ATT and mobile-generated signals. We document that privacy-driven restrictions can materially weaken the financial information environment by increasing forecast errors, reducing price informativeness, and impairing market efficiency.

This unintended consequence highlights the need for policymakers to carefully balance data privacy with market transparency. More broadly, firms and market participants that once thrived in a data-rich environment may be more susceptible to this adverse trend, and one promising avenue for future research is to study the distributional consequences for investors and firms. A further direction is to examine the long-term dynamics of data production if free digital products originally sustained by data monetization alter their business models or exit the market, potentially shrinking the ecosystem that creates rich user-generated, surveillance-like data.

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, 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.

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*.

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.

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 (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.

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.

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

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.

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

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.

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

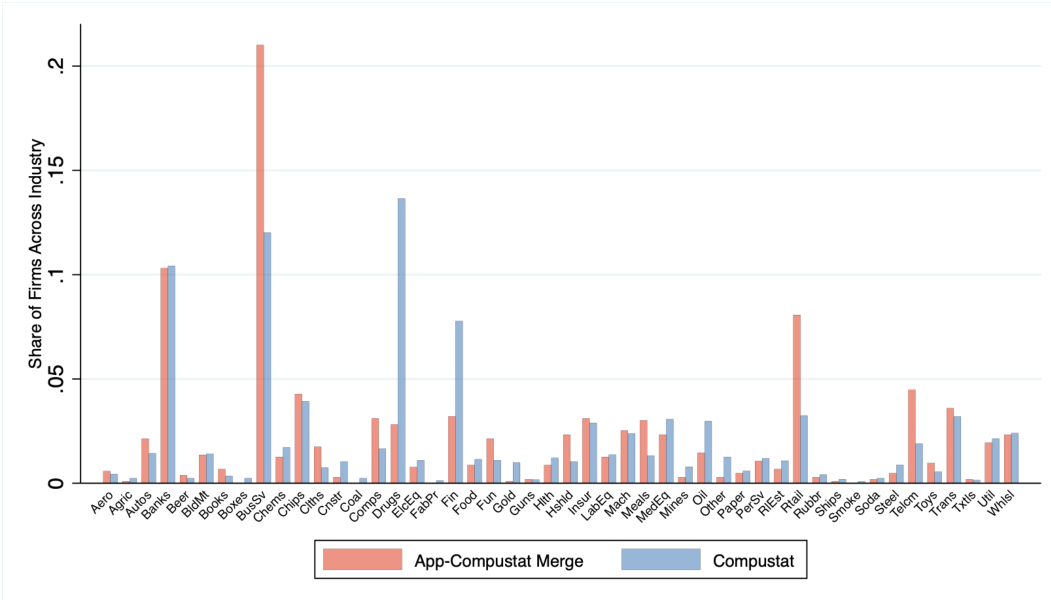
Statista (2024b). Number of available apps in the Apple App Store from 2008 to July 2023. https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/. Accessed August 02, 2024.

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.

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

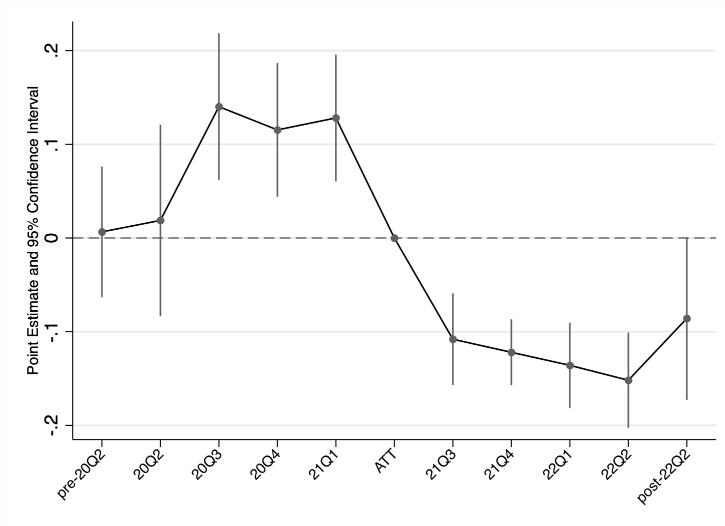


(a) Industry Distribution

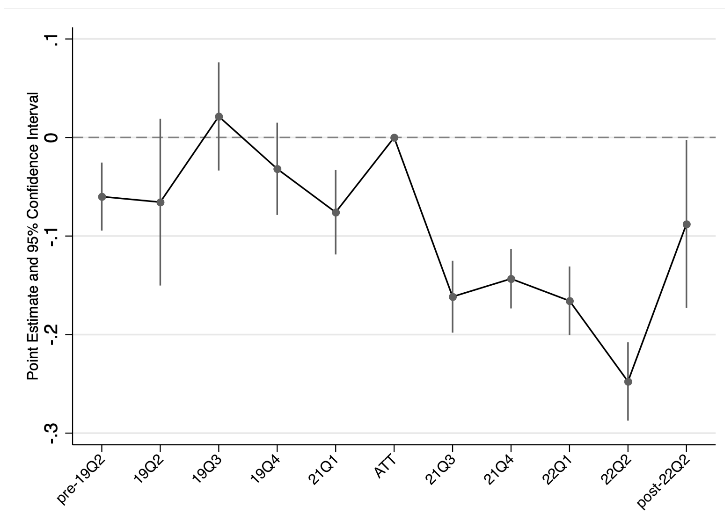
(b) Share of App-Owning Firms Across Industries

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: App Downloads and Firm Performance Around ATT



(a) Sales Growth



(b) Random Walk-Based SUE

Figure 2 plots the estimated coefficients for the quarterly time dummies from Equation (3). Dummy variables are included for each quarter from 2019Q2 through 2022Q2, while quarters before 2019Q2 and after 2022Q2 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 3: Fund Picking Ability at Various Horizons Around ATT

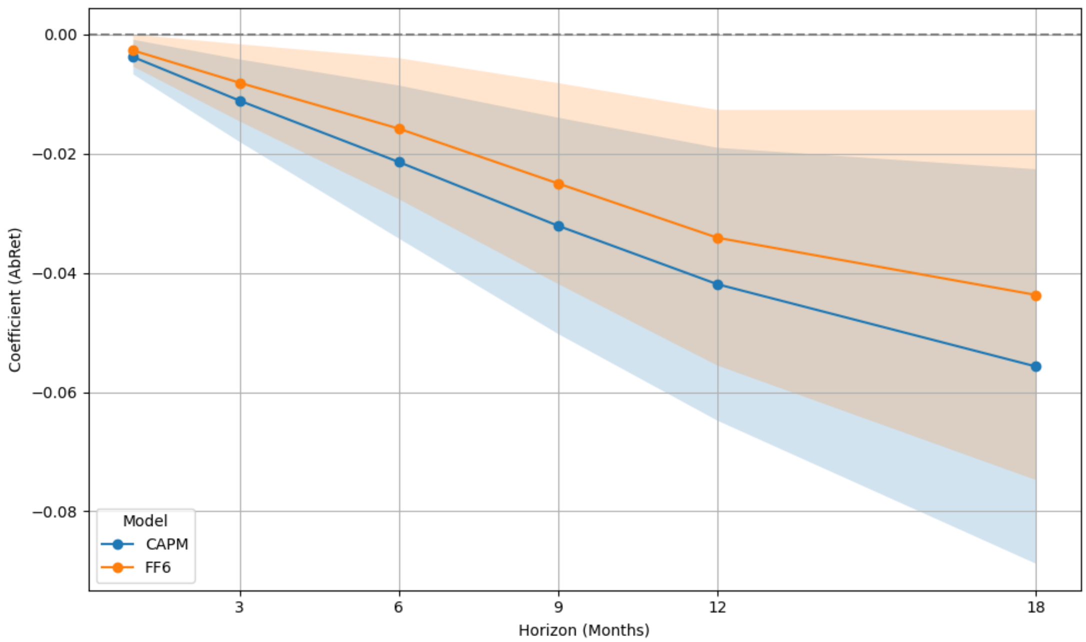


Figure 3 displays the estimated coefficients of $ATT \times Exposed \times Specialized$ as specified in Equation (7). Each dot represents an estimate based on a different *picking* measure, constructed at horizons ranging from 1 to 18 months ahead in increments of 3 to 6 months.

Figure 4: Fund Picking Ability at 9-Month Horizon Around ATT

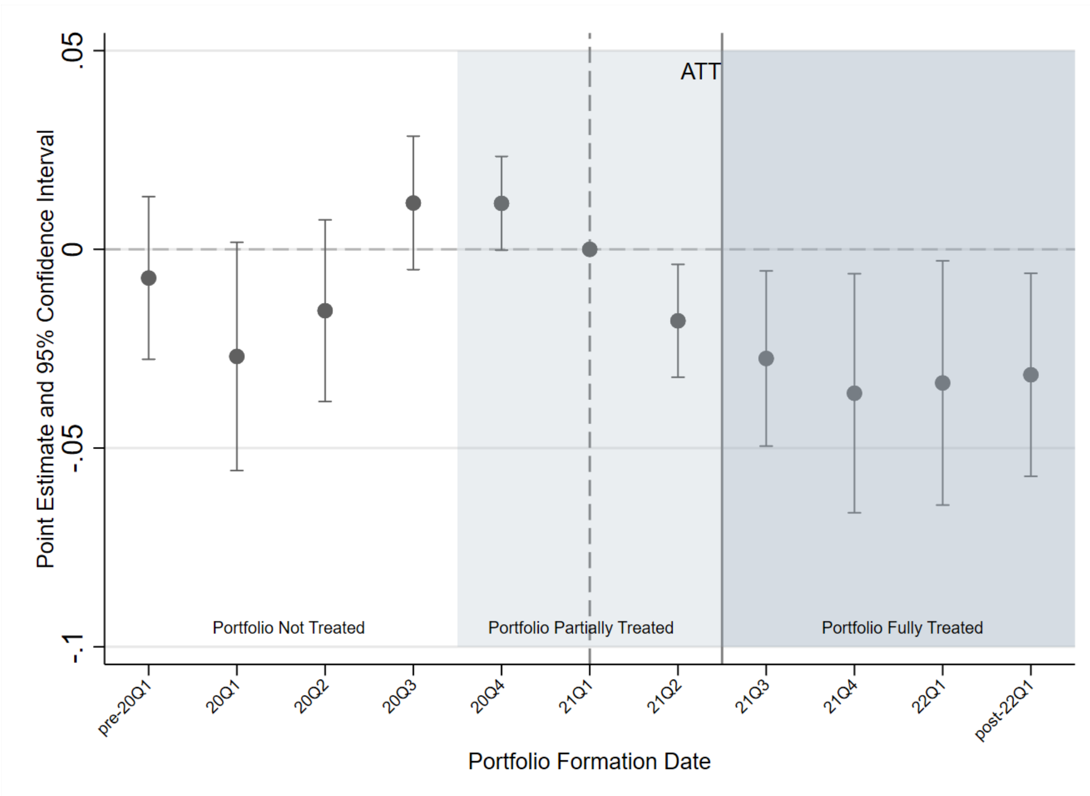


Figure 4 shows the dynamic triple-difference estimates of ATT's impact on fund picking ability. Specifically, the coefficients trace how the gap in picking ability evolves between specialized and non-specialized funds, for exposed versus non-exposed stocks, over time. The picking measure is a 9-month forward-looking CAR estimated using the Fama–French 6-factor model. The reference quarter is 2021Q1, defined as the quarter of the portfolio formation date. Shading highlights how ATT enters the 9-month CAR window: the light-shaded region marks periods where portfolios are partially exposed to ATT (i.e., the portfolio constructed at the beginning of 2020Q4 includes post-ATT months in May–June 2021); the dark-shaded region marks periods where portfolios are fully exposed (i.e., the entire 9-month window falls after ATT); and the unshaded region marks pre-ATT portfolios that contain no post-ATT months. We estimate coefficients on quarter dummies spanning 2020Q1 through 2022Q1, where all periods prior to 2020Q1 are grouped into one group and all periods after 2022Q1 into another.

Figure 5: Analyst Forecast Errors and Usage of Mobile Signals Around ATT

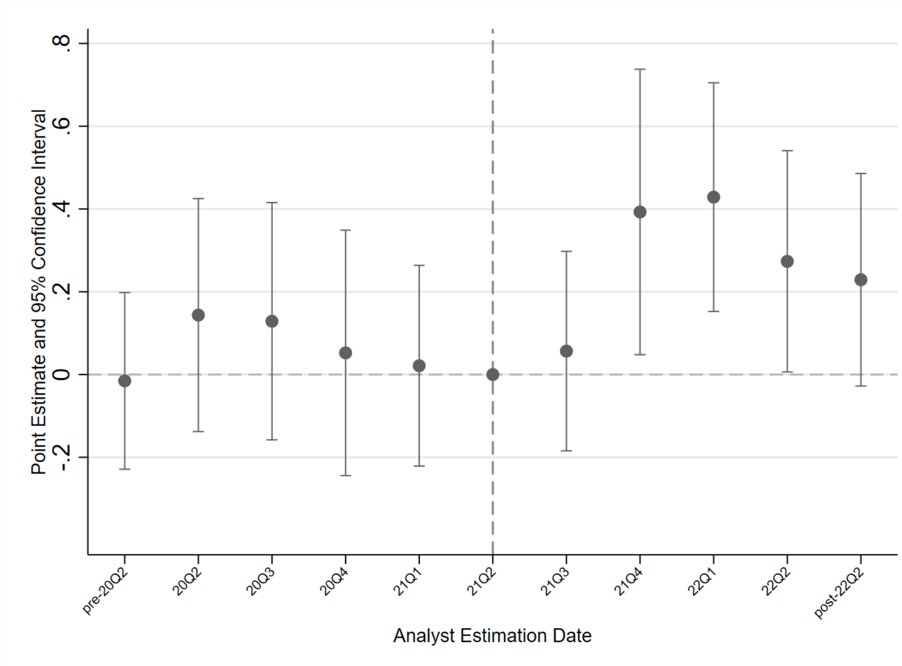


Figure 5 plots the dynamic effect of ATT on the gap in analyst forecast errors between analysts who differ in their reliance on mobile signals. The reference quarter is 2022Q2, which is the quarter of ATT implementation. We estimate coefficients on quarter dummies spanning 2020Q2 through 2022Q2, where all periods prior to 2020Q2 are grouped into one group and all periods after 2022Q2 into another.

Table 1: Summary Statistics — Firms

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE (random walk)	-0.061	5.65	-1.92	-0.36	0.09	0.51	1.89	15,964
CAR - CAPM	-0.059	10.64	-11.74	-5.27	-0.02	5.02	11.51	15,964
CAR - FF6	-0.066	10.44	-11.44	-5.10	-0.05	4.80	11.20	15,964
Sales growth	8.595	30.33	-14.23	-1.63	6.85	18.44	35.92	16,038
Size (in log)	22.956	2.07	20.41	21.46	22.91	24.35	25.74	16,001
Cash/Assets	17.813	17.86	1.92	4.59	11.61	24.92	44.53	15,994
Tangibles/Assets	20.712	21.65	1.28	4.44	11.95	30.66	57.24	15,699
Debt/Assets	32.653	24.25	3.67	13.39	30.52	45.94	63.31	15,861
Market-to-Book	1.648	6.89	0.03	0.14	0.35	0.77	1.47	15,981
Downloads (in log)	12.593	2.52	9.47	10.87	12.56	14.21	15.80	16,038
MAU (in log)	12.846	2.66	9.63	11.10	12.83	14.48	16.14	16,038
DAU (in log)	11.387	2.98	7.66	9.38	11.48	13.34	14.99	16,038
Engagement	0.269	0.12	0.12	0.18	0.26	0.34	0.41	16,038
Bid-ask spread	3.363	1.80	1.64	2.11	2.88	4.10	5.74	13,417
Post EA vol. (FF6)	1.789	1.30	0.75	1.00	1.43	2.13	3.12	13,354
Specialized analysts (keyword usage)	0.626	0.48	0.00	0.00	1.00	1.00	1.00	16,038
Specialized analysts (retail & service focus)	0.380	0.49	0.00	0.00	0.00	1.00	1.00	16,038
Specialized funds	0.253	0.43	0.00	0.00	0.00	1.00	1.00	16,038

Panel B. No-App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE (random walk)	0.918	11.97	-3.83	-0.79	0.11	1.19	5.25	84,697
CAR - CAPM	-0.295	11.86	-13.50	-5.59	0.00	4.95	12.33	89,739
CAR - FF6	-0.042	11.69	-12.87	-5.23	0.00	5.06	12.54	89,739
Sales growth	9.200	50.59	-27.16	-4.61	7.39	22.24	50.00	76,055
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
Specialized analysts (keyword usage)	0.228	0.42	0.00	0.00	0.00	0.00	1.00	91,125
Specialized analysts (retail & service focus)	0.120	0.33	0.00	0.00	0.00	0.00	1.00	91,125
Specialized funds	0.190	0.39	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: Summary Statistics — Fund Characteristics

Variable	Specialized	mean	std	10%	25%	50%	75%	90%	count
Fund age	N	5.280	0.79	4.14	4.88	5.45	5.78	6.06	181,290
Fund age	Y	5.210	0.77	4.09	4.74	5.41	5.72	6.02	56,860
Fund size	N	6.127	1.95	3.53	4.70	6.13	7.47	8.68	181,290
Fund size	Y	5.849	1.99	3.27	4.31	5.80	7.34	8.64	56,860
Expense ratio	N	0.009	0.00	0.01	0.01	0.01	0.01	0.01	181,290
Expense ratio	Y	0.010	0.00	0.01	0.01	0.01	0.01	0.01	56,860
Turnover	N	0.577	0.49	0.14	0.25	0.44	0.75	1.15	181,290
Turnover	Y	0.502	0.45	0.11	0.22	0.38	0.62	1.00	56,860
Flow growth	N	-0.005	0.04	-0.03	-0.01	-0.01	0.00	0.02	181,290
Flow growth	Y	-0.004	0.04	-0.03	-0.01	-0.01	0.00	0.02	56,860
Flow vol	N	0.034	0.06	0.01	0.01	0.02	0.03	0.07	181,290
Flow vol	Y	0.035	0.06	0.01	0.01	0.02	0.04	0.07	56,860
Style-mkt	N	0.982	0.12	0.84	0.92	0.98	1.05	1.12	181,290
Style-mkt	Y	0.984	0.12	0.83	0.91	0.99	1.06	1.13	56,860
Style-SMB	N	0.222	0.38	-0.18	-0.08	0.10	0.53	0.79	181,290
Style-SMB	Y	0.196	0.36	-0.19	-0.08	0.09	0.44	0.76	56,860
Style-HML	N	0.069	0.25	-0.25	-0.09	0.06	0.24	0.40	181,290
Style-HML	Y	0.043	0.26	-0.28	-0.14	0.04	0.22	0.38	56,860
Style-MOM	N	-0.037	0.15	-0.22	-0.12	-0.04	0.05	0.14	181,290
Style-MOM	Y	-0.042	0.15	-0.24	-0.14	-0.04	0.05	0.15	56,860
Style-RMW	N	-0.027	0.25	-0.34	-0.16	-0.00	0.12	0.25	181,290
Style-RMW	Y	-0.029	0.28	-0.38	-0.18	-0.01	0.14	0.29	56,860
Style-CMA	N	-0.135	0.30	-0.53	-0.32	-0.12	0.06	0.23	181,290
Style-CMA	Y	-0.161	0.32	-0.58	-0.36	-0.15	0.05	0.23	56,860

Table 2 reports summary statistics on time-varying fund characteristics by fund specialization as measured in Section 5.1.1. 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.

Table 3: Summary Statistics — Fund Stock Picking Ability

Variable	Specialized	Exposed	mean	std	10%	25%	50%	75%	90%	count
Picking (CAPM)	N	N	-0.024	0.171	-0.168	-0.066	-0.010	0.019	0.101	90,645
Picking (CAPM)	Y	N	-0.030	0.200	-0.216	-0.093	-0.019	0.037	0.135	28,430
Picking (CAPM)	N	Y	-0.015	0.172	-0.156	-0.055	-0.004	0.029	0.110	90,645
Picking (CAPM)	Y	Y	-0.016	0.206	-0.198	-0.081	-0.010	0.049	0.153	28,430
Picking (FF6)	N	N	0.001	0.141	-0.107	-0.032	0.000	0.032	0.105	90,645
Picking (FF6)	Y	N	0.001	0.170	-0.145	-0.053	0.000	0.053	0.148	28,430
Picking (FF6)	N	Y	0.003	0.133	-0.094	-0.029	0.000	0.032	0.102	90,645
Picking (FF6)	Y	Y	0.006	0.161	-0.120	-0.046	0.000	0.051	0.141	28,430

Table 3 reports summary statistics on fund stock-picking ability, aggregated by stock exposure to ATT using simple averages at the fund–month level. Funds are further partitioned by their specialization in mobile-generated alternative data, as defined in Section 5.1.1. The stock-picking measure is constructed from 9-month future idiosyncratic returns, estimated using both the CAPM and the Fama–French 6-factor models.

Table 4: 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
AFE	0.300	0.55	0.02	0.04	0.11	0.28	0.71	90,358
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
AFEA	0.384	0.63	0.02	0.06	0.15	0.39	1.00	163,400
PMEAFE	-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 4 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 that do not include firms with major apps (i.e., apps with quarterly downloads consistently below 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 four days of the report issuance date.

Table 5: App Downloads and Firm Performance Around ATT

Decile	Sales Growth		SUE (random walk)	
	(1)	(2)	(3)	(4)
Downloads	0.085**	0.104***	0.031***	0.050***
	(0.04)	(0.03)	(0.01)	(0.01)
ATT $\times$ Downloads	-0.101***	-0.123***	-0.064***	-0.090***
	(0.04)	(0.04)	(0.02)	(0.02)
Industry-Quarter FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Observations	16,038	15,534	16,314	15,508
R-sq	0.242	0.288	0.170	0.184

Table 5 presents estimation results from Equation (2) using the app sample. 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, calculated as in Equation (1)). 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 $t - 4$. 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.

Table 6: 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
Stock 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 6 presents the estimation results of Equation (4), 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 7: Fund Picking Ability at 9-Month Horizon by Stock Exposure Around ATT

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	-0.032***	-0.032***	-0.025***	-0.025***
	(0.009)	(0.009)	(0.009)	(0.009)
Exposed × Specialized	0.015***	0.015***	0.010**	0.010**
	(0.005)	(0.005)	(0.005)	(0.005)
ATT × Exposed	-0.061***		-0.055***	
	(0.008)		(0.008)	
Exposed	0.029***		0.021***	
	(0.004)		(0.005)	
Fund-Month FE	Y	Y	Y	Y
Exposed-Month FE	N	Y	N	Y
Observations	238,150	238,150	238,150	238,150
R-sq	0.700	0.710	0.570	0.580

Table 7 reports the estimation results of Equation (7). The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the nine-month horizon $[t+1, t+9]$. 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 5.1.1. 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 controls for Exposed×Month interactions. 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.

Table 8: 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)	(0.07)	(0.07)
Keyword usage intensity	-0.089**	-0.095**		
	(0.04)	(0.04)		
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(Broker 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 8 reports the estimation results of Equation (8) 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 9: 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 9 reports the estimation results of Equation (9) 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 10: Information Friction Around ATT
– by Firm Exposure to Specialized Funds and Analysts

Exposure based on	Fund Specialization		Analyst Keyword Usage		Analyst Retail & Service Focus	
	(1) Bid-ask spread	(2) Post EA vol. (FF6)	(3) Bid-ask spread	(4) Post EA vol. (FF6)	(5) Bid-ask spread	(6) Post EA vol. (FF6)
ATT × Exposure	0.283***	0.154**	0.412***	0.294***	0.724***	0.396***
	(0.10)	(0.06)	(0.10)	(0.06)	(0.20)	(0.08)
Exposure	-0.231***	-0.194***	-0.140	-0.119*	-0.229	-0.205*
	(0.08)	(0.06)	(0.10)	(0.07)	(0.17)	(0.12)
Firm FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Observations	14,286	14,217	14,286	14,217	14,286	14,217
R-sq	0.620	0.453	0.621	0.453	0.622	0.453

Table 10 reports the estimation results of Equation (10) 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) is above the 75^{th} percentile in terms of the share of market capitalization held by specialized funds (Columns 1–2), (ii) is referenced in keyword-based reports in more than 50% of quarters (Columns 3–4), or (iii) is covered by specialized analysts in more than 50% of quarters (Columns 5–6). Specialized funds are defined as those that incorporate mobile signals into their trading strategies (Section 5.1.1). Keyword-referencing reports are defined as those mentioning any mobile-related keywords listed in Table B.1. Specialized analysts are defined as those who primarily cover firms in the retail & services industries—specifically, analysts for whom the share of such firms among their coverage exceeds the 90^{th} percentile among all analysts covering the same firm in the same quarter. All regressions include industry-quarter fixed effects and 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, double-clustered at the industry and quarter levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

A Variable Definition

Table A.1: Variable Definition

Variable Name	Definition
Firm-Level Variables	
SUE (consensus)	The difference between actual earnings per share (EPS) and the consensus forecast, 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. (CAPM)	Standard deviation of abnormal returns over trading days 6 to 28 following the earnings announcement, where abnormal returns are estimated using the CAPM. Winsorized at the 1% level in both tails.
Post-EA Vol. (FF6)	Standard deviation 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 (keyword usage)	Indicator equal to one if the firm is covered by specialized analysts—defined as those who reference app keywords in their reports in more than 50% of pre-ATT quarters.
Exposure to specialized analysts (retail & service)	Indicator equal to one if the firm is covered by specialized analysts—defined as those primarily covering retail and service industries (SIC codes 5200–5999 or 7000–8999)—in more than 50% of pre-ATT quarters. Specialization is defined as being above the 90$^{\text{th}}$ percentile in terms of the share of forecasts issued for these industries.
Exposure to specialized funds	Indicator equal to one if, prior to ATT, the firm is above the 75$^{\text{th}}$ 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	Average stock-picking ability across a group of stocks (exposed or unexposed), where stock-level picking ability is defined as the product of 9-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$^{\text{th}}$ percentile of the distribution. The specialization measure is defined as the product of (i) the within fixed-effect adjusted R^2 from Equation (5) 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 portfolio holdings 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. 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

A Pop-Up Notification Mandated by ATT **A-3**

B Analyst Reports from LSEG Workspace **A-4**
- B.1 Scraping, Cleaning, and Merging . . . A-4
- B.2 Keyword List . . . A-9
- B.3 Examples of Mobile Data Usage . . . A-10

C Predictive Power of Mobile Metrics on Firm Performance: Robustness **A-12**

D Fund-Level Analysis **A-15**
- D.1 Determining the Benchmark Portfolio . . . A-15
- D.2 Construction of Fund Specialization . . . A-17
- D.3 Additional Robustness . . . A-18

E Analyst-Level Analysis: Robustness **A-21**

List of Appendix Figures

- A.1 **An Example Prompt Mandated by Apple's App Tracking Transparency Policy** A-3
- B.1 **Search Query for a Ticker Bucket in LSEG Workspace** . . . A-7
- B.2 **Keyword Match Panel and Snippets in LSEG Workspace** . . . A-7
- B.3 **Report Overview Panel with Attributes in LSEG Workspace** . . . A-8
- B.4 **Usage of Mobile Data (Sensor Tower) in Analyst Reports – An Example** . . . A-10
- B.5 **Usage of Mobile Data (Apptopia) in Analyst Reports: An Example** . . . A-11
- C.1 **Mobile Metrics and Firm Performance Around ATT – Robustness** . . . A-13
- D.1 **Distributions of Regression Statistics by Fund Specialization** . . . A-17
- D.2 **Fund Picking Ability Around ATT – Robustness: Benchmark portfolio, Horizons, Specialized Funds** . . . A-20

List of Appendix Tables

- B.1 List of Mobile Data Keywords . . . A-9
- C.1 Additional Mobile Metrics and Firm Performance Around ATT – MAU and Engagement A-12
- C.2 App Downloads and Firm Performance Around ATT – Excluding COVID-19 Period . . . A-14
- D.1 Fund Picking Ability at 9-Month Horizon Around ATT – Excluding COVID-19 Period . . . A-18

D.2 Fund Picking Ability at 9-Month Horizon Around ATT – Placebo Stock-Level ATT Exposure . . . A-19
E.1 Analyst Forecast Errors Around ATT – Absolute Forecast Error . . . A-21
E.2 Analyst Forecast Errors Around ATT – Placebo . . . A-22
E.3 Analyst Forecast Errors Around ATT – Excluding COVID-19 Period . . . A-23
E.4 Analyst Forecast Errors – Switcher vs. Non-Switcher . . . A-24

A Pop-Up Notification Mandated by ATT

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

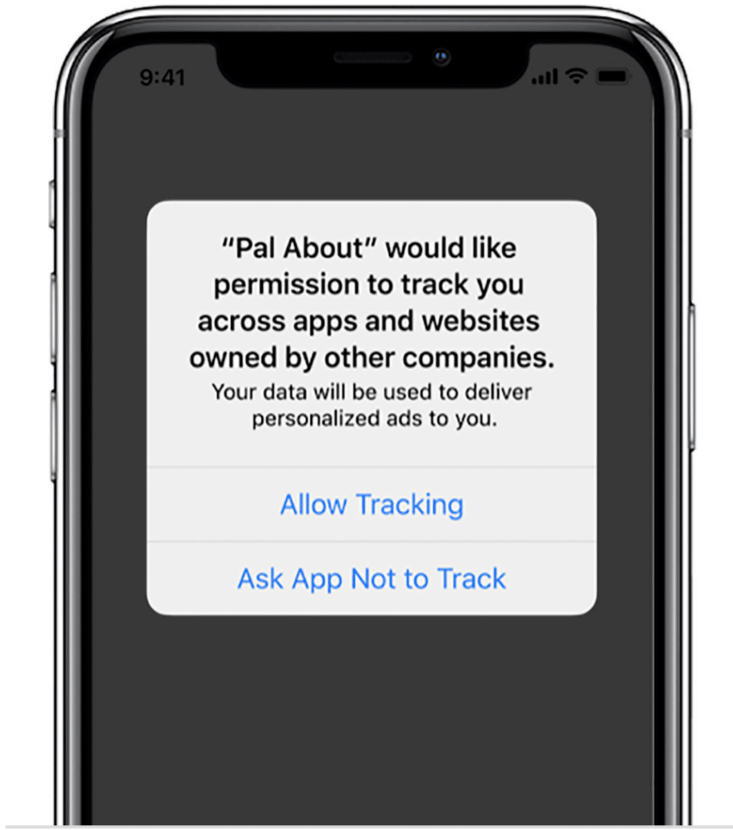


Figure A.1 presents an example of a prompt required by the ATT policy. By default, users are opted out of data sharing, and firms may share data only if users explicitly consent by clicking "*Allow Tracking*". Source: https://www.branch.io/resources/blog/what-happens-to-idfas-if-you-stop-showing-the-att-prompt/.

B Analyst Reports from LSEG Workspace

B.1 Scraping, Cleaning, and Merging

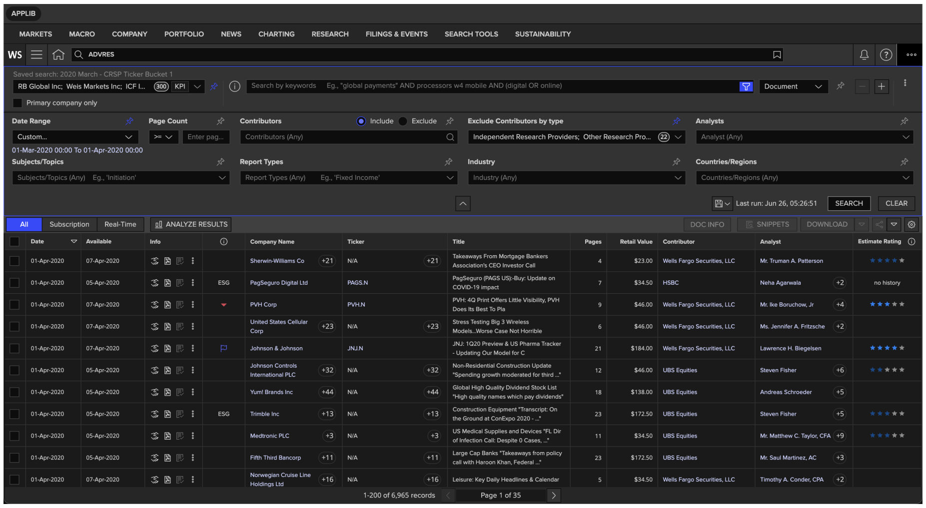
This section discusses how we obtain analyst reports from LSEG Workspace and link them to CRSP and I/B/E/S.

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.

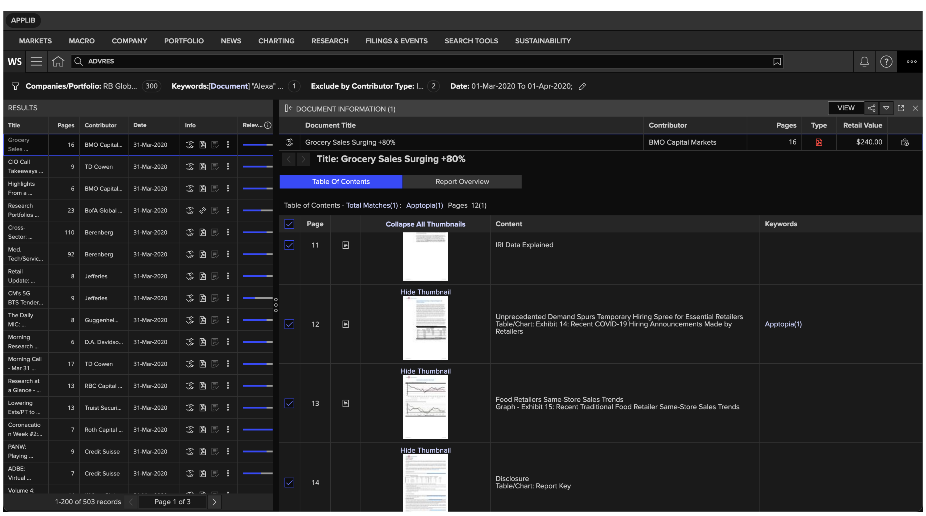
Mapping with I/B/E/S To integrate the LSEG Workspace scraped keywords data and reports metadata with standardized institutional identifiers, we developed a matching framework that linked broker and analyst names to I/B/E/S. In the first stage, we matched LSEG contributor names to I/B/E/S brokerage names. 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

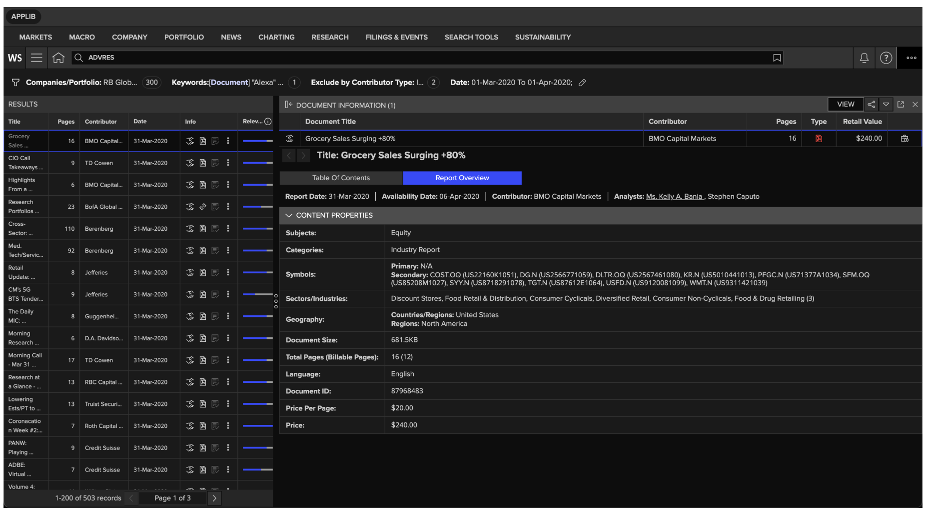


Figure B.3 presents detailed metadata for a selected broker report, including contributor, analyst(s), publication dates, page count, document size, research category, and geographic focus.

B.2 Keyword List

Table B.1: List of Mobile Data Keywords

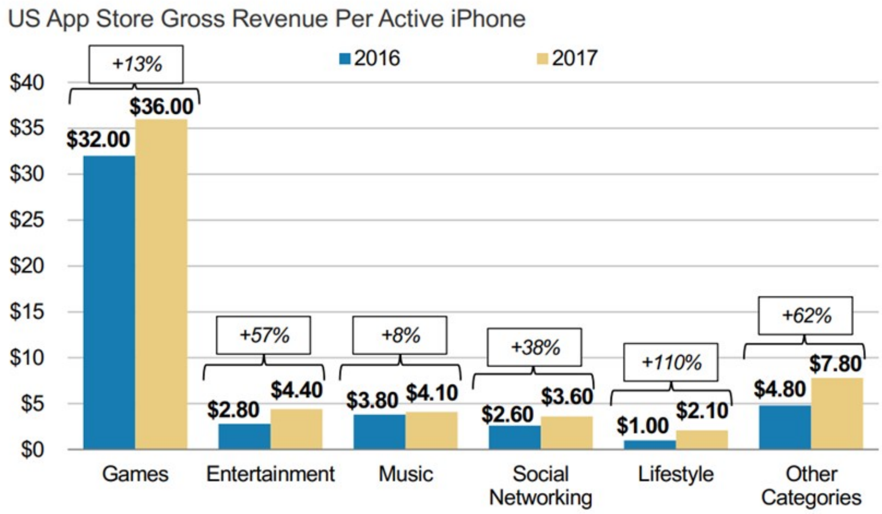
active user (8443)	alexa (10782)	app activity (340)	app annie (3578)
app data (2740)	app engagement (693)	app payment (391)	app performance (283)
app purchase (1091)	app tracking (970)	app usage (5138)	app use (184)
appannie (2009)	Apptopia (9818)	brand sentiment (495)	churn rate (7925)
comscore (19362)	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)

Table B.1 lists the keywords used to identify analyst reports that reference alternative data. Keywords highlighted in gray are both frequent and relevant to app performance, representing either mobile data providers or commonly used performance metrics. The number in brackets indicates how many times the corresponding word was mentioned in reports issued between 2017 and 2023.

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.

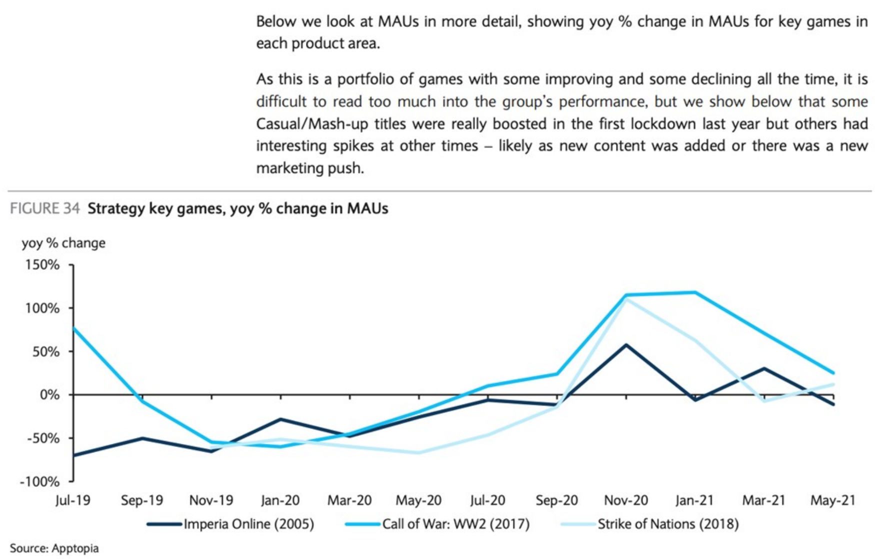


Source: Sensor Tower Store Intelligence, Tech Crunch, Morgan Stanley Research

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: Robustness

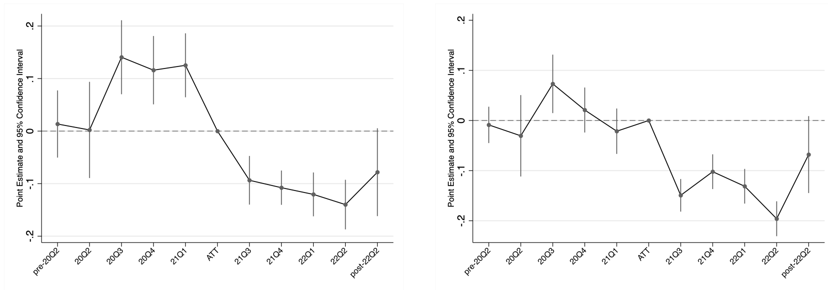
Table C.1: Additional Mobile Metrics and Firm Performance Around ATT – MAU and Engagement

Deciles	Sales Growth			SUE		
	(1)	(2)	(3)	(4)	(5)	(6)
MAU	0.093***			0.050***		
	(0.03)			(0.01)		
Engagement		1.537**			0.460	
		(0.60)			(0.28)	
L1.Downloads			0.099***			0.048***
			(0.03)			(0.01)
ATT × MAU	-0.108***			-0.090***		
	(0.03)			(0.02)		
ATT × Engagement		-2.433***			-0.830*	
		(0.56)			(0.47)	
ATT × L1.Downloads			-0.135***			-0.091***
			(0.04)			(0.02)
Industry-Quarter FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
Observations	15,534	15,534	15,348	15,508	15,508	15,363
R-sq	0.287	0.287	0.293	0.184	0.182	0.185

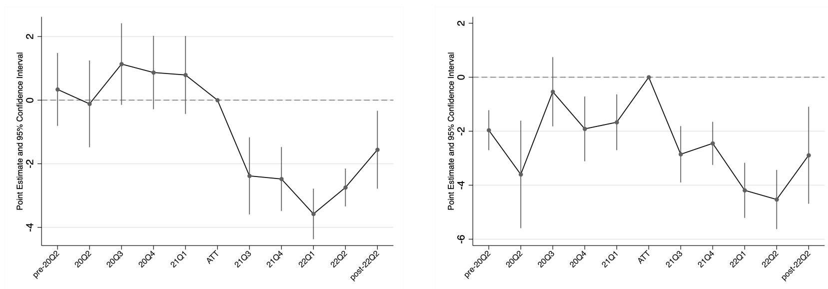
Table C.1 presents estimation results from Equation (2) using the app sample. We consider two additional mobile performance metrics, MAU and user engagement ratio. The variable *MAU* denotes the natural logarithm of total monthly active users for a firm's apps in a given quarter. *Engagement* is defined as the ratio of DAU to MAU. *L1.Downloads* is the natural logarithm of downloads from previous quarter. 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 in Equation (1)). 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 $t-4$. In all 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.

Figure C.1: Mobile Metrics and Firm Performance Around ATT – Robustness

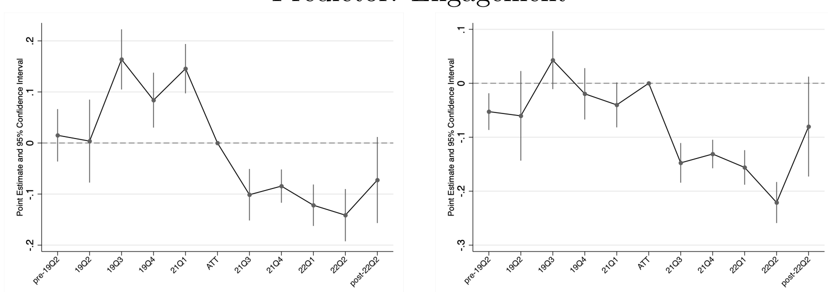
Outcome: Sales Growth | **Outcome: SUE**



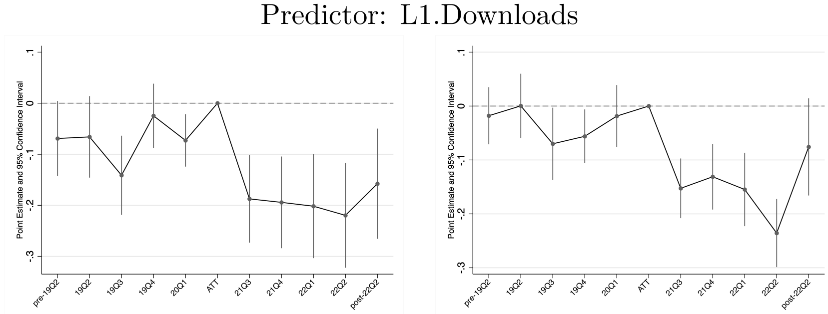
Predictor: MAU



Predictor: Engagement



Predictor: L1.Downloads



Excl. COVID period

Figure C.1 plots the estimated coefficients for the time dummy variables from Equation (3), 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.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 $\times$ 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 (2) 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, calculated as in Equation (1)). 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 Fund-Level Analysis

D.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:

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

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 $B = 7$ 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
- w_{it}^b is the weight of stock i in benchmark b

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

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

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 relies on this fund-specific benchmark portfolio.

D.2 Construction of Fund Specialization

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

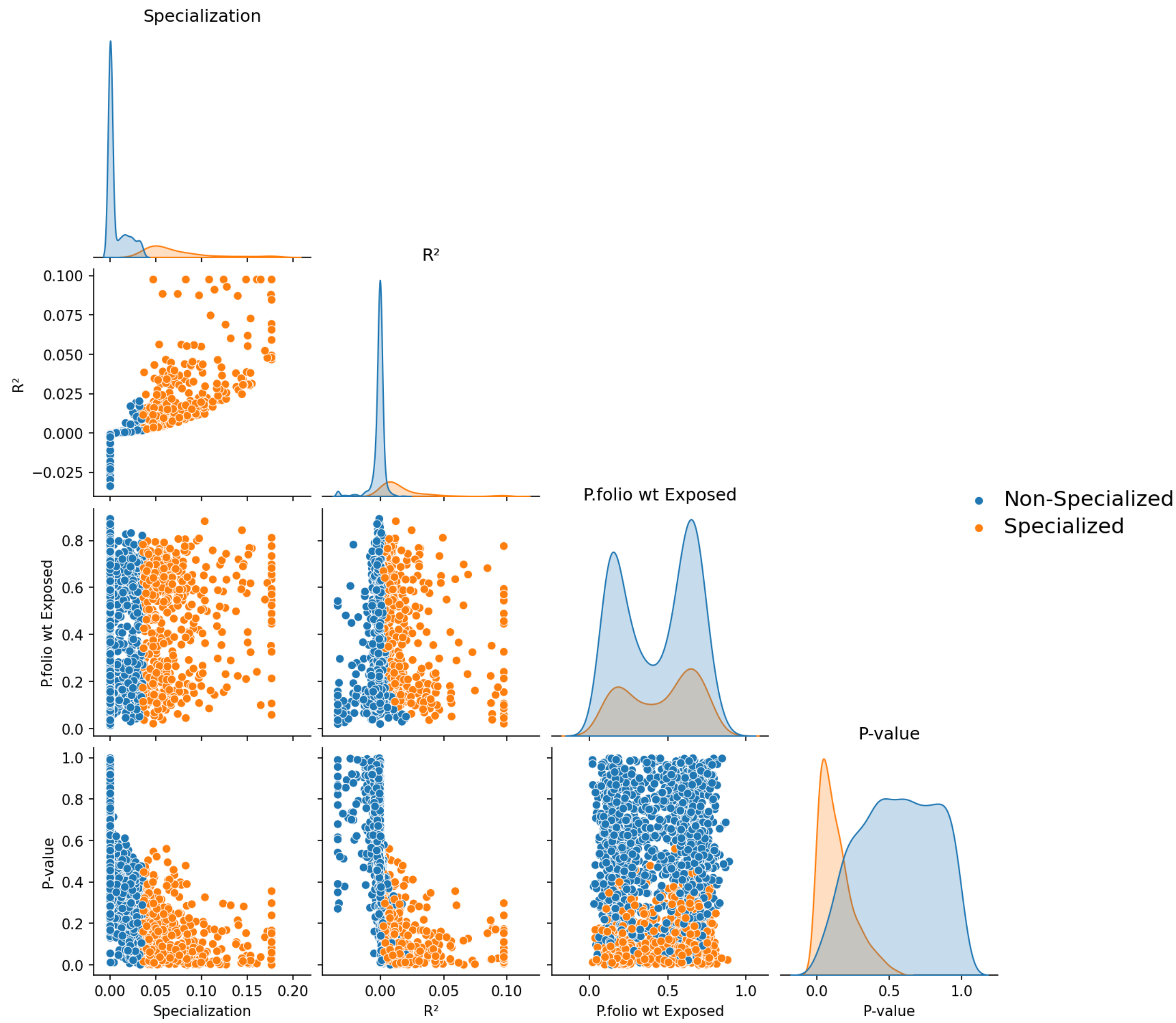


Figure D.1 displays pairwise distributions between regression statistics and components of the fund specialization measures: (i) the composite measure (portfolio weight in exposed stocks × within–fixed effect adjusted R^2, both measured prior to ATT), (ii) within–fixed effect adjusted R^2, (iii) portfolio weight in exposed stocks, and (iv) p-value of the fund-level coefficient of abnormal downloads from Equation (5), 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.

D.3 Additional Robustness

Table D.1: Fund Picking Ability at 9-Month Horizon Around ATT – Excluding COVID-19 Period

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × Exposed × Specialized	-0.036***	-0.036***	-0.027***	-0.026***
	(0.009)	(0.009)	(0.008)	(0.008)
Exposed × Specialized	0.019***	0.019***	0.012**	0.012**
	(0.006)	(0.006)	(0.005)	(0.005)
ATT × Exposed	-0.067***		-0.060***	
	(0.008)		(0.008)	
Exposed	0.035***		0.026***	
	(0.005)		(0.005)	
Fund×Month FE	Y	Y	N	Y
Exposed× Month FE	N	Y	N	Y
Observations	174,574	174,574	174,574	174,574
R-sq	0.66	0.67	0.57	0.58

Table D.1 reports the estimation results of Equation (7), excluding 18 COVID-relevant months from 2019m7 to 2021m3. The future idiosyncratic return used to construct *Picking* is defined as the compounded abnormal return over the 9-month horizon $[t+1, t+9]$, 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, 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$^{\text{th}}$ percentile of the distribution, as defined in Section 5.1.1. 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 controls for Exposed×Month interactions. 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.

Table D.2: Fund Picking Ability at 9-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	0.009	0.009	0.015*	0.015*
	(0.009)	(0.009)	(0.009)	(0.009)
ATT × Exposed	-0.003		0.006	
	(0.008)		(0.006)	
Exposed × Specialized	0.000	0.000	-0.004	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)
Exposed	-0.001		-0.006**	
	(0.003)		(0.003)	
Fund×Month FE	Y	Y	N	Y
App-owning × Month FE	N	Y	N	Y
Observations	237,966	237,966	237,966	237,966
R-sq	0.690		0.580	0.590

Table D.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 9-month horizon $[t+1, t+9]$. 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. For each exposure group, stock-level picking measures are then aggregated to the fund level using simple averages. 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 adds the interaction terms Exposed×Month. Standard errors, double-clustered at the fund and month levels, are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

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

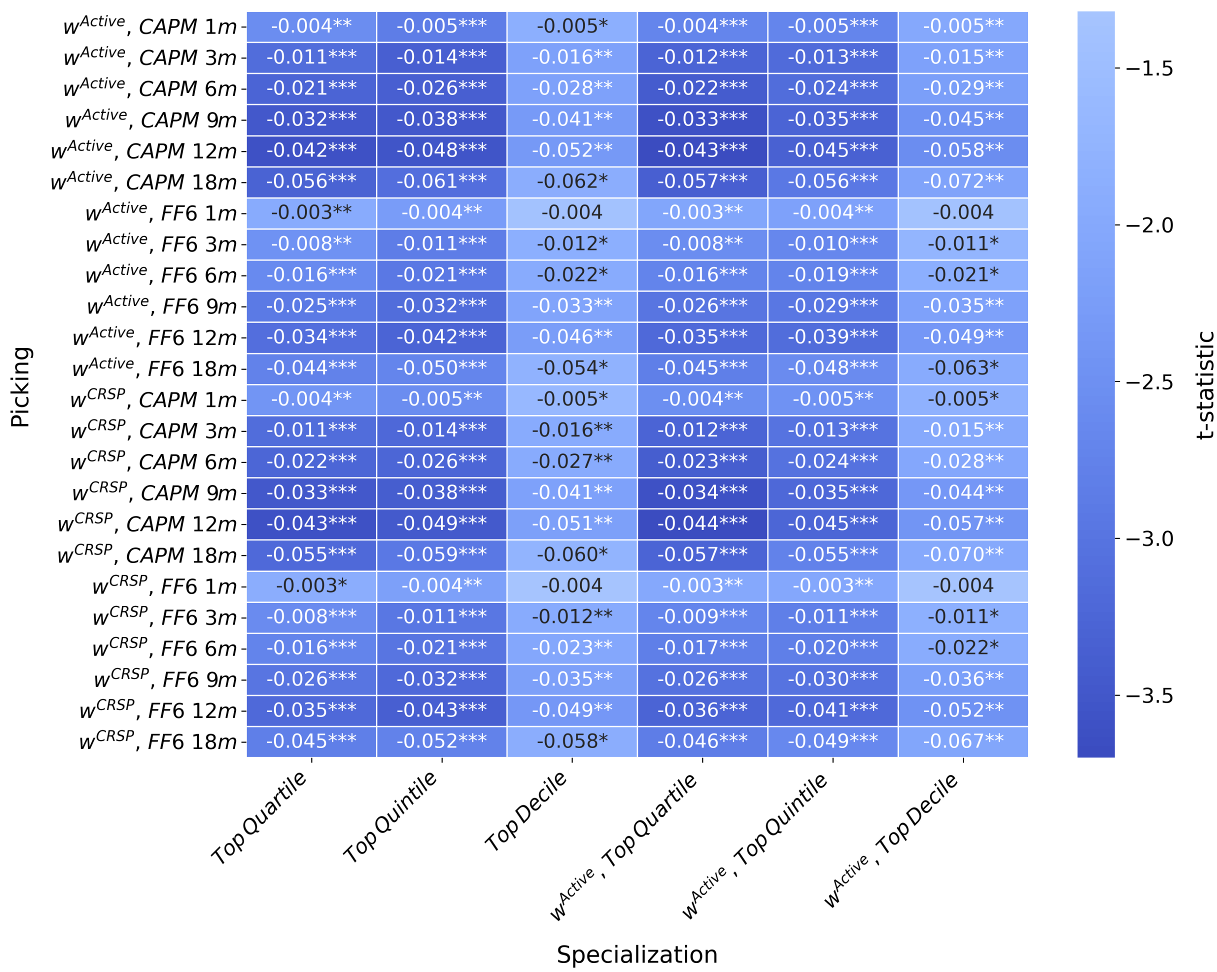


Figure D.2 reports estimates of the coefficient on the triple-interaction term ATT × Exposed × Specialized from Equation (7). The vertical axis lists alternative stock-picking measures, defined using different benchmark portfolios (CRSP market vs. the closest passive benchmark funds) and horizons for 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 Analyst-Level Analysis: Robustness

Table E.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(Broker size)		0.001		-0.003
		(0.00)		(0.00)
Constant	0.300***	0.286***	0.298***	0.616**
	(0.00)	(0.01)	(0.00)	(0.25)
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 E.1 reports the estimation results of Equation (8) 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 E.2: Analyst Forecast Errors Around ATT – Placebo

	PMAFE			
	(1)	(2)	(3)	(4)
ATT × Keyword usage intensity	0.044	0.046	0.027	0.032
	(0.07)	(0.07)	(0.08)	(0.08)
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(Broker size)		-0.005		0.000
		(0.00)		(0.01)
Constant	-0.024***	-0.116***	-0.024***	0.751*
	(0.00)	(0.02)	(0.00)	(0.44)
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 E.2 reports the estimation results of Equation (8) 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 E.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**
	(0.10)	(0.10)	(0.09)	(0.09)
Keyword usage intensity	-0.134*	-0.138*		
	(0.07)	(0.07)		
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
		(0.01)		(0.01)
Constant	-0.026***	-0.089**	-0.031***	0.323
	(0.01)	(0.04)	(0.00)	(0.60)
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 E.3 reports the estimation results of Equation (8), 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 E.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
	(0.13)	(0.12)	(0.21)	(0.18)
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)	(0.04)	(0.11)
ln(#Firms covered)	-0.009	0.007	0.003	-0.015
	(0.01)	(0.02)	(0.01)	(0.03)
ln(Broker size)	0.010	-0.011	0.001	-0.007
	(0.01)	(0.02)	(0.01)	(0.02)
Constant	1.071	-1.121	1.355**	-0.709
	(0.74)	(1.23)	(0.57)	(1.45)
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 E.4 presents the estimation results of Equation (8) 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.