Data Privacy Regulations and Financial Information Production*

Simona Abis† Bo Bian‡ Huan Tang§

Abstract

This paper investigates the impact of privacy-motivated data sharing restrictions on financial markets, using Apple's App Tracking Transparency (ATT) policy as a laboratory. We show that ATT, by limiting the sharing of user data across apps, significantly reduces the predictive power of mobile-generated data, leading to increased forecast errors and decreased price informativeness. Firms reliant on third-party data experience a greater decline in information quality, which amplifies market forecast errors and reactions to earnings announcements. Our findings underscore the critical role of alternative data in financial markets and the broad implications of data sharing restrictions on market efficiency.

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

JEL Codes: G14, G12, O16

*We thank Allen Hu and Xiaoyun Yu for their helpful comments and suggestions. We thank Rohan Joseph and Bolin Xu for excellent research assistance. This version: March 17, 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

Recent years have been characterized by an explosion in the collection, sharing, and usage of alternative data by firms and financial market participants. On the one hand, the benefits of this trend are clear: alternative data, coupled with powerful big data technologies (Machine Learning-ML, Artificial Intelligence-AI), enables economic agents to improve their predictive abilities (Begenau et al., 2018; Dugast and Foucault, 2018; Farboodi et al., 2022; Veldkamp and Chung, 2024). On the other hand, it raises serious privacy concerns, which lead governments and private companies to impose restrictions on data harvesting and sharing. These restrictions have the arduous objective of safeguarding individuals' privacy, while limiting the potential negative impact on economic innovation and growth. In an ever more interconnected digital economy, though, it is often hard to assess the full ramifications of privacy-led restrictions.

In this paper we provide evidence of the far reaching implications of privacy-motivated data sharing restrictions, by highlighting the role of financial markets in amplifying their impact. We focus on Apple's App Transparent Tracking (ATT) policy, which restricts the ability to share user identifiable information across apps and with third parties, on devices using Apple's iOS operating system. We show that this restriction significantly worsens the information environment of firms which track/share user data through apps, decreasing the predictive power of mobile-generated signals over future firm performance. Since market participants learn from mobile-generated data (e.g., mobile advertising and app usage, foot-traffic), the decline in its predictive power leads to an increase in forecast errors and a decrease in price informativeness. Effects are heterogeneous, based on the reliance of firms and market participants on third party data, increasing the policy's distortionary effect.

We focus on the mobile apps market for a few reasons. First, it has become one of the fastest growing realities for businesses in the digital economy. In 2023 total direct revenues of the US app market are estimated at 153.04bn and are projected to grow to 230.44bn by 2027 (Statista, 2024a); and those estimates do not include indirect revenues such as the sale

of user data. Second, the usage of mobile apps, all residing in a single device, has opened the possibility of collecting more detailed information about customer preferences, by connecting behaviors observed through different activities to a single customer, which firms use to drive revenues through better targeting. Third, being able to collect user data across mobile apps has fueled a large industry of alternative data providers, who aggregate information obtained across a variety of apps to sell it to firms and financial market participants. Financial market participants, in turn, use mobile generated alternative data as leading indicators of future firm performance, incorporating it into prices through their trades.

The usage of alternative data by market participants has rapidly grown in recent years. Industry experts estimate that, as of 2021, *"about 50% of investment managers are currently using some form of alternative data, with another quarter planning to start incorporating alternative data in the next 12 months"* (Tingley, 2021). While, Chi et al. (2023) document that, by 2018-2019, 28% of financial analysts use alternative data in making their forecasts. According to AlternativeData.org, mobile usage, geo-location, credit-debit card transactions, e-mail/consumer receipts and social media sentiment are five of the nine categories of alternative data most used by market participants (AlternativeData.org, 2019); all of which should contain data collected through mobile apps.

Data sharing permissions are particularly relevant for capital markets, as the usage of third party data that was acquired without the appropriate permissions might constitute insider trading.[1] Eagle Alpha, a leading company in alternative data procurement for financial firms, says they have observed an increased involvement of the legal and compliance teams of financial firms in alternative data procurement. In particular, they highlight a growing concern over changes in the terms and conditionsfor data sharing and usage on the part of Big Tech companies (e.g., Meta, Apple, Google, and Amazon), which *"can have ripple effects on data quality, security, and governance"* (Ryan, 2023).

[1] A chief example is App Annie, a third party data provider which in 2021 was charged with securities fraud by the SEC for selling data about mobile app usage to hedge funds without the appropriate data sharing permissions (SEC, 2021).

Despite the increased practitioners' attention to this issue, there is little academic evidence for how privacy-motivated data sharing restrictions might impact capital markets. We aim to fill that gap by using Apple's App Transparent Tracking (ATT) policy as a shock to data sharing. ATT, which was rolled out worldwide in April 2021, introduced the requirement for explicit consent to share user identifiable information with other apps or data brokers on iOS devices. It is estimated that in the US, in the first month around ATT's introduction, only 4% of users allowed tracking, that number only increased to 18% one year after its introduction (Balasubramanian, 2022). Since iOS holds more than 50% market share in the US mobile operating systems market (comScore, 2024), ATT represents a large and sudden shock to firms' ability to track users across apps and to share identifying information (e.g., geolocation) with data brokers. For those reasons, ATT is an ideal shock for us to study the implications of reduced data sharing ability on firms and financial markets.

We proceed by formulating detailed hypotheses of how ATT should impact firms' information environment, and hence, capital markets. We are able to test them empirically by levering a granular dataset of mobile app advertisement, usage, and data policies, jointly with analyst reports and firm characteristics.

First, we hypothesize that, post-ATT, the reduced ability of firms to share mobile data with third parties should lead to a deterioration in the quality of information available to market participants. If market participants utilize information obtained from mobile signals in their forecasts, this should cause a decrease in price informativeness. That decrease should be greatest among firms which share data with third parties, as they should experience the greatest decrease in mobile data availability post-ATT. Additionally, we should observe a greater decrease in informativeness for those firms that have a greater concentration of market participants which utilize mobile data in making their forecasts.

We test that hypothesis by studying the link between earnings surprises (SUE) and cumulative abnormal returns (CAR) around earnings announcements. A vast literature has studied price reactions to earning announcements. The basic premise of those studies is

that the more information is anticipated by market participants and incorporated into the price prior to the announcement, the smaller should be the reaction of prices to the information disclosed in the announcement itself (e.g., Kim and Verrecchia (1991), Demski and Feltham (1994)). If ATT causes a significant decrease in the information available to market participants, we should observe an increase in the reaction of prices to announcements. Indeed, we show a significant increase in the relationship between SUE and CAR around earnings announcements post-ATT. That increase is concentrated among firms which share data with third parties, or that are followed by financial analysts specialized in alternative data. The economic magnitude of those effects is significant. The baseline reaction of CAR to an increase in SUE of one standard deviation is 74.0*bps*; which increases to 99.2*bps* for user tracking/data sharing firms post-ATT, a 34% increase relative to the baseline effect. Similarly, that reaction increases from 58.1*bps* to 128.5*bps* for firms followed by analysts specialized in alternative data, more than doubling the baseline effect. To further validate our results, we show that there is no change in the response of prices to earnings announcements for firms that do not possess a mobile app, even when they are followed by analysts specialized in alternative data.

Next, we hypothesize that the deterioration in price informativeness should come from an increase in forecast errors by market participants. If mobile-collected data becomes less precise, it should contain a lower signal to noise ratio. When market participants utilize those noisier signals to update their priors about firms' payoffs, it should lead them to noisier posteriors and hence, to greater forecast errors.

To test that hypothesis, we study changes in analysts' forecast errors around ATT. We focus on analysts as we can directly observe the history of their forecasts. That is a more direct test of our hypothesis than, for instance, studying portfolio decisions by asset managers, as it more cleanly isolates forecast errors from other portfolio effects. Additionally, for analysts, we are able to collect the data sources utilized in formulating their predictions from the text of their reports, allowing us to better identify exposure to the shock. We start

by constructing different analyst-level metrics of specialization in alternative data. First, we consider analysts following an above average number of data intensive firms as being specialized in alternative data.[2] Second, 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.[3] Then, for each forecast, we compute forecast errors relative to other analysts making forecasts for the same firm in the same quarter. We show that, post-ATT, relative forecast errors of analysts specialized in alternative data increase, so to bring an analyst from the median to the $55^{th} - 57^{th}$ percentile, depending on the metric used. Importantly, our most comprehensive regression specification includes both analyst and firm-quarter fixed-effects, allowing us to absorb any firm or analyst specific variation in forecast errors, while isolating the impact of ATT on alternative data precision. Interestingly, when excluding analyst-level fixed-effects, we observe that analysts mentioning the usage of mobile generated data in their reports present *lower* relative forecast errors pre-ATT, of an absolute magnitude similar to their *higher* relative forecast errors post-ATT. That suggests a more subtle interpretation of results–i.e., that the usage of mobile data gave specialized analysts an advantage relative to their peers pre-ATT, while that advantage disappears with the loss in predictive power of mobile-generated data post-ATT. We further validate that interpretation by showing that ATT does not impact the forecast errors of analysts specialized in alternative data, when they make forecasts for firms that should not be impacted by ATT.

Finally, we focus on providing direct evidence for a possible mechanism underlying our results. Existing studies have shown that ATT directly impacted firms' ability to link customer data across apps and hence, their ability to drive revenues through targeted advertising (Kraft et al., 2023). We hypothesize that a reduction in targeting ability should imply that firms should change their advertising strategy to accommodate less precise targeting and that, as a consequence, the volume of traffic driven to apps should be noisier, coming from

[2]The full list of data intensive firms considered is found in Appendix Table A.4
[3]A full list of keywords searched can be found in Appendix Table A.2

some target customers and some less valuable users, so it should lose predictive power over app engagement and firm performance.

To establish that firms change their advertising strategy as a response to ATT in a manner that is consistent with a loss in targeting ability, we rely on empirical and theoretical evidence in the marketing literature. According to marketing theories (e.g., Iyer et al. (2005)), the loss of targeting abilities in an environment with a low marginal cost of advertising (as is the mobile ads market) should lead to more homogeneous and spread out advertising. Additionally, various empirical marketing studies show that dynamic advertisements, such as videos, are more effective at engaging customers and raising brand awareness, particularly when ads are not targeted (e.g., (Bruce et al., 2017)). In line with that evidence, we observe a sudden decrease in ad publishers concentration post-ATT, indicating a shift towards more spread out advertising. Additionally, we document a significant change in the mix of ad types. Post-ATT the share of the more engaging video ads increases; while, conditionally of utilizing static ads, their size decreases, so reducing their marginal cost.

Next, we show that the reduced capacity to target high valued customers leads to noisier and less quality traffic to apps, such that app traffic becomes less tightly linked to the number of loyal customers who consistently interact with the app, and eventually to firm performance. We do so by documenting that metrics of mobile app traffic become less predictive of app engagement, measured as the ratio between daily active users (DAU) and monthly active users (MAU), as well as losing predictive power over firm performance, measured as sales growth or earning surprises (SUE). As expected, the loss in predictability is concentrated in firms which share data with third parties. Economically those effects are significant, for instance, for user tracking/data sharing firms, pre-ATT, an increase in downloads of 1 standard deviation leads to an increase in SUE (sales growth) of 1.16 (17.15%); while post-ATT the same increase in downloads only leads to an increase in SUE (sales growth) of 0.56 (12%). That loss in predictability is financially relevant, if constructing a trading strategy based on those signals. Indeed, between 2017 and 2021 a portfolio which buys high

download stocks and short sells low download stocks registers an overall CAR of 35%-120%, depending on portfolio construction; while no additional gains are registered on that strategy post-ATT.

Mobile app usage is one of the main mobile-collected alternative data sources sold to market participants by various alternative data providers (e.g., the previously mentioned App Annie, AlternativeData.org (2019)). Hence, the loss in predictive power of this data over future firm performance could partly explain the deterioration in forecast errors and price informativeness previously documented. Of course, mobile usage data is not the only alternative data affected by ATT. Any dataset containing user-identifying information collected through mobile apps should experience a deterioration in quality post-ATT. For instance, ATT restricts the ability to share user geo-location with data brokers without explicit consent. As such, it should also decrease the quality of alternative datasets measuring foot traffic to retail establishments thought mobile-collected geo-locations. The magnitude of the capital markets effects previously documented should be interpreted as a compounded effect of the deterioration of all impacted mobile-collected datasets utilized by market participants.

Related Literature.

Our study examines the impact of privacy restrictions on capital markets through the quality as well as amount of alternative data. On the one hand, we contribute to the growing body of research on privacy in economics and finance, especially the work focusing on 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).[4] On the other hand, our paper connects with the research on the rising usage of alternative data in capital markets. Many different types of alternative data are studied by existing work, such as satellite images (Zhu, 2019; Katona et al., 2024; Ke, 2023), customer and employer reviews (Huang, 2018; Green et al., 2019), real-time and granular sales information

[4]See Johnson (2022) for a review of the literature examining GDPR.

(Froot et al., 2017; Blankespoor et al., 2022; Dichev and Qian, 2022; Jin et al., 2022; Du and 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 literatures, documenting the significant spillover effects that privacy regulations can have on the capital markets.

More specifically, we advance the understanding of how mobile-generated data influences capital markets by building on existing studies that explore its usage by market participants. For example, Chen et al. (2024) shows that mobile app performance serves as a leading performance indicator, though investors and analysts only partially incorporate this data into their trading decisions and revenue forecasts. Similarly, Jin et al. (2022) uses 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 market participants, we extend this research by examining the impact of privacy regulations on its usefulness. By focusing on ATT, we provide causal evidence of a negative event affecting the use of alternative data. More importantly, ATT represents a unique and nuanced shock that degrades the quality of alternative data due to the *interconnectedness* of the data economy. This mechanism, which drives our main findings, goes beyond the direct positive shocks to data access typically explored in previous literature.

Our work also adds to the literature on analyst forecasts, with a particular focus on their use of alternative data. The proliferation of alternative data vendors (Grennan and Michaely, 2021) and the explosion of consumer surveillance via mobile technologies have created new information channels for financial analysts. Applying textual analysis to analysts' written reports, Chi et al. (2023) show that analysts draw insights from alternative data and generate more accurate earnings forecasts. Consistent with their findings, we provide granular evidence of how analyst' forecast errors are related to the use of mobile-collected alternative

datasets. While the growing reliance on alternative data marks a significant shift in how market participants gather, analyze, and utilize information, it also raises important questions about data privacy, ethical considerations, and the implications of relying on pervasive consumer surveillance to inform financial decisions. This tradeoff highlights our deviation from existing work – we are the first to document that privacy regulations emerging in the era of big data can have unintended consequences on the accuracy of analyst forecast and the broader information environment. In a similar vein, recent work investigates the challenges posed by the rise of big data in the financial market. Dessaint et al. (2024) explore the tradeoff between long-term and short-term informativeness, documenting that equity analysts' long-term forecasts may deteriorate as they shift their attention to short-term information in alternative data. Bonelli and Foucault (2023) argue that big data may displace high-skill workers in finance as it undermines active mutual funds' stock picking abilities.

Finally, we contribute to the growing body of literature on the effects of privacy regulations on targeted marketing. We provide novel evidence related to empirical and theoretical studies in the marketing literature. Using millions of survey responses, Goldfarb and Tucker (2011) study how privacy regulation in the EU reduces the effectiveness of personalized advertising. Johnson et al. (2020) examine how the AdChoices program, which allows consumers to opt out of online behavioral advertising, affect ad targeting. Using the same laboratory as ours, Kraft et al. (2023) document a reduction in the share of trackable (versus untrackable) Apple traffic and ad impressions following ATT. Our evidence is broadly consistent with these studies. While existing papers focus on the loss of targeting ability, we further connect privacy regulations to a loss in predictive power when considering mobile-generated data as leading indicators of future firm performance. This allows us to extend this literature to the broader capital market and study valuation implications.

2 Institutional Background

Our laboratory features a major privacy policy shock that limits data sharing via mobile devices. A vast set of alternative datasets, increasingly used by financial market participants,

is sourced through mobile devices and it is hence, exposed to this shock.

2.1 The Mobile App Market

The development and usage of mobile applications experienced a fast growth in the last decade (Statista, 2024b,c), driven by an increase in internet penetration and the vast adoption of smartphones (Statista, 2024d,c). Moreover, technological advancements (ML, AI, Augmented Reality) enable fast innovation in the services provided through apps, fueling their omnipresence in day to day activities (Forbes, 2024).

For those reasons, apps represent a growing source of revenue for firms. In the US alone, the total direct revenues generated through the mobile apps market are estimated at $153.04bn$ in 2023, and are projected to grow to $230.44bn$ by 2027. The vast majority of those revenues (76% in 2023) are linked to mobile advertising (Statista, 2024a). Of course, those estimates ignore indirect benefits of mobile apps, as the possibility of acquiring detailed user data, which can be used to optimize business strategies and sold to third parties.

The two leading mobile operating systems are Google's Android and Apple's iOS. Globally, iOS accounts for around 25% of market share, while it reaches over 50% in the US (StatCounter, 2024; comScore, 2024). Apps are downloaded to mobile devices through integrated stores: Google Play for Android devices and the App Store for iOS devices.

2.2 Data Sharing and ATT

Both Android and iOS traditionally adopted an opt-out policy with respect to data sharing. That implies sharing by default a device specific identifier, along with other identifying information (e.g., geolocation), across apps and with third parties (e.g., data brokers). That allows linking all activities done on a single device, which can then be analyzed for better targeting. Tracking customer behavior across apps also enables precise computation of ad performance, as advertisers can directly monitor user behavior after interacting with an ad.

On April 26^{th}, 2021, with the release of iOS 14.5, Apple globally introduced the App Tracking Transparency (ATT) policy, which was accompanied by an extensive privacy-

motivated campaign to educate users regarding data sharing through their mobiles (iFans, 2021; Apple, 2021). ATT requires that on all devices utilizing iOS 14.5 or later, apps obtain explicit consent to share the device specific Identifier for Advertisers (IDFA), or any other identifiable information (e.g., location data or e-mail lists), with other apps or data brokers (Apple, 2024b). Practically, after any iOS user upgrades their operating system to iOS 14.5, the first time they enter an app that wishes to shares data with third parties, they are prompted with a pop-up message asking whether they would like to allow the app to *"track your activities across other companies' apps and websites"*, further specifying that *"your data will be used to deliver personalized ads to you"*. Users who do not wish to grant or deny permissions one app at the time, also have the option of disabling user tracking/data sharing for all apps at once, in their phone's settings (Apple, 2024a).

According to estimates provided by Flurry, a leading provider of mobile analytics, within the first month of introduction, only 11% of users globally, and 4% in the US, opted for allowing user tracking. Those rates increased to 25% of users globally and 18% in the US after one year of the policy's introduction (Balasubramanian, 2022).

Considering the large market share of iOS devices in the US (above 50%), the low opt-in rate to user tracking/data sharing (4%-18%), and the fast roll-out of the policy, ATT represents a large and sudden shock to firms' ability to link user activities across apps and to share detailed user data with third parties. For those reasons, ATT represents an ideal laboratory for us to study the impact of data sharing restrictions on data quality and capital markets.

2.3 The Alternative Data Market

The market for alternative data has expanded rapidly in recent years, and it is being used by a wide range of financial market participants. By searching the text of analyst reports, Chi et al. (2023) document that by 2009/2010 11% of financial analysts mention using alternative data, that number increased to 28% by 2018/2019. According to Greenwich Associates, a data analytics firm working with asset managers, as of 2021 (the year of ATT introduction),

more than 50% of asset managers already used alternative data and another 25% planned to start using it within the following 12 months (Tingley, 2021). More recent survey evidence shows that this trend continues growing, in 2023 only 18% of surveyed hedge funds mention they are not currently using alternative data and do not plan to do so in the next 6-12 months (Moss et al., 2023).

The usage of alternative data might pose risks for financial participants. In a 2017 report, Deloitte identified regulatory risk and data risk (including provenance and adherence to privacy regulation) as two of the main risks in the adoption of alternative data by asset managers (Deloitte, 2017). Those remain primary concerns, as corroborated by recent survey evidence in which hedge funds reported that "*increasing regulatory scrutiny*" (50%) and "*cost and time associated with vetting vendors and the quality, integrity, and reliability of the data they sell*" (38%) were their top concerns with respect to alternative data usage (Moss et al., 2023).

The relevance of those concerns was validated by the recent case of App Annie, a popular alternative data provider for mobile usage data. In September 2021, the SEC charged App Annie with securities fraud for selling mobile usage data to hedge funds, without obtaining the appropriate data sharing permissions (SEC, 2021). Eagle Alpha, a leading consulting firm in the procurement of alternative data for asset managers, reports that the case of App Annie has heightened the attention of the asset management industry to the legal aspects of alternative data procurement (Ryan, 2023). In fact, alternative data that does not comply with privacy regulations/restrictions might be considered as Material Non-Public Information, the usage of which would constitute insider trading. In particular, Eagle Alpha highlights concerns over changes in the privacy policies and terms and conditions of Big Tech companies (e.g., Meta, Apple, Google, and Amazon) as these might *"have ripple effects on data quality, security, and governance"*. Given those concerns, it is of primary relevance to study how changes in the privacy policies of tech companies impact capital markets.

To assess the relevance of mobile collected alternative data for capital markets, it is useful

to consider the mapping of the top 100 alternative data providers in the institutional investment space (AlternativeData.org, 2019). AlternativeData.org identifies nine main types of alternative data: app usage, credit/debit card, e-mail/consumer receipts, geo-location, sentiment/public data, satellite, weather, web data, and web traffic. Of these, app usage is entirely collected through mobile apps, geo-location is predominantly collected through mobile apps, and credit/debit card transactions is partly collected through mobile apps. For other categories, mobile-collected data represents a less dominant data source, even if it might still be relevant. For instance, many users access their e-mails or digital wallets through mobile apps, potentially impacting data collection in the e-mail/consumer receipts category. Similarly, users often access social media platforms, forums, and news outlets through apps, from which sentiment data is generally sourced. Hence, at least five of the nine categories of alternative data most used by asset managers should be affected by ATT. That makes ATT an ideal shock to study the impact of privacy-motivated data sharing restrictions on capital markets.

3 Hypotheses Development

Figure 1 graphically summarizes the logical links in our study. Firms collect data about user behavior through apps. That can be first party data (i.e., directly collected within the app) or third party data (i.e., obtained from third parties). Firms can share some or all of that data with third parties (e.g., data brokers or firms like Facebook), either for a price or in exchange for services (e.g., data analysis or targeting insights). Firms (either directly or through partners like Facebook) analyze first party and third party data, often through AI-models, to draw better insights about their users and improve engagement/targeting. A key targeting tool at firms' disposal is mobile advertising. Customer targeting then drives traffic (downloads, MAU, DAU) to a firm's app. The quality of that traffic is linked to customer engagement and ultimately to firm performance. At various stages of that process firms can share more data with third parties, as for instance mobile advertising or app usage information. Alternative data providers acquire data about users (e.g, geo-location or transactions),

mobile advertising, and app usage, either from firms or from data brokers. They then use statistical models, often AI-driven, to analyze that unstructured big data and produce structured datasets which can be sold to financial market participants. Market participants use signals obtained from that alternative data (e.g., foot-traffic or app downloads), in combination with other signals, to update their priors about firms' future payoffs. They trade on those insights, so driving the price of securities. Finally, price informativeness and liquidity have a feedback effect to firms' real outcomes–e.g., through their cost of capital.

We next formulate hypotheses for how data sharing restrictions might impact the linkages in the above structure. We particularly focus on the ripple effects that restrictions at the app level can have on the quality of alternative data, and ultimately on capital markets.

First, we focus on firms' information environment and price efficiency. ATT reduces firms' ability to share identifying information with third parties. That should impact the quantity and quality of alternative data collected through mobiles. If mobile-collected data was informative about future payoffs, its deterioration should worsen firms' information environment. According to the Efficient Market Hypothesis (EMH), prices should reflect all available information about future payoffs. Hence, a worsening in firms' information environment should cause prices to become less efficient, or else less informative about the intrinsic value of securities. That reduction in price informativeness should be concentrated among firms for which information extracted from mobile-collected data was incorporated into prices pre-ATT. Firm level heterogeneity should have two main drivers: (1) the relevance of mobile-collected third party data for the payoffs of each security; (2) the usage of mobile collected data by market participants in pricing each security.

Hypothesis 1 (H1): *ATT should cause a reduction in price informativeness. That reduction should be concentrated among firms that share mobile-generated data with third parties and/or are followed by market participants that use mobile-generated alternative data.*

Second, we dig deeper into the price formation mechanism. The channel through which ATT should cause a decrease in price informativeness is the forecasting action of market par-

ticipants. Indeed, it is through their trading that information gets impounded into prices. Market participants use signals about future payoffs (some of which might be mobile generated) to update their priors and generate posteriors. Their posteriors then determine trading behavior. According to Bayes law, posterior precision is directly proportional to signal precision. ATT should decrease the precision of mobile-generated signals, inducing more volatile (less precise) posteriors. This should lead to greater forecast errors post-ATT. Similarly to the previous hypothesis, forecast errors should increase for: (1) market participants who use mobile-generated alternative data in forming their forecasts; (2) the forecasts about firms for which mobile-collected third party data is relevant in determining future payoffs.

Hypothesis 2 (H2): *ATT should cause an increase in forecast errors. That increase should be concentrated among market participants who use mobile-generated alternative data in making their forecasts, and should be more relevant for the forecasts about firms that share mobile-generated data with third parties.*

Finally, we focus on potential underlying mechanisms for the decrease in quality of mobile-generated alternative data. ATT should impact the quality of mobile-collected data through one or more of the following related channels: (1) given the high opt-out rates, it should reduce the volume of alternative data available; (2) since many alternative data providers (e.g., App Annie) use AI-powered models to generate firm-level estimates from unstructured data, it should cause those estimates to become less precise and/or less representative of the overall population of users; (3) given the impossibility to use precise user information, firms' targeting ability through mobile advertising should deteriorate, causing mobile advertising and app usage to become less informative of future engagement and firm performance; (4) mobile data usage, mobile ad spend, geo-location and/or transaction data represent inputs for other alternative datasets purchased by market participants (e.g., yipitDATA), which might also experience a deterioration in quality.

To formulate concrete hypotheses for how those mechanisms materialize, we focus on app usage data; one of the main mobile-collected alternative datasets used by market participants.

Hypothesis 3 (H3): *ATT should decrease the quality of mobile usage data.*

As previously highlighted, ATT reduces firms' ability to connect data obtained across different apps to a single user, and to share data with third parties (e.g., Facebook). That should impair firms' ability to draw precise user preferences and hence, to engage and target users optimally. One of the main channels used by firms to target users is mobile advertising. If ATT decreased the ability of firms to precisely target customers through mobile advertising, we should observe changes in their advertising strategy that are consistent with a lower targeting ability. In the marketing literature, the theoretical work of Iyer et al. (2005) shows that a decrease in targeting ability, when the marginal cost of advertising is low (as in the mobile ads market), causes firms to produce more homogeneous ads that they spread out to a broader audience. As a result, they experience a decrease in advertising efficiency. Additionally, various empirical studies (e.g., Li and Bukovac (1999), Dahlén (2001), Hong et al. (2007), Bruce et al. (2017)) show that optimal ad types change depending on targeting ability and advertising objective. In particular, Bruce et al. (2017) show that dynamic ads, like videos, are better at engagement and creating brand awareness, particularly when ads are not directly targeted to specific users (e.g., re-targeting).

Subhypothesis 3.1 (H3.1): *ATT should cause a change in mobile advertising strategy consistent with a reduction in targeting ability: more homogeneous ads, displayed to a wider audience, with lower marginal cost, and a greater share of videos.*

If ATT reduces firms' ability to reach their target audience (e.g., through targeted mobile ads), then, as suggested by Bruce et al. (2017), the efficiency of their marketing strategy should be reduced. We hypothesize that the reduction in marketing efficiency post-ATT should affect the quality of traffic driven to apps. Indeed, if firms produce more homogeneous advertisements which are spread out to a broader audience, some of the customers they attract to their apps might not be high-value customers, who consistently engage with the app and drive performance. Hence, ATT should make app usage a noisier predictor of app

engagement and ultimately of firm performance. In other words, it should decrease the signal to noise ratio in measures of app usage as predictors of future firm payoffs. That effect should be aggravated by the fact that app usage metrics available for purchase are not observed but are the product of AI-estimates by alternative data providers. If the amount of data shared with them is reduced post-ATT, their estimates of app usage might also become less precise about true app usage, increasing noise in those metrics.

Subhypothesis 3.2 (H3.2): *ATT should cause app usage metrics to become worse predictors of app engagement and firm performance.*

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 ATT introduction. We lever a granular dataset of the mobile apps market (mobile ads, app usage and data sharing policies), which we merge to firm financials and analyst forecasts. That allows us to compute measures of ATT exposure at the firm and analyst level, which we use to test the impact of ATT on capital markets.

4.1 Mobile Apps

We acquire information about mobile app usage (e.g., downloads, active users, sessions length) 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 financial tickers globally. They then sell that structured dataset back to firms and financial market participants. We acquire mobile usage data specifically from AppTopia as, since 2019, it is featured on the Bloomberg terminal. Hence, their mobile usage estimates are broadly available to financial market participants. Throughout our analysis, we consider combined app usage across all apps belonging to a given firm.

We additionally acquire mobile advertisements data from AppTopia. We obtain both creatives (images, GIFs, videos) and ads metadata. Ads metadata includes information about each creative; e.g., format, size, the period in which it was active, and the channels through which it was displayed (networks, platforms). This is a newer product form AppTopia and it is not currently available on Bloomberg. Mobile advertisement data is available for a smaller set of firms than the more comprehensive app usage data.[5]

Finally, we scrape information about apps' data policy from their official page on Apple's App Store. As of September 2020, Apple requires each app to disclose their policies with respect to data collection and usage. We focus on a section called *"Data Used to Track You"* (*Data sharing*) in which firms must disclose if they collect any type of user information for the purposes of user tracking (i.e., linking it across apps and/or sharing it with third parties).

4.2 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 collect information about the data sources utilized in forming forecasts from the full text of each report, based on the keyword lists in Chi et al. (2023). Importantly, we append two keyword lists: one focused on app usage and the other on alternative data providers likely affected by ATT.[6] We search for specific terms in these lists (e.g., MAU, downloads, geo-location, foot traffic) and also scan the text of each report for mentions of alternative data providers.[7]

We construct measures of analyst specialization in alternative data using both datasets. Our first proxy is based on the I/B/E/S data. We start by identifying a list of big data companies.[8] Then, for every quarter, we define as specialized (*"Specialized Analyst (0/1)"*)

[5]Section 4.4 provides further details on sample differences.

[6]Following Dessaint et al. (2024), we start from JP Morgan's Alternative Data Handbook. We then manually identify the data products likely affected ATT.

[7]Appendix Table A.1 contains a full list of the keywords and alternative data providers searched.

[8]The full list is found in Appendix Table A.4. The list is compiled from J.P. Morgan US Internet Guidebook.

those analysts that follow an above average number of big data firms.[9] Our other proxy is based on the data sources mentioned in analyst reports. Every quarter we define as specialized (*"Keywords Usage (0/1)"*) those analysts who mention at least one keyword related to the usage of mobile-generated alternative data in their reports.[10] Whereas we are able to construct the first metric for all analysts present in the I/B/E/S dataset, we restrict the second to analysts for whom we are able to identify more than 100 reports over 2019-2020 in the LSEG dataset and whose names we can match to I/B/E/S.

4.3 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.

4.4 Summary Statistics

Tables 1-3 report summary statistics for all variables and samples used in the analysis.[11]

Table 1 shows firm-level summary statistics based on different data samples. Panel A displays statistics for the 802 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 420 unique firms in the *"Ads Sample"*. This is the subset of the *"App Sample"* for which we have mobile advertising data. Finally, Panel C displays statistics for

[9] The average number of big data firms followed is computed conditional on following at least one. The full list of big data firms considered can be found in Appendix Table A.4

[10] The full list of keywords searched can be found in Appendix Table A.2

[11] All continuous variables are winsorized at the 0.5% level.

the 6,443 unique firms in the *"No-app Sample"*. This includes all observations of firms that do not have an app. All three panels are similar in terms of firm size and firm characteristics, with the exception that firms with apps (Panels A and B) tend to have more cash relative to firms without apps (Panel C). Looking at mobile-related variables (Panels A and B), we observe that the "Ads Sample" (Panel B) includes firms with more app downloads (median of 1,760,811 vs. 281,144 per quarter), more likely to be followed by analysts specialized in alternative data (32% vs. 17%), and more likely to track user data (64% vs. 37%). Hence, firms in the *"Ads Sample"* are more likely to be impacted by ATT.

Table 2 displays statistics for the *"App Sample"*, split by whether firms track users/share user data (Panel A) or not (Panel B). Looking at firm characteristics: size is similar across the two samples, while firms that track/share user data tend to have more cash, more debt, and less tangibles. Regarding mobile characteristics, data sharing firms tend to have significantly more downloads (median of 1,259,702 vs. 120,287 per quarter) and are more likely to be followed by specialized analysts (26% vs. 12%).

Finally, Table 3 displays analyst-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). The *"App Sample"* (placebo sample) includes forecasts from 2,977 (2,150) unique analysts, belonging to 274 (320) unique brokerage firms. The *"App Sample"* (Panel A) is more likely to contain analysts specialized in alternative data (8.2% vs. 1.4%) and firms are covered by the same analyst for a longer period (12.78 vs. 10.97 quarters). The two samples are otherwise similar in terms of the number of days between forecast and announcement, overall analyst experience, the number of firms covered, and the size of the brokerage firms.

5 Empirical Analysis

5.1 Price Efficiency

We start our empirical analysis by testing hypothesis 1; i.e., that price efficiency decreases post-ATT, particularly so for firms that share data with third parties, or that are followed by market participants who make use of alternative data.

We start by analyzing changes in short-term price informativeness, 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 investors using private information. Mobile-collected alternative data (e.g., app usage, or foot-traffic) could be used as leading indicators of future firm performance by sophisticated investors, allowing them to anticipate some of the information disclosed in announcements. Through their trading, they would then incorporate that information into prices, determining a smaller reaction of prices to announcements. If ATT reduces the quality of those signals, the ability of sophisticated investors to anticipate the information contained in earning announcements should also be reduced. Hence, we expect an *increase* of the reaction of prices to earnings announcements post-ATT. That would indicate a *decrease* in short-term price informativeness.

Similarly to Zhu (2019), we 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 to control for differences in public information. We measure SUE by assuming 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 $(t-4)$.[12]

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

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

The relation between CAR and SUE should then denote the pre-announcement information set of privately-informed investors relative to that of naive investors, basing their estimates on past earnings. We don't measure SUE relative to analyst forecasts, as analysts might also use mobile-generated alternative data, invalidating our test. The draw down of our approach is that SUE might not identify surprises relative to all payoff relevant public information. To obviate to that issue, we further condition the slope's measurement on whether firms are followed by financial analyst specialized in the use of alternative data. More specifically, we run the following regression:

$$CAR_{it} = \alpha + \beta_1 SUE_{it} + \beta_2 Specialized \times SUE_{it} + \beta_3 Post^{ATT} \times SUE_{it} + \boldsymbol{\beta_4} Post^{ATT} \times Specialized_{it} \times SUE_{it} + \beta_5 Specialized_{it} + \beta_6 Post^{ATT} \times Specialized_{it} + \Gamma X_{it-4} + \theta_i + \iota_t + \epsilon_{it} \quad (2)$$

where i indicates firms, t year-quarters, CAR is the cumulative abnormal return relative to a factor model (CAPM or Fama-French-Carhart 6-factor model) over trading days [0,+10] post earning announcements; $Post^{ATT}$ is an indicator variable with a value of 1 for all quarters after ATT; $Specialized$ is an indicator variable equal to 1 if the firm is followed by at least one analyst specialized in alternative data in quarter t;[13] SUE is measured as in Eq. 1, X is a vector of control variables measured in quarter $t-4$, namely size, and cash, tangibles and debt over assets; we include firm (θ_i) and year-quarter (ι_t) fixed-effects, and cluster standard errors at the firm level.

The coefficient of interest is $\boldsymbol{\beta_4}$, indicating the differential reaction of CAR to SUE post-ATT for firms followed by analysts specialized in alternative data, relative to those that are not. Looking at $\boldsymbol{\beta_4}$ allows us to control for public information, while isolating changes in information sets pre- and post-ATT due to changes in the quality of alternative data. Indeed, if investors use analyst forecasts in forming their expectations of future payoffs, analysts'

[13] We proxy for each analyst's specialization in alternative data using different approaches, outlined in section 4.2.

information sets should be impounded into prices. We expect to observe a positive and significant $\hat{\beta}_4$, indicating a greater price reaction to information released in announcements, and a deterioration in the information set of investors utilizing alternative data, post-ATT.

Table 4 reports the outcome of that regression for different samples, control variables, and CAR construction methods. Across specifications $\hat{\beta}_4$ is always positive and both economically and statistically significant. To assess economic significance, let's consider our most comprehensive specification in Column (6), which includes the full App Sample, firm-level control variables, and in which CAR is measured in excess of a Fama-French-Carhart 6-factor model. The baseline slope pre-ATT is 0.099 ($\hat{\beta}_1$, significant at 1% level), indicating that an increase in SUE of one standard deviation is related to an increase in CAR of 58.1*bps* (0.099×5.87). For firms followed by analysts specialized in alternative data, pre-ATT, that slope is not statistically distinguishable from zero ($\hat{\beta}_1 + \hat{\beta}_2 = 0.099 - 0.188 = -0.089$; not significant).[14] 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 ($\hat{\beta}_3$, not significant). For firms followed by analysts specialized in alternative data, instead, post-ATT an increase in SUE of one standard deviation is linked to an increase in CAR of 128.5*bps* ($(0.099 - 0.188 + 0.308) \times 5.87$), more than doubling the baseline effect.[15] That indicates a large deterioration in the information environment of firms followed by analysts specialized in alternative data post-ATT.

To further highlight the mechanism through which the decrease in short-term price informativeness operates, we substitute *Specialized* in Eq. 2 with an indicator variable equal to 1 if firms share/track user data obtained through mobile apps (*Data sharing*).[16] We expect the coefficient $\hat{\beta}_4$ to be positive and statistically significant, indicating that the deterioration in information environment comes from firms directly affected by a decrease in tracking

[14] We run the following test: $H_0 : \hat{\beta}_1 + \hat{\beta}_2 = 0$, which cannot be rejected.
[15] We run the following test: $H_0 : \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 = 0$, which is rejected with 1% confidence.
[16] We construct that variable as outlined in Section 4.1.

ability post-ATT. Table 5 reports the results of that regression. The coefficient $\hat{\beta}_4$ is positive and both economically and statistically significant across all samples and specifications. Considering again the most comprehensive specification in column (6), the baseline effect is 0.126 ($\hat{\beta}_1$, significant at the 1% level); i.e., an increase in SUE of one standard deviation increases CAR of 74.0*bps* (0.126×5.87). For firms which share data with third parties there is no significant reaction of CAR to SUE pre-ATT ($0.126 - 0.187 = -0.06$, not significant).[17] Post-ATT, there is no change in slope for firms that do not share/track user data ($\hat{\beta}_3$, not significant). For firms that share/track user data through mobile apps, instead, an increase in SUE of one standard deviation post-ATT is linked to an increase in CAR of 99.2*bps* ($(0.126 - 0.187 + 0.230) \times 5.87$), a 34% increase relative to the baseline effect.[18]

As a final validation of our results, we repeat the analysis in Table 4 in the *No App Sample*, i.e., including all firms that do not possess a mobile app.[19] The results are reported in Table 6. 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. In other words, pre- and post-ATT, mobile-collected data could not be used as a leading indicator of those firms' performance (independently from analyst specialization). Hence, the deterioration in mobile-collected data post-ATT does not affect the amount of information impounded into prices for those firms.

5.2 Forecast Errors

Next we test Hypothesis 2; i.e., that forecast errors of market participants using alternative data should increase post-ATT, particularly so for their forecasts of firms that share/track user data through mobile apps.

In testing that hypothesis, we focus on the forecast errors of financial analysts, as they

[17] We run the following test: $H_0 : \hat{\beta}_1 + \hat{\beta}_2 = 0$, which cannot be rejected.

[18] We run the following test: $H_0 : \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 = 0$, which is rejected with 1% confidence.

[19] We cannot repeat the specification as in Table 5 as, not possessing an app, none of the firms in this sample can share/track user data through mobile apps.

disclose periodic earnings forecasts for each firm independently. That allow us to more directly observe changes in forecast errors than, for instance, if studying asset allocation decisions of other market participants. Indeed, looking at asset allocation decisions, would make it harder to isolate forecast errors from other portfolio effects. Additionally, each forecast by a financial analyst is accompanied by a textual report in which sources and methods are disclosed. By searching for mentions of alternative data providers or metrics in those reports, we are able to further identify which financial analysts make use of alternative data. That allows us to construct a direct test our hypothesis.

For each quarter and each firm-analyst pair we compute Proportional Mean Absolute Forecast Errors (PMAFE), as Absolute Forecast Errors (AFE–i.e., the 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. We then run the following regressions:

$$PMAFE_{jit} = \alpha + \beta_1 Specialized_{jt} + \boldsymbol{\beta_2} Post^{ATT} \times Specialized_{jt} + \Gamma X_{jt} + \zeta_d + \theta_{it} + \iota_j + \kappa_b + \epsilon_{jit} \tag{3}$$

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; *Specialized* is an indicator variable equal to 1 if the analyst is specialized in alternative data. $Post^{ATT}$ is an indicator variable with a value of 1 for all quarters after ATT; X is a list of control variables, 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 the following set of fixed-effects: estimation date (ζ_d); firm-year-quarter (θ_{it}), analyst (ι_j), and brokerage firm (κ_b); standard errors are clustered at the firm and estimation month levels.

The coefficient of interest is $\boldsymbol{\beta_2}$. We expect $\hat{\beta}_2$ to be positive and significant, indicating an increase in forecast errors post-ATT for analysts specialized in alternative data. The inclusion of firm-year-quarter fixed-effects allows us to difference out any time-varying factors

at firm level in driving forecast accuracy around ATT and identify the effect of changes in analysts' access to alternative data. We further include analyst fixed-effect to control for unobservables at the analyst level that might influence forecast errors.

Table 7 reports the outcome of those regressions for the forecasts of firms in the *App Sample*. Columns (1-3) report results when identifying specialized analysts as those following an above average number of big data firms (*Specialized Analyst (0/1)*, while Columns (4-6) report results when identifying specialized analysts as those mentioning the usage of mobile generated data in their reports (*Keywords Usage (0/1)*).[20] The coefficient $\hat{\beta}_2$ is positive and significant in all specifications, independently of the inclusion of control variables, or of the full set of fixed-effects reported above. Columns (3) and (6) report the most restrictive specification, as described in Eq. 3, when identifying specialized analysts based on the *Specialized Analyst (0/1)* and *Keywords Usage (0/1)* metrics respectively. Looking at Column (3)(Column (6)), $\hat{\beta}_2$ is equal to 0.046 (0.051), and is significant at the 5% level. Economically, an increase in PMAFE of 0.046 (0.051) would bring an analyst who is at the median in terms to forecast errors to the 55^{th} (57^{th}) percentile. Column (4) provides further insights on the interpretation of results. When precisely identifying analyst specialized in the usage of mobile-generated data (*Keyword Usage (0/1)*), end excluding analyst-level fixed-effects, we observe that pre-ATT those analyst presented *lower* relative forecast errors than their peers ($\beta_1 = -0.041$–significant at the 1% level). Their advantage, though, fully disappears post-ATT ($\beta_2 = 0.040$–significant at the 5% level). That leads us to conclude that, whereas the usage of mobile-generated data provided a forecasting advantage to analysts utilizing it pre-ATT, the loss in predictive power of that data post-ATT fully erodes that advantage, making the forecasts of analysts using mobile-generated data not statistically different from those of analysts not using it.

To further validate our interpretation of $\boldsymbol{\beta_2}$, 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-

[20] Variables are constructed as highlighted in Section 4.2.

grained Fama-French 48 industries which doesn't contain firms with major apps. That choice is voted 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 Table 8. Once firm-year-quarter and analyst fixed-effects are included, $\hat{\beta}_2$ is not statistically different from zero. 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.

5.3 Alternative Data Quality

Finally, we provide empirical evidence for a possible mechanism underlying our results by testing Hypothesis 3; i.e., that the reduction in firms' targeting ability reduces the signal to noise ratio in mobile-app-collected alternative data, used by market participants.

We start by providing evidence consistent with Hypothesis 3.1; i.e., that firms affected by ATT change their mobile advertising strategies in a manner consistent with a loss in targeting ability. We run this test on the *Ads Sample*, the subset of the *App Sample* for which we were able to acquire mobile advertising data from Apptopia. As described in Section 4.4, this sample contains apps with more downloads, a greater rate of data sharing, and overall more data collection through mobiles. Hence, we consider this sample as being mostly affected by ATT. The advertisement data is at the creative (image, gif, video) level. Hence, for each firm-quarter we aggregate creative-level data to construct firm-level measures of advertisement activity. As discussed in Section 3, based on theoretical and empirical evidence in the marketing literature, we expect affected firms post-ATT to spread their mobile ads to a broader audience (measured using publisher concentration through a Herfindal index - HHI); while decreasing the marginal cost of advertising (measured as the percentage of static ads which is large – $Large\%$),[21] and increasing the share of videos (measured as the percentage of video ads relative to total ads – $Video\%$). We do not expect firms to reduce

[21] A banner is considered large if its horizontal dimension is above the 75^{th} percentile.

the overall number of posted ads (measured as the total number of active ads – $Number$), consistent with an environment with a low marginal cost of advertising.

We test those hypotheses by running the following regressions:

$$AdChar_{it} = \alpha + \mathbf{\Gamma}\Theta + \iota_i + \epsilon_{it} \tag{4}$$

where i indicates firms, t year-quarters; $AdChar = [Number,\ HHI,\ Video\%,\ Large\%]$; $\Theta = \left[\theta_{-5},\ \theta_{-4}, \cdots, \theta_{6},\right]$ is a vector of time dummies, the last element correspond to all quarters on or after quarter 6 post-ATT, other elements are quarter t dummies, for $t = -5, \ldots, 5$; $t = 0$ is the quarter of ATT announcement; $\mathbf{\Gamma} = \left[\gamma_{-5}, \gamma_{-4}, \cdots, \gamma_{6}\right]^T$ is the vector of regression coefficients to be estimated; we include firm fixed-effects (ι_i).

The coefficients of interest are $\boldsymbol{\gamma_t}$ for $t = -5, \ldots, 6$. We don't expect to observe changes in $\hat{\gamma}_t$ post-ATT ($t \geq 1$) when the dependent variable is the number of active ads ($Number$); while we expect $\hat{\gamma}_t$ for $t \geq 1$ to be significantly lower than $\hat{\gamma}_t$ for $t \leq -1$ when the dependent variable is publisher concentration (HHI) or the percentage of large static ads ($Large\%$), and significantly greater when the dependent variable is the percentage of video ads ($Video\%$).

Figure 2 displays the coefficients $\hat{\gamma}_t$ for $t = -5, \ldots, 6$. Panel 2a displays those coefficients when the dependent variable is the number of active ads. As expected, the number of active ads does not decrease following the ATT announcement. Instead, it continues increasing for an additional quarter, before stabilizing. Panel 2b shows those coefficients when the dependent variable is publisher concentration (HHI). We observe a sudden decrease in publisher concentration starting the quarter after the ATT announcement, while there are no clear pre-trends. This indicates that, as expected, affected firms start spreading mobile ads to a broader audience after ATT. Panels 2c and 2d display $\hat{\gamma}_t$ coefficients when the dependent variable is the share of videos ($Video\%$) or the share of large static ads ($Large\%$), respectively. As expected, those Figures display a stark change in mobile-ads composition starting the quarter after the ATT announcement. The share of videos significantly increases

(with no pre-trends), and the share of large static ads significantly decreases (with a trend reversal). Those results point towards a sudden change in mobile advertising strategy right after the ATT announcement, which is consistent with a reduced targeting ability.

Finally, we test hypothesis 3.2; i.e., that app traffic becomes a worse predictors of app engagement and of firm performance post-ATT. We do so in two steps, first, we establish that app traffic becomes a worse predictors of app engagement. Then, we show that app traffic becomes less predictive of firm performance, and that this leads to a loss in profitability of a long-short trading strategy sorting portfolios on downloads.

In our first empirical test we relate app-level measures of user traffic and user engagement. We measure user traffic as the natural logarithm of the number of downloads ($Downloads$), of monthly Active Users (MAU), and of daily active users (DAU); while we measure user engagement as the ratio of DAU and MAU ($Engage$). That indicates the portion of users utilizing the app on a monthly bases that also engages with it daily. We consider engagement as a proxy for the effectiveness of user targeting at the intensive margin; i.e., user value conditional on app usage. We measure the predictive power of app traffic over app engagement one week ahead. We utilize a weekly frequency for this analysis as both the dependent variable and the predictors are high frequency, and it is reasonable to believe that traffic should predict engagement over a shorter time window. We test for changes in the predictive power of app traffic on app engagement by running the following regressions:

$$Engage_{it} = \alpha + \mathbf{\Gamma}\Theta \times Traffic_{it-1} + \iota_i + \theta_t + \epsilon_{it} \mid DS = 0 \text{ or } DS = 1 \quad (5)$$

where i indicates apps, t weeks; $Traffic = [Downloads,\ MAU,\ DAU]$; $Engage = \frac{DAU}{MAU}$; $DS = Datasharing$; $\Theta = \left[\theta_{t-5},\ \theta_{-4}, \cdots, \theta_8,\ \theta_9\right]$ is a vector of time dummies, the last element corresponds to all weeks on or after week 9 post-ATT, other elements are week t dummies, for $t = -5, \ldots, 8$; $t = 0$ is the week of ATT announcement; $\mathbf{\Gamma} = \left[\gamma_{-5}, \gamma_{-4}, \cdots, \gamma_9\right]^T$ is the vector of coefficients to be estimated; we include app (ι_i) and week (θ_t) fixed-effects;

standard errors are clustered at the app and week level; coefficients are estimated separately for the sub-samples of firms sharing user data through mobile apps ($DS = 1$), and not ($DS = 0$).

The coefficients of interest are $\boldsymbol{\gamma_t}$ for $t = -5, \ldots, 9$, indicating the sensitivity of engagement in week t to app traffic in week $t-1$. We expect the predictive power of traffic on engagement to decrease post-ATT. Hence, we expect $\hat{\gamma}_t$ for $t \geq 1$ to be significantly lower than $\hat{\gamma}_t$ for $t \leq -1$, particularly so for apps that share/track user data ($DS = 1$).

The panels of Figure 3a display $\hat{\gamma}_t$ for $t = -5, \ldots, 9$ for the two sub-samples, for each of the app traffic proxies. We observe that the predictive power of traffic on engagement starts decreasing approximately 4 weeks after ATT, and it continues to decrease for all weeks following, particularly so for apps that track/share user data. The long-term effect $\hat{\gamma}_9$ is lower for both firm groups, but it is significantly more negative for firms that track/share user data through apps. The spillover to all firms should be related to the fact that, just like market participants, we observe app usage as estimated by AppTopia. The reduction in the amount/quality of data that apps are able to share with third parties, should cause the estimation models of alternative data providers to become less predictive across the board.

Next we study changes in predictability of app traffic over firm performance.[22] We do so by relating future 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 Eq. 1, in quarter t to downloads in quarter $t-1$. We do so by running the following regression:

$$Performance_{it} = \alpha + \beta_1 Downloads_{it} + \beta_2 Data\ Sharing_i \times Downloads_{it} +$$
$$\beta_3 Post^{ATT} \times Downloads_{it} + \boldsymbol{\beta_4} Post^{ATT} \times Data\ Sharing_i \times Downloads_{it} +$$
$$\beta_5 Data\ Sharing_i + \beta_6 Post^{ATT} \times Data\ Sharing_i + \beta_7 Downloads_{it-1} + \Gamma X_{it-4} + \iota_i + \theta_t + \epsilon_{it} \quad (6)$$

[22] We report results only based on downloads as it is the most commonly used metric, and it is highly correlated with other app traffic metrics. In unreported tables we verify that results are consistent when measuring traffic as MAU or DAU.

where i indicates firms, t year-quarters; $Downloads_{it}$ is the natural logarithm of total downloads for apps of firm i in quarter t; $Performance = [Sales\ Growth,\ SUE]$; $Datasharing$ is an indicator variable equal to 1 if any app of firm i tracks/shares user data; $Post^{ATT}$ is an indicator variable equal to 1 for all quarter after ATT announcement; X is a vector of control variables measured in quarter $t-4$, namely size, and cash, tangibles and debt over assets; we include firm fixed-effects (ι_i), and year-quarter fixed-effects (θ_t); standard errors are clustered at the firm level.

The coefficient of interest is $\boldsymbol{\beta_4}$, indicating the differential sensitivity of performance to downloads post-ATT, for firms that share/track user data through apps, relative to the ones that don't. We expect $\hat{\beta}_4$ to be negative and statistically significant, indicating a decrease post-ATT in the predictive power of downloads in quarter t on performance in the same quarter, for firms that share/track user data through mobile apps.

Table 9 reports the outcome of those regressions using Sales Growth (Columns 1 and 2) or SUE (Columns 3 and 4) as dependent variables. For both dependent variables, $\hat{\beta}_3$ is not statistically significant, while $\hat{\beta}_4$ is negative and statistically significant, indicating a decrease in the predictive power of downloads on performance post-ATT, only for firms that track/share user data through apps. To evaluate the economic magnitude of that result, let's consider the most comprehensive specifications in Columns 2 and 4, which include firm-level control variables. $\hat{\beta}_1$ is positive and statistically significant at the 1% level in both Columns, indicating that pre-ATT downloads predict both Sales Growth and SUE. An increase in downloads of 1 standard deviation increases sales growth of 8.65% (2.50×0.036) and SUE of 0.67 (2.50×0.266), 11.4% of its standard deviation (0.67/5.87). Pre-ATT, the predictive power of downloads on firm performance is even greater for firms that track/share user data through apps ($\hat{\beta}_2$ positive and significant at the 1%(10%) level in Column 2(4)). An increase in downloads of 1 standard deviation increases sales growth of 16.75% ($(0.036+0.031)\times 2.50$)), and SUE of 1.14 ($(0.266+0.188)\times 2.50$), 19.4% of its standard deviation(1.14/5.87). That is consistent with firms tracking users/sharing user data through apps being better at targeting

customers pre-ATT, so making app traffic more likely to come from performance-driving customers. Post-ATT, there are no differences in the predictive power of downloads over firm performance for firms that do not track/share user data through apps ($\hat{\beta}_3$ not statistically significant). For firms that do, instead, we observe a significant reduction in predictive power ($\hat{\beta}_4$ negative and significant at the 5% level). A 1 standard deviation increase in downloads increases sales growth of 11.75% $((0.036 + 0.031 - 0.02) \times 2.50)$, and SUE of 0.548 $((0.266 + 0.188 - 0.235) \times 2.50)$, 9.3% (0.548/5.87) of its standard deviation. For sales growth that means a loss of almost all of the pre-ATT additional predictive power for firms that track/share user data through apps; while for SUE post-ATT the net effect for those firms is even lower than the pre-ATT baseline effect ($\hat{\beta}_1$). That is consistent with firms tracking users/sharing user data through apps losing targeting ability post-ATT, such that downloads are less linked to traffic from performance-driving customers. The pattern highlighted in Table 9 is consistent with results in Section 5.1. There, we document that, for firms that track/share user data through apps, short-term price informativeness is greater pre-ATT, while it decreases below the baseline post-ATT. We conjecture that this pattern should be driven by a greater signal precision of mobile-collected data for user tracking/data sharing firms pre-ATT, with a subsequent significant decline in its usefulness post-ATT. We have now provided direct evidence of that pattern for one of the mobile-collected datasets most used by market participants.

Finally, to highlight the economic implications of that loss in predictability 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.[23]

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

Figure 4 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 indicates that app downloads become a financially-irrelevant sorting variable. Taken independently, the loss in profitability of this trading strategy could be linked to a greater use of downloads as a leading indicator by market participants which, hence, would already be incorporated into prices. Considering this result jointly with the reduction in price informativeness documented in Section 5.1 and the loss in the predictive power of downloads reported above, leads us to conclude that the most plausible explanation for this result is a decreased signal to noise ratio of downloads in predicting future payoffs.

Of course, app usage, and downloads in particular, are not the only mobile-generated signal that should be impacted by ATT. As previously discussed, any dataset containing identifiable data collected from mobile apps should have registered a loss in predictive power post-ATT. Hence, the effects of ATT on price efficiency and forecast errors previously documented should be considered as cumulative, due to a deterioration in quality of all affected alternative datasets used by market participants.

6 Concluding Remarks

The increasing adoption of privacy regulations, such as Apple's App Transparent Tracking (ATT) policy, presents a complex tradeoff between safeguarding individual privacy and sustaining the flow of alternative data crucial for capital markets. Our study provides empirical evidence of how privacy-motivated data sharing restrictions can significantly degrade

the information environment within financial markets, leading to increased forecast errors and reduced price informativeness. These findings highlight the broad and potentially unintended economic consequences of privacy regulations in an increasingly interconnected digital economy.

Implications for Policy Makers Our study underscores the delicate balance policymakers must strike between protecting individual privacy and preserving the utility of alternative data in financial markets. As the use of alternative data continues to expand among financial participants, privacy initiatives like Apple's ATT policy can have far-reaching effects on market efficiency and investor behavior. This trend is expected to continue, as seen with Google's ongoing phase-out of third-party cookies in Chrome starting in 2024. The increasing regulatory scrutiny and the risks associated with data provenance, as highlighted by recent cases like that of App Annie, illustrate the potential ripple effects of data protection initiatives.

Policymakers need to consider these implications carefully. While the protection of personal data is essential, regulations should be designed to minimize disruptions to the data-driven insights that are increasingly vital to market participants. This may involve developing standardized frameworks for anonymized data sharing that ensure compliance with privacy laws while still providing valuable information to investors.

Future work Future research should explore the long-term impacts of privacy regulations on market efficiency, particularly as firms adapt to new data environments. Additionally, investigating the unequal access to alternative data among different types of investors, and how this disparity might be exacerbated by privacy restrictions, could provide valuable insights. Understanding which industries or market participants are most affected by data sharing limitations will help in crafting more targeted policy interventions. Lastly, examining technological advancements that could preserve data utility while ensuring privacy will be crucial in navigating the evolving landscape of data-driven financial markets.

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 (2021). A day in the life of your data. https://www.apple.com/au/privacy/docs/A_Day_in_the_Life_of_Your_Data.pdf. Accessed August 02, 2024.

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.

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., 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.

Bruce, N. I., B. Murthi, and R. C. Rao (2017). A dynamic model for digital advertising: The effects of creative format, message content, and targeting on engagement. *Journal of Marketing Research 54*(2), 202–218.

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 (2023). The use and usefulness of big data in finance: Evidence from financial analysts. *Nanyang Business School Research Paper* (22-01).

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.

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

Dahlén, M. (2001). Banner advertisements through a new lens. *Journal of Advertising Research 41*(4), 23–30.

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.

Dugast, J. and T. Foucault (2018). Data abundance and asset price informativeness. *Journal of Financial economics 130*(2), 367–391.

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*.

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

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

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.

Hong, W., J. Y. Thong, and K. Y. Tam (2007). How do web users respond to non-banner-ads animation? the effects of task type and user experience. *Journal of the American Society for Information Science and Technology 58*(10), 1467–1482.

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

iFans (2021). Privacy on iphone | tracked | apple. https://www.youtube.com/watch?v=fjf7c-O5GyQ. Youtube. Accessed August 02, 2024.

Iyer, G., D. Soberman, and J. M. Villas-Boas (2005). The targeting of advertising. *Marketing Science 24*(3), 461–476.

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.

Johnson, G. A., S. K. Shriver, and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science 39*(1), 33–51.

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

Kraft, L., B. Skiera, and T. Koschella (2023). Economic impact of opt-in versus opt-out requirements for personal data usage: the case of appleâs app tracking transparency (att). *Available at SSRN 4598472*.

Li, H. and J. L. Bukovac (1999). Cognitive impact of banner ad characteristics: An experimental study. *Journalism & Mass Communication Quarterly 76*(2), 341–353.

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.

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.

StatCounter (2024). Market share of mobile operating systems worldwide from 2009 to 2024, by quarter. https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/. Statista. Accessed August 02, 2024.

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). Number of smartphone users worldwide from 2014 to 2029. https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world. Accessed August 02, 2024.

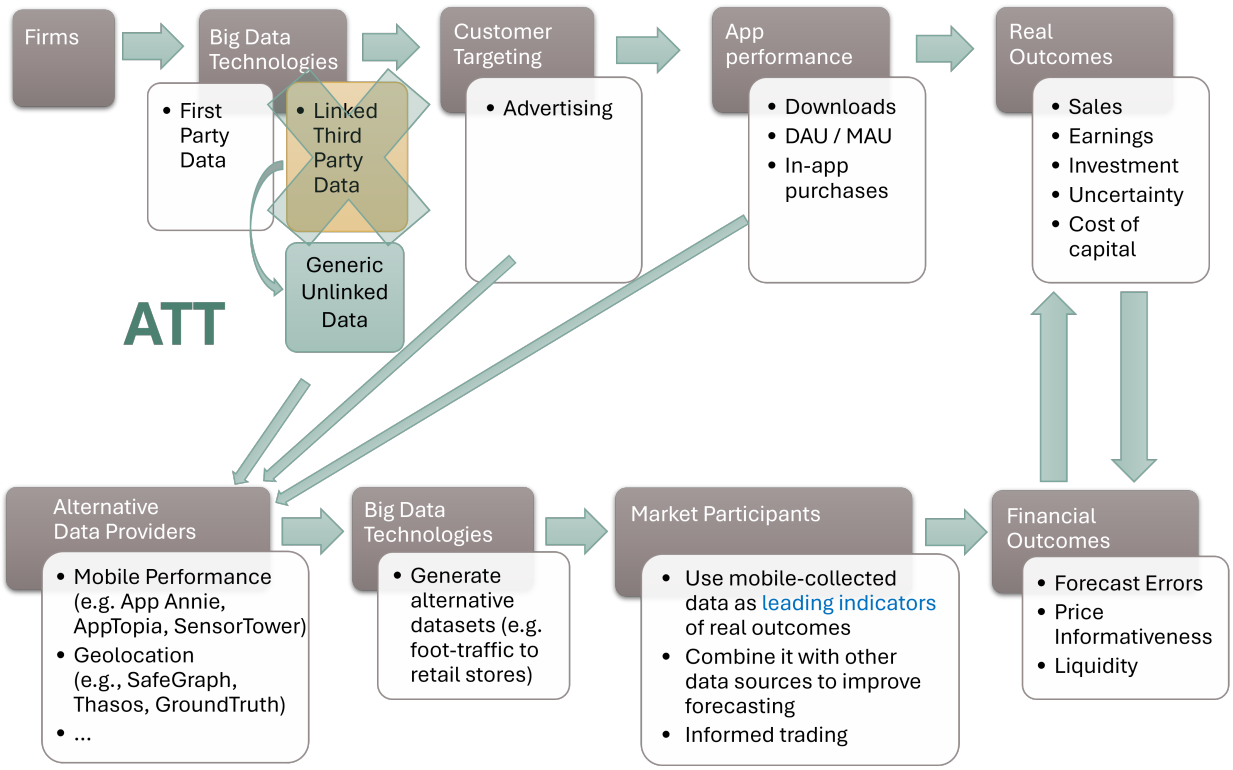
Statista (2024d). Percentage of population using the internet in the united states from 2000 to 2024. https://www.statista.com/statistics/209117/us-internet-penetration/. Accessed August 02, 2024.

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

Veldkamp, L. and C. Chung (2024). Data and the aggregate economy. *Journal of Economic Literature 62*(2), 458–484.

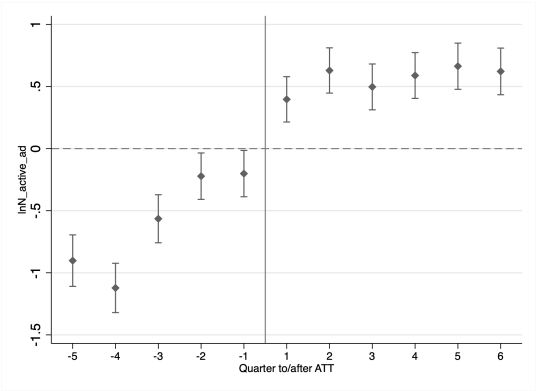
Zhu, C. (2019). Big data as a governance mechanism. *The Review of Financial Studies 32*(5), 2021–2061.

Figure 1: Mechanism

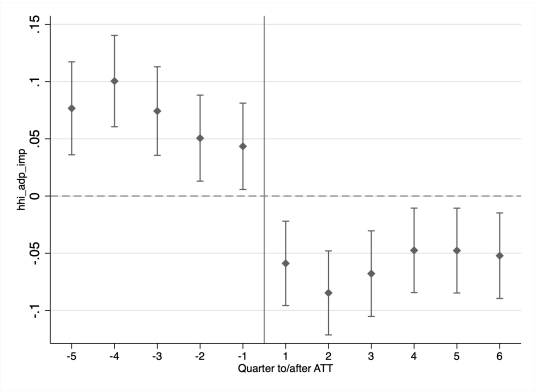


NOTE.— Figure 1 illustrates the logical links in our study. Firms collect data about user behavior through apps. That can be first party data (i.e., directly collected within the app) or third party data (i.e., obtained from third parties). Firms can share some or all of that data with third parties (e.g., data brokers or firms like Facebook), either for a price or in exchange for services (e.g., data analysis or targeting insights). Firms (either directly or through partners like Facebook) analyze first party and third party data, often through AI- models, to draw better insights about their users and improve engagement/targeting. A key targeting tool at firms' disposal is mobile advertising. Customer targeting then drives traffic (downloads, MAU, DAU) to a firm's app. The quality of that traffic is linked to customer engagement and ultimately to firm performance. At various stages of that process firms can share more data with third parties, as for instance mobile advertising or app usage information. Alternative data providers acquire data about users (e.g, geo-location or transactions), mobile advertising, and app usage, either from firms or from data brokers. They then use statistical models, often AI-driven, to analyze that unstructured big data and produce structured datasets which can be sold to financial market participants. Market participants use signals obtained from that alternative data (e.g., foot-traffic or app downloads), in combination with other signals, to update their priors about firms' future payoffs. They trade on those insights, so driving the price of securities.

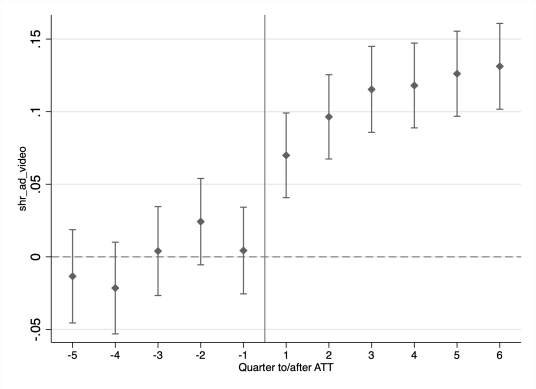
Figure 2: Advertising Activity around ATT



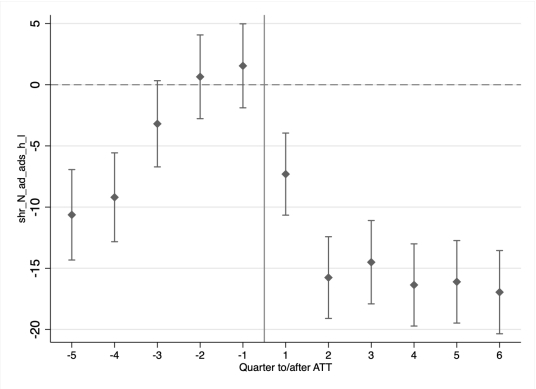
(a) Number of active ads



(b) Concentration of Ad Publishers



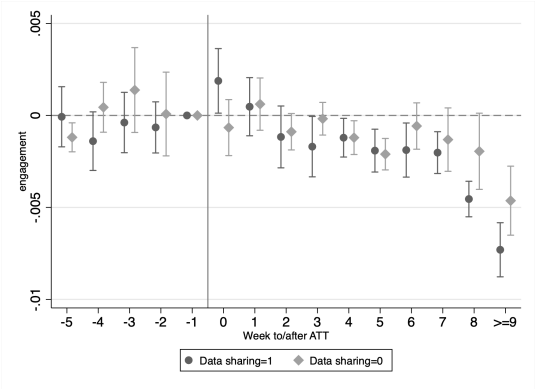
(c) Share of video ads



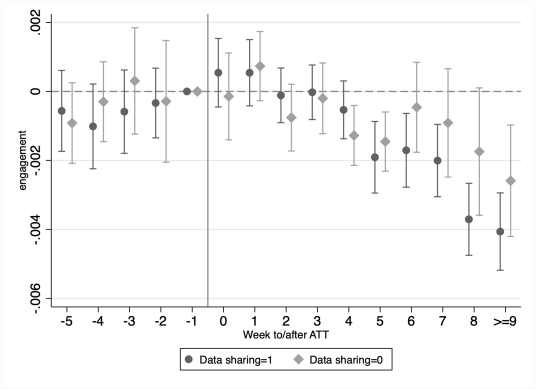
(d) Share of large ads

NOTE.—Figure 2 presents the estimated coefficients of the time dummies, as specified in Equation (4). We include dummies for each quarter from $t = -5$ to $t = 6$, with $t = 0$ (the quarter of ATT implementation, 2021Q2) serving as the reference group. The outcome variables in panels (a) through (d) are: (a) the number of active ads (log-transformed), (b) publisher concentration measured by the Herfindahl-Hirschman Index (HHI), (c) the share of video ads, and (d) the share of large static ads, respectively.

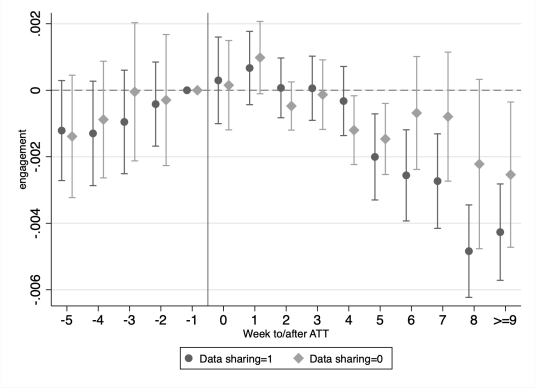
Figure 3: App Traffic vs. Engagement around ATT



(a) x=ln(Downloads)



(b) x=ln(DAU)



(c) x=ln(MAU)

NOTE.—Figure 3 plots the estimated coefficients of the time dummies, as specified in Equation (5). We include dummies for each week from $t = -5$ to $t = 8$, with $t \geq 9$ representing all weeks from week 9 onward post-ATT. The reference group is $t = 0$, corresponding to the week of ATT implementation (April 26, 2021). The outcome variables in panels (a) through (c) are downloads, DAU, and MAU, all in natural logarithms.

Figure 4: CAR of a Long-Short Porfolio Sorted on App Downloads



NOTE.— Figure 4 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-Carhart 6-factor model.

Table 1: Summary Statistics Firms

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE	-0.001	5.87	-1.95	-0.37	0.10	0.53	1.99	16,352
CAR - CAPM	-0.050	10.67	-11.68	-5.14	0.00	4.92	11.47	16,892
CAR - FF6	-0.049	10.46	-11.38	-4.98	0.00	4.69	11.19	16,892
Sales growth	0.086	0.30	-0.14	-0.02	0.07	0.18	0.36	16,069
L4.Size	22.868	2.12	20.23	21.35	22.81	24.27	25.73	16,118
L4.Cash/Assets	18.412	18.80	1.87	4.53	11.67	25.61	47.69	16,110
L4.Tangibles/Assets	20.383	21.54	1.26	4.33	11.62	29.82	56.69	15,766
L4.Debt/Assets	30.877	23.63	2.49	11.78	28.25	44.50	61.12	15,970
Specialized analyst (0/1)	0.174	0.38	0.00	0.00	0.00	0.00	1.00	16,964
Data sharing (1/0)	0.367	0.48	0.00	0.00	0.00	1.00	1.00	16,964
Downloads	12.565	2.50	9.45	10.86	12.55	14.19	15.73	16,964
MAU	12.814	2.62	9.59	11.09	12.81	14.46	16.05	16,964
DAU	11.353	2.94	7.64	9.35	11.47	13.32	14.89	16,964
Engagement	0.268	0.12	0.12	0.18	0.26	0.34	0.40	16,964

Panel B. Ad Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE	-0.173	6.08	-2.70	-0.52	0.09	0.68	2.73	3,397
CAR - CAPM	-0.225	11.30	-12.81	-6.06	0.00	5.34	12.42	3,544
CAR - FF6	-0.190	10.99	-12.34	-5.81	0.00	5.09	11.95	3,544
Sales growth	0.104	0.39	-0.21	-0.03	0.08	0.24	0.49	3,331
L4.Size	23.200	1.98	20.74	21.72	23.03	24.52	26.18	3,344
L4.Cash/Assets	21.313	19.83	2.62	6.11	14.38	31.02	52.44	3,344
L4.Tangibles/Assets	24.447	22.43	2.00	6.02	16.01	40.17	61.14	3,327
L4.Debt/Assets	39.289	25.49	8.73	19.77	37.31	52.15	72.54	3,314
Specialized analyst (0/1)	0.324	0.47	0.00	0.00	0.00	1.00	1.00	3,566
Data sharing (1/0)	0.640	0.48	0.00	0.00	1.00	1.00	1.00	3,566
Downloads	14.523	1.99	12.22	13.24	14.38	15.71	17.34	3,566
MAU	14.858	2.12	12.47	13.53	14.65	16.03	17.69	3,566
DAU	13.753	2.25	11.20	12.35	13.64	15.01	16.80	3,566
Engagement	0.362	0.11	0.24	0.30	0.36	0.41	0.46	3,566
Active ads	5.483	2.83	1.61	3.14	5.64	7.63	9.07	3,566
Ads concentration	0.492	0.26	0.20	0.30	0.44	0.65	0.93	3,566
Video ads	17.209	23.40	0.00	0.00	6.62	26.34	50.00	3,566
Large banner	35.141	25.06	0.00	14.37	35.28	51.22	64.67	3,566

Panel C. No-App Sample

	mean	sd	p10	p25	p50	p75	p90	count
SUE	0.918	11.95	-3.79	-0.78	0.11	1.16	5.17	86,132
CAR - CAPM	-0.289	11.86	-13.48	-5.59	0.00	4.96	12.35	91,182
CAR - FF6	-0.041	11.68	-12.88	-5.24	0.00	5.06	12.53	91,182
Sales growth	0.093	0.50	-0.27	-0.04	0.07	0.22	0.50	77,586
L4.Size	20.678	2.03	17.97	19.29	20.77	22.05	23.18	81,458
L4.Cash/Assets	23.013	27.81	1.19	3.40	10.05	31.77	74.48	81,379
L4.Tangibles/Assets	21.542	25.90	0.53	2.03	10.05	31.68	67.73	77,247
L4.Debt/Assets	26.423	24.70	0.04	4.81	21.33	41.86	59.34	80,777
Specialized analyst (0/1)	0.019	0.14	0.00	0.00	0.00	0.00	0.00	92,719

NOTE.—Panel A includes all firms that have an app, for which we have advertising data; Panel B includes all firms with an app; Panel C includes all firms without an app.

Table 2: Summary Statistics Firm by Data Sharing Policy

Panel A. Data-sharing Firms

	mean	sd	p10	p25	p50	p75	p90	count
SUE	-0.081	5.57	-2.48	-0.54	0.09	0.65	2.36	5,956
CAR - CAPM	-0.221	10.85	-12.25	-5.48	-0.00	4.89	11.52	6,195
CAR - FF6	-0.204	10.63	-12.03	-5.38	-0.00	4.80	11.38	6,195
Sales growth	0.082	0.33	-0.17	-0.03	0.06	0.18	0.39	5,824
L4.Size	22.986	2.10	20.40	21.51	22.81	24.27	26.07	5,872
L4.Cash/Assets	18.909	18.22	1.93	4.91	12.80	27.44	47.22	5,865
L4.Tangibles/Assets	22.734	21.02	1.89	5.62	15.46	35.61	57.15	5,753
L4.Debt/Assets	35.127	26.32	3.87	15.56	31.63	48.88	70.64	5,837
Specialized analyst (0/1)	0.260	0.44	0.00	0.00	0.00	1.00	1.00	6,227
Downloads	14.102	2.18	11.35	12.77	14.05	15.51	16.92	6,227
MAU	14.411	2.30	11.57	13.07	14.32	15.85	17.35	6,227
DAU	13.147	2.49	9.96	11.73	13.14	14.67	16.20	6,227
Engagement	0.318	0.12	0.19	0.25	0.31	0.37	0.43	6,227

Panel B. Non-data-sharing Firms

	mean	sd	p10	p25	p50	p75	p90	count
SUE	0.044	6.04	-1.62	-0.29	0.10	0.48	1.79	10,396
CAR - CAPM	0.049	10.56	-11.40	-4.96	0.00	4.94	11.41	10,697
CAR - FF6	0.041	10.37	-10.97	-4.74	0.00	4.64	11.14	10,697
Sales growth	0.088	0.29	-0.13	-0.01	0.07	0.19	0.35	10,245
L4.Size	22.801	2.12	20.11	21.28	22.81	24.26	25.56	10,246
L4.Cash/Assets	18.127	19.12	1.83	4.35	11.10	24.64	48.19	10,245
L4.Tangibles/Assets	19.032	21.71	1.10	3.74	10.41	25.26	56.27	10,013
L4.Debt/Assets	28.428	21.55	1.83	9.96	26.50	41.71	55.88	10,133
Specialized analyst (0/1)	0.123	0.33	0.00	0.00	0.00	0.00	1.00	10,737
Downloads	11.674	2.22	9.02	10.15	11.70	13.17	14.44	10,737
MAU	11.888	2.34	9.15	10.36	11.95	13.45	14.72	10,737
DAU	10.313	2.66	7.14	8.47	10.37	12.14	13.67	10,737
Engagement	0.239	0.11	0.11	0.16	0.23	0.31	0.38	10,737

NOTE.—Panel A includes all firms that have an app, and share user data with third parties; Panel B includes all firms that have an an app, and do not share user data with third parties.

Table 3: Summary Statistics Analysts

Panel A. App Sample

	mean	sd	p10	p25	p50	p75	p90	count
Specialized analyst (0/1)	0.082	0.27	0.00	0.00	0.00	0.00	0.00	151,101
AFE	0.204	0.28	0.01	0.03	0.09	0.23	0.57	151,101
PMEAFE	-0.025	0.53	-0.71	-0.35	-0.03	0.20	0.63	151,101
Forecast age	3.535	1.02	2.08	2.77	3.89	4.44	4.51	151,101
Analyst/Firm experience	12.779	9.08	1.10	5.23	11.51	19.37	26.37	151,101
Analyst experience	18.313	8.65	7.81	11.58	17.88	25.23	30.38	151,101
#Firms covered	2.570	0.56	1.79	2.30	2.64	2.94	3.18	151,101
Broker size	3.548	0.93	2.20	3.04	3.85	4.23	4.53	151,101

Panel B. Placebo Sample

	mean	sd	p10	p25	p50	p75	p90	count
Specialized analyst (0/1)	0.014	0.12	0.00	0.00	0.00	0.00	0.00	393,200
AFE	0.165	0.24	0.01	0.03	0.08	0.19	0.42	393,200
PMEAFE	-0.021	0.55	-0.73	-0.36	-0.03	0.23	0.67	393,200
Forecast age	3.663	0.89	2.40	3.09	4.01	4.43	4.51	393,200
Analyst/Firm experience	10.972	8.67	0.47	3.76	9.28	16.38	24.27	393,200
Analyst experience	17.793	8.65	7.38	10.96	17.12	24.74	30.13	393,200
#Firms covered	2.621	0.58	1.79	2.30	2.71	3.00	3.26	393,200
Broker size	3.479	0.93	2.20	2.94	3.71	4.16	4.51	393,200

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).

Table 4: Price Informativeness by Analyst Specialization

	CAR - CAPM			CAR - FF6		
	(1)	(2)	(3)	(4)	(5)	(6)
SUE	0.098***	0.119***	0.101***	0.098***	0.122***	0.099***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ATT × SUE	0.022	-0.020	-0.013	0.020	-0.024	-0.001
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
Specialized × SUE		-0.162***	-0.179***		-0.188***	-0.188***
		(0.06)	(0.06)		(0.05)	(0.06)
ATT × Specialized × SUE		0.285***	0.304***		0.307***	0.308***
		(0.09)	(0.09)		(0.09)	(0.09)
Specialized		-0.243	-0.423		-0.129	-0.324
		(0.44)	(0.44)		(0.43)	(0.43)
ATT × Specialized		-0.689	-0.398		-0.449	-0.168
		(0.49)	(0.51)		(0.49)	(0.50)
L4.Size			-1.664***			-1.400***
			(0.32)			(0.32)
L4.Cash/Assets			0.008			0.002
			(0.01)			(0.01)
L4.Tangibles/Assets			0.010			0.004
			(0.02)			(0.02)
L4.Debt/Assets			-0.001			-0.003
			(0.01)			(0.01)
Constant	-0.042	0.049	37.690***	-0.042	0.014	31.916***
	(0.08)	(0.11)	(7.36)	(0.08)	(0.10)	(7.23)
Firm FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y
Observations	16,344	16,344	15,543	16,344	16,344	15,543
R-sq	0.062	0.063	0.065	0.060	0.061	0.062

NOTE.—Table 4 reports the estimation results of Equation (2) based on the app sample. CAR is measured using the CAPM in Columns 1-3 and the Fama-French-Carhart 6-factor model in Columns 4-6. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT announcement. The variable *Specialized* is an indicator equal to 1 if the firm is followed by at least one analyst specialized in alternative data during quarter t. SUE is calculated as specified in Equation (1), assuming that earnings per share (EPS) follow a seasonal random walk. The model includes firm fixed effects (θ_i) and year-quarter fixed effects (ι_t), with standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Price Informativeness by Firms' Data Sharing Policy

	CAR - CAPM			CAR - FF6		
	(1)	(2)	(3)	(4)	(5)	(6)
SUE	0.098***	0.155***	0.130***	0.098***	0.152***	0.126***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ATT × SUE	0.022	-0.087**	-0.062	0.020	-0.059	-0.025
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
Data sharing × SUE		-0.211***	-0.195***		-0.200***	-0.187***
		(0.04)	(0.04)		(0.04)	(0.04)
ATT × Data sharing × SUE		0.327***	0.286***		0.257***	0.230***
		(0.07)	(0.07)		(0.06)	(0.07)
Data sharing		0.779	0.907		0.934	1.008
		(0.74)	(0.75)		(0.73)	(0.74)
ATT × Data sharing		-0.920**	-1.003**		-0.960**	-1.002***
		(0.38)	(0.39)		(0.38)	(0.39)
L4.Size			-1.679***			-1.398***
			(0.32)			(0.31)
L4.Cash/Assets			0.011			0.004
			(0.01)			(0.01)
L4.Tangibles/Assets			0.013			0.007
			(0.02)			(0.02)
L4.Debt/Assets			-0.001			-0.002
			(0.01)			(0.01)
Constant	-0.042	-0.201	37.610***	-0.042	-0.252	31.454***
	(0.08)	(0.28)	(7.33)	(0.08)	(0.27)	(7.21)
Firm FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y
Observations	16,344	16,344	15,543	16,344	16,344	15,543
R-sq	0.062	0.064	0.066	0.060	0.062	0.063

NOTE.—Table 5 reports the estimation results of Equation (2), where we subsitute *Specialized* with *Datasharing*, an indicator variable equal to 1 if firms share/track user data obtained through mobile apps. We report the results based on app sample. CAR is measured using the CAPM in Columns 1-3 and the Fama-French-Carhart 6-factor model in Columns 4-6. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT announcement. *SUE* is calculated as specified in Equation (1), assuming that earnings per share (EPS) follow a seasonal random walk. The model includes firm fixed effects (θ_i) and year-quarter fixed effects (ι_t), with standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Price Informativeness by Analyst Specialization Placebo

	CAR - CAPM			CAR - FF6		
	(1)	(2)	(3)	(4)	(5)	(6)
SUE	0.074***	0.073***	0.083***	0.068***	0.067***	0.077***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
ATT × SUE	-0.015**	-0.016**	-0.006	-0.004	-0.005	0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Specialized × SUE		0.145**	0.120*		0.130*	0.113
		(0.07)	(0.07)		(0.07)	(0.07)
ATT × Specialized × SUE		-0.039	-0.100		-0.025	-0.101
		(0.09)	(0.11)		(0.09)	(0.10)
Specialized		-0.558	-0.523		-0.417	-0.337
		(0.58)	(0.58)		(0.57)	(0.58)
ATT × Specialized		0.179	0.518		0.150	0.435
		(0.77)	(0.80)		(0.76)	(0.80)
L4.Size			-0.689***			-0.577***
			(0.14)			(0.14)
L4.Cash/Assets			-0.010*			-0.014**
			(0.01)			(0.01)
L4.Tangibles/Assets			-0.001			-0.002
			(0.01)			(0.01)
L4.Debt/Assets			0.000			0.005
			(0.00)			(0.00)
Constant	-0.322***	-0.313***	14.255***	-0.077*	-0.069*	12.180***
	(0.04)	(0.04)	(3.01)	(0.04)	(0.04)	(2.97)
Firm FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y
Observations	85,838	85,838	75,160	85,838	85,838	75,160
R-sq	0.101	0.101	0.101	0.094	0.094	0.094

NOTE.—Table 6 reports the estimation results of Equation (2) based on the placebo firms, defined as firms that do not have an app. CAR is measured using the CAPM in Columns 1-3 and the Fama-French-Carhart 6-factor model in Columns 4-6. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT announcement. The variable *Specialized* is an indicator equal to 1 if the firm is followed by at least one analyst specialized in alternative data during a given quarter. SUE is calculated as specified in Equation (1), assuming that earnings per share (EPS) follow a seasonal random walk. The model includes firm and year-quarter fixed effects, with standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Analysts' Excess Forecast Errors

	Excess Forecast Error: PMAFE					
	Specialized Analyst			Keyword Usage		
	(1)	(2)	(3)	(4)	(5)	(6)
Specialized Analyst (0/1)	-0.016	-0.021	-0.020			
	(0.01)	(0.02)	(0.02)			
ATT × Specialized Analyst (0/1)	0.058***	0.046**	0.046**			
	(0.02)	(0.02)	(0.02)			
Keyword Usage (0/1)				-0.041***		
				(0.01)		
ATT × Keyword Usage (0/1)				0.040**	0.049**	0.051**
				(0.02)	(0.02)	(0.02)
Forecast age			-0.001			-0.001
			(0.01)			(0.01)
Analyst/Firm experience			-0.001**			-0.006
			(0.00)			(0.01)
Analyst experience			0.000			
			(0.00)			
#Firms covered			-0.003			-0.011
			(0.01)			(0.02)
Broker size			-0.003			-0.035
			(0.01)			(0.03)
Constant	-0.026***	-0.025***	0.011	-0.004	-0.040***	0.156
	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)	(0.06)
Firm-Quater FE	Y	Y	Y	Y	Y	Y
Estimation date FE	Y	Y	Y	Y	Y	Y
Analyst FE	N	Y	Y	N	Y	Y
Brokerage FE	N	Y	Y	N	Y	Y
Observations	151,371	151,101	151,101	44,961	44,960	44,960
R-sq	0.041	0.093	0.093	0.225	0.248	0.248

NOTE.— Table 7 reports the estimation results of Equation (3) based on the app sample. $PMAFE$, or Proportional Mean Absolute Forecast Errors, is Absolute Forecast Errors (AFE), the difference between the analyst's forecast and actual earnings scaled by the mean absolute forecast error across all analysts making forecasts for the same firm in that quarter. We computed $PMAFE$ for each quarter and each firm-analyst pair. The variable ATT is an indicator equal to 1 for all quarters following the ATT announcement. The variable *Specialized* is an indicator equal to 1 if the firm is followed by at least one analyst specialized in alternative data during a given quarter. To construct $Keyword\ Usage$, we consider analysts who issued more than 100 reports over 2019-2020 and whose names we can match to I/B/E/S, ensuring that this variable is constant at the analyst level. $Keyword\ Usage$ is an indicator set to 1 if the analyst mentioned app-related keywords more than 10 times, as listed in Table A.2. The results for *Specialized* and $Keyweord\ Usage$ are reported in Columns 1-3 and Columns 4-6, respectively. Regression controls include the age of the forecast, analyst experience (since first issuing forecasts and in relation to each firm), number of firms covered, and size of the brokerage firm. In Columns 1 and 3, we include firm-quarter and estimate-date fixed effects. In all other columns, we further include analyst and brokerage 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.

Table 8: Analysts' Excess Forecast Errors Placebo

	Excess Forecast Error: PMAFE					
	Specialized Analyst			Keyword Usage		
	(1)	(2)	(3)	(4)	(5)	(6)
Specialized analyst (0/1)	-0.020	0.011	0.012			
	(0.02)	(0.02)	(0.02)			
ATT × Specialized analyst (0/1)	0.047*	0.021	0.021			
	(0.03)	(0.03)	(0.03)			
Keyword Usage (0/1)				-0.023**		
				(0.01)		
ATT × Keyword Usage (0/1)				0.006	-0.010	-0.009
				(0.02)	(0.02)	(0.02)
Forecast age			0.015**			0.010
			(0.01)			(0.01)
Analyst/Firm experience			-0.000			-0.000
			(0.00)			(0.00)
Analyst experience			0.000			
			(0.00)			
#Firms covered			-0.005			-0.000
			(0.00)			(0.01)
Broker size			-0.017*			-0.001
			(0.01)			(0.03)
Constant	-0.021***	-0.022***	-0.001	-0.015***	-0.026***	-0.052
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.11)
Firm-Quater FE	Y	Y	Y	Y	Y	Y
Estimation date FE	Y	Y	Y	Y	Y	Y
Analyst FE	N	Y	Y	N	Y	Y
Brokerage FE	N	Y	Y	N	Y	Y
Observations	393,594	393,200	393,200	70,468	70,468	70,468
R-sq	0.026	0.062	0.062	0.259	0.277	0.277

NOTE.— Table 8 reports the estimation results of Equation (3) 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). $PMAFE$, or Proportional Mean Absolute Forecast Errors, is Absolute Forecast Errors (AFE), the difference between the analyst's forecast and actual earnings scaled by the mean absolute forecast error across all analysts making forecasts for the same firm in that quarter. We computed $PMAFE$ for each quarter and each firm-analyst pair. The variable ATT is an indicator equal to 1 for all quarters following the ATT announcement. The variable *Specialized* is an indicator equal to 1 if the firm is followed by at least one analyst specialized in alternative data during a given quarter. To construct $Keyword\ Usage$, we consider analysts who issued more than 100 reports over 2019-2020 and whose names we can match to I/B/E/S, ensuring that this variable is constant at the analyst level. $Keyword\ Usage$ is an indicator set to 1 if the analyst mentioned app-related keywords more than 10 times, as listed in Table A.2. The results for *Specialized* and $Keyweord\ Usage$ are reported in Columns 1-3 and Columns 4-6, respectively. Regression controls include the age of the forecast, analyst experience (since first issuing forecasts and in relation to each firm), number of firms covered, and size of the brokerage firm. In Columns 1 and 3, we include firm-quarter and estimate-date fixed effects. In all other columns, we further include analyst and brokerage 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.

Table 9: App Downloads and Firm Performance by Data Sharing Policy

	(1) Sales growth	(2) Sales growth	(3) SUE	(4) SUE
Downloads	0.040***	0.036***	0.253**	0.266***
	(0.01)	(0.01)	(0.10)	(0.10)
Data sharing × Downloads	0.029**	0.031***	0.214*	0.188*
	(0.01)	(0.01)	(0.12)	(0.11)
ATT × Downloads	-0.002	0.004	0.077	0.079
	(0.01)	(0.01)	(0.06)	(0.06)
ATT × Data sharing × Downloads	-0.018*	-0.020**	-0.246**	-0.235**
	(0.01)	(0.01)	(0.11)	(0.10)
Data sharing	-0.395**	-0.402***	-2.869*	-2.473
	(0.17)	(0.15)	(1.61)	(1.54)
ATT × Data sharing	0.259**	0.260**	3.206**	2.961**
	(0.13)	(0.12)	(1.54)	(1.41)
L.Downloads	-0.011*	-0.006	-0.175**	-0.180**
	(0.01)	(0.01)	(0.07)	(0.07)
L4.Size		-0.228***		-0.936***
		(0.02)		(0.24)
L4.Cash/Assets		0.004***		0.015**
		(0.00)		(0.01)
L4.Tangibles/Assets		-0.004***		-0.065***
		(0.00)		(0.02)
L4.Debt/Assets		0.004***		0.050***
		(0.00)		(0.01)
Constant	-0.287***	4.878***	-1.453*	19.722***
	(0.07)	(0.46)	(0.77)	(5.38)
Firm FE	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y
Sample	full	full	full	full
Observations	13,484	13,051	13,671	13,087
R-sq	0.229	0.282	0.069	0.074

NOTE.—Table 9 reports the estimation results of Equation (6) based on the app sample. The outcome variable is *Sales Growth* in Columns 1-2 and *SUE* in Columns 3-4. The *Downloads* represents the natural logarithm of total downloads for a firm's apps in a given quarter. *Data Sharing* is an indicator equal to 1 if any app of a firm tracks or shares user data. The variable *ATT* is an indicator equal to 1 for all quarters following the ATT announcement. Control variables include firm size, and cash, tangibles, and debt over assets, all measured in quarter $t-4$. All columns include firm and year-quarter fixed effects, with standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

For Online Publication:

Appendix to "Data Privacy Regulations and Financial Information Production"

Table A.1: List of Alternative Data Categories and Keywords

App Tracking Keywords
app tracking
app tracking transparency
ATT
california consumer privacy act
CCPA
consumer tracking
cookie
data collection
data privacy
data protection
data regulation
data security
developer guidelines
GDPR
general data protection regulation
identifier for advertisers
identifying information
IDFA
nutrition label
personal data
personal information
privacy framework
privacy label
privacy law
privacy nutrition label
privacy policy
privacy protection
privacy regulation
right to delete
right to non-discrimination
right to opt-out
targeted advertising
user data
user permission
user privacy
user tracking

Sentiment Keywords
brand sentiment
cms data
consumer sentiment
customer rating
customer review
customer satisfaction rating
customer satisfaction trend
facebook analysis
facebook data
facebook like
facebook post
facebook track
facebook user
guest sentiment
instagram data
instagram engagement
instagram follower
internet world stats
investing analytics
medicare plan finder
merchant centric
net sentiment
netbase
online review
proper insights
ratings on tripadvisor
review analytics
scoring released by cms
sentiment analysis
sentiment data
social media analysis
social media engagement
social media follower
stars rating
tracking on twitter
tripadvisor ratings
twitter analysis
twitter data
twitter purchase intent
twitter sentiment
web analytics
web mining
web scraping
yelp

Employee Keywords
earnings call transcript
indeed.com
job posting
job trend
mining of earnings calls
online hiring
text analytics
trascriptlytics
web analytics
web scraping
web mining

Geospatial Keywords
branch network model
branch rationalization tool
demographic analysis
foursquare
geospatial
market quality analysis
store overlap

Satellite Image Keywords
satellite image
orbital insight
remote sensing metrics
parking lot fill rate
parking lot traffic
proprietary satellite data
rs metrics
satellite analysis
satellite image
traffic analysis

Web Traffic Keywords
baidu analysis
baidu data
baidu search data
baidu search index
comscore
daily traffic
google search analysis
google trend
google-searched
iphone monitor
iphone tracker
scrapehero
scrape hero
search interest
search trend
search volume
smartphone tracker
thinknum
traffic monitor
web hit activity
web search
traffic analysis

Transaction Keywords
mobile transaction
device transaction
smartphone transaction
mobile app transaction
app transaction
mobile payment
digital wallet
mobile commerce
in-app transaction
in-app purchase
app payment

Point-of-Sale Keywords
1010data
1010 data
airbnb listing
compared online prices
discount tracker
financial rate monitor
first data spendtrend
footlocker
footwear scrapes
hotel tracker
hotel tracking
listing monitor
mastercard advisors
nielson
online price survey
proprietary data
point-of-sale
point of sale
price comparisons
price intelligence
price monitoring
price observations
pricing monitor
pricing study
pricing tracker
property listing
sg2
spend tracker
standard media index
superdata
vehicle listing
web analytics
web mining
web scraping
zillow
user transaction
mobile financial
transaction log
user spending
user purchase
digital transaction
digital purchase
contactless transaction
contactless payment
app purchase
mobile e-commerce
transaction activity

App Usage Keywords
active user
jiguang
similarweb
similar web
mau
dau
app use
app usage
downloads
active session
app activity
app engagement
engagement rate
retention rate
churn rate
in app time
daily sessions
monthly sessions
user interaction
session duration
session length
user cohort
app performance
user behavior

Other Keywords
dodge
dun & bradstreet
edmunds
entgroup
evidence lab macro
home improvement tracker
innovata
one click retail
ookla
opensignal
root metrics
str data
wait time monitor
lower end spending
m2m
macro-to-micro
network traffic lab
nowcast

Table A.2: List of Frequently Used Mobile Data Keywords

Alexa	dau
App Annie	downloads
App data	geospatial
AppAnnie	iResearch
Apptopia	mau
Comscore	point-of-sale
Envestnet	proprietary data
First Data	search interest
OTAS	search volume
QuestMobile	similar web
SafeGraph	similarweb
Sensor Tower	superdata
SensorTower	targeted advertising
Share This	thinknum
Trust Data	user behavior
UnionPay	user data
active user	user engagement
app usage	user spending
comscore	yelp
	zillow

Table A.3: List of Alternative Data Vendors

Advan Research Corporation
Advan
Alphamatician
AppAnnie
App Annie
BMLL Technologies Limited
BrandWatch
Brand Watch
Data Provider
Dataminr
DataPulse
Data Pulse
ExtractAlpha
Extract Alpha
GNIP
Heckyl
Here Technologies
LikeFolio
Like Folio
Percolata
Premise Data Corporation
ShareThis
Share This
ShopperTrak
Shopper Trak
SuperFly
Super Fly
TheySay
They Say
TickData
Tick Data
Wall Street Horizon
YipitData
Yipit Data
CargoMetrics
Cargo Metrics
ListenFirst
Listen First
MariTrace
Mari Trace
Orbital Insight
Informal Financial Intelligence
M&A Portal
OTAS
Owl Analytics
Refinitiv
Sigmai
Sustainalytics
Sandalwood Advisors
AllTheRooms
All The Rooms
FIRST TO INVEST
Social Market Analytics
Thinknum
Eagle Alpha
Stocksnips Inc
SecondMeasure
Second Measure
Thasos
VertoAnalytics
Verto Analytics

Kayrros
THDataCapital
TH Data Capital
Comlinkdata
Nowcast
Brickstream
Enigma
Gildata
MixRank
Mix Rank
Towergate Informatics
TrustData
Trust Data
Unacast
DigitalGlobe
Digital Globe
Buddy
Lenddo
Envestnet Yodlee
Yodlee / Envestnet
Yodlee
Envestnet
Bridg
iResearch
RootMetrics
Root Metrics
TMT Analysis
AreaMetrics
Area Metrics
Big One Lab
DataStreamX
Data Stream X
Foursquare
Four Square
Reveal Mobile
Skydeo
Uber Media
Accern
Alexa
AlphaSense
Alpha Sense
AYLIEN Limited
Data Provider
ENGAGE Research
Heckyl
Inferess
Percolata
Repustate
Scutify
Sentifi AG
SentifiAG
SESAMm
Social Market Analytics
SumZero
Sum Zero
TheySay
They Say
Wikimapia
Orbital Insight
Rakuten Intelligence
RetailNext

Retail Next
Sandalwood Advisors
ListenFirst
Listen First
Panvista Analytics
Placemeter
YipitData
Yipit Data
Eagle Alpha
Big Byte Insights
MarketCheck
Second Measure
FIRST TO INVEST
Social Market Analytics
Edison Software
Alt&A Advisory
Alt & A Advisory
Alt and A Advisory
Toyo Keizai
Nowcast
Brickstream
Enigma
Suzy
GB. Office for National Statistics
VertoAnalytics
Verto Analytics
FactorTrust
Factor Trust
ARM Insight
Comscore
Cuebiq
AreaMetrics
Area Metrics
Associated Press
GfK Boutique Research
Lota Data
Reveal Mobile
Sharablee
Tutela
TalkingData
Talking Data
QuestMobile
Quest Mobile
ARM Insight
Affinity Solutions
First Data
MasterCard Advisors
Master Card Advisors
UnionPay
Union Pay
PlaceIQ
Place IQ
SafeGraph
Safe Graph
SensorTower
Sensor Tower
Appdata
App data
Apptopia
Pathmatics

Table A.4: List of Big Data Companies

PERMNO	Company Name	PERMNO	Company Name
14593	APPLE COMPUTER INC	15850	MATCH GROUP INC
20190	AIRBNB INC	89393	NETFLIX INC
84788	AMAZON COM INC	20130	NERDY INC
13106	ANGIES LIST INC	19455	OPENDOOR TECHNOLOGIES INC
86783	BOOKING HOLDINGS INC	17782	SPOTIFY TECHNOLOGY S A
20471	BUMBLE INC	21125	ZIPRECRUITER INC
17315	CARDLYTICS INC	18485	PINTEREST INC
14273	CHEGG INC	19196	PORCH GROUP INC
18727	CHEWY INC	18940	PELOTON INTERACTIVE INC
14240	CRITEO S A	16829	REDFIN CORP
20189	DOORDASH INC	16932	ROKU INC
18063	EVENTBRITE	20441	ROVER GROUP INC
86356	EBAY INC	17038	STITCH FIX INC
17880	EVERQUOTE INC	16595	SNAP INC
90808	EXPEDIA INC	13168	TRIPADVISOR INC
18077	FARFETCH LTD	16535	TRIVAGO
18064	FRONTDOOR INC	18576	UBER TECHNOLOGIES INC
18768	FIVERR INTERNATIONAL LTD	22610	VACASA INC
19679	GOODRX HOLDINGS INC	21120	VIMEO INC
90319	GOOGLE INC	20659	VIZIO HOLDING CORP
78840	IAC INTERACTIVECORP	21593	XOMETRY INC
18452	LYFT INC	13318	YELP INC
19859	MEDIAALPHA INC	87075	ZIFF DAVIS INC
13407	META PLATFORMS	12927	ZILLOW INC