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

Simona Abis† Bo Bian‡ Huan Tang§

Abstract

Data-rich environments have been shown to improve market efficiency and benefit firms with extensive digital footprints. Yet as investors increasingly rely on granular behavioral data, growing privacy concerns have prompted regulations that risk disrupting these informational advantages. We study this tension through the lens of Apple's App Tracking Transparency (ATT) policy, which sharply restricted the sharing of user-level data across mobile applications. Exploiting ATT as a natural experiment, we find that it significantly reduces the predictive power of mobile-generated signals for firm performance and investor behavior. Mutual funds that had previously exhibited superior stock-picking ability for mobile-data-exposed firms lose this advantage post-ATT, and analyst forecasts referencing mobile data become less accurate and elicit weaker market reactions. These disruptions are accompanied by a decline in price efficiency, particularly for firms most exposed to mobile-data-informed analysis. Our findings provide causal evidence that privacy regulations can impair financial information flows and reduce market efficiency.

Keywords: Mobile Apps, Data Sharing, Privacy, Price Informativeness, Forecast Errors, Earnings Surprises, Digital Targeting

JEL Codes: G14, G12, O16

*We thank Zhi Da, Mengqiao Du, Olivier Dessaint, Thierry Foucault, Allen Hu, Wei Jiang, Connie Mao, Ye Mao, Xiaoyun Yu, and seminar/conference participants at Indiana University, University of Colorado Boulder, and Penn/NYU Conference on Law and Finance for their helpful comments and suggestions. We thank Quinn Danielson, Rohan Joseph, and Bolin Xu for excellent research assistance. This version: July 2, 2025.

†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

A growing body of research shows that data-rich environments improve financial market efficiency, facilitate capital allocation, and disproportionately benefit large firms with deeper digital footprints and longer data histories (Bai et al., 2016; Begenau et al., 2018; Farboodi et al., 2022). However, this narrative overlooks a rising counterforce: regulatory uncertainty around data privacy and ownership. As financial markets grow more reliant on granular behavioral and platform-generated data, firms become increasingly exposed to policy shifts that constrain data access. This creates a tension: the same data infrastructure that enhances market-based information about firms may also subject them to new sources of regulatory risk. We examine this tension by studying Apple's App Tracking Transparency (ATT) policy, a major privacy intervention that reshaped the flow of mobile-generated signals to financial markets. Our analysis traces how this disruption affected the production and use of financial information in capital markets.

Unlike the commonly studied, technology-driven shocks that expand data availability, ATT is not merely their inverse. The inverse of a data expansion would be a data contraction, but ATT does not cause the removal of any existing dataset. Rather, it represents a permanent change in how user-level data are shared and aggregated across the digital ecosystem. Introduced in 2021, the policy restricts firms' access to device-specific identifiers on iOS, effectively severing the cross-app tracking of mobile users that underpins the production and distribution of mobile-generated alternative data. Hence, ATT preserves all existing datasets but decreases their signal-to-noise ratio. That is an important distinction in the presence of switching frictions. We leverage this policy shock to causally assess how data sharing restrictions affect the quality of financial information. We find that ATT reduces the predictive content of mobile-generated signals, increases forecast errors among market participants, and weakens price efficiency, demonstrating that privacy regulations can impose meaningful frictions on financial information flows.

Mobile-generated data are a particularly relevant focus because they constitute a sig-

nificant share of the alternative data landscape and are directly affected by ATT. Among the nine most widely used alternative data categories by institutional investors (AlternativeData.org, 2019), at least five—including mobile usage, geolocation, and social media sentiment—rely in full or in part on mobile platforms. The ease of data sharing in the mobile environment is a double-edged sword. It has made mobile-generated data central to the alternative data economy, widely used by institutional investors and sell-side analysts (Tingley, 2021; Chi et al., 2024). However, the heavy reliance of these signals on user tracking across apps makes them particularly vulnerable to privacy regulations like ATT, which restrict data sharing and user tracking. In practice, opt-in rates were low (just 4–18% in the U.S. Balasubramanian (2022)), and with iOS comprising over 50% of the U.S. market (comScore, 2024), the policy sharply curtailed access to mobile-collected data, providing a valuable setting to examine how data sharing restrictions affects investor behavior and market outcomes.

To assess the consequences of this data restriction for financial markets, we examine how ATT reduced the informativeness of mobile-generated signals and its downstream effects on investor behavior and price formation. Our empirical strategy combines detailed app-level activity data with firm fundamentals, mutual fund portfolio holdings, and sell-side analyst forecasts. We first test whether ATT weakened the predictive power of mobile signals for firm performance, and then trace how this disruption affected trading behavior, forecast accuracy, and ultimately, price efficiency.

We begin by hypothesizing that ATT's restrictions on user-level data sharing should reduce the predictive power of mobile-generated signals for firm fundamentals. Consistent with the hypothesis, we show that app traffic (e.g., downloads) entirely loses its predictive power over sales growth and earnings surprises (SUE) after ATT. The loss in predictability is relevant for portfolio construction. Between 2017 and 2021 a portfolio which buys stocks with high abnormal downloads and shorts stocks with low abnormal downloads registers an overall Cumulative Abnormal Return (CAR) of 35%-120%, depending on portfolio construction;

while no additional gains are registered on that strategy post-ATT.

Taken in isolation, the loss of profitability from the above strategy could be attributed to an *increase* in the use of mobile signals by market participants, such that prices already fully reflect the information embedded in those signals. However, this explanation would not explain why abnormal returns flatten exactly when ATT takes effect. To further corroborate our interpretation, we provide more direct evidence on the usage of such signals by examining mutual fund holdings. Specifically, we measure changes in the portfolio weight of each stock that result from buying or selling decisions. We then estimate predictive regressions to test whether abnormal app downloads predict these trading decisions. We find that, prior to ATT, mutual fund buying decisions are positively correlated with past abnormal downloads, whereas this correlation becomes statistically indistinguishable from zero post-ATT, indicating a *decrease*, rather than an *increase*, in the use of these signals by market participants.

Next, we hypothesize that if market participants rely on these signals in forming their forecasts, then a loss in their predictive power should result in greater forecast errors. Specifically, less precise mobile-collected data should exhibit a lower signal-to-noise ratio. When market participants use such noisier signals to update their priors about firms' future payoffs, the resulting posteriors should also be noisier, ultimately leading to larger forecast errors. To test this hypothesis, we examine changes in both mutual funds' stock-picking abilities and financial analysts' forecast errors around the implementation of ATT.

Since we do not directly observe mutual funds' forecasts, we infer their information sets by studying their portfolio choices. More specifically, we build stock-level measures of picking ability as the co-variance between (1) the weight each stock represents in a fund's portfolio, in excess of the weight it represents in the fund's benchmark and (2) that stock's future idiosyncratic return (6-month ahead in our baseline specification). A higher co-variance is indicative of a higher forecasting ability. For each fund-month, we aggregate stock-level picking abilities as their mean or sum across two distinct groups of stocks: those that possess

an app and hence, should be affected by ATT (*exposed*) and those that do not (*unexposed*). Finally, we implement a difference-in-differences (DiD) analysis by studying the gap in stock picking ability between *exposed* and *unexposed* stocks before and after ATT.

We find that pre-ATT, funds have a significantly higher picking ability over exposed stocks, while the opposite is true post-ATT. The effect is economically large. When summing stock picking abilities over the stocks within each group, we find that prior to ATT, the picking ability over exposed stocks is approximately one quartile higher than that over *unexposed* stocks, while post-ATT, this effect is more than reversed. When aggregating using the average stock-picking ability, we obtain qualitatively similar results; however, the magnitude of the pre-ATT advantage for exposed stocks is smaller (approximately half a quartile), and the post-ATT decline is even larger. This indicates that the loss in forecasting ability is partially mitigated through re-balancing: pre-ATT, funds appear to have a competitive advantage in selecting *exposed* stocks, which they overweight in their portfolios; post-ATT, they lose that advantage and consequently underweight those stocks. That result is consistent with the predictions in both Kacperczyk et al. (2014) and Dessaint et al. (2024), where capacity constrained investors optimally allocate their attention (and hence their investment) based on the marginal value or on the cost of information. We then repeat our baseline analysis by constructing a picking measure for different horizons of future idiosyncratic returns. Consistent with the findings of Dessaint et al. (2024), we observe that the sharpest decline occurs for forecasts up to 12 months ahead, suggesting that mobile-generated data, like other forms of alternative data, are more valuable for short-term predictions.

To further test our hypothesis across different types of market participants, we complement the mutual fund analysis with an investigation of sell-side financial analysts, whose forecast histories and information sources are directly observable. We start by constructing different analyst-level metrics of specialization in alternative data. First, for a subset of analysts, we are able to directly identify those using mobile-generated data by searching for mentions of mobile-based indicators (e.g., daily active users) and mobile data providers (e.g.,

App Annie) in the text of their reports.[1] We then define "*Keywords usage intensity*", as the ratio of an analyst's reports that contain app keywords to the total number of reports that analyst has written. Second, for each quarter, we define an analyst as specialized if they primarily cover firms in the retail and services industries, whose performance is particularly dependent on app traffic (Bian et al., 2021). Then, for each forecast, we compute forecast errors relative to other analysts making forecasts for the same firm in the same quarter. Importantly, we include firm-quarter fixed effects to absorb any firm-specific shocks, and our regression thus isolates the impact of ATT by comparing analysts covering the same firm-quarter but differing in their reliance on mobile-generated alternative data.

We show that, pre-ATT the relative forecast errors of analysts specialized in alternative data are lower than those of non-specialized analysts, while these errors significantly increase post-ATT. We further validate this interpretation by showing that ATT has no impact on the forecast errors of specialized analysts when they issue forecasts for firms unlikely to be affected by ATT. Interestingly, the increase in Post-ATT forecast errors (sufficient to bring an analyst from the median to the 57th percentile) is larger than their pre-ATT advantage. That indicates that they not only lose their predictive edge post-ATT, but their forecasts become on average worse than those of analysts that do not use alternative data. To shed light on this result, we split analysts into two categories: *switchers*, those with an above-median decline in "keyword usage intensity" post-ATT, and *non-switchers*. We show that *switchers* merely lose their competitive advantage post-ATT, while non-switchers, by continuing to rely on now-uninformative signals, perform much worse than analysts who do not consider alternative data in their forecasts. This highlights another reason why ATT is not just the reverse of increased data availability. When analysts face limits on time or resources, they may find it hard to shift attention from familiar but weakened signals, especially when better alternatives are costly or unclear. As a result, part of their attention remains tied up, which can worsen the negative impact of lower data quality. Finally, we show that the increase in

[1] A full list of keywords searched can be found in Appendix Table B.1.

forecast errors of analysts who utilize alternative data happens sharply around ATT, with little pre-trends, strengthening our causal interpretation.

To connect analyst forecast errors with broader market effects, we also examine how the content of individual reports shapes immediate price reactions upon release. This report-level analysis allows us to directly test whether the market processes and responds to specific report content, as well as how this responsiveness changes following ATT. We find that, prior to ATT, reports that reference more app keywords trigger significantly stronger market reactions, measured in cumulative abnormal returns. However, this differential response disappears following the implementation of ATT. This finding further supports the interpretation that the information content of analyst reports declined following ATT.

Our final 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.

We begin by examining short-term price efficiency through the lens of how earnings surprises (SUE)—the difference between actual earnings and the consensus forecast, scaled by price—translate into stock price reactions. By examining the sensitivity of stock reactions around earnings announcements to SUE, this approach implicitly controls for the information set of the median analyst and allows us to assess how much private, incremental information about earnings is already incorporated into prices before the earnings release. To proxy for firms' pre-ATT reliance on alternative data, we use firm-level coverage by analysts specialized in such data. We find that, prior to ATT, the sensitivity of CAR to SUE is about one-quarter lower for these firms, revealing a better information environment. Post-ATT, however, this sensitivity increases significantly, indicating that the quality of private signals relative to the median analyst's information set deteriorates. This effect is absent among firms without mobile apps, supporting the role of ATT in driving these results.

To more generally assess changes in firms' information environments outside of the nar-

row earnings announcement window, we also 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 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 uncertainties and investor disagreements. Together, these results provide evidence that ATT impaired the flow of alternative data into financial markets, weakening price efficiency and amplifying informational frictions at the firm level.

Admittedly, mobile usage data is not the only form of alternative data affected by ATT. Any mobile-collected dataset containing user-identifiable information, such as geo-location data used to measure foot traffic, is likely to deteriorate in quality post-ATT. Thus, the capital market effects we document should be seen as the compounded impact of reduced quality across all affected mobile-collected datasets used by market participants.

Related Literature Our study investigates how privacy restrictions affect capital markets through their impact on both the quality and availability of alternative data. On one hand, we contribute to the growing literature on privacy in economics and finance, particularly studies examining the effects of privacy regulations such as GDPR (Goldberg et al., 2019; Janssen et al., 2022; Jia et al., 2021; Aridor et al., 2020; Bessen et al., 2020; Demirer et al., 2024; Peukert et al., 2022), CCPA (Doerr et al., 2023; Bae et al., 2023), and ATT (Kesler, 2022; Bian et al., 2023; Kraft et al., 2023).[2]

On the other hand, our paper connects with a rich body of research examining the rising use of alternative data in capital markets. Prior work has explored a diverse range of data sources, including satellite imagery (Zhu, 2019; Katona et al., 2024; Ke, 2023), customer and employer reviews (Huang, 2018; Green et al., 2019), real-time and granular sales data (Froot et al., 2017; Blankespoor et al., 2022; Dichev and Qian, 2022; Jin et al., 2022; Du and

[2]See Johnson (2022) for a comprehensive review of the literature on GDPR.

Qian, 2024), EDGAR web traffic and Google searches (Chen et al., 2020; Da et al., 2011), robo-journalism and media synthesis (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), and cloud data records (Chang and Da, 2022).

Our main contribution is to bridge these two strands of literature by showing that privacy regulations, motivated by growing concerns over data use, can generate significant ripple effects on financial markets. In doing so, we highlight a largely overlooked perspective: that the very data intensity which improves market efficiency potentially invites regulatory constraints, creating a paradox in which informational advantage triggers informational erosion. Our paper is the first to show that data privacy regulations can unintentionally degrade the quality of forecasts and the broader information environment.

More specifically, we advance the understanding of how mobile-generated data influences capital markets by leveraging ATT as a natural experiment. Prior work has documented the benefits of mobile-generated data for market participants, showing that such data can serve as a valuable input in forecasting firm performance and guiding investment decisions. For example, Chen et al. (2024) show that mobile app performance serves as a leading indicator of firm earnings, though financial markets only partially incorporate this data, resulting in predictable forecast errors and excess returns. Similarly, Jin et al. (2022) use GPS location data from customers' mobile devices to develop new measures of shopping behavior and discuss their valuation implications. Building on the premise that mobile-generated data is valuable to financial actors, we examine what happens when access to such data is curtailed. We provide causal evidence that ATT degraded the quality of mobile-generated signals by disrupting the broader interconnected network through which alternative data is aggregated, shared, and monetized across the digital economy (Bian et al., 2024).

Our work also contributes to the literature on the evolving use of alternative data by different market participants. The proliferation of alternative data vendors (Grennan and Michaely, 2021) and the rise of mobile-based consumer surveillance have created new infor-

mation channels for analysts. Applying textual analysis to analyst reports, Chi et al. (2024) show that analysts draw insights from alternative data and generate more accurate earnings forecasts. Dessaint et al. (2024) highlight the tradeoff between short- and long-term informativeness, documenting that analysts' long-term forecasts may deteriorate as they shift toward short-term signals embedded in alternative data. Turning to fund managers, Bonelli and Foucault (2023) use the availability of satellite imagery as a positive data shock and argue that big data may displace high-skill workers in finance by weakening active funds' stock-picking edge. While positive shocks typically introduce new, sometimes niche datasets, ATT permanently impaired a widely adopted information set relied upon by a broad set of firms and investors. This contrast is important: whereas positive shocks may create temporary informational gains, ATT offers a rare opportunity to examine whether market participants recognize the reduced informational content of degraded data, whether effective substitutes exist, and whether effort and attention are reallocated or stickiness persists in data usage.

2 Institutional Background

Our setting features a major privacy policy shock, Apple's App Tracking Transparency (ATT), that restricts data sharing via mobile devices. A wide range of alternative datasets, increasingly relied upon by financial market participants, is sourced through mobile apps and therefore directly affected by this policy. In what follows, we discuss the scale of the mobile app market, the unique characteristics of mobile-generated data, the ATT policy itself, and broader regulatory concerns in the alternative data ecosystem.

2.1 The Mobile App Market

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).[3]

2.2 Data Sharing and ATT

The growing popularity of mobile apps has established them as a key channel for collecting user-level data. The sharing and aggregation of such data have become increasingly widespread, enabled by persistent device-specific identifiers—such as the Identifier for Advertisers (IDFA) on iOS and the Google Advertising ID on Android—which are unique to the mobile ecosystem.[4] These identifiers enable firms to link user activity across apps and companies, facilitating user tracking, behavioral profiling, and the optimization of advertising campaigns, thereby contributing to customer acquisition and retention (Bian et al., 2024). 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-party aggregators (e.g., data brokers), 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 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.

Although ATT was announced ten months prior to implementation, its rollout had a

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

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

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).

Given iOS's dominant share of over 50% in the U.S. market, the low opt-in rates, and the fast global roll-out, ATT constituted a substantial shock to firms' ability to link user behavior across apps and to share granular data externally. Crucially, as a first-of-its-kind data policy that had adverse effects on firm performance, its impact were unanticipated by financial markets, as documented by Bian et al. (2021).[5]

2.3 Regulatory Concerns in Alternative Data Markets

Despite its growth, the adoption of alternative data raises 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). Mobile data is especially vulnerable to such concerns. Of the nine major categories of alternative data used by institutional investors (AlternativeData.org, 2019), at least five—including app usage, geo-location, and credit/debit card data—rely heavily on mobile collection. Many other categories, such as sentiment and consumer receipts, are at least partially dependent on mobile app access.

A prominent example came in 2021, when the SEC charged App Annie, a major provider of mobile data, with securities fraud for misusing confidential mobile data and selling it to

[5]Unlike ATT, which directly restricts access to device-level identifiers critical for linking user behavior across mobile apps, GDPR and CCPA primarily impose disclosure and consent requirements without technically disabling cross-platform tracking by default. As a result, GDPR and CCPA tend to shift compliance procedures rather than immediately disrupting the underlying data infrastructure, leading to less direct and far-reaching effects on firm performance. Figure C.9 of Bian et al. (2021) documents an immediate and sharp decline in the cumulative abnormal returns (CAR) for app-owning firms following the implementation of ATT, while CARs remained consistently around zero prior to the event—consistent with a lack of investor attention or anticipation. The 30-day CAR is approximately -2% to -3%.

trading firms without proper disclosure or consent (SEC, 2021). According to Eagle Alpha, this case heightened industry awareness around the legal boundaries of alternative data use (Ryan, 2023).

As ATT directly restricts the sharing and aggregation of all mobile-user-level data, it constitutes a salient and far-reaching regulatory shock to the alternative data ecosystem, making it a natural experiment for studying the impact of privacy-driven data restrictions on capital markets.

3 Hypotheses Development

Figure 1 graphically summarizes the logical links in our study. Firms collect user data through apps, either as first-party data—such as browsing and purchase activity collected directly within the app—or as third-party data, including location information obtained from external sources. This data may be shared with third parties, such as data brokers or platforms like Google Analytics, in exchange for payment or services like analytics and targeting. Firms, either directly or through partners like Google Analytics, analyze both first-party and third-party data—often using AI models—to gain insights about their users and improve targeting. A key targeting tool at firms' disposal is mobile advertising. Customer targeting then drives traffic (e.g., downloads, monthly active users or MAU, daily active users or DAU) to a firm's app. The quality of that traffic is linked to customer retention and, ultimately, to firm performance.

While the direct impact of ATT on firm performance is well documented (e.g., Kesler, 2022; Bian et al., 2021, 2024; Kraft et al., 2023), its broader effects on financial markets remain underexplored. Alternative data providers collect user-level signals—such as app usage, transactions, and geolocation—from firms or data brokers and transform them into structured datasets for investors. These signals inform expectations about firm fundamentals and drive trading, influencing asset prices and, ultimately, firm decisions such as investment and financing.

We now formulate hypotheses on how data-sharing restrictions may disrupt the informa-

tion flow outlined above, beginning with ATT's impact on the quality of mobile-generated alternative data. As discussed in Section 2.2, high opt-out rates reduce data volume, while limits on data sharing undermine the precision of data for customer profiling. With less granular data, firms' targeting becomes less effective, diminishing the predictive power of app usage and advertising expenditure for future performance. Similarly, other investor-facing datasets built on mobile inputs may also decline in quality. Anchoring our predictions on app usage data, we note that pre-ATT, app downloads were predictive of revenue due to effective targeting. Post-ATT, however, downloads may increasingly reflect untargeted, low-quality traffic that contributes less to sales and earnings, weakening the signal-to-noise ratio. Since app usage metrics are typically produced using machine learning, reductions in data volume or representativeness may further degrade model accuracy. This reasoning gives rise to the first prediction.

Hypothesis 1. *ATT makes app usage metrics worse predictors of firm's cash flows and stock performance.*

Second, we examine how reduced signal quality affects financial market participants. Investors incorporate alternative data into prices by updating their beliefs and trading accordingly. If ATT lowers the precision of mobile-based signals, it should lead to noisier posterior beliefs and weaker decisions. As a result, we expect diminished stock-picking performance by fund managers and larger forecast errors among analysts post-ATT. We test these predictions using fund holdings and analyst forecasts.

Hypothesis 2a. *ATT reduces the stock-picking performance of fund managers, with the effect concentrated among stocks whose performance was previously predictable using mobile-generated alternative data.*

Hypothesis 2b. *ATT leads to an increase in analyst forecast errors, with the effect concentrated among those who previously relied on mobile-generated alternative data in forming their forecasts.*

Finally, we examine the impact of ATT on firms' information environment and price efficiency. The ripple effects of ATT are expected to degrade the information environment. According to the Efficient Market Hypothesis (EMH), asset prices reflect all available information about future payoffs. When the information environment worsens, prices are likely to become less efficient, i.e., less informative about firm fundamentals. This decline in price efficiency should be most pronounced for firms whose valuations previously incorporated insights derived from mobile-collected data.

Hypothesis 3. *ATT worsens the information environment, with the effect concentrated among firms followed by market participants who rely on mobile-generated alternative data.*

4 Data and Measurement

We focus on publicly listed companies, for the period of 2017 to 2023, which encompass a few years before and after the implementation of ATT. We leverage a granular dataset of the mobile apps market, which we merge with firm financials, mutual fund holdings, and analyst forecasts. That allows us to compute measures of exposure to ATT at the firm and analyst level, which we use to test the impact of ATT.

4.1 Mobile Apps

We acquire information about mobile app usage (e.g., downloads, active users) from Apptopia. Apptopia is an alternative data provider which collects, structures and models data about the mobile apps market. Apptopia acquires mobile usage data from apps they partner with, they combine it with data from a users panel, and with publicly available data (e.g., App Store rankings). They then use AI-powered proprietary models to estimate app usage metrics for over 7 million apps. Among those apps, they provide estimates for over 3,000 publicly listed firms worldwide. They then sell this structured dataset to companies seeking industry insights and to financial market participants seeking price signals. We source mobile-usage data from Apptopia, whose estimates have been accessible via Bloomberg Terminal since 2019 and are therefore widely available to financial market participants. Our

analysis aggregates usage across all apps associated with each firm.

4.2 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 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 we 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. Finally, we gather control variables for the funds in our final sample by aggregating observations across all share classes of each fund. Age is based on the first offer date of the oldest share class, size is obtained as the sum of the TNA of all share classes, while other variables (fees, returns, turnover) are weighted by lagged TNA. We measure flows as the percentage change in TNA, not due to returns, and we compute flow volatility using 12 months rolling windows. Fund style is then constructed as the average exposure of stocks held to Fama-French 6 factors.

4.3 Analyst Forecasts

We collect analyst forecasts information from I/B/E/S. We keep forecasts made within 90 days of each earnings announcement, keeping the last forecast for each analyst-broker pair. Additionally, we use the LSEG dataset 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. 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 the ATT, such as "satellite image" and "dun & bradstreet." For each firm in our app sample introduced in Section 4.5, 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. The full procedure is described in Appendix B.

Among the keywords listed in Appendix Table B.1, we further highlight those that are particularly susceptible to ATT and frequently cited by analysts, with frequency counts reported in parentheses: *mau* (62,480), *dau* (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 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. We provide illustrative report excerpts in Appendix Section B.3. We use this restricted set of keywords (henceforth, "app keywords") to identify analysts who are highly specialized in app data.

We construct measures of pre-ATT analyst specialization in mobile-generated data using both the I/B/E/S and LSEG-scrapped keyword datasets. Our first proxy is based on the keywords mentioned in analyst reports. We define the intensity of app keyword usage (*"Keywords usage intensity"*), which is the ratio of an analyst's reports that contain app keywords to the total number of reports that analyst has written.[6] Our second proxy is based on I/B/E/S data. We define an analyst as specialized if they primarily cover firms in the retail and services industries (SIC code ranges: 5200–5999 or 7000–8999), whose performance is particularly dependent on app traffic (Bian et al., 2021). Specifically, we classify an analyst

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

as specialized if the share of forecasts issued on retail and services firms in their forecast portfolio exceeds the 90th percentile among all analysts in the pre-ATT period. We define firm-level exposure to specialized analysts as an indicator variable equal to one if the firm is followed by at least one specialized analyst in at least half of the quarters prior to ATT.

4.4 Firm Financials

We obtain quarterly accounting information from Compustat. 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 mil; and, if I/B/E/S forecasts are available, we require earning announcement dates in Compustat and I/B/E/S not to differ by more than one day. Finally, we obtain financial market data (price, market capitalization, returns) from CRSP, and factors from Ken French data library.

We construct two sets of variables to measure the quality of the information environment at firm-level. The first is the average bid-ask spread in a quarter. The daily spread is calculated as the difference between the closing ask and bid prices, divided by the mid-price, and then averaged over all trading days in the quarter. It serves as a proxy for information asymmetry and trading frictions in the firm's information environment. The second variables are post-earnings announcement volatility. Following Loughran and McDonald (2014) and Kim et al. (2023), this variable captures the volatility, i.e., the root-mean-squared error of *abnormal returns* over the window spanning trading days 6 to 28 after the earnings announcement date. Abnormal returns are estimated using two alternative models: (i) the Capital Asset Pricing Model (CAPM) and (ii) the Fama-French 6-factor (FF6) model. This measure reflects the residual uncertainty or disagreement in the market following the initial incorporation of earnings information and is interpreted as a proxy for noise in the information environment.

A complete list of variables and their definitions used in the analysis is provided in Table A.1.

4.5 Summary Statistics

Tables 1-2 report summary statistics for the main variables and samples used in the analysis.[7]

Table 1 shows firm-level summary statistics based on different data samples. Panel A displays statistics for the 809 unique firms in the *"App sample"*, the main sample of interest in our analysis. That includes all observations of firms which have an app with on average at least 1000 daily active users throughout our sample period (2017-2023). This ensures that we ignore a long tail of unused apps that are unlikely to be relevant for firms. Panel B displays statistics for the 5,970 unique firms in the *"No-app Sample"*. This includes all observations of firms that do not have an app. The median firm in the two samples are comparable in terms of size and firm characteristics, with the exception that firms with apps tend to have a higher debt-to-asset and market-to-book ratio. Not surprisingly, firms with an app are more likely to be followed by specialized analysts—defined based on keyword usage—at a rate of 63.5%, compared to 24.5% for firms without an app. Looking at mobile-related variables, the median firms have 290,140 app downloads and a 379,611 MAU per quarter.

Table 2 displays analyst-level and report-level statistics for their forecasts of firms in the *"App Sample"* (Panel A) or for their forecasts of a placebo set of firms (Panel B). Here, the placebo sample consists of forecasts for 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 above the sample median). As such, firms in the placebo sample may be also followed by specialized analysts. The *"App Sample"* (placebo sample) includes forecasts from 2,981 (4,369) unique analysts, belonging to 274 (309) unique brokerage firms. The average analyst in the *"App Sample"* (Panel A) use app keywords in a higher fraction of reports (6.4% vs. 2.8%) than in the *"Placebo Sample"* and cover the same firm for a longer period (13.8 vs. 11.6 quarters). The two samples are otherwise similar in terms of the number of days between forecast and announcement ("Forecast age"), the number of firms covered, and

[7]Summary statistics for the fund-level control variables are reported in Appendix Table D.1, as they are not shown in the main regression tables.

the size of the brokerage firms.

Finally, Panel C of Table 2 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. The average report contains 0.11 app keyword per page, has 19.9 pages, and sells at $196.

5 Test of Hp 1: Decreased Quality of Mobile-Generated Alternative Data

We start by testing Hypothesis 1—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 a previously profitable long–short strategy based on download rankings loses its edge.

5.1 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 SUE—as in Equation (1), in quarter t to downloads in the same quarter. In constructing $\text{SUE}_{it}^{random\ walk}$, we assume that earnings per share (EPS) follow a seasonal random walk (e.g., Livnat and Mendenhall (2006)). So, the best estimate of EPS for stock i in quarter t is EPS the same quarter of the previous year $(t-4)$.[8] 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.

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

We estimate the following regression:

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

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

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 is an indicator variable equal to 1 for all quarter after ATT announcement; X'_{it-4} is a vector of control variables measured in quarter $t-4$, namely size, and cash, tangibles, debt over assets, and market-to-book ratios. 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.

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. According to Hypothesis 1, we expect $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$.

Table 4 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. To evaluate the economic magnitude of that result, 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 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.[9]

[9] 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.

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:

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

where we replace ATT 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.

We report the predictive power of mobile data based on downloads as it is the most widely used metric, and it is highly correlated with other app traffic metrics that are frequently referenced in analyst reports. In Appendix Table C.1 and Appendix Figure C.1, we verify that results are robust to alternative measures of traffic, such as *MAU* or *Engagement*, the latter defined as the ratio between DAU and MAU. These two alternative measures were mentioned by analysts for 62,480 and 15,254 times based on 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.

5.2 Portfolio Sorting

Next, to highlight the implications of that loss in predictability for financial markets, we perform a portfolio sorting exercise. Every month we sort stocks into quintiles based on on their abnormal downloads the previous month, defined as the downloads (in log) in that month minus the average downloads (in log) in the twelve preceding months. We then construct a long-short portfolio which every month buys the top quintile of stocks and short-sells the bottom one. Finally, for each month, we compute the risk-adjusted returns of long-short portfolio using the risk exposure estimated over the past twelve months.[10]

Figure 3 reports cumulative abnormal returns (CAR) for the long-short portfolio over our sample period, for different portfolio weighting methods namely: (1) value weighted portfolios sorted by downloads (green dotted); (2) equal weighted portfolios sorted by downloads, excluding micro-caps (blue solid); (3) MAU-weighted portfolios sorted by download, excluding micro-caps (yellow dashed). The vertical line indicates the month of ATT announcement. All three series present an upwards trend pre-ATT, indicating that investing in portfolios sorted by downloads leads to significant CAR between 2017 and 2021 (35% using method 1, 80% using method 2, and 120% using method 3). Post-ATT no trend can be detected and the portfolio does not register any additional CAR until the end of our sample period. That, taken jointly with the loss in predictive power of mobile signals on sales and earnings, likely indicates that app downloads become an uninformative sorting variable.

5.3 Predicting Trading Activities

One may argue that the decline in profitability of this trading strategy could be attributed to a broader adoption of downloads as a leading indicator by market participants, rendering such information already incorporated into prices. That, though, would not explain why abnormal returns flatten exactly when ATT takes effect. To provide more direct evidence,

[10]The risk adjustment is based on a Fama-French 6-factor model.

we supplement the portfolio sorting exercise with fund holdings data to show that fund managers move away from using mobile signals in guiding trading decisions post-ATT.

To measure trade-induced changes in portfolio weights, we focus on mutual fund holdings. We decompose changes in portfolio weights into two components driven by passive asset returns and active trading actions, respectively. For each fund f, stock i, and time t, we compute a hypothetical *no-trade* portfolio that adjusts weights only by asset returns, and compare it to the actual portfolio at t. Omitting the subscript f for fund, for a portfolio at time $t-h$ with holding horizon h, we define the hypothetical no-trade weights as:

$$w_{i,t}^{\text{noT,h}} = \frac{w_{i,t-h} \cdot (1 + \text{cumret}_{i,t}^{(h)})}{\sum_j w_{j,t-h} \cdot (1 + \text{cumret}_{j,t}^{(h)})},$$

where $w_{i,t-h}$ is portfolio weight at the end of calendar month $t-h$, and $\text{cumret}_{i,t}^{(h)}$ is the cumulative return of stock i during period $t-h$ to t.[11] Trade-induced changes are then computed as the difference between actual and no-trade weights:

$$D_{i,t}^{h} = w_{i,t} - w_{i,t}^{\text{noT,h}}.$$

We then relate the trade-induced weight differences to the abnormal downloads in the previous months used in the portfolio analysis above, by estimating the following equation:

$$D_{ift}^{h} = \alpha + \theta_{if} + \iota_t + \sum_{k=1}^{h} \beta_k \cdot \text{Abnormal Downloads}_{i,t-k} + \sum_{k=1}^{h} \delta_k \cdot \text{Abnormal Downloads}_{i,t-k} \times \text{ATT} + \Gamma_1 X'_{ft} + \epsilon_{ift}, \quad (4)$$

where $\text{Abnormal Downloads}_{i,t-k}$ represent abnormal downloads measured at the end of month $t-k$. We set $h=3$ to reduce noise in the weight changes as monthly holdings are not always available. We include fund-by-stock fixed effects (θ_{if}) to control for all time-invariant

[11] We compute cumulative return as $\text{cumret}_{i,t}^{(h)} = \exp\left(\sum_{\tau=t-h+1}^{t} \text{logret}_{i,\tau}\right) - 1$, with $\text{logret}_{i,\tau} = \log(1 + R_{i,\tau})$.

characteristics specific to each fund-stock pair, such as a fund's baseline preference or informational advantage regarding a particular stock, as well as any persistent stock-level fundamentals or fund mandates. Month fixed effects (ι_t) are also included to account for time-varying shocks that are common across all stocks and funds in a given month, such as macroeconomic conditions, market-wide sentiment, or regulatory announcements. Together, these fixed effects help isolate the within fund-stock and within-month variation in trading activity that can plausibly be attributed to to variations in abnormal app usage. X_{ft} is a vector of time-varying fund characteristics, including age, size, expense ratio, turnover ratio, style, flow, and flow volatility. Standard errors are triple clustered at the fund, stock, and calendar month levels.

We separately examine positive and negative trades to capture potential asymmetries in how mutual fund managers respond to app traffic signals. Because short-selling is typically constrained for mutual funds and holdings are bounded below at zero, negative signals may not necessarily lead to large reductions in positions. Accordingly, we expect $\hat{\beta}_k > 0$ and $\hat{\delta}_k < 0$, with larger absolute magnitudes for positive trades.

Table 5 reports the results. Consistent with app downloads becoming an uninformative sorting variable post-ATT, an increase in abnormal downloads in month $t-1$ predicts positive trades in month t before ATT, but this predictive power disappears entirely afterward (Column 1). The effect on negative trades is not statistically significant (Column 2), likely due to short-selling constraints and the fact that holding data are truncated at zero. Earlier lags of abnormal downloads from $t-2$ and $t-3$ do not predict trades. Including time-varying fund characteristics has little impact on the point estimates (Columns 3 and 4).

6 Test of Hp 2: ATT Decreases the Forecasting Ability of Market Participants

So far, our results provide strong evidence that the signal-to-noise ratio of mobile-generated data declined after ATT. A natural next step is to ask how this weaker information set affects the forecasting ability of market participants, who serve as the conduit through which signals are incorporated into prices. In this section, we focus on fund managers and their stock-

picking ability, as well as sell-side analysts who issue earnings forecasts.

6.1 Funds' stock-picking ability

We test Hypothesis 2a, which posits that fund managers experience a decline in their stock-picking ability, with the effect concentrated among stocks whose returns were previously predictable using mobile-generated alternative data.

We construct a measure of stock-level picking ability to assess whether mutual funds systematically overweight (or underweight) stocks with positive (or negative) informational signals. If a fund disproportionately allocates capital to stocks with strong forward-looking performance indicators (e.g., expected returns), this suggests deliberate and informed selection skill. Specifically, for each fund f, stock i, and month t, we define the stock-level picking measure as:

$$\text{Picking}_{ift}^{s,h} = 100 \cdot \Delta w_{ift} \cdot \text{Return}_{i[t+1,t+h]}^{s}, \tag{5}$$

where $\text{Return}_{i[t+1,t+h]}^{s}$ is the stock-level cumulative idiosyncratic 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. Because mutual funds vary substantially in their investment styles, we determine the relevant benchmark for each fund from a set of passive benchmark portfolios 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). We consider 7 passive funds from the Vanguard family as possible benchmarks: broad, large-cap, mid-cap, small-cap, growth, value, and momentum. The procedure is described in detail in Appendix D.1.

Since we are interested in changes in picking ability over firms whose performance was predictable using mobile data, we define as exposed stocks for which an app was available in at least half of the quarters before ATT. The remainder of the stocks held constitute the comparison group. For each fund-month, we aggregate stock-level picking abilities within the exposed ($e = 1$) and unexposed ($e = 0$) groups by taking the sum or the mean. Let

$\mathcal{I}_{fte}$ denote the set of stocks held by fund f in month t that have exposure e, we then define the total picking measure $\text{Picking}_{fte}^{\text{sum}}$ and the average picking measure $\text{Picking}_{fte}^{\text{mean}}$ for future idiosyncratic return s at horizon h as follows:

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

Our measures are similar to those of Kacperczyk et al. (2014), they, though, aggregate stock-level picking ability across the whole portfolio by summing over all stocks held. Bonelli and Foucault (2023) also utilize a similar measure, but they only consider stock-level ability without any aggregation. Our approach is in between the above two.[12] Our two methodologies for aggregating picking abilities (i.e., sum and mean) over exposed and unexposed firms allow us to speak to both stock-level changes in abilities, as well as to a reallocation effect. Indeed, changes in the average picking measure capture deviations in portfolio weights conditional on holding a given stock, changes in the total picking measure additionally reflect shifts in portfolio allocation between the exposed and unexposed groups. To sharply identify ATT-induced changes in stock-picking ability, we restrict the sample to the period from July 2020 to January 2022, covering nine months before and after the ATT implementation in April 2021.[13] Table 3 reports summary statistics for the 28,134 fund-month observations. The average $\text{Picking}^{\text{sum}}$ over the sample period is higher for unexposed stocks than for exposed ones, although the unexposed group also exhibits greater variation. In contrast, $\text{Picking}^{\text{mean}}$ is comparable across the two groups.

We estimate the changes in the picking ability of mutual funds around ATT using the following regression specification:

$$\text{Picking}_{fte}^{s,h} = \alpha + \beta_1 \cdot \text{Exposed}_{fe} + \beta_2 \cdot \text{ATT}_t \times \text{Exposed}_{fe} + \Gamma_1 X'_{ft} + \delta_f + \lambda_t + \varepsilon_{fte}, \quad (6)$$

[12] Both Kacperczyk et al. (2014) and Bonelli and Foucault (2023) use the market portfolio as a benchmark for all funds. Our results are robust to using this alternative benchmark portfolio, as reported in Appendix Table D.3.

[13] Our results are robust to using longer sample periods.

where ATT_t is an indicator variable that equals 1 for all months after April 2021. We include fund and month fixed effects (δ_f, λ_t). In our baseline specification, we set $h = 6$ and compound the daily abnormal returns over a six-month horizon $[t+1, t+6]$. Standard errors are double-clustered at fund and month level.

The coefficient of interest is β_2, which captures the change in fund stock-picking ability for exposed stocks relative to unexposed stocks around the ATT. We expect $\hat{\beta}_2 < 0$.

Table 6 presents estimation results, where the dependent variable is the total picking ability, Picking^{sum}, computed using cumulative abnormal returns (CAR) estimated either from the CAPM (Columns 1–2) or the Fama-French 6-factor model (Columns 3–4). The estimated coefficient of the interaction term, $\hat{\beta}_2$, is negative and statistically significant in all specifications, consistent with a decline in the total stock picking ability after ATT for exposed relative to unexposed stocks.

We use Column 3, where the future performance indicator is based on CAR estimated using the FF6 model, to assess the economic magnitude. The coefficient $\hat{\beta}_1 = 0.910$ implies that, prior to ATT, the stock picking ability of funds in exposed stocks is approximately one quartile higher than in unexposed stocks (eg, from the 50th to the 75th percentile according to the summary statistics reported in Table 3). After ATT, the relationship reverses: funds become relatively worse at selecting exposed stocks. Their ability in exposed stocks falls by $\hat{\beta}_2 = -2.403$, with a net effect of $\hat{\beta}_1 + \hat{\beta}_2 = -1.493$, sliding back of more than one quartile. Assuming that the pre-ATT advantage in selecting exposed stocks is entirely driven by mobile-generated alternative data, our findings imply that fund managers reallocate effort from exposed to unexposed stocks.[14]

Indeed, total stock-picking ability reflects two components of portfolio rebalancing: the inclusion or exclusion of stocks in the portfolio (extensive margin) and adjustments in portfolio weights conditional on a stock being held (intensive margin). To disentangle these effects,

[14]If fund managers simply cease using mobile signals without reallocating effort across stock groups, we would expect $\beta_1 + \beta_2 = 0$, that is, fund managers would exhibit similar selection ability for exposed and unexposed stocks after the ATT.

Appendix Table D.2 presents regression results based on the average picking ability for each group, Picking^{mean}, which isolates the intensive margin. The estimate of β_1 is positive and statistically significant at 10%, indicating that the superior picking ability for exposed stocks prior to ATT is not solely driven by their greater representation in portfolios. Rather, fund managers are more skilled at selecting the average exposed stock compared to the average non-exposed stock. Importantly, the magnitude of that advantage is smaller at the intensive margin (approximately half a quartile). In contrast, the decline in picking ability is even more pronounced for the subset of exposed stocks that managers continue to hold post-ATT (more than a quartile).[15] The fact that total picking ability is relatively higher than average picking ability for exposed stocks pre-ATT and it deteriorates less post-ATT further supports the reallocation interpretation: fund managers actively increase their allocation to exposed stocks when they have a competitive advantage in picking them and drop them from their portfolios following the implementation of ATT, when that competitive advantage disappears. That interpretation is consistent with the model predictions in both Kacperczyk et al. (2016) and Dessaint et al. (2024), where capacity constrained investors optimally decide which information to pay attention to (and hence which stocks to invest in) based on its marginal value or its cost.

Next, we evaluate the timing of the decline in fund stock-picking ability. To do so, we estimate a dynamic version of Equation (6), replacing ATT_t with a series of monthly time indicators spanning the 9 months before and after April 2021. Specifically, we include indicators $\mathbb{1}\{t = k\}$ for each month $k \in \{-9, -8, \ldots, 9\}$, where $t = 0$ corresponds to the month of the ATT and serves as the reference category.

$$\text{Picking}^{s,h}_{fte} = \alpha + \sum_{k=-9,\ k\neq 0}^{9} \delta_k \cdot \mathbb{1}\{t = k\} \cdot \text{Exposed}_{fe} + \Gamma_1 X'_{ft} + \delta_f + \lambda_t + \varepsilon_{fte}. \quad (7)$$

Figure 4 visualizes the estimated coefficients δ_k for $k \in \{-9, -8, \ldots, 9\}$. A sharp decline is observed at month $t = -4$, which is expected, as the six-month forward-looking idiosyncratic

[15] From $\hat{\beta}_1 = 0.019$ to $\hat{\beta}_1 + \hat{\beta}_2 = 0.019 - 0.067 = -0.048$.

return used to compute $\text{Picking}_{fte}^{s,h}$ begins to incorporate post-ATT months. Specifically, at $t = -4$, two post-ATT months are included in the future return window, during which firms experienced a negative and unanticipated impact on their stock market performance (Bian et al., 2021). The decline in picking ability continues as the share of post-ATT months increases in the forward-looking return window. By $t = 0$ (April 2021), fund portfolio rebalancing becomes directly affected by the diminished quality of mobile-generated signals, leading to a further drop in stock-picking performance.[16] This pattern reveals the deterioration in fund performance due to the lower informativeness of mobile-generated signals. Moreover, the persistence of the deterioration in picking ability beyond $t = 2$ suggests that fund managers were unable to fully substitute the degraded mobile-generated signal with alternative data sources.

Finally, we examine whether the decline in stock-picking ability is concentrated at shorter horizons. Dessaint et al. (2024) show that alternative data tends to be oriented toward short-term signals. This implies that the impact of the ATT on fund stock-picking ability should be most pronounced at near-term horizons.

To test this prediction, we estimate Equation (6) across a range of horizons h, spanning from 1 to 36 months ahead in 3-to-6-months increments, using the same specification as Column 4 of Table 6. We extend this analysis to the full sample period from 2017 to 2023 in order to ensure sufficient coverage for longer-horizon signals. The coefficients of interest, $\hat{\beta}_2$, estimated from each of the 12 regressions, are plotted in Figure 5, where the x-axis denotes the horizon of the forward-looking idiosyncratic return used to construct the picking measure. Each line in the figure represents cumulative abnormal returns estimated under a different factor model: the CAPM, the Fama-French-Carhart (FFC) four-factor model, and the Fama-French 6-factor model.

Since longer-horizon future returns embed shorter-horizon performance, we expect the estimated coefficients to be negative at all horizons, unless fund managers can fully offset

[16] Our regression specification in Equation (6) does not adjust for the decline observed during the pre-ATT period, and therefore likely underestimates the true effect observed for the period during months [-4,2].

the effect by significantly improving their ability to select stocks with strong long-term performance. This is precisely what we find: $\hat{\beta}_2$ is negative and statistically significant across all horizons. Moreover, the magnitude of $\hat{\beta}_2$ decreases monotonically up to the 12-month horizon, it further decrease with a smaller slope up to 24 months, and then flattens out. This pattern is consistent with the interpretation that mobile-generated signals are most informative about short-term stock performance (i.e., within a 12-month window).

6.2 Sell-side Analysts

Earnings Forecast Errors We now turn to testing Hypothesis 2b, focusing on the forecast errors of financial analysts. Compared to the analysis of fund managers, examining analyst forecasts offers two advantages. First, unlike fund trading, particularly that of hedge funds, 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 effects of the ATT and isolate variation attributable to differential analyst usage of alternative data.

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.[17]

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

We then run the following regressions:

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

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 4.3. 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 control for changes in firms' earnings volatility around ATT, which may affect forecast accuracy. This helps isolate the effect of varying degrees of alternative data usage across analysts. Additionally, we include analyst-firm fixed effects to account for the average forecasting 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 the 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 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, indicating that analysts who use alternative data experience an increase in forecast errors following the ATT.

Table 7 reports regression results for analyst forecasts of firms in the *App Sample*. 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, regardless of the inclusion of control variables or

fixed effects.

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 20% increase in the share of analyst reports referencing any of app keywords is associated with a 0.047 increase in PMAFE (0.233×0.20), which would move an analyst at the median of the forecast error distribution to approximately the 57^{th} percentile post-ATT. 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 which doesn't contain firms with major apps. That choice 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 statistically indifferent from zero across specifications. That indicates that the forecasts 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.

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 the ATT ("switchers"), defined as those with an above-median decline in keyword usage intensity from the pre-ATT to post-ATT period, to the "non-switchers". Regression results reported in Appendix Table E.3 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, denoted $\mathbb{1}\{t = k\}$ for $k \in [-4, 4]$. We further define $\mathbb{1}\{t < -4\}$ and $\mathbb{1}\{t > 4\}$ to capture 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 period.

$$\begin{aligned}\text{PMAFE}_{ijt} = \alpha + \delta_{-5} \cdot \mathbb{1}\{t < -4\} \cdot \text{Keyword usage}_j + \sum_{k=-4, k\neq 0}^{4} \delta_k \cdot \mathbb{1}\{t = k\} \cdot \text{Keyword usage}_j \\ + \delta_5 \cdot \mathbb{1}\{t > 4\} \cdot \text{Keyword usage}_j + \Gamma X_{jt} + \theta_{it} + \iota_{ij} + \kappa_d + \varepsilon_{ijt} \quad (9)\end{aligned}$$

Figure 6 visualizes the estimated δ_k coefficients. Forecast errors begin to rise in the second quarter following the ATT and remain elevated through the third and fourth quarters. Importantly, no evidence of a pre-trend is observed, supporting the parallel trends assumption.

The results from the dynamic DiD specification and the placebo test strengthen our interpretation that the ATT causally reduced the predictive value of alternative data used by analysts.

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 the ATT on report-level market reactions by estimating the following regression:

$$CAR_{ikt}^{adj,\ 3d} = \alpha + \beta_1 \cdot \text{Keyword per page}_{kt} + \boldsymbol{\beta_2} \cdot \text{ATT} \times \text{Keyword per page}_{kt} + \Gamma X_{kt} + \theta_{it} + \iota_{br,i} + \kappa_d + \epsilon_{ikt}, \tag{10}$$

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. 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 are available on LSEG within four days of the report issuance date, the last requirement ensuring that the event-specific market response captures the full effect of the report. 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.

The estimation results are presented in Table 8, 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 the ATT, analyst reports citing more app keywords triggered stronger market reactions. In contrast, following the ATT, there is no longer a differential market response based on keyword intensity. Control variables behave as expected: more expensive reports and shorter reports (conditional on price) are associated with larger market reactions.

7 Test of Hp 3: ATT Worsens Firms' Information Environment

As ATT negatively affects the information production by analysts and limits fund managers' ability to trade prices toward fundamentals, we expect price efficiency to deteriorate at the firm level. In this section, we test hypothesis 3; i.e., that price efficiency decreases post-ATT, particularly so for firms followed by market participants who rely on mobile-generated data.

We start by analyzing changes in short-term price efficiency, before and after ATT. Motivated by theory (e.g., Kim and Verrecchia (1991), Demski and Feltham (1994)), we expect greater price reactions to earnings announcements when the information released in the announcement is novel, and not anticipated by sophisticated market participants using private information, such as mobile-generated alternative data. We follow the approach by Zhu (2019) to assess the information content of earnings announcements by analyzing changes in the slope coefficient of cumulative abnormal returns (CAR) and surprise earnings (SUE), pre- and post-ATT. Looking at that slope, rather than at changes in abnormal returns per se, allows us to control for differences in public information. We measure SUE as the difference between actual earnings and consensus earnings forecast, scaled by stock price, as follows:

$$SUE_{it}^{concensus} = \frac{EPS_{i,t} - EPS_{it}^{concensus}}{P_{it}}. \tag{11}$$

The relation between CAR and SUE should then denote the pre-announcement information set of privately-informed investors relative to that of unsophisticated investors, basing their estimates on consensus forecast.[18] We further examine whether the slope depends on whether firms are followed by financial analyst specialized in the use of alternative data. More specifically, we run the following regression:

$$\begin{aligned}\text{CAR}_{it} = \alpha + \beta_1 \cdot \text{SUE}_{it}^{concensus} + \beta_2 \cdot \text{ATT}_t \times \text{SUE}_{it}^{concensus} + \beta_3 \cdot \text{Exposure}_i \times \text{SUE}_{it}^{concensus} + \\ \beta_4 \cdot \text{ATT}_t \times \text{Exposure}_i \times \text{SUE}_{it}^{concensus} + \beta_5 \cdot \text{Exposure}_i + \beta_6 \text{ATT}_t \times \text{Exposure}_i + \\ \Gamma_1 X_{it-4} + \Gamma_2 X_{it-4} \times \text{SUE}_{it} + \theta_{sic2,t} + \epsilon_{it}, \quad (12)\end{aligned}$$

where i indicates firms, t year-quarters, CAR is the cumulative abnormal return relative to either CAPM or FFC 6-factor model over trading days [0,+10] post earning announcements; $Exposure$ is an indicator variable equal to 1 if the firm is followed by at least one analyst specialized in alternative data in quarter t, using the two proxies defined in Section 4.3. SUE is measured as in Equation (11), X_{it-4} is a vector of control variables measured in quarter $t-4$, namely size, and cash, tangibles, debt-to-assets ratio and market-book ratio; $X_{it-4} \times SUE_{it}$ captures the differential impact of firm characteristics on the slope estimates. We include industry-by-quarter ($\theta_{sic2,t}$) fixed-effects, and double-cluster standard errors at the industry and quarter level.

The coefficient of interest is β_4, which captures the differential sensitivity of CAR to SUE post-ATT for firms followed by analysts specialized in alternative data, relative to firms that are not. We expect a positive and significant $\hat{\beta}_4$, indicating a stronger price reaction to earnings surprises and a deterioration in the informativeness of forecasts based on mobile-generated alternative data following the ATT.

Table 9 reports regression results across various exposure definitions, sets of control vari-

[18] Post-ATT, the median forecast, particularly those issued by specialized analysts, may exhibit a lower signal-to-noise ratio, but the CAR–SUE sensitivity still captures the extent to which private information, relative to that signal, is priced in. A higher CAR–SUE sensitivity post-ATT would imply that less private information from sophisticated investors is incorporated into pre-announcement prices.

ables, and CAR construction methods. *Exposure* is based on coverage by specialized analysts in more than 50% of quarters prior to the ATT. In Columns 1–4, specialized analyst is defined as an indicator equal to 1 if they write any analyst report referencing app keywords. In Columns 5–8, specialized analyst are defined as those focused on retail and service industries (SIC 5200–5999 or 7000–8999). As discussed in Section 4.3, the keyword-based exposure measure captures only a subset of analysts who issue reports alongside their forecasts. Therefore, we additionally consider an analyst specialization measure based on industry focus to broaden coverage in our firm-level analysis.

Across specifications, $\hat{\beta}_4$ is always positive. To assess economic significance, consider the most comprehensive specification in Column 8, which uses industry focus to measure analyst specialization and examines CAR measured in excess of a Fama-French 6-factor model. The baseline slope pre-ATT is 1.838 ($\hat{\beta}_1$, significant at 1% level), indicating that a one-percentage-point increase in SUE is related to an increase in CAR of 1.838 p.p. For firms followed by analysts specialized in alternative data, pre-ATT, that slope is -0.423 p.p. smaller ($\hat{\beta}_3 = -0.423$, significantly at 5% level). That indicates a better information environment for firms with analysts specialized in alternative data pre-ATT. Post-ATT, there is no difference in the reaction of CAR to SUE for firms that are not followed by analysts specialized in alternative data ($\hat{\beta}_2$, not significant). For firms followed by analysts specialized in alternative data, instead, post-ATT, a one-percentage-point increase in SUE is linked to an increase in CAR of 0.683 p.p, offsetting the baseline advantage in information environment.[19] That indicates a large deterioration in the information environment of firms followed by analysts specialized in alternative data post-ATT.

Results are similar when we use keyword usage to measure analyst specialization in Columns 1–4, with the exception that $\hat{\beta}_3$ is not statistically significant in any specification. This may be driven by the partial coverage of analysts in the LSEG report sample, which only includes those who issue reports in addition to forecasts. We additionally conduct a

[19] We run the following test: $H_0 : \hat{\beta}_2 + \hat{\beta}_4 = 0$, which cannot be rejected at 10%.

placebo test, by repeating the analysis in Appendix Table 9 in the *No App Sample*, i.e., including all firms that do not possess a mobile app. In none of the specifications $\hat{\beta}_4$ is statistically significant. That indicates that the information environment of firms followed by analysts specialized in alternative data does not deteriorate post ATT, if those firms do not possess a mobile app.

While the CAR-SUE sensitivity analysis sheds light on *short-term price efficiency* by capturing how strongly market prices react to earnings surprises during a short window, it primarily reflects the short-term informativeness of the earnings release relative to the consensus forecast. To more comprehensively assess the quality of the information environment for exposed firms post-ATT, we complement this analysis with two additional measures: bid-ask spreads and post-earnings announcement volatility (with residual returns estimated either based on CAPM or Fama-French 6-factor models), as defined in Section 4.4. These measures allow us to examine, respectively, the degree of *ex ante information asymmetry* and the extent of *post-disclosure residual uncertainty* in how markets process earnings releases. To do so, we estimate the following specification:

$$\text{Info Friction}_{it} = \alpha + \beta_1 \cdot \text{Exposed}_i + \beta_2 \cdot \text{ATT}_t \times \text{Exposed}_i + \Gamma_1 X_{it-4} + \theta_{sic2,t} + \epsilon_{it}, \quad (13)$$

where i indexes firms and t indexes year-quarters. We include industry-by-quarter fixed effects and cluster standard errors at the industry and quarter levels.

The results, presented in Table 10, show that firms with exposure to specialized analysts exhibit lower information asymmetry prior to ATT ($\hat{\beta}_1 < 0$). However, this advantage is more than reversed following ATT ($\hat{\beta}_1 + \hat{\beta}_2 > 0$, significant in Columns 1–4), consistent with increased uncertainty, investor disagreement, and greater opacity in the firm's information environment. One caveat to this analysis is that we cannot fully rule out the possibility that the observed effect is partly driven by a direct impact of ATT on firms' earnings volatility, as discussed in Section 3. Nevertheless, the overall pattern of results supports the interpretation

that ATT, by reducing the quantity and quality of alternative data, impaired the information environment of firms.

8 Conclusion

The rise of privacy regulations, such as Apple's App Tracking Transparency (ATT) policy, presents a tradeoff between protecting individual privacy and preserving the flow of alternative data critical to financial markets. Using ATT as a natural experiment, our study provides the first empirical evidence that privacy-driven data restrictions can significantly weaken the information environment by raising forecast errors, reducing price informativeness, and impairing capital market efficiency.

These findings underscore the need for policymakers to carefully data privacy with market transparency. As alternative data becomes increasingly embedded in financial decision-making, regulations should aim to safeguard privacy without unduly disrupting access to informative signals. Future research should examine the long-run effects of such policies, the distributional consequences across investor types, and technological solutions that preserve both data utility and privacy in an evolving digital economy.

References

AlternativeData.org (2019). Alternative data stack. https://alternativedata.org/alternative-data-stack/. Accessed August 04, 2024.

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 (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. Technical report, National Bureau of Economic Research.

Bae, D., R. Mayya, and T. Nian (2023). Privacy regulation and its unintended consequence on consumption behaviors: Evidence from CCPA.

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*.

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

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

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

Bonelli, M. and T. Foucault (2023). Displaced by big data: Evidence from active fund managers. *Available at SSRN 4527672*.

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

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*.

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.

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.

Demski, J. S. and G. A. Feltham (1994). Market response to financial reports. *Journal of Accounting and Economics 17*(1-2), 3–40.

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.

Evans, R. B. (2010). Mutual Fund Incubation. *The Journal of Finance 65*(4), 1581–1611. Publisher: [American Finance Association, Wiley].

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.

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. Johnson, and S. Shriver (2019). Regulating privacy online: The early impact of the GDPR on european web traffic & e-commerce outcomes. *Available at SSRN 3421731*.

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.

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

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. Stubben, and K. Ton (2022). Customer shopping behavior and the persistence of revenues and earnings. *Available at SSRN 3744417*.

Johnson, G. (2022). Economic research on privacy regulation: Lessons from the GDPR and beyond.

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

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved Actions of Mutual Funds. *The Review of Financial Studies 21*(6), 2379–2416. Publisher: [Oxford University Press, Society for Financial Studies].

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

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*.

Kim, O. and R. E. Verrecchia (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research 29*(2), 302–321.

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

Kraft, L., B. Skiera, and T. Koschella (2023). Economic impact of opt-in versus opt-out requirements for personal data usage: the case of apple's app tracking transparency (att). *Available at SSRN 4598472*.

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.

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.

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.

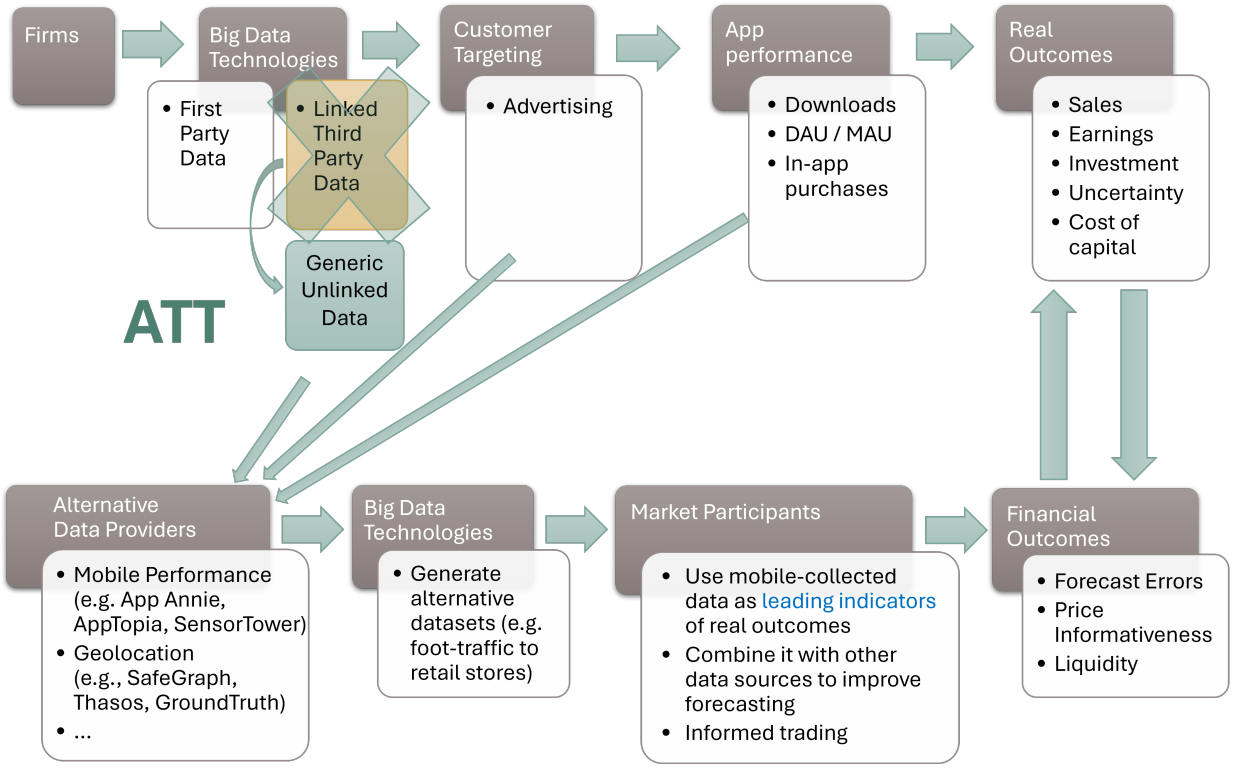
Statista (2024b). Number of available apps in the apple app store from 2008 to july 2023. https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/. Accessed August 02, 2024.

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.

Tingley, B. (2021). Fisd alternative data council: A guide to alternative data. Technical report, Greenwich Associates.

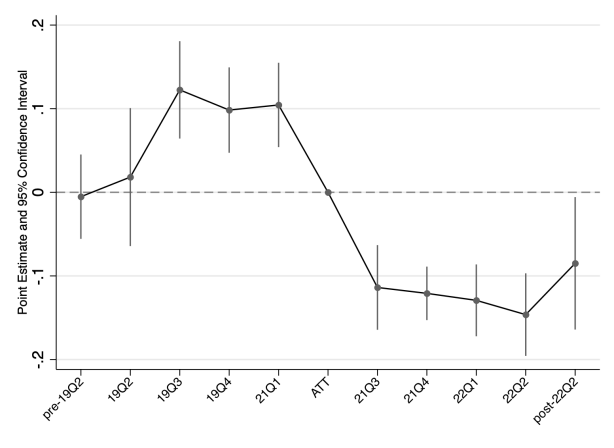
Zhu, C. (2019). Big data as a governance mechanism. *The Review of Financial Studies 32*(5), 2021–2061.

Figure 1: Illustration for the Mechanism underlying ATT's Impact

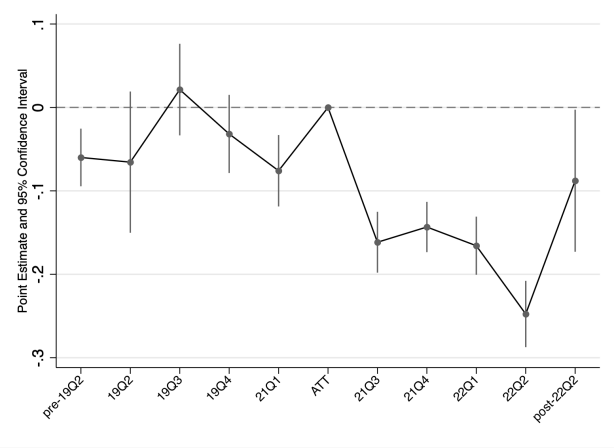


NOTE.— Figure 1 graphically summarizes the logical links in our study. Firms collect user data through apps, either as first-party data—such as browsing and purchase activity collected directly within the app—or as third-party data, including location information obtained from external sources. This data may be shared with third parties, such as data brokers or platforms like Google Analytics, in exchange for payment or services like analytics and targeting. Firms, either directly or through partners like Google Analytics, analyze both first-party and third-party data—often using AI models—to gain insights about their users and improve targeting. A key targeting tool at firms' disposal is mobile advertising. Customer targeting then drives traffic (e.g., downloads, monthly active users or MAU, daily active users or DAU) to a firm's app. The quality of that traffic is linked to customer retention and, ultimately, to firm performance. Alternative data providers collect user-level signals—such as app usage, transactions, and geolocation—from firms or data brokers and transform them into structured datasets for investors. These signals inform expectations about firm fundamentals and drive trading, influencing asset prices and, ultimately, firm decisions such as investment and financing.

Figure 2: App Downloads and Firm Performance around ATT



(a) Sales Growth



(b) Random Walk-based SUE

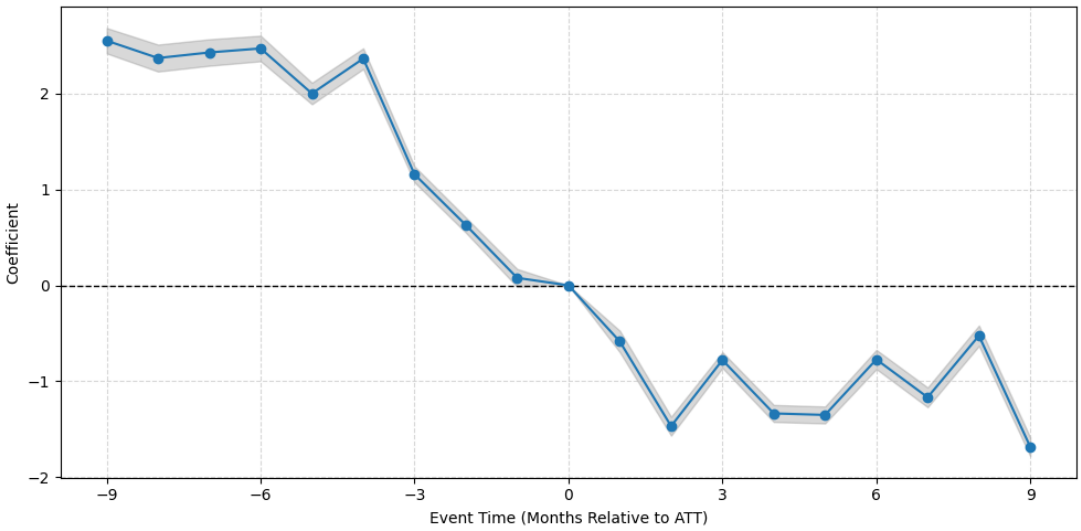
NOTE.— Figure 2 plots the estimated coefficients for the quarterly time dummies from Equation (3). Dummy variables are included for each quarter from 2019Q2 through 2022Q2, while quarters before 2019Q2 and after 2022Q2 are grouped into respective indicators. The reference category is 2021Q2, corresponding 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: CAR of a Long-Short Portfolio Sorted on App Downloads



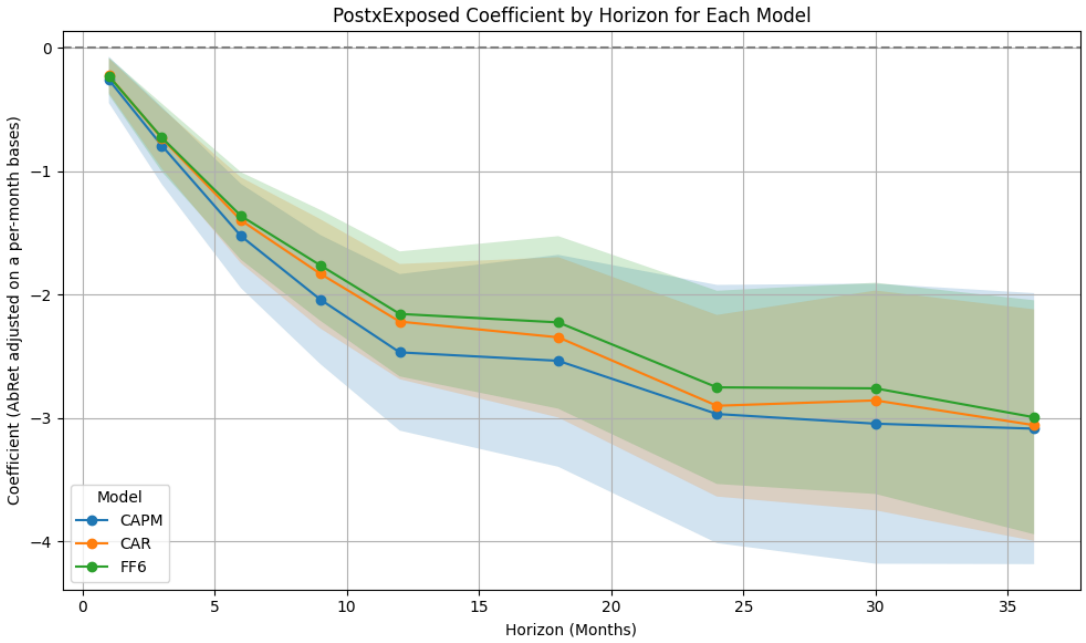
NOTE.— Figure 3 plots the cumulative abnormal returns (CAR) for the long-short portfolio over our sample period, using different portfolio weighting methods: (1) value-weighted portfolios sorted by downloads (blue solid), (2) equal-weighted portfolios sorted by downloads, excluding micro-caps (green dotted), and (3) MAU-weighted portfolios sorted by downloads, excluding micro-caps (yellow dashed). The vertical line marks the month of the ATT announcement. Each month, we sort stocks into quintiles based on their abnormal downloads from the previous month, defined as the log of downloads in that month minus the average log of downloads over the preceding twelve months. We then construct a long-short portfolio that buys the top quintile of stocks and short-sells the bottom quintile each month. Finally, for each month, we compute the risk-adjusted returns of the long-short portfolio, using the risk exposure estimated over the past twelve months. The risk adjustment is based on the Fama-French 6-factor model.

Figure 4: Fund Picking Ability at 6-month Horizon Around ATT



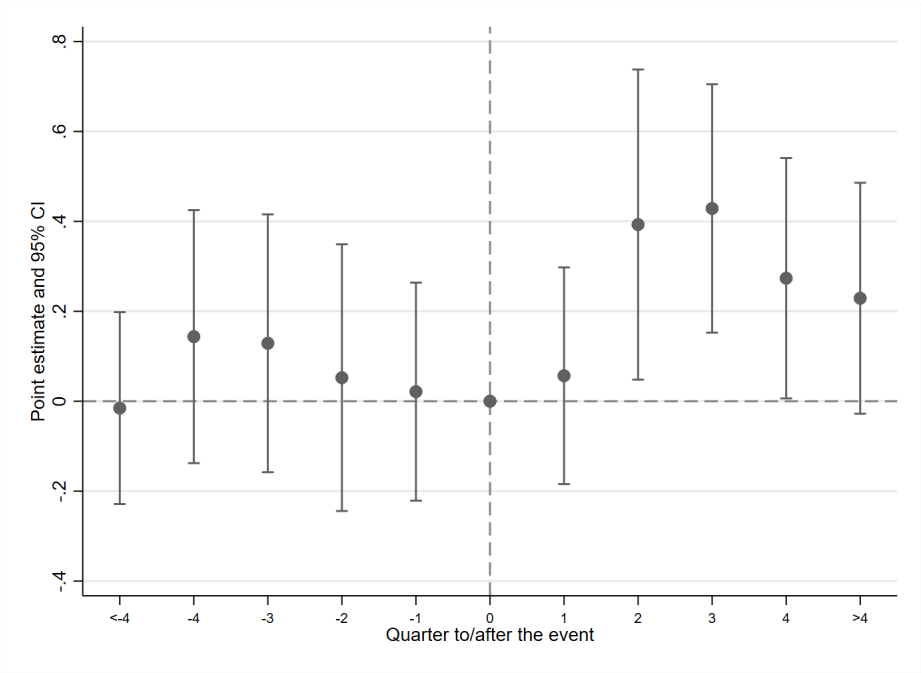
NOTE.— Figure 4 plots the estimated coefficients of *Exposed* interacted with monthly time dummies, as specified in Equation (7). The sample period spans from July 2020 to January 2022, covering nine months before and after the ATT month of April 2021. We include dummies for each month from $t = -9$ to $t = 9$. The reference period is $t = 0$, corresponding to the month of ATT implementation (April, 2021). The picking measure is constructed using a 6-month forward-looking CAR estimated using Fama-French 6-factor model.

Figure 5: Fund Picking Ability at Various Horizons Around ATT



NOTE.— Figure 5 displays the estimated coefficients of *Exposed* × *ATT* as specified in Equation (6). Each dot represents an estimate based on a different *picking* measure, constructed at horizons ranging from 1 to 36 months ahead in increments of 3-to-6 months. The variable *ATT* equals 1 beginning in the month following the ATT implementation in April, 2021.

Figure 6: Analyst Forecast Errors and Usage of Mobile Signals Around ATT



NOTE.— Figure 6 plots the estimated coefficients of the time dummies, as specified in Equation Equation 9. We include dummies for each quarter from $t = -4$ to $t = 4$, with $t \geq 4$ ($t \leq -4$) representing all quarters before quarter $t = -4$. The reference group is $t = 0$, corresponding to the quarter of ATT implementation (2021Q2).

Table 1: Summary Statistics — Firms

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE (randow walk)	-0.020	6.05	-1.98	-0.38	0.09	0.53	2.00	17,854
SUE (concensus)	0.092	2.14	-0.30	-0.01	0.07	0.26	0.76	16,543
CAR - CAPM	-0.053	10.77	-11.80	-5.20	0.00	4.95	11.62	18,425
CAR - FF6	-0.043	10.56	-11.51	-5.01	0.00	4.72	11.34	18,425
Sales growth	8.586	30.36	-14.29	-1.76	6.88	18.52	36.11	17,553
Size	22.808	2.13	20.14	21.29	22.76	24.24	25.65	18,471
Cash/Assets	18.294	18.72	1.90	4.59	11.66	25.30	46.72	18,462
Tangibles/Assets	20.512	21.68	1.21	4.29	11.67	30.24	57.23	18,126
Debt/Assets	32.102	24.32	3.21	12.33	29.81	45.67	63.31	18,298
Market-to-Book	1.739	7.27	0.04	0.14	0.35	0.78	1.49	18,445
Bid-ask spread	3.523	1.95	1.68	2.16	2.98	4.31	6.18	15,428
Post-EA vol. (CAPM)	2.055	1.48	0.84	1.12	1.62	2.48	3.72	15,373
Post EA vol. (FF6)	1.876	1.38	0.77	1.02	1.48	2.25	3.36	15,373
Specialized analysts (keyword usage)	0.635	0.48	0.00	0.00	1.00	1.00	1.00	18,512
Specialized analysts (retail & service focus)	0.375	0.48	0.00	0.00	0.00	1.00	1.00	18,512
Downloads	12.607	2.52	9.47	10.88	12.58	14.24	15.83	17,068
MAU	12.860	2.66	9.62	11.12	12.85	14.51	16.16	17,068
Engagement	0.269	0.12	0.12	0.18	0.26	0.34	0.40	17,068

Panel B. No-App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE (randow walk)	0.918	11.97	-3.83	-0.79	0.11	1.19	5.25	84,697
SUE (concensus)	-0.034	3.36	-1.35	-0.22	0.07	0.42	1.43	65,367
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	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. (CAPM)	2.741	1.99	0.99	1.37	2.15	3.47	5.17	74,944
Post EA vol. (FF6)	2.582	1.92	0.92	1.27	1.99	3.24	4.93	74,944
Specialized analysts (keyword usage)	0.245	0.43	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

NOTE.— 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 — 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

NOTE.— Panel A includes all firms that have an app; Panel B include 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 3: Summary Statistics — Fund Stock Picking Ability

	mean	sd	p10	p25	p50	p75	p90	count
$Picking^{sum}$ (unexposed)	-0.204	2.587	-3.450	-1.217	-0.022	1.015	2.534	28,134
$Picking^{sum}$ (exposed)	-0.431	1.925	-2.573	-1.463	-0.429	0.571	1.728	28,134
$Picking^{mean}$ (unexposed)	-0.003	0.136	-0.112	-0.039	-0.000	0.032	0.095	28,134
$Picking^{mean}$ (exposed)	-0.016	0.129	-0.105	-0.044	-0.007	0.013	0.061	28,134

NOTE.— This table reports summary statistics on fund stock-picking ability, aggregated based on stock exposure to the ATT. We use 6-month future idiosyncratic returns (estimated using FF 6-factor model) to construct the picking measure. The sample period spans from July 2020 to January 2022, covering nine months before and after the ATT month of April 2021.

Table 4: 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 × Downloads	-0.101***	-0.123***	-0.064***	-0.090***
	(0.04)	(0.04)	(0.02)	(0.02)
Constant	4.870***	6.637***	5.305***	5.473***
	(0.39)	(0.76)	(0.11)	(0.28)
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

NOTE.— Table 4 presents estimation results from Equation (2) using the app sample. The dependent variable is *Sales Growth* in Columns 1–2 and *SUE*, i.e., earnings surprises based on a random walk model, calculated as in Equation (1), in Columns 3–4. 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 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 5: App Downloads and Changes in Portfolio Weights Around ATT

Direction of Trades	Positive	Negative	Positive	Negative
	(1)	(2)	(3)	(4)
L1.Abnormal Downloads	0.0043**	0.0007	0.0040**	0.0001
	(0.0017)	(0.0014)	(0.0017)	(0.0014)
L2.Abnormal Downloads	0.0015	-0.0000	0.0013	0.0001
	(0.0012)	(0.0018)	(0.0012)	(0.0017)
L3.Abnormal Downloads	-0.0012	-0.0009	-0.0010	-0.0013
	(0.0018)	(0.0015)	(0.0018)	(0.0015)
ATT × L1.Abnormal Downloads	-0.0051**	0.0012	-0.0041*	0.0018
	(0.0024)	(0.0025)	(0.0023)	(0.0026)
ATT × L2.Abnormal Downloads	-0.0015	-0.0020	-0.0015	-0.0020
	(0.0018)	(0.0019)	(0.0017)	(0.0020)
ATT × L3.Abnormal Downloads	0.0007	-0.0012	0.0006	-0.0002
	(0.0027)	(0.0022)	(0.0027)	(0.0022)
Constant	0.2039***	-0.2037***	0.2547***	-0.2712***
	(0.0001)	(0.0000)	(0.0888)	(0.0762)
Fund × Stock	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Fund Controls	N	N	Y	Y
Observations	2,861,868	2,879,601	2,782,026	2,793,741
R-sq	0.49	0.46	0.50	0.47

NOTE.— Table 5 presents the estimation results of Equation (4), where the outcome variable is the trade-induced weight changes for stock i, fund f, and month t are calculated as the difference between the actual and no-trade weights over the three-month window from $t-2$ to 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 three lags of abnormal downloads as explanatory variables, measured at the end of months $t-1$, $t-2$, and $t-3$, 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. The odd columns include fund × stock and month fixed effects. The even columns add time-varying fund characteristics (including size, flows, turnover, style exposures, and fees) along with their interactions with ATT. Standard errors are triple-clustered at the fund, month, and stock levels and are reported in parentheses. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% levels, respectively.

Table 6: Fund Picking Ability at 6-Month Horizon by Stock Exposure Around ATT

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
ATT × App-owning firms	-1.250***	-1.229***	-2.403***	-2.401***
	(0.35)	(0.35)	(0.32)	(0.32)
App-owning firms	-0.048	-0.058	0.910***	0.907***
	(0.24)	(0.24)	(0.29)	(0.29)
Constant	0.580***	19.437**	-0.198**	0.723
	(0.09)	(7.56)	(0.08)	(4.15)
Fund FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Fund Controls	N	Y	N	Y
Observations	57,784	56,256	57,784	56,256
R-sq	0.36	0.44	0.28	0.33

NOTE.— Table 6 reports the estimation results of Equation (6). The future idiosyncratic return used to construct the *Picking* is defined as the compounded abnormal return over the 6-month horizon $[t+1, t+6]$. The sample period spans from July 2020 to January 2022, covering nine months before and after the ATT month of April 2021. In Columns 1–2 (resp. 3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. Stock-level picking measures are aggregated to the fund level using simple averages. Exposed stocks are defined as those that operated an app for more than 50% of the quarters prior to the ATT. The variable *ATT* is an indicator equal to 1 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 and month fixed effects; the second adds time-varying fund characteristics (size, flows, turnover, style exposures, and fees) and the interaction terms between fund characteristics and *ATT*. 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.

Table 7: 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)
Constant	-0.025***	-0.086***	-0.030***	0.870
	(0.00)	(0.03)	(0.00)	(0.54)
Firm-Quater 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

NOTE.— Table 7 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 1 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, and brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimate-date fixed effects. Columns 3 and beyond 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 8: 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)
Constant	0.483***	0.200	0.488***	0.260
	(0.02)	(0.29)	(0.02)	(0.28)
Firm-Quater 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

NOTE.— Table 8 reports the estimation results of Equation (10) 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 computed direction-adjusted CAR during the three-day window following the report issuance date. In Columns 1–2 (resp. 3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. The variable *ATT* is an indicator equal to 1 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 (Regression controls include 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 9: Market Responses to Earnings Surprises Around ATT – by Firm Exposure to Specialized Analysts

Exposure based on	Keyword Usage				Retail & Service Focus			
	(1) CAPM	(2) CAPM	(3) FF6	(4) FF6	(5) CAPM	(6) CAPM	(7) FF6	(8) FF6
SUE (concensus)	0.536***	1.774**	0.570***	1.899**	0.609***	1.686**	0.626***	1.838*
	(0.17)	(0.76)	(0.19)	(0.84)	(0.15)	(0.79)	(0.16)	(0.91)
ATT × SUE (concensus)	-0.074	-0.199	-0.122	-0.234	-0.067	-0.173	-0.083	-0.181
	(0.16)	(0.22)	(0.19)	(0.23)	(0.15)	(0.18)	(0.16)	(0.18)
Exposure × SUE (concensus)	-0.023	-0.079	-0.101	-0.130	-0.248	-0.417**	-0.295*	-0.423**
	(0.16)	(0.28)	(0.17)	(0.28)	(0.19)	(0.16)	(0.17)	(0.19)
ATT × Exposure × SUE (concensus)	0.293	0.501	0.420**	0.635**	0.415**	0.589**	0.490**	0.683**
	(0.19)	(0.32)	(0.18)	(0.27)	(0.18)	(0.26)	(0.20)	(0.27)
Exposure	0.564*	0.720**	0.527	0.732**	0.449**	0.359	0.475***	0.394
	(0.31)	(0.32)	(0.31)	(0.32)	(0.19)	(0.37)	(0.15)	(0.32)
ATT × Exposure	-0.662	-0.928**	-0.739**	-0.979**	-0.583	-0.529	-0.486	-0.458
	(0.41)	(0.44)	(0.33)	(0.36)	(0.45)	(0.60)	(0.41)	(0.59)
Constant	-0.274**	2.118	-0.238*	2.795**	-0.142*	1.802	-0.179**	2.516*
	(0.12)	(1.49)	(0.14)	(1.28)	(0.08)	(1.48)	(0.08)	(1.31)
Industry-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
ATT × Controls	N	Y	N	Y	N	Y	N	Y
Observations	16,416	15,213	16,416	15,213	16,416	15,213	16,416	15,213
R-sq	0.111	0.121	0.106	0.116	0.111	0.120	0.106	0.115

NOTE.— Table 9 reports the estimation results of Equation (12) based on the report sample. The dependent variable, CAR, is measured in a 10-day window following earnings announcements, using either the CAPM or the Fama-French 6-factor model. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT implementation in 2021Q2. The variable *Exposure* is an indicator equal to 1 if the firm has any keyword-referencing reports (Columns 1–4) or is covered by specialized analysts (Columns 5–8) in more than 50% of the quarters prior to ATT. 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 pre-ATT. *SUE* is defined as the earnings surprise relative to the consensus forecast. All regressions include industry-quarter fixed effects. The even-numbered columns additionally 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, reported in parentheses, are double-clustered at the industry and quarter levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Information Friction Around ATT
– by Firm Exposure to Specialized Analysts

Exposure based on	Keyword Usage			Retail & Service Focus		
	(1) Bid-ask spread	(2) Post-EA vol. CAPM	(3) Post EA vol. FF6	(4) Bid-ask spread	(5) Post-EA vol. CAPM	(6) Post EA vol. FF6
ATT × Exposure	0.430***	0.350***	0.317***	0.724***	0.453***	0.396***
	(0.11)	(0.07)	(0.06)	(0.20)	(0.10)	(0.08)
Exposure	-0.212*	-0.177**	-0.166**	-0.229	-0.218*	-0.205*
	(0.11)	(0.08)	(0.08)	(0.17)	(0.13)	(0.12)
Constant	13.185***	7.947***	7.689***	13.326***	8.030***	7.770***
	(0.74)	(0.61)	(0.59)	(0.72)	(0.59)	(0.58)
Industry-Quarter FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
Observations	14,286	14,217	14,217	14,286	14,217	14,217
R-sq	0.621	0.470	0.453	0.622	0.469	0.453

NOTE.— Table 10 reports the estimation results of Equation (13) based on the app sample. We examine three measures of information friction: average bid-ask spread during the quarter, volatility of daily stock returns following earnings announcements, where we defined volatility as the root-mean-squared error of the residuals from a CAMP or the Fama-French 6-factor model. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT implementation in 2021Q2. The variable *Exposure* is an indicator equal to 1 if the firm has an above-median share of exposure to keyword-referencing reports (Columns 1–4) or is covered by specialized analysts (Columns 5–8) prior to ATT. 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. *SUE* is defined as the earnings surprise relative to the consensus forecast. 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, reported in parentheses, are double-clustered at the industry and quarter levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

A Variable Definition

Table A.1: Variable Definition

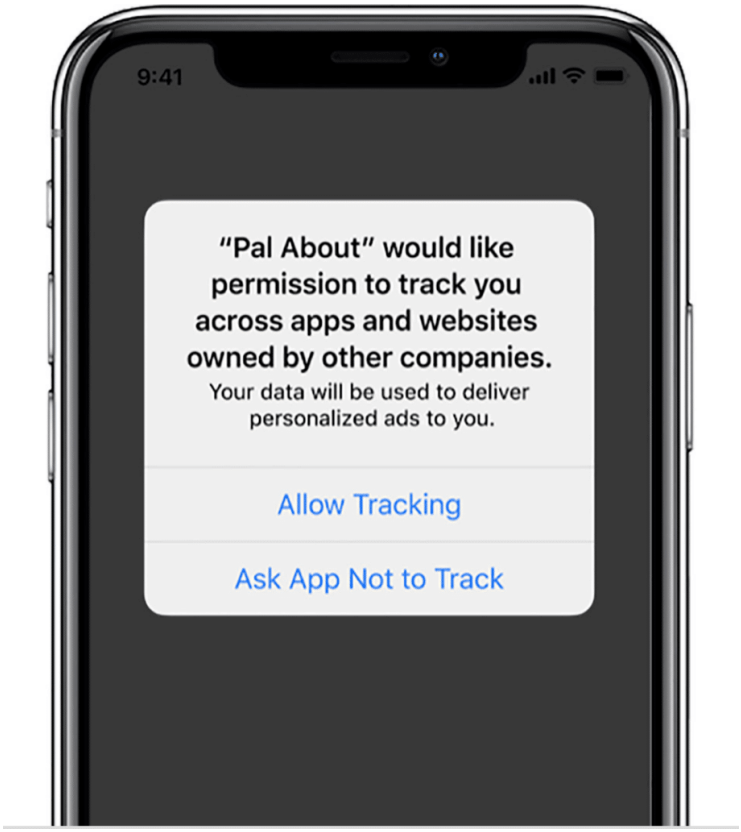
Variable Name	Definition
Firm-level Variables	
SUE (consensus)	The difference between actual earnings per share (EPS) and the consensus forecast, scaled by stock price. Values are expressed in percentage points and winsorized at the 1% level in both tails.
CAR – CAPM	Cumulative abnormal return in a 10-day 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 six-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 six-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 issuing forecasts for these industries in over the 90th percentile of analysts.
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 d 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 six-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	
$Picking^{sum}$ (unexposed)	Total stock-picking ability summed across unexposed stocks, where stock-level picking ability is defined as the product of 6-month forward-looking cumulative abnormal returns (estimated using the Fama-French six-factor model) and portfolio weight deviations from a benchmark. Unexposed stocks are those from firms that do not own an app in more than 50% of pre-ATT quarters. Variable is multiplied by 100.
$Picking^{sum}$ (exposed)	Total stock-picking ability summed across exposed stocks, using the same method as above. Exposed stocks are those from firms that own an app in more than 50% of pre-ATT quarters. Variable is multiplied by 100.
$Picking^{mean}$ (unexposed)	Average stock-picking ability across unexposed stocks, calculated as the mean of stock-level picking ability (as defined above). Variable is multiplied by 100.
$Picking^{mean}$ (exposed)	Average stock-picking ability across exposed stocks, calculated similarly. Variable is multiplied by 100.
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:

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

A Pop-up Notification Mandated by ATT

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



NOTE.–This figure 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 Scrapping, 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.

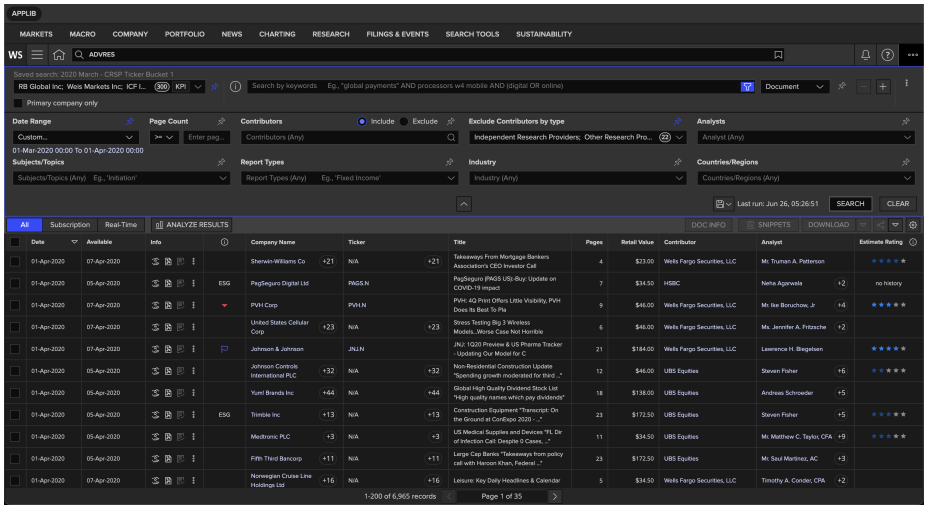
Scraping Metadata To build a complete index of analyst reports from LSEG Workspace, we then developed a dedicated scraping pipeline focused on extracting report-level metadata. While the keyword scraping pipeline targeted documents mentioning pre-specified terms, this

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, relying on Selenium in a headless Firefox environment to interact with the JavaScript-based and iframe/shadow DOM-heavy ADVRES interface. 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 behaviour 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 a 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

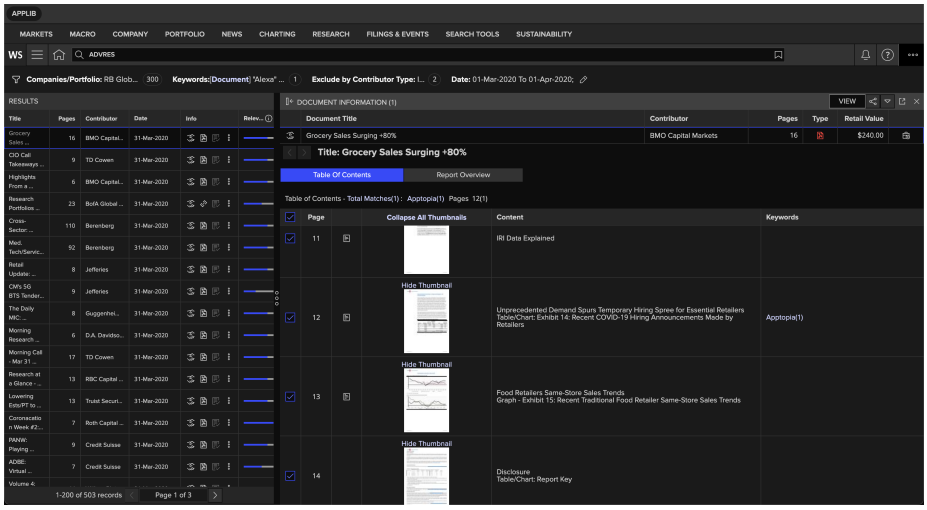
duplicate analyst entries for alternative broker-ESTIMID combinations. These adjustments were necessary because some analysts in LSEG were tied to outdated or unmatched broker IDs in I/B/E/S. By inspecting clusters of analysts repeatedly linked to the same mismatched ID, we reassigned them to more accurate 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



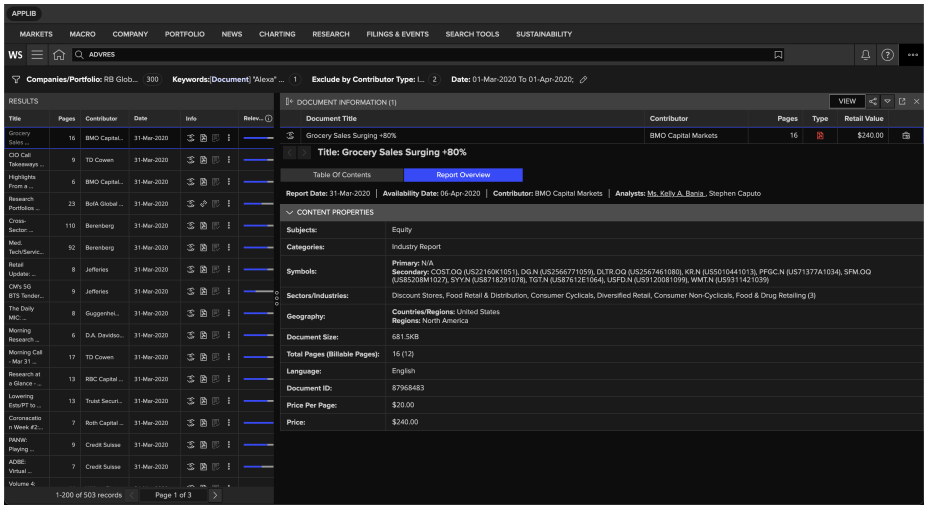
NOTE.–This figure shows the use of a saved search query configuration in the Advanced Research (ADVRES) portal, where a predefined set of 300 tickers is selected for scraping along with contributor-type and date range filters.

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



NOTE.–This figure 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



NOTE.–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

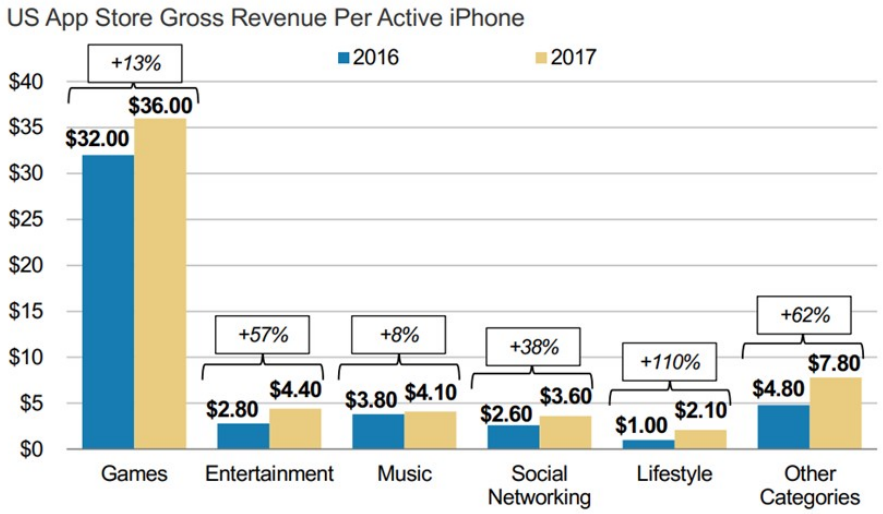
active user (8443)	alexa (10782)	app activity (340)	app annie (3578)
app data (2740)	app engagement (693)	app payment (391)	app performance (283)
app purchase (1091)	app tracking (970)	app usage (5138)	app use (184)
appannie (2009)	Apptopia (9818)	brand sentiment (495)	churn rate (7925)
comscore (19362)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

Note.–This table 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.

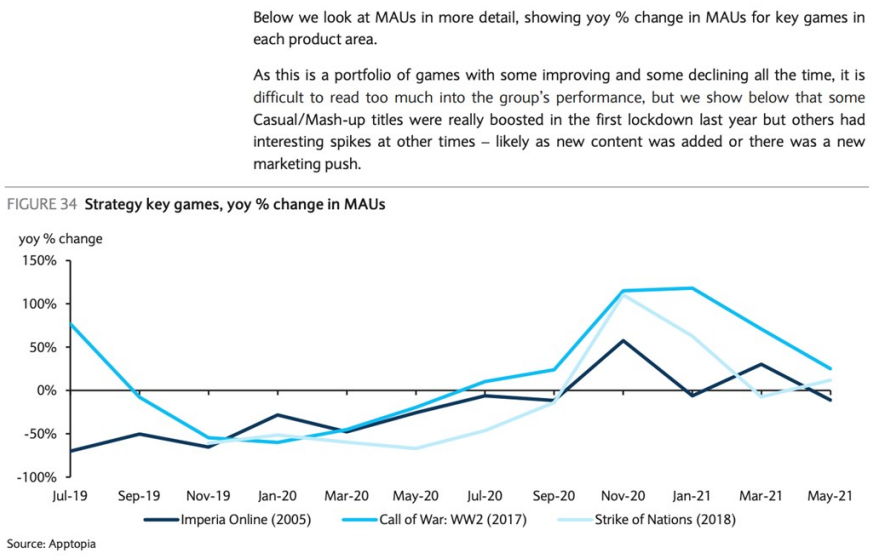


Source: Sensor Tower Store Intelligence, Tech Crunch, Morgan Stanley Research

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.

NOTE.–This excerpt 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

NOTE.–This excerpt 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.

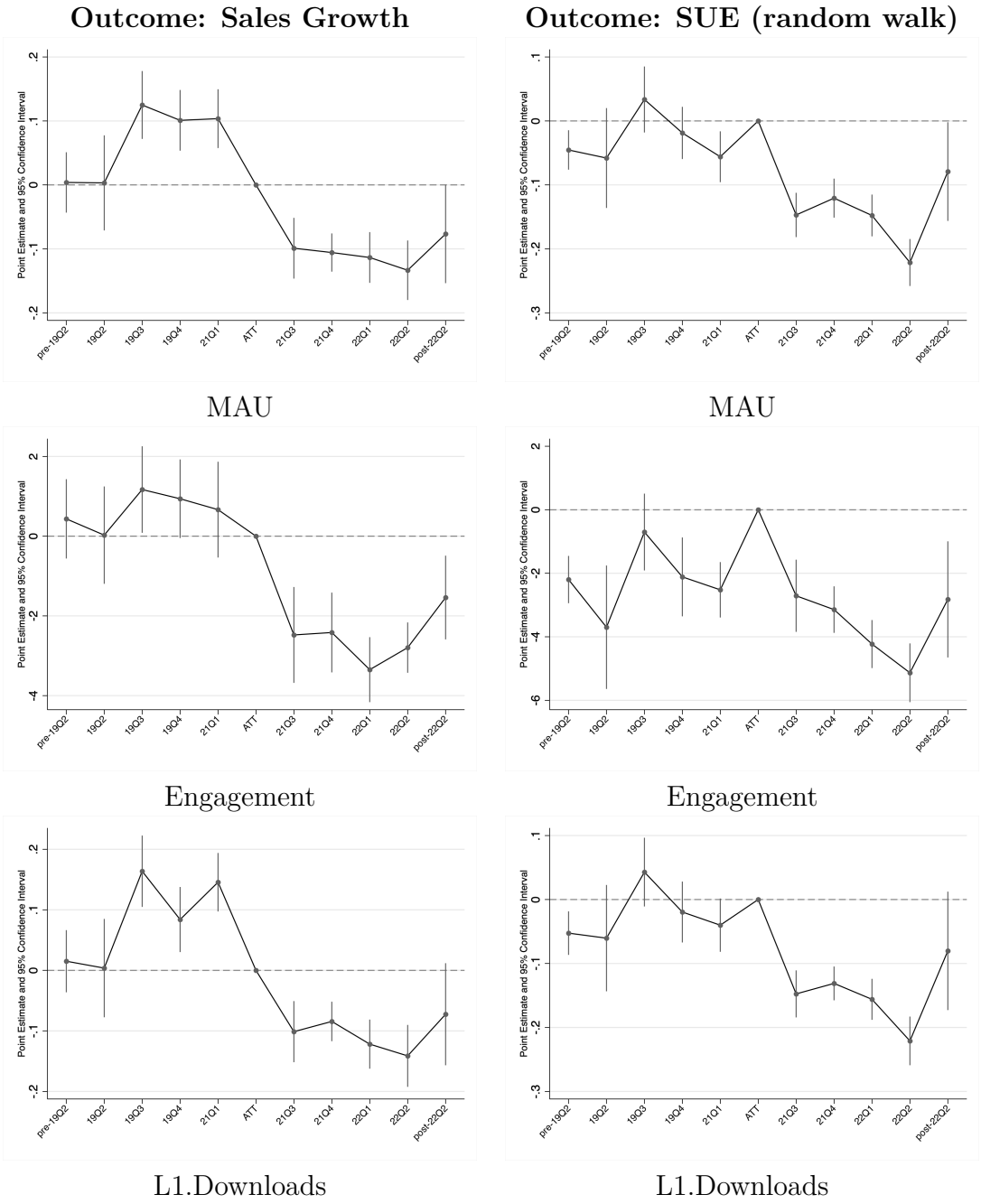
C Predictive Power of Alternative Mobile Metrics on Firm Performance

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

Deciles	Sales Growth			SUE (random walk)		
	(1)	(2)	(3)	(4)	(5)	(6)
MAU	0.093***			0.042***		
	(0.03)			(0.01)		
Engagement		1.571**			0.446	
		(0.57)			(0.42)	
L1.Downloads			0.112***			0.047**
			(0.04)			(0.02)
ATT × MAU	-0.108***			-0.066***		
	(0.03)			(0.01)		
ATT × Engagement		-3.049***			-0.220	
		(0.76)			(0.60)	
ATT × L1.Downloads			-0.145***			-0.065***
			(0.03)			(0.02)
Constant	6.700***	5.531***	5.803***	5.570***	5.273***	5.233***
	(0.79)	(1.01)	(1.03)	(0.27)	(0.21)	(0.23)
Industry-Quarter FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
Observations	15,534	6,888	6,805	15,508	6,855	6,789
R-sq	0.287	0.261	0.266	0.183	0.154	0.155

NOTE.— 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 number of DAU scaled by that of MAU. *L*1.*Downloads* is the natural logarithm of downloads from previous quarter. The indicator variable *ATT* The dependent variable is *Sales Growth* in Columns 1–3 and *SUE*, i.e., earnings surprises based on a random walk model, calculated as in Equation (1), in Columns 4–6. Both variables are expressed in deciles. *ATT* equals 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 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: Additional Mobile Metrics and Firm Performance Around ATT – MAU and Engagement



NOTE.–Figure 2 plots the estimated coefficients for the time dummy variables from Equation (**??**), where the mobile metrics are MAU, user engagement ratio, and L1.Downloads. *MAU* denotes the natural logarithm of total monthly active users for a firm's apps in a given quarter. *Engagement* is defined as the number of DAU scaled by that of MAU. *L*1.*Downloads* is the natural logarithm of downloads from previous quarter. 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. 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.

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 (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:

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

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 (e.g., 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:

$$\text{AS}_{fbt} = 0.5 \sum_i \left| w_{ift}^{\text{Fund}} - w_{it}^{b} \right|$$

Where:

- w_{ift}^{Fund} 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:

$$b_{ft}^* = \arg\min_{b \in \mathcal{B}} \text{AS}_{fbt}$$

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 Summary statistics and Robustness Checks

Table D.1: Summary Statistics: Fund Characteristics

	mean	sd	p10	p25	p50	p75	p90	count
Fund age	5.326	0.758	4.234	4.898	5.509	5.802	6.075	28,142
Fund size	6.236	1.975	3.608	4.797	6.213	7.619	8.849	28,142
Expense ratio	0.009	0.003	0.005	0.008	0.009	0.011	0.013	28,142
Turnover	0.543	0.462	0.130	0.230	0.420	0.700	1.110	28,142
Flow growth	-0.003	0.034	-0.028	-0.013	-0.005	0.003	0.020	28,142
Flow vol.	0.034	0.050	0.005	0.009	0.017	0.035	0.077	28,142
Style-mkt	1.009	0.118	0.873	0.938	1.003	1.073	1.156	28,142
Style-SMB	0.210	0.377	-0.197	-0.084	0.106	0.473	0.760	28,142
Style-HML	0.110	0.278	-0.262	-0.080	0.109	0.322	0.476	28,142
Style-MOM	-0.003	0.159	-0.199	-0.103	-0.008	0.088	0.199	28,142
Style-RMW	-0.077	0.257	-0.407	-0.225	-0.049	0.090	0.217	28,142
Style-CMA	-0.237	0.305	-0.658	-0.423	-0.207	-0.027	0.115	28,142

NOTE.— This table reports summary statistics on time-varying fund characteristics. The sample period spans from July 2020 to January 2022, covering nine months before and after the ATT month of April 2021.

Table D.2: Fund Picking Ability at 6-Month Horizon by Stock Exposure Around ATT – Average Across Stocks

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
App-owning firms	0.005	0.005	0.019*	0.018*
	(0.00)	(0.00)	(0.01)	(0.01)
ATT × App-owning firms	-0.029***	-0.028***	-0.067***	-0.065***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.011***	0.533*	-0.003	-0.070
	(0.00)	(0.28)	(0.00)	(0.23)
Fund FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Fund Controls	N	Y	N	Y
Observations	57,784	56,256	57,784	56,256
R-sq	0.170	0.220	0.180	0.200

NOTE.— Table D.2 reports the estimation results of Equation (6), where the outcome variable is the average picking ability $\text{Picking}_{fte}^{mean}$. The future idiosyncratic return used to construct the *Picking* is defined as the compounded abnormal return over the 6-month horizon $[t, t+6]$. In Columns 1–2 (resp. 3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. Stock-level picking measures are aggregated to the fund level using simple averages. Exposed stocks are defined as those that operated an app for more than 50% of the quarters prior to the ATT. The variable *ATT* is an indicator equal to 1 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 and month fixed effects; the second adds time-varying fund characteristics (size, flows, turnover, style exposures, and fees) and the interaction terms between fund characteristics and *ATT*. Standard errors are double-clustered at the fund and month levels. *t*-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table D.3: Fund Picking Ability at 6-Month Horizon by Stock Exposure Around ATT – CRSP Market Portfolio as Benchmark

Future Signal used in Picking	CAR - CAPM		CAR - FF6	
	(1)	(2)	(3)	(4)
App-owning firms	-0.165	-0.177	1.112***	1.108***
	(0.36)	(0.35)	(0.32)	(0.32)
ATT × App-owning firms	-1.085**	-1.065**	-2.684***	-2.683***
	(0.49)	(0.49)	(0.35)	(0.35)
Constant	0.635***	23.666**	-0.189*	1.700
	(0.12)	(8.28)	(0.09)	(4.51)
Fund FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Fund Controls	N	Y	N	Y
Observations	57,784	56,256	57,784	56,256
R-sq	0.330	0.420	0.290	0.350

NOTE.— Table 6 reports the estimation results of Equation (6). The future idiosyncratic return used to construct the *Picking* is defined as the compounded abnormal return over the 6-month horizon $[t, t+6]$, using the CRSP market portfolio as the benchmark portfolio. The sample period spans from July 2020 to January 2022, covering nine months before and after the ATT month of April 2021. In Columns 1–2 (resp. 3–4), abnormal returns are computed using the CAPM (Fama-French 6-factor) model. Stock-level picking measures are aggregated to the fund level using simple averages. Exposed stocks are defined as those that operated an app for more than 50% of the quarters prior to the ATT. The variable *ATT* is an indicator equal to 1 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 and month fixed effects; the second adds time-varying fund characteristics (size, flows, turnover, style exposures, and fees) and the interaction terms between fund characteristics and *ATT*. 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.

E Analyst-level analysis: Robustness Checks

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

	AFE/Absolute Actaul			
	(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-Quater 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

NOTE.— Table E.1 reports the estimation results of Equation (8) based on the app sample. The dependent variable is AFE (Absolute Forecast Error), the absolute forecast error, i.e., 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 1 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, and brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimate-date fixed effects. Columns 3 and beyond 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-Quater 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

NOTE.— 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 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 1 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, and brokerage size, along with their interactions with *ATT*. Columns 1–2 include firm-quarter and estimate-date fixed effects. Columns 3 and beyond 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 – 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-Quater 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

NOTE.— Table E.3 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-switchers' 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 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 1 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, and brokerage size, along with their interactions with ATT. Columns 1–2 include firm-quarter and estimate-date fixed effects. Columns 3 and beyond additionally include analyst-firm fixed effects. Standard errors are double-clustered at the firm and estimation month levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

F Firm-level analysis: Robustness Checks

Table F.1: Market Responses to Earnings Surprises – by Firm Exposure to Specialized Analysts, Placebo

Exposure based on	Keyword Usage				Retail & Service Focus			
	(1) CAPM	(2) CAPM	(3) FF6	(4) FF6	(5) CAPM	(6) CAPM	(7) FF6	(8) FF6
SUE (concensus)	0.459***	1.363**	0.460***	1.333**	0.445***	1.379**	0.452***	1.334**
	(0.07)	(0.65)	(0.07)	(0.62)	(0.07)	(0.64)	(0.07)	(0.61)
ATT × SUE (concensus)	0.017	0.049	0.007	0.041	0.054	0.085*	0.035	0.068
	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Exposure × SUE (concensus)	-0.120	-0.101	-0.070	-0.057	-0.030	-0.038	-0.011	-0.018
	(0.12)	(0.11)	(0.10)	(0.09)	(0.24)	(0.19)	(0.21)	(0.17)
ATT × Exposure × SUE (concensus)	0.154	0.159	0.097	0.107	-0.048	-0.025	-0.058	-0.037
	(0.16)	(0.16)	(0.14)	(0.14)	(0.22)	(0.18)	(0.20)	(0.15)
Exposure	0.312	0.304	0.426*	0.417*	0.301	0.294	0.227	0.221
	(0.27)	(0.27)	(0.24)	(0.24)	(0.29)	(0.30)	(0.32)	(0.32)
ATT × Exposure	-0.162	-0.171	-0.380	-0.388	-0.739	-0.752	-0.555	-0.568
	(0.33)	(0.35)	(0.33)	(0.34)	(0.70)	(0.72)	(0.65)	(0.67)
Constant	0.334	0.429	1.769	1.859*	-0.183	-0.062	1.155	1.269
	(1.35)	(1.35)	(1.06)	(1.08)	(1.34)	(1.34)	(1.09)	(1.08)
Industry-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
ATT × Controls	N	Y	N	Y	N	Y	N	Y
Observations	53,833	53,833	53,833	53,833	53,833	53,833	53,833	53,833
R-sq	0.073	0.075	0.065	0.066	0.073	0.075	0.064	0.066

NOTE.— Table F.1 reports the estimation results of Equation (12) based on the placebo firms, defined as firms that do not have an app. The dependent variable, CAR, is measured in 1 10-day window following earnings announcements, using either the CAPM or the Fama-French 6-factor model. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT implementation in 2021Q2. The variable *Exposure* is an indicator equal to 1 if the firm has any keyword-referencing reports (Columns 1–4) or is covered by specialized analysts (Columns 5–8) in more than 50% of the quarters prior to ATT. Keyword-referencing reports are defined as those mentioning any mobile-related keywords listed in **??**. 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. *SUE* is defined as the earnings surprise relative to the consensus forecast. All regressions include industry-quarter fixed effects. The even-numbered columns additionally 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, reported in parentheses, are double-clustered at the industry and quarter levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.