CONSUMER SURVEILLANCE AND FINANCIAL FRAUD

Bo Bian
Michaela Pagel
Huan Tang

WORKING PAPER 31692

NBER WORKING PAPER SERIES

CONSUMER SURVEILLANCE AND FINANCIAL FRAUD

Bo Bian
Michaela Pagel
Huan Tang

Working Paper 31692
http://www.nber.org/papers/w31692

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2023

We thank Simona Abis, Hunt Allcott, Tania Babina, Nathan Blascak, Samuel Kruger, Kai Li, Markus Mobius, David Rothschild, Jonathan Wallen, Constantine Yannelis, and Anthony Lee Zhang for valuable comments as well as seminar and conference participants at the Chicago Conference on Empirical Finance, Future of Financial Information Conference, Household Finance Brownbag Series, IMF, Microsoft Research, NBER SI (Household Finance), University College London, UW-Madison Junior Finance Conference, and University of Innsbruck. Niels Wagner provided outstanding research assistance. We also thank Amber Howe from the Federal Trade Commission for providing the financial fraud data in response to our freedom of information act (FOIA) request. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Consumer Surveillance and Financial Fraud
Bo Bian, Michaela Pagel, and Huan Tang
NBER Working Paper No. 31692
September 2023
JEL No. G5

ABSTRACT

Companies near constantly surveil their customers to collect, analyze, and profit from their private information. A prevailing concern is that the market for private data and security breaches expose consumers to financial fraud. In this study, we exploit Apple's App Tracking Transparency (ATT) policy, which greatly limited the tracking and sharing of personal information on the iOS platform, providing a major shock to the data industry. Using a difference-in-differences design and granular variations in iOS user shares across the US, we find that if 10% more people disallow tracking, the number of financial fraud complaints in the average zip code decreases by approximately 3.21%. We then show that the effects are concentrated in complaints related to lax data security and privacy, identified using keyword searches and machine learning on complaint narratives, and in complaints about firms that engage in intensive consumer surveillance and lack data safeguards. Our evidence quantifies one of the main consumer costs of lax data security standards.

Bo Bian
UBC Sauder School of Business
2053 Main Mall, Vancouver, BC V6T 1Z2
Vancouver
Canada
bo.bian@sauder.ubc.ca

Michaela Pagel
Washington University
Olin Business School
John Simon Hall 242
1 Brookings Drive
St. Louis, MO 63130
and NBER
mpagel@wustl.edu

Huan Tang
The Wharton School,
University of Pennsylvania
3620 Locust Walk
2300 Steinberg-Dietrich Hall
Philadelphia, PA 19104
huan.ht.tang@gmail.com

1 Introduction

We live in a commercial surveillance economy. People use many technologies, more or less essential to everyday life, that are enabled to near constantly surveil their private lives. At the same time, the different forms of commercial surveillance practices might be very opaque to consumers, which limits the effectiveness of consumer consent and public scrutiny. The collecting, tracking, sharing, and selling of private data in today's internet economy may expose people's identities to hackers and thieves and may have heightened the risks and stakes of deception, manipulation, and other abuses by fraudsters. In response to these risks, the European Union (EU) strengthened its data protection standards in 2018, and the state of California (CA) followed suit in 2020. At the federal level in the United States, the Federal Trade Commission (FTC) and Consumer Financial Protection Bureau (CFPB) are currently considering new rules to protect people's privacy and enhance data security.

Lax data privacy measures, security breaches, and the market for personal data can enable financial fraudsters to target and harm consumers. Law enforcement and federal agencies receive over one million fraud complaints each year, with more than half of the consumers filing complaints reporting financial losses. The FTC estimates that approximately 10% of U.S. adults fall victim to fraud each year, and consumers reported losing over $5.8 billion to fraud in 2021 alone. The National White Collar Crime Center estimates a prevalence rate of 17% (Huff et al., 2010), and a survey conducted by DeLiema et al. (2017) found that half of the respondents reported victimization by one or more major categories of fraud.

In this paper, we ask whether an industry-led initiative that majorly limited the tracking and sharing of personal information across apps and websites reduced financial fraud. Our findings are useful for answering the broader question of whether or not we should implement new regulations, rules, or other regulatory alternatives concerning the ways in which companies collect, aggregate, protect, use, analyze, and retain consumer data, as well as transfer, share, sell, or otherwise monetize that data.

Specifically, we investigate the impact of Apple's Tracking Transparency (ATT) policy on

financial fraud. We utilize the ATT policy as a source of variation in consumer surveillance practices. The ATT policy, introduced by Apple on April 26, 2021, required all apps to obtain explicit user permission before tracking users across apps or websites owned by other companies. By default, users were opted-out of tracking, i.e., Apple would not provide apps and websites with user identifiers any more unless they obtained permission. This policy significantly limited the tracking of users in the mobile app market, and as of February 2022, only 18% of app users allowed tracking among those who were asked for permissions.[1]

We thus consider the implementation of ATT as an event that enhances data security and privacy standards. We then exploit the fact that ATT only affects iOS users, but not Android users. To capture variations in the exposure to ATT, we leverage detailed foot traffic data from Safegraph to calculate zip-code-level shares of iPhone users out of all smartphone users. This allows us to examine the effects of the limited tracking of consumers in zip codes with different shares of iPhone users.

In our analysis, we focus on two outcome variables: the number of financial fraud complaints and the amount of money lost due to fraud, obtained from the CFPB and FTC. Our results demonstrate that limiting the tracking and sharing of personal information has a significant impact on reducing financial fraud. Specifically, our analysis of CFPB complaints shows that a 10% increase in the share of Apple users in a zip code leads to a 2.63% reduction in the number of financial fraud complaints. Accounting for the 82% opt-out rate of ATT, this translates to a 3.21% reduction in financial fraud complaints due to the tracking limits. We also establish that areas with high and low iOS shares experience similar pre-treatment trends in the likelihood and number of financial fraud complaints.

Moreover, we investigate how the treatment effect varies across different demographic groups. Our results show that the treatment effect is stronger for minorities, women, and younger people, suggesting that these groups are more vulnerable to fraud. Our findings

[1]See source: https://www.flurry.com/blog/att-opt-in-rate-monthly-updates/. The ATT policy was a major shock to the data industry. The entire industry moved away from using "MIDs" (mobile IDs) to email addresses as the primary way to link and track users across databases from apps and websites.

therefore contribute to the ongoing discussion on the distributional effects of new regulations and rules that enhance consumer data protection and privacy.

To address the concern that not all CFPB complaints relate to financial fraud originating from lax data security, we classify the complaints into more or less relevant cases using two methods. First, we conduct simple keyword searches in the issue, subissue, and consumer narrative fields for indicators such as fraud, scam, or identity theft. Second, we employ a machine learning method that generates a likelihood of any given complaint being related to financial fraud caused by data security issues. We then examine the correlation between the keyword search and machine learning scores and find that they generate consensus about which are the most and least relevant complaints. According to the scores, complaints in certain product categories (e.g., credit reporting, debt collection) are most likely to be relevant whereas others (e.g., student loan, mortgages) are least likely to be relevant. We then rerun our main specifications using complaints in the top- and bottom-ranked product categories. We find a strong and statistically significant effect for top-ranked categories, consisting of highly relevant cases. ATT reduced the number of financial fraud complaints about credit reporting and debt collection in a zip code by 2.48% and 0.61%, respectively, when it has 10% more iOS users. In contrast, we detect no treatment effect for the bottom-ranked product categories.

We then use a number of different sampling criteria and regression specifications to verify the robustness of our main results. Furthermore, we perform a range of placebo tests. First, we use pre-ATT months as placebo treatment dates and find no significant effect on financial fraud. Second, we conduct a randomization test by shuffling the iOS shares at the zip-code level, and we observe that the estimates in our main specification are far below the empirical distributions of the placebo coefficients obtained with 1,000 random samples.

The CFPB database contains complaints about *financial* products and services that individuals or other agencies submit to the CFPB. To complement the analysis using CFPB data, we also obtain additional data via a freedom-of-information act request (FOIA) from

the FTC. The FTC data allows us to identify complaints in the categories most likely affected by data privacy: internet services, data security, and identity theft. While the FTC sample does not contain zip code information for privacy-preserving reasons, we confirm at the state-city-month level that ATT is more effective in reducing financial fraud in cities with a higher share of pre-treatment iOS users. Another advantage of the FTC data is that it contains information on monetary losses. We estimate that the reduction in tracking reduces money lost in all complaints by 4.7% and money lost reported in internet and data security complaints by 40.1%.

To establish a stronger link between consumer surveillance and fraud, we recognize that the effect of ATT may vary depending on firms' surveillance practices and data security measures. Therefore, we use Apple's privacy nutrition labels and Google's data safety forms to identify firms that are more likely to expose their customers data to fraudsters. In December 2020, Apple introduced privacy labels on its App Store product pages. Similarly, in July 2022, Google launched data safety forms on its Google Play platform. Both platforms require firms to disclose, in a standardized manner, the types of data collected from users, whether the data is shared with third parties, and how it is used, such as for third-party advertising. Google's data safety form also requires disclosure of data security practices, including whether the user data is encrypted during transit.

We first identify the subset of complaints about specific financial companies. We then collect data on which companies offer an active mobile app in either the Apple App Store or Google Play Store. In turn, we collect the privacy labels and safety forms of these companies by scraping their app store pages. We find that approximately 26% of financial companies listed in the CFPB complaints database own an app, and 11% of them collect and share user data with third parties, such as data brokers, other websites, and advertising networks. Our results indicate that the effect of ATT on consumer complaints is more pronounced for companies that are active in the app market, share user data with third parties, or do not encrypt user data in transit. Specifically, compared to companies without an app, those with

at least one app experience a 1.1-percentage-point (or 5.5% of the mean value) reduction in the likelihood of receiving complaints and a 3.9% decline in the number of complaints. Moreover, conditional on having an app, companies that share data with third parties or do not encrypt data in transit experience a 1.7 (0.8) unit drop in the number of complaints per thousand downloads.

Last, we supplement the firm-level analysis using comprehensive data on cyber incidents from Advisen.[2] We find that following ATT, companies with an app are 42% less likely to experience cyber incidents, compared to firms without an app. This effect is stronger when we focus on cyber events that were caused by unauthorized data collection or disclosures, malicious data breaches, or identity theft, and when the events resulted in violations of the Fair Debt Collection Practices and Fair Credit Reporting Acts. Importantly, these two specific regulations target debt collection and credit reporting, the two product categories most susceptible to fraud as shown above.

These findings lend support to the notion that ATT has indeed reduced the adverse impact of firms' data sharing, vulnerability, and breach risks on consumers. Moreover, consistent with expectations, the effect at the firm level is particularly pronounced in product categories that are likely related to data issues, while there is no discernible effect in less-relevant product categories.

Overall, our results provide compelling evidence in favor of industry-led regulations aimed at constraining consumer surveillance practices. Going forward, efforts to limit data transfers across firms, such as Google's plan to phase out third-party cookies in Chrome by 2024, are likely to generate similarly positive social benefits.

Literature review

Our study contributes to the emerging literature on financial fraud by examining its prevalence and impact. We also situate our work within the context of recent research on the effects of data privacy regulations.

[2]The same data is also used by insurance companies to assess cyber security risks.

Our study complements the descriptive analysis conducted by Burnes et al. (2020), who investigate the risk and protective factors of identity theft victimization in the US. Their findings highlight the importance of individual-level behaviors, such as online purchasing frequency and data protection practices, in determining the likelihood of identity theft. They also note the association between identity theft and data breaches in the government and corporate sectors. Similarly, Guabudeanu et al. (2021) explore the trade-off between financial fraud detection and privacy intrusions in survey data. Finally, Khan et al. (2015) provide a comprehensive survey of the security challenges, threats, and vulnerabilities of the mobile ecosystem, as well as suggestions for enhancing mobile security and defending against data privacy threats.

In addition to the literature we have discussed, there are two related areas of research that focus on specific subsets of financial fraud. The first literature examines cases of financial fraud involving elderly victims. The second literature focuses on financial fraud specifically committed by financial advisers or investment managers.

Several studies have focused on financial fraud committed against elderly individuals. Carlin et al. (2020) analyze the impact of new regulations aimed at combating elder financial abuse by deputizing financial professionals across different US states. Their findings suggest that such policies can be effective in reducing financial fraud against elderly individuals. DeLiema et al. (2020) develop and fielded a module on investment and prize/lottery fraud in the 2016 Health and Retirement Study, specifically targeting individuals aged 50 and older. Their analysis revealed the incidence of investment and prize/lottery fraud, as well as prospective risk factors associated with these types of financial fraud. Another study by Alves and Wilson (2008) focus on the disproportionate impact of telemarketing fraud on older adults, particularly those who are socially isolated. Their data was obtained from a questionnaire completed by twenty-eight older adult telemarketing fraud victims, assessing variables related to vulnerability to telemarketing fraud. They found that victims were often male, divorced/separated, college-educated, and between the ages of 60 and 70. Age and

marital status were significantly associated with loneliness, which may exacerbate the risk of financial fraud.

DeLiema (2018) distinguishes between two forms of financial victimization that target older adults: elder financial exploitation by individuals in positions of trust, and elder fraud by predatory strangers. Their research highlights the different patterns and risk factors associated with these two types of financial fraud. Another study by DeLiema et al. (2012) recruit Spanish-speaking Latino older adults aged 66 and over living in low-income communities in Los Angeles to assess the frequency of various types of abuse and neglect, including financial exploitation. The authors found high rates of victimization among this vulnerable population. James et al. (2014) examine the correlates of susceptibility to scams in a cohort of 639 community-dwelling older adults without dementia. They found that susceptibility was positively associated with age and negatively associated with income, cognition, psychological well-being, social support, and literacy. Lichtenberg et al. (2013) examine the national prevalence of older adults who report having been victims of fraud, created a population-based model for predicting fraud, and investigated how fraud is experienced by the most psychologically vulnerable older adults. Their research drew on data from the 2008 Leave Behind Questionnaire, a component of the Health and Retirement Study, which involved 4,400 participants.

Our paper is also related to the literature focusing on financial fraud committed by financial advisers and investment managers. Dimmock and Gerken (2012) investigate the predictability of investment fraud using a panel of mandatory disclosures filed with the Securities Exchange Commission (SEC). They find that disclosures related to past regulatory and legal violations, conflicts of interest, and monitoring have significant power to predict fraud. Deliema et al. (2020) use telephone interviews to profile victims of investment fraud. Their research provides insight into the characteristics of individuals who are most vulnerable to investment fraud and may help identify effective strategies for preventing victimization. Egan et al. (2019) document the widespread extent of misconduct among financial advisers

and the associated labor market consequences. Their research highlights the need for greater regulatory oversight and better industry standards to protect consumers and improve the integrity of financial markets.

Our paper adds to to recent work that examine the effects of data privacy regulation. Existing research has linked GDPR to European web traffic (Goldberg et al., 2019), the entry and exit of apps (Janssen et al., 2021), VC financing (Jia et al., 2021), and the ability of firms to collect and monetize consumer data (Aridor et al., 2020; Bessen et al., 2020; Peukert et al., 2021). Babina et al. (2022) show that open banking policies spur investments into FinTech startups. A couple of recent studies focus on Apple's privacy initiatives, including the privacy label policy and the App Tracking Transparency (ATT) policy. Bian et al. (2021) show that Apple's privacy labels lead to a 14% weekly download reduction and a 15% decline in revenue from user subscriptions and in-app purchases for iPhone users (using Android users as the control group). In addition, the ATT policy leads to an immediate negative stock market reaction for public firms with apps. Kesler (2022) show that the ATT framework implemented by Apple leads more apps to become paid apps and turn to in-app purchases as an alternative revenue source.

2 Institutional Background and Data

In this section, we provide background information on the data privacy regulation implemented by Apple. In turn we will also describe the outcome and explanatory variables that we construct from various data sources.

2.1 App Tracking Transparency policy

In June 2020, Apple announced its plans to move to a new version of its iOS operating system, iOS 14, and starting from the release of iOS 14.5 on April 26, 2021, it introduced a new privacy feature that required all apps to ask for explicit user permission before tracking users across apps or websites owned by other companies. This feature, dubbed "App Tracking Transparency (ATT)", grants users greater and easy control over their personal data. An

example of the prompt notification is provided in Appendix Figure A.1. By default, a user is opted-out of tracking. Industry reports suggest that the vast majority of users, upon seeing the notification, did not opt-in for tracking. The opt-in design of the ATT framework significantly reduced the amount of personal data shared across firms, websites, advertising networks, and data brokers. Bian et al. (2021) document a sharp and negative stock market reactions on firms owning an active iOS apps around the implementation of ATT, corroborating its substantial impact on the data economy (see Appendix Figure A.2).

Although the ATT policy only applies to mobile users, it has implications for commercial surveillance and fraud among the general population due to the prevalence of smartphones. According to Comscore (2019), smartphones account for 70% of all digital media time in the US. The majority of the time consumers spend on their smartphones, they use mobile apps. In the US, mobile phone time spent on apps is 88% (eMarketer, 2020), and this percentage is increasing year by year. Given the ultimate importance of the mobile app industry, consumer surveillance via mobile apps is essential in terms of data collection in today's internet economy.

We use the implementation of the ATT policy as an exogenous shock to the sharing, selling, or leaking detailed consumer data merged from different sources on the iOS platform. We argue that the policy reduced the value and availability of data for fraudsters. Below, we also introduce two other disclosure policies, Apple's Privacy Label and Google's Data Safety form, that allow us to extract information about firms' data collection and security practices.

Apple's Privacy Nutrition Label and Google's Data Safety form Apple's ATT policy constitutes an arguably exogenous shock on consumer surveillance practices, different firms are impacted to a different degree depending on their previous consumer surveillance intensities. Apple's "nutrition" privacy labels and Google's Safety form allow us to measure the heterogeneities in firms' data collection and security practices pre-policy.

On December 14th, 2020, Apple implemented a requirement for all developers to provide information about their data practices in a standardized and user-friendly format. Developers who fail to comply with this policy face the risk of having their future app updates rejected by Apple's app store. Appendix Figure A.3 (Panel a) provides an example of these privacy labels. The privacy labels contain four types of information. First, there are three major data categories: "Data Used to Track You", "Data Linked to You", and "Data Not Linked to You". "Data Used to Track You" refers to data that an app collects and shares with other companies' apps and websites. If an app doesn't collect any data, it will be labeled as "Data Not Collected". Second, under the "Data Linked to You" and "Data Not Linked to You" categories, app developers disclose how they use personal data. There are a total of 6 data uses listed in the figure. Third, the labels include information about the data types that the firm collects, with a total of 14 types. Each data type can be associated with any of the 6 purposes of data use mentioned earlier. Lastly, within each data type, there are specific data items listed, totalling 32 items.

Similarly, Google launched a data safety form for Android apps, starting in July 2021, to disclose privacy and security practices in a concise manner. The structure of Google's safety form is similar to Apple's privacy labels but includes additional information on data security practices, such as encryption, adherence to security standards, and user data deletion requests. The set of information available in the Google's Safety form is illustrated in Appendix Figure A.3. Firms have to disclose whether the firm is committed to follow the Play Families Policy, whether the firm received an independent security review, whether data is encrypted in transit, and whether users can request their data to be deletion.[3] has been independently reviewed for conformance with a global security standard, and allows user to request the deletion of their data.

Based on these mandatory disclosures, we construct three measures to assess a firm's data vulnerability and breach risks. Firms are classified as having a high data vulnerability

[3]See Google's official explanations on the three states of data encryption, at rest, in transit, and in use, here: https://cloud.google.com/docs/security/encryption-in-transit.

if they have an active app in either Apple's app store or Google Play, share information with third parties, and do not encrypt data in transit.

We then match companies in the Consumer Financial Protection Bureau (CFPB) complaint database with app developers. The CFPB pre-processes company names in the complaint database to minimize errors before releasing the complaints to the public.[4] To obtain a list of financial companies as app developers, we first pull the universe of finance apps from Sensor Tower and extract the developers' identifying information (name and website).[5] Using fuzzy name matching and extensive manual checks, we map companies in the CFPB complaint database to app developers.

2.2 Exposure to ATT: share of iPhone and Android users

Because the ATT policy only affects iOS users, we measure treatment intensity using the share of iPhone users at the zip code level. We construct this variable using data from Safegraph, a company that tracks foot traffic using GPS location data from mobile devices. This data has information on daily visits of 6 million points of interest across the country. For each point of interest, Safegraph reports a rich set of information, including time-invariant information such as brand (if the POI belongs to a brand that can be identified), NAICS code, postal code, and time-varying information, such as monthly visit or visitor counts. Crucial for our study, Safegraph reports the number of visitors that use Android vs. iOS devices. Safegraph aims for a representative sample. Li et al. (2023) documents a near-perfect correlation (r>0.97) between the number of sampled device and census population at the county level, and only minor sampling biases among demographic categories such as age, gender, and moderate income (-0.05 to +0.05).

For the purpose of our analysis, we aggregate all visits made to retail and grocery stores (identified by the two-digit NAICS code 44) and financial institutions (identified by the two-

[4]Consumers can select a company from a pre-defined list when submitting the complaint form. If the company is not listed, consumers will be directed to complete contact information for the company. See for details: https://www.consumerfinance.gov/complaint/.

[5]The developer name is publicly available for all apps listed in the Apple's app store.

digit NAICS code 52) based on the device operating system (iOS or Android) and zip code. This aggregation covers the period from January 2019 to June 2022, providing a comprehensive view of foot traffic trends over time. We specifically focus on foot traffic to retail locations as they represent the majority of visits, and any potential operating-system-specific bias is relatively limited compared to other types of locations such as workplaces or hospitals. We expect that the share of iOS users at these general-purpose retail locations is representative of the iOS share within the corresponding zip code. Although our primary focus is on retail locations, we also include banks and other financial institutions in our analysis due to our interest in understanding financial fraud patterns. However, it is important to note that the foot traffic to these financial institutions is relatively small compared to retail locations. Consequently, excluding these institutions has minimal impact on the measurement.

In our analysis, we primarily utilize the pre-ATT average iOS user share for each zip code, rather than employing a time-varying measure. The reason is to mitigate potential confounding factors that could arise from the treatment itself. For instance, one could argue that the introduction of Apple's privacy initiatives might lead to changes in the popularity of iPhones or attract a different population of users over time. Moreover, the accuracy of foot traffic measurement by Safegraph could be influenced by the implementation of ATT. As Safegraph relies on location tracking to collect data from users, the introduction of ATT might impact the ability to precisely capture foot traffic information.

2.3 Financial fraud data

We construct our outcome variables based on financial fraud incidents reported in two datasets: CFPB complaint filings and FTC fraud reports. We provide a description of each dataset below.

CFPB complaint filings When individuals believe they have been victims of financial fraud, they have the option to submit a complaint through various government agency websites. All received complaints are eventually forwarded to the Federal Trade Commission

(FTC) and collected in the Consumer Sentinel Network. Several government websites, such as the Consumer Financial Protection Bureau (CFPB), direct individuals to the FTC's complaint form. If the complaints submitted by individuals mention any "financial services or products", they are forwarded to the CFPB. Additionally, the CFPB provides its own complaint submission template on their website. When the CFPB receives complaints it forwards them to the respective companies involved and publishes them in a publicly available dataset known as the Consumer Complaint Database. This database has been widely utilized by researchers to investigate various aspects of financial fraud. For instance, Haendler and Heimer (2021) employed this dataset to examine racial disparities in restitution for disputed financial services.

When filing a complaint with the CFPB, individuals have the option to select a "product" category from a list of 18 pre-defined categories and an "issue" from a list of 165 pre-defined issues. They can also provide more specific information by selecting a "subproduct" or "subissue" if applicable. The major product categories include "credit reporting", "debt collection", and "mortgage". Common issues reported include "Incorrect information on your report", "Problem with a credit reporting company's investigation into an existing problem", "Improper use of your report", "Attempts to collect debt not owed", "Written notification about debt", "Communication tactics", "Problem with a purchase shown on your statement", "Took or threatened to take negative or legal action", "Unable to get your credit report or credit score", and "Fraud or scam".

Furthermore, individuals have the option to provide a narrative statement describing their case. If they choose to share this statement publicly, the CFPB publishes it in the complaint database after taking steps to remove personal information. By analyzing the combination of these fields, including product, issue, subproduct, subissue, and narrative statements, we gain insights into the specific details and circumstances of each complaint related to financial fraud incidents.

The CFPB highlights the most prevalent categories of financial fraud and the different

tactics used by fraudsters to victimize consumers as follows:

- Identity theft: In this type of fraud, perpetrators acquire personal information from individuals and use it to assume their identities. They may open credit cards and bank accounts in the victims' names and make unauthorized purchases.
- Credit and debit card fraud: This involves the unauthorized use of credit or debit card information. Fraudsters may obtain the card details through various means and then attempt to deceive victims by offering to lower their credit card interest rates or provide other advantageous terms. These fraudulent activities often occur through imposter scams conducted via phone calls.
- Debt collection fraud: Fraudsters engage in this type of fraud by attempting to collect unpaid bills, regardless of their validity. They may use deceptive tactics to coerce individuals into making payments.
- Mortgage scams: This type of fraud targets distressed homeowners. Scammers may pose as legitimate entities, such as mortgage assistance programs or financial institutions, and exploit homeowners' vulnerable situations to extract money from them. These scams can also involve imposter scams conducted through phone calls.

Not all financial fraud complaints are directly linked to the collection and misuse of personal information by thieves or hackers. The "issues" or "subissues" field in the CFPB data may not explicitly distinguish between relevant and less relevant categories specifically related to fraud arising from lax data privacy regulations. Each reported issue has the potential to be relevant to financial fraud. For instance, the presence of "Incorrect information on your report" could indicate a situation where a fraudster has applied for a credit card while pretending to be the account holder. Similarly, "Attempts to collect debt not owed" could be facilitated by collecting the phone number and loan information of someone.

Given the complexity and evolving nature of financial fraud, it is crucial to consider the broader context of data privacy regulations and their impact on fraud prevention. The CFPB

data does not explicitly delineate the role of lax data privacy regulations in specific fraud incidents. To determine the relevance of complaints to data privacy issues, we employ two approaches using the consumer narrative field, which is available for 40% of the complaints in our dataset.

First, we conduct keyword searches based on the subproduct, issue, subissue, and narratives complaint fields. We compile a list of keywords related to data privacy, such as "incorrect", "fraud", "theft", "identity", and "data breach". If any of these keywords appear in the relevant fields, we assign an indicator variable with a value of one. This approach allows us to identify complaints that potentially involve data privacy concerns.

Second, we utilize a machine learning method called zero-shot learning (ZSL) to assess the likelihood of a narrative being related to fraud arising from data issues. The advantage of ZSL is that it does not require manual annotations and can identify relevant patterns automatically. For detailed information on the ZSL algorithm, please refer to Appendix C. The output of this algorithm is a continuous likelihood score indicating the relevance of a complaint to data-related fraud, relative to other complaint types such as mistakes or unresponsiveness.

Since the narrative-based likelihood score is only available for a subset of complaints with consumer narratives, we extrapolate the scores at the product category level to identify categories that are more or less relevant to data privacy issues. We classify categories with higher average scores as more relevant. Appendix Table C.1 presents the mean and standard deviation of complaint-level scores by product category. Two patterns are worth noting. First, both the keyword search method and the machine learning approach generate meaningful variations in the average scores at the product level, allowing us to distinguish between more and less relevant cases. For example, the highest and lowest scores generated by the keyword search method are 0.82 and 0.30, respectively, while the highest and lowest scores generated by the ZSL method are 0.53 and 0.16, respectively. Second and more importantly, the scores generated by these two methods exhibit a high correlation at the

tails, indicating a consensus on the most relevant and irrelevant complaints. Both methods consistently rank "Credit reporting" and "Debt collection" as the most relevant categories, while "Student loan" and "Mortgage" receive the lowest scores, suggesting lower relevance to data privacy issues.

FTC fraud reports In addition to the CFPB database, we obtain data on fraud complaints from the Federal Trade Commission (FTC) through a freedom of information act request. The FTC collects the largest set of fraud reports filed by consumers. These reports can be submitted directly to the FTC or collected and forwarded by the 45 members of the Consumer Sentinel Network. The financial fraud complaints in the CFPB database thus represent a subset of all the fraud reports submitted to the FTC.

When individuals file a complaint with the FTC, they provide various details about the fraud incident, including personal information, information about the fraud itself (such as the contact involved, the amount requested, the amount of money lost), whether the fraud was resolved, and their suspicions regarding the identity of the fraudster. An example of a complaint form can be found in Appendix B.

According to the FTC, consumers filed approximately 1.4 million fraud reports in 2020, covering 20 different product and service categories. The top fraud categories reported are imposter scams or related to online shopping and internet services. Clearly, not all of the fraud complaints are directly related to data breaches or privacy issues. For our research purposes, we focus specifically on the categories of privacy, data security, and cyber threats, as well as internet services, as these are likely the most relevant complaints to us. It's important to note that the FTC database includes not only fraud reports but also a separate category for identity theft, which is sourced using a different online form.

The nature of the FTC data has a few implications for our empirical analysis. First, the FTC de-identified the complaint-level data by removing consumer names, zip codes, narratives about the incidents, and information about the alleged fraudsters. As a result,

our regressions with the FTC data are conducted at the self-reported city-state level instead of the more detailed zip code level. Second, since the narratives are not available in the FTC data, we cannot rely on them to identify more relevant cases. Consequently, our main analysis is primarily based on the data from the Consumer Financial Protection Bureau (CFPB), while the FTC data serves as supplementary evidence that complements the main analysis in two important ways. To start with, the FTC data is a comprehensive source of information on all reported fraud incidents, covering a broader range of products and services, including internet services, data security, and cyber fraud, which are directly impacted by Apple's App Tracking Transparency (ATT) policy but may not be forwarded to the CFPB. This allows us to gain a more complete view of the effects of ATT on fraud incidents. Additionally, the FTC data reports information on the financial losses resulting from data-driven fraud, enabling us to quantify the economic impact of these incidents.

2.4 Summary statistics

Table 1 presents summary statistics for the variables used in our regression analysis. The main regression sample consists of a balanced panel at the zip code level, spanning from January 2019 to June 2022. Zeros were filled in for zip codes without any reported complaints.

In terms of the likelihood of having any complaint in a zip code, approximately 31% of zip codes have at least one complaint in any given month. The mean number of complaints per 1,000 people in a zip code per month is 0.07, indicating that around 7 people out of every 100,000 file a complaint in a given month. Aggregating across the entire US, consumers file an average of 36,936 complaints per month or around 443,000 complaints per year. This accounts for approximately 10% of the total complaints compiled by the FTC, highlighting the significant contribution of CFPB as a data source for FTC consumer complaint reports.[6]

Figure 1a displays the number of CFPB complaints per 1,000 residents for each zip code in the US, providing a visual representation of the spatial variation in complaint rates.

[6]The annual number of CFPB complaints in 2019-2021 are 277,325, 444,347, and 496,018, respectively. The Consumer Sentinel Network data book provides statistics on the top data contributors: https://www.ftc.gov/system/files/ftc_gov/pdf/CSN%20Annual%20Data%20Book%202021%20Final%20PDF.pdf.

Figure 1b illustrates the total number of complaints per month over the entire sample period, showing temporal variations in complaint volume.

Panel a of Table 1 also presents summary statistics for the outcome variables and firms' data practices at the firm-month level. The dataset includes 6,458 unique financial institutions, with an average firm receiving at least one complaint every 5 months and an average of 1.6 complaints per month. Approximately 26% of the complaints are related to firms that own at least one app, which is consistent with the fraction of unique firms that own an app (1,674 out of 6,458). Among these app-owning firms, the average number of complaints per 1,000 app downloads is around 3. Regarding data sharing practices, 11% of app-owning firms share data with third parties, while 1% of them do not encrypt data in transit. [7]

Table 1b presents summary statistics for the FTC complaints sample. The likelihood of having at least one complaint at the city-month level is 12% for internet services and 1% for privacy, data security, and cyber threats. These categories are particularly relevant for our analysis as they are likely to be directly affected by data protection initiatives such as ATT.

The table also displays summary statistics for the iOS share. It is important to note that the iOS share differs between the CFPB and FTC samples due to the aggregation level (zip code vs. city). Using foot traffic data, we find that the average iOS share across US zip codes is 0.46. According to Statista, the iOS share aggregated over the entire US fluctuates around 50% during the period of 2019-2022.[8]

3 The Impact of the Privacy Regulation on Financial Fraud

3.1 Regression specification

Our main regression specification is:

$$Fraud_{z,t} = \alpha_z + \alpha_{county,t} + \beta iOS\ Share_z \times Post\ Policy_t + \varepsilon_{z,t} \tag{1}$$

[7] Please note that the number of observations decreases when normalizing the number of complaints by app downloads, as data on app downloads is currently available for relatively popular apps, and the collection of download data for all app-owning firms is still ongoing.

[8] See https://www.statista.com/statistics/266572. Additionally, DeviceAtlas documents large variations in iOS share across US states. See https://deviceatlas.com/blog/mobile-os-popularity-by-us-state.

The outcome variable, $Fraud_{z,t}$, is constructed by aggregating CFPB complaints to each zip-month. We consider multiple variables, including an indicator for any complaint in a zip-month, a winsorized count of complaints, the number of complaints per 1,000 residents, and the logarithm of one plus the number of complaints. To address concerns raised in Cohn et al. (2022), we prioritize the use of the complaint indicator and the population-scaled count as our preferred outcome variables.

To capture the variation in exposure to ATT, we use the variable $iOS\ Share_{z,t}$, which represents the average pre-treatment iOS share of users at the zip code level. This variable remains constant for each zip code since it is based on pre-treatment data. The treatment event indicator, $Post\ Policy_t$, takes a value of one starting from May 2021, signifying the month when the ATT policy took effect. Since our analysis includes high-dimensional fixed effects, the independent regressors, $iOS\ Share_z$ and $Post\ Policy_t$, are absorbed.

To account for time-invariant characteristics that contribute to financial fraud, we include zip code fixed effects. Additionally, we incorporate county by year-month fixed effects, denoted as $\alpha_{county,t}$, to control for time-varying confounders at the county level. These confounders may include region-specific data regulations, local fraud news, or local economic development. The inclusion of granular fixed effects helps address any concerns about time-varying measurement error in financial fraud complaints within a county. We cluster the standard errors by state, adopting the most conservative approach. In our robustness checks, we explore alternative sets of fixed effects and clustering choices.

3.2 Baseline results

Table 2 presents the regression results based on our preferred specifications. We observe a significant and negative coefficient on the interaction term $iOS\ Share_z \times Post\ Policy_t$. This finding indicates that, within a given county and month, zip codes with a higher proportion of iOS users experience a decline in financial fraud following the implementation of the ATT policy, compared to zip codes with a lower iOS user share.

The effects of ATT are economically meaningful. A zip code with a 10% higher iOS user

share exhibits a 0.65% decrease in the likelihood of experiencing at least one complaint in any given month (Column 1). This reduction corresponds to a 2.1% decrease relative to the pre-ATT average likelihood of complaints, which was 31%.

Considering the count of complaints and the count normalized by population (Columns 2 and 3), a zip code with a 10% higher iOS user share sees a decrease of 0.17 complaints and a 0.93 reduction in the number of complaints per 1,000 capita. Additionally, when examining the log-transformed outcome variable (Column 4), we find an estimated coefficient of -0.263. This implies that a zip code with 10% more iOS users experiences a 2.63% decrease in the monthly number of complaints. Given the observed tracking opt-out rate of 82%, our results suggest that if 10% of mobile app users were to disallow data tracking, financial fraud could be reduced by approximately 3.21% (calculated as 2.63% divided by 0.82). These findings highlight the effectiveness of ATT in enhancing privacy and reducing financial fraud.

3.3 Dynamics, placebo tests, and robustness checks

A key identifying assumption is parallel trends, i.e., in the absence of the ATT policy, the intensity of financial fraud would have followed similar trends in both high-iOS-share and low-iOS-share zip codes. To validate this assumption, we employ two approaches. First, we visually examine the pre-ATT trends in financial fraud in zip codes with different levels of exposure to the shock. Second, we perform two sets of placebo tests in which we 1) randomly assign treatment intensity to zip codes and 2) estimate the treatment effect using placebo event dates. This subsection describes these tests in detail and presents evidence supporting our identifying assumption. We also conduct robustness checks using an extensive list of alternative specifications.

3.3.1 Dynamics

While we compare within the same county-month, high-iOS-share zip codes may differ from low-iOS-share zip codes in various dimensions that could affect the levels and trends of financial fraud. For example, ownership of Apple products predicts higher income and better education (Bertrand and Kamenica, 2018), which can lead to reduced financial fraud among

iOS users over time. To rule out alternative explanations, we examine pre-trends in financial fraud and plot dynamic DiD coefficients.

In Figure 2, we present the results. To reduce estimation error, we group together all three months within a corresponding quarter. The analysis covers a total of nine quarters before the introduction of the ATT policy and four quarters after its implementation.[9] We define quarter -1 as the quarter immediately preceding the implementation (2021Q1), which serves as the benchmark quarter. Quarters prior to $t = -4$ are combined into a single period.

We first examine the impact of the ATT policy on fraud complaints using an indicator variable as the outcome measure. The dynamic DiD coefficients confirm that the effect of ATT only becomes apparent after the policy's implementation. Prior to the introduction of ATT, the coefficients for all quarters are not statistically different from zero. However, there is a clear negative post-trend, suggesting a decline in fraud complaints following the policy's implementation in zip codes with larger shares of iOS users. The same pattern can be observed for an alternative outcome measure that captures the intensity of fraud complaints, as shown in panel b of Figure 2.

Additionally, we divide all zip codes into three terciles based on their pre-policy iOS shares and compare the trajectories of fraud complaints between the top and bottom terciles around the implementation of ATT. In Figure 3, we present the estimated coefficients over time for the interaction term between relative policy timing indicators and the top tercile indicator. Regardless of the outcome variable used in the regression, we find evidence consistent with parallel trends before the policy implementation and a negative and significant effect after ATT, indicating a reduction in fraud complaints in zip codes with higher iOS shares.

We also estimate the treatment effect separately for each individual month, spanning 27 months before ATT and 13 months afterwards. For ease of visualization, we group the months before $t = -10$ (or after $t = 10$) together. Figure D.1 and Figure D.2 depict the dynamic treatment effects at a monthly frequency. Consistent with our previous findings, we

[9] Due to the introduction of the ATT policy in 2021Q2 and the end of our sample period in June 2022, we have data for four quarters after the event.

consistently observe insignificant coefficients before the introduction of ATT. However, the point estimates for the post-event months are mostly negative and significant, particularly when using the number of complaints per 1,000 residents as the outcome variable. This suggests that the ATT policy is associated with a decline in financial fraud complaints in the months following its implementation. It is worth noting that the estimations become noisier when the outcome variable is a dummy indicating the existence of complaints. Nonetheless, the point estimates for the post-ATT period tend to be more negative compared to the pre-ATT period.

Overall, the consistent patterns of insignificant coefficients before ATT and negative and significant coefficients after ATT provide reassurance that the observed decline in financial fraud complaints after the introduction of ATT is unlikely to be driven by differential trends among zip codes with varying levels of exposure to the policy.

3.3.2 Placebo tests

To further strengthen the validity of our findings, we conduct placebo tests in both the time series and cross-section to demonstrate that our results are not driven by factors unrelated to ATT.

In the time series placebo test, we estimate our main regression specification by replacing the actual event month with 18 placebo treatment months between July 2019 and January 2022. We maintain a consistent 4-month post-event window for all placebo estimates. Figure 4 presents the 18 rolling-window estimates for both the likelihood of having any complaint in a zip code and the number of complaints per 1,000 residents. In both subfigures of Figure 4, we observe a kink at the beginning of 2021. Following this kink, the estimated coefficients abruptly drop and become significantly negative over time. These patterns align with the timing of the actual ATT policy implementation: from January 2021 onwards, the four-month post-event window encompasses the actual post-treatment months that start from April 2021. This supports our argument that the decline in fraud complaints is attributable to ATT. Unless there are other events that perfectly coincide with

ATT and generate a similar effect, we can confidently attribute the observed decline in fraud complaints to the implementation of ATT.

In some specifications, the placebo DiD coefficients were slightly negative in April 2020, which coincide with the first wave of economic impact payments issuance ("EIP") in the US. These payments were associated with an uptick in fraud that might have affected Android users more than iOS users. Given that the second and third waves of EIP were issued shortly before ATT (December 2020 and March 2021), our estimates may be contaminated by the impact of EIP. In the next section, we address this concern by leveraging the product category associated with each complaint.

The cross-sectional placebo tests provide further evidence to support the robustness of our main findings. By randomizing the matches between iOS shares and zip codes, we generate 1,000 placebo samples, each with a different mapping between iOS shares and zip codes. By estimating the DiD coefficients using these placebo samples, we can assess the likelihood of obtaining our main results purely by chance.

Figure 5 displays the histogram of the estimated DiD coefficients from all the placebo samples, with the actual point estimate from the real sample marked in red as a benchmark. In both panels, we observe that the actual point estimate falls to the far left of the distribution of placebo point estimates and is significantly different from the lowest placebo point estimate. This indicates that our main results are unlikely to be driven by peculiarities or unaccounted correlation structures of the data.

3.3.3 Robustness checks

We then conduct additional robustness checks to ensure the reliability of our results. These tests include alternative specifications to address measurement error in iOS shares, additional fixed effects to control for potential confounding factors, and the use of a Poisson regression model for count data. Despite these variations, our main findings remain consistent and robust, providing strong support for the relationship between ATT and the reduction in financial fraud complaints.

Table E.1 presents results from alternative regression specifications and samples. Column 1 presents the baseline result for comparison. In Column 2, we replace county-by-month fixed effects with state-by-month fixed effects. The results remain largely similar. In Column 3, we cluster standard errors at the zip code level. Again, the results are consistent with the baseline.

To address potential measurement error in iOS shares, we group them into categories in Column 4. Specifically, we create a dummy variable for the top tercile of iOS shares, while excluding the middle tercile to emphasize the contrast between high and low treatment intensity areas. When considering the number of complaints per 1,000 residents as the outcome variable (panel b), the coefficient on the interaction term indicates that ATT reduces the number of complaints per thousand residents by 0.025, which is approximately 30% of the mean value, relative to zip codes in the bottom tercile.

In Column 5 of Table E.1, we exclude years before 2020 to address potential measurement issues related to the coverage of foot traffic data. By focusing on more recent data when the coverage of Safegraph has been steadily increasing, we obtain stronger results. Columns 6 and 7 take a different approach by excluding zip codes with low foot traffic. We drop zip codes with fewer than 100 or 1,000 visits to retail locations in the corresponding zip code-month, representing the 5th and 15th percentile of the distribution, respectively. This approach helps improve the precision of regression estimates by mitigating potential measurement error in areas with low foot traffic.

Figure 6 provides a visual comparison of the results from the robustness analyses, including additional specifications. Here, we also divide zip codes into different groups based on their pre-treatment iOS device share and incorporate group-specific linear trends in the regression. This stricter setting allows for the possibility of pre-existing trends. Interestingly, the estimated coefficient remains highly similar to that in the baseline specification.

Finally, Figure E.1 presents the effects of ATT on financial fraud complaints for different regression samples and specifications, using Poisson regression models. The top row of the

figure corresponds to the baseline Poisson estimate, using the same sample and fixed effects as in Table 2 and considering the number of complaints at the zip code level (winsorized at 1%) as the outcome variable. The figure shows that while the point estimates vary across different specifications, the relative magnitudes of the effects remain similar to those obtained using OLS regressions. This robustness analysis further supports our main findings and indicates that the observed impact of ATT on reducing financial fraud complaints is not specific to a particular choice of the empirical model.

3.4 Heterogeneity across demographics

Table 3 examines the heterogeneous impact of ATT across different demographic groups. The demographic variables are constructed using zip code level data from the 2020 census release. The results reveal that the effects of ATT are stronger among communities (zip codes) with a higher share of black or Asian people, as well as a higher share of teenagers or women. This suggests that minorities, females, and relatively young or old populations tend to be more susceptible to financial fraud triggered by personal data sharing.

There are two potential explanations for this pattern. First, these user groups may have a heavy online presence and lower privacy awareness, resulting in more personal data sharing and potential data leakage. Second, conditional on data sharing and leakage, fraudsters may specifically target these populations as they may be more vulnerable to fraud.

The findings imply that privacy regulations, such as the ATT policy, can be beneficial for high-fraud-risk groups, including certain marginalized populations. By reducing the exposure of personal data through tracking opt-outs, privacy regulations can help protect these vulnerable populations from financial fraud.

4 Economic Mechanism

To explore the economic mechanism underlying the impact of ATT, we focus on specific types of CFPB complaints that are more likely to reflect data-privacy-related issues. We also analyze the FTC fraud report data that contain additional product categories that are

susceptible to fraud driven by lax data security and data breaches. Specifically, we expect internet services and data security to be most influenced by the implementation of ATT.

4.1 Breakdown by types of complaints

To identify data-driven fraud incidents in the CFPB data, we employ keyword search and machine learning techniques as described in Section 2. This allows us to classify consumer complaints into cases that are more closely related to data breach, abuse, or misuse, and those that are less relevant to such issues. These two methods generate consensus over the most and least relevant fraud categories. Among all the CPBF products categories, "Credit reporting" and "Debt collection" are found to be the most relevant, while "Mortgage" is the least relevant. We estimate our main regression specification separately for these three categories and report the results in Table 4.

Consistent with the hypothesis that ATT reduces data-driven financial fraud, we find negative and statistically significant effects on complaints within the top two fraud categories (Panels a. and b.). The magnitude of the effects is comparable to that observed in the full sample of complaints. Specifically, following the implementation of ATT, a 10% increase in the share of iOS users in a zip code is associated with a 0.98% decrease in the probability of having a complaint related to credit reporting (Column 1). Moreover, the number of complaints declines by 2.95% (Column 4). Similarly, for complaints related to debt collection, there is a statistically significant 0.49% reduction in the probability of having a complaint (Column 1) and a 0.82% decline in the number of complaints (Column 4).

In contrast to the significant effects observed in the most relevant fraud categories, ATT has an insignificant and close to zero impact on complaints related to irrelevant products, such as mortgages (Panel c of Table 4).

To further refine our analysis using the continuous score generated by ZSL, we focus on the top two fraud categories and drop complaints with low ZSL scores. This refinement ensures that the remaining complaints are more likely to be driven by data issues. Specifically, we exclude the bottom three deciles of the likelihood distribution, as approximately 70-80% of

complaint cases in the top two categories are deemed relevant based on keyword search. For the least relevant fraud category, we drop complaints in the top two deciles of the likelihood distribution. After this refinement, the remaining complaints in the top (bottom) categories are even more (less) related to data issues. Complaints without narratives are automatically excluded from the analysis. We re-estimate our main specification and find qualitatively similar results, which are reported in Table 5.

Addressing the concern of EIP disbursement related to COVID-19, we also leverage heterogeneities across product categories. Complaints related to EIP disbursement are typically associated with checking or savings accounts and prepaid cards, as indicated in the CFPB Complaint Bulletin on COVID-19 issues. To assess the impact of EIP on differentiating the responses of Android and iOS users to ATT, we re-run the regression separately for these two product categories. The results, reported in Appendix F, show economically small coefficients on the interaction term between iOS shares and the post-ATT indicator. The magnitudes of the effects on these categories are 1/10 of the baseline effects. This suggests that the differential responses to ATT in high versus low iOS share zip codes are largely not driven by EIP-related fraud incidents.

Overall, the refined analyses provide consistent evidence that ATT primarily affects fraud complaints related to credit reporting and debt collection, while its impact on complaints about irrelevant products, such as mortgages, is negligible. Furthermore, the lack of economically meaningful effects in product categories associated with EIP-related complaints suggests that the observed differential responses to ATT are not driven by the economic impact payment program.

4.2 Effects on data-driven fraud based on FTC reports

As explained in Section 2, our analysis of financial fraud complaints initially focused on the subset of complaints filed with CFPB. However, to provide additional evidence, we turn to the FTC complaint data, which includes all consumer fraud reports collected by members of the Consumer Sentinel Network.

Using the FTC complaint data, we provide two aspects. First, we utilize detailed category information to identify highly relevant fraud reports, related to internet services and data security. These particular complaints are not processed through CFPB and are therefore absent from the CFPB data. Second, in addition to the number of complaints, we incorporate the self-reported dollar loss available in the FTC data, enabling us to quantify the impact of ATT on reducing financial losses among consumers.

Our analysis is conducted at the city-month level and incorporates city and state by year-month fixed effects. We cluster the standard errors at the state level. We present the estimated effects on complaint cases concerning internet services and data security in Table 6. Due to the lack of reliable and complete city-level population information, we examine two outcome variables: an indicator denoting whether there is at least one report about a specific product in a city-month, and the logarithm of the number of reports. Columns 1-2 report the results for internet services, while Columns 3-4 pertain to data security.

Our results reveal that ATT significantly reduces fraud incidents in these relevant categories. At the extensive margin, a 10% increase in iOS users in a city is associated with a 0.23% decrease in the probability of local residents submitting at least one fraud complaint about internet services. Similarly, for data security, a 10% increase in iOS users corresponds to a 0.10% decrease in the probability of local residents submitting at least one fraud complaint about data security. At the intensive margin, the number of fraud complaints in these categories decreases by 0.34% and 0.08%, respectively. The smaller point estimates in these specific categories are expected as complaints related to internet services and data security are generally less frequent. These findings emphasize that ATT not only reduces data-driven fraud incidents within financial product categories, but also has broader effects in cutting fraud reports pertaining to internet services and data security.

We proceed by examining the impact of ATT on reducing the financial losses associated with fraud incidents. The results can be found in Table 7. In Columns 1 and 2, we focus on a subsample of all reports that include positive dollar losses. The estimates indicate that a

10% increase in local iOS share is associated with a 0.2% decrease in the likelihood of having financial loss in a city, as well as a 0.68% decrease in the number of such reports after ATT. Columns 3 and 4 utilize the logarithm of one plus the dollar loss as the outcome variable. While ATT does not show a significant effect when examining the entire universe of FTC fraud reports, the monetary loss resulting from incidents highly exposed to data issues is estimated to decrease by 4% after ATT if the city has a 10% higher iOS share.

In Appendix Table G.1, we analyze the impact of ATT on another prominent report category in the FTC database—identity theft. Due to the lack of information on identity theft types, this sample serves primarily to verify our main findings, as data breaches often lead to identity theft.[10] The difference-in-differences estimates indicate that after ATT, cities with a higher share of iOS users experience a decline in the number of identity theft reports relative to cities with a lower share. The effect is not statistically significant at the extensive margin, likely due to the high prevalence of identity theft reports, with most cities having at least one report filed in any given month.[11]

Additionally, we show in Appendix Table G.2 that our findings are qualitatively similar when we scale the number of fraud incidents and identity theft reports by city-level population. Note that the sample size is significantly reduced, mainly because city names in the FTC data are self-reported and noisy.[12] Among the subsample of cities that we match to the census and foot traffic data, we find a significant and negative effect of ATT on the number of reports per capita, related to data security products and identity theft. In addition, the number of fraud cases with dollar losses is reduced. The impact on internet service related fraud is negative but insignificant.

[10] FTC maintain an internal categorization of identity theft reports, including credit card fraud, loan lease fraud, and bank fraud. This information was not shared with us.

[11] It is worth noting that identity theft reports account for approximately 50% of all fraud reports, totalling 1.4 million in 2020. More information on the top report types can be found in the following link: https://public.tableau.com/app/profile/federal.trade.commission/viz/FraudReports/FraudFacts, under the "Top Reports" tab.

[12] "City" is not a unambiguous administrative unit across US. In different datasets, city may correspond to different administrative divisions within a particular state or region.

5 Firm Exposure to Data Breaches and Cyber Risks

The analyses in the previous section exploit variations in a locality's exposure to the ATT policy. In this section, we explore variations in firms' exposures to the ATT policy. We utilize information from Apple's data privacy nutrition labels and Google's safety forms to identify firms that are more likely to expose their customers to fraud. We then examine two set of outcomes variables: CFPB fraud complaints and cyber incidents. These analyses provide additional evidence supporting the connection between the privacy policy change and the reduction in financial fraud complaints related to lax data security standards.

As explained in Section 2, we construct three simple measures to capture a firm's vulnerability to data breaches and risks of user information leakage based on the two platform-mandated disclosures. Specifically, we consider a firm to be vulnerable if it has an active app in either Apple's app store or Google Play, shares information with third parties, and does not encrypt data during transit.

We present evidence that the effectiveness of ATT in reducing consumer complaints varies depending on the data security of the company. The results, shown in Tables 8, indicate that ATT has a stronger impact on reducing consumer complaints about a specific company when the company has an iOS app, shares data with third parties, or does not encrypt user data in transit.

Comparing firms with and without an app, we find that firms with at least one app are 1.1% less likely to receive complaints in a given month after the implementation of ATT. This reduction represents a 5.5% decline compared to the sample mean of 20%. Focusing on the intensive margin, these firms experience a 3.9% decline in the number of monthly complaints compared to firms that are less exposed to the policy.

In Columns 3-4, we narrow down the analysis to firms with apps and normalize the number of complaints by the size of the firms' user bases, measured by the number of app downloads. Column 3 reveals that following ATT, the number of complaints per 1,000 new downloads decreased by 1.7, which represents a more than 50% decline compared to the

sample mean of 3.14 complaints per thousand downloads. Similarly, Column 4 shows that companies that do not encrypt data in transit experience a decrease of 0.8 in the number of complaints per thousand downloads.

We further analyze the impact of ATT on consumer complaints at the firm level, broken down by product categories, and report the results in Table 9. Consistent with our expectations, we find that the effect of ATT is statistically significant only for the most relevant product categories, such as credit reporting and repair services, as well as debt collection. In contrast, the number of complaints in unrelated categories is not affected by the policy. These findings further confirms the economic mechanism: the reduction in consumer complaints is due to ATT limiting the amount and scope of personal data subject to breaches and exploitations.

Cyber incidents Does ATT reduce firms' exposure to cybersecurity risks? Answering this question would allow us to provide additional evidence of the mechanism. We do so by leveraging data on cyber incidents from Advisen. This dataset covers more than 90,000 cyber events between 2000 and 2023, all collected from publicly verifiable sources, including government websites, keyword-based searches, and official court and litigation sources.[13] We use this data to identify the presence of cyber incidents for each firm in each month during the sample period, 2020-2022.

Furthermore, the Advisen data provides rich information about each incident, such as its cause and whether it led to violations of specific regulations. This allows us to focus the analysis on cyber incidents that are more exposed to ATT and are more likely to result in financial fraud. A detailed description of these incident-level variables is provided in Appendix H.

We use four indicators as our outcome variables. First, an indicator variable for whether the firm was exposed to any cyber incident in a given month. Second, whether the cyber

[13]For more information on Advisen's data sources, see:https://www.advisenltd.com/wp-content/uploads/2017/01/cyber-risk-data-methodology.pdf

incident was the result of malicous breaches or privacy violations, such as unauthorized data collection and disclosure by the firm. Third, whether the cyber incident violated regulations concerning consumer protection. Lastly, whether the cyber incidents violated the Fair Debt Collection Practices or the Fair Credit Reporting Acts, which are the two most cited regulations in CFPB complaint narratives. These regulations also correspond to the top two fraud product categories.

Our findings offer evidence consistent with ATT reducing firms' cyber incidents. The results are reported in Table 10. Across all columns, the estimates are negative and significant at the 1% level. The economic magnitudes are substantial: following ATT, companies experience a 4.7-percentage-point reduction in the likelihood of experiencing cyber incidents compared to firms without the app. This represents 42% of the baseline likelihood. The relative effects are more pronounced on cyber incidents that were more likely to be influenced by ATT: a reduction of 49% for incidents resulting from privacy risks, 79% for incidents leading to violations of pro-consumer regulations, and 75% for incidents resulting in violations of regulations related to debt collection and credit reporting.

6 Conclusion

In this paper, we took a first step in analyzing and quantifying the impact of lax data privacy regulations on financial fraud. By leveraging the implementation of Apple's App Tracking Transparency (ATT) policy, which restricts the tracking and sharing of personal information on the iOS platform, and utilizing variations in iOS device share across different localities in the US, we have demonstrated that ATT had a substantial effect in reducing financial fraud.

To the best of our knowledge, our study is the first to provide empirical evidence on the beneficial impact of privacy regulations in mitigating financial fraud. These findings contribute to the growing body of research that underscores the importance of data privacy regulation. Furthermore, we encourage future efforts to implement measures that restrict data transfer across firms, as exemplified by Google's plan to phase out third-party cookies in Chrome by 2024. Such initiatives hold the potential to further reduce the occurrence of

financial fraud.

Financial institutions are also taking proactive measures to monitor transactions for money laundering, fraud, and other illicit activities (Levinson, 2008), while the adoption of two-factor authentication systems is becoming more widespread. These ongoing and future endeavors to restrict data transfer across industries are expected to generate significant positive societal benefits beyond the scope of financial fraud prevention.

References

Alves, L. M. and S. R. Wilson (2008). The effects of loneliness on telemarketing fraud vulnerability among older adults. *Journal of elder abuse & neglect 20*(1), 63–85.

Aridor, G., Y.-K. Che, and T. Salz (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. NBER Working Paper 26900, Columbia University, Massachusetts Institute of Technology.

Babina, T., G. Buchak, and W. Gornall (2022). Customer data access and fintech entry: Early evidence from open banking. *Available at SSRN*.

Bertrand, M. and E. Kamenica (2018). Coming apart? cultural distances in the United States over time. Technical report, National Bureau of Economic Research.

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

Burnes, D., M. DeLiema, and L. Langton (2020). Risk and protective factors of identity theft victimization in the united states. *Preventive medicine reports 17*, 101058.

Carlin, B. I., T. Umar, and H. Yi (2020). Deputizing financial institutions to fight elder abuse. Technical report, National Bureau of Economic Research.

Cohn, J. B., Z. Liu, and M. I. Wardlaw (2022). Count (and count-like) data in finance. *Journal of Financial Economics 146*(2), 529–551.

DeLiema, M. (2018). Elder fraud and financial exploitation: Application of routine activity theory. *The Gerontologist 58*(4), 706–718.

DeLiema, M., M. Deevy, A. Lusardi, and O. S. Mitchell (2020). Financial fraud among older americans: Evidence and implications. *The Journals of Gerontology: Series B 75*(4), 861–868.

DeLiema, M., Z. D. Gassoumis, D. C. Homeier, and K. H. Wilber (2012). Determining prevalence and correlates of elder abuse using promotores: Low-income immigrant latinos report high rates of abuse and neglect. *Journal of the American Geriatrics Society 60*(7), 1333–1339.

DeLiema, M., G. R. Mottola, and M. Deevy (2017). Findings from a pilot study to measure financial fraud in the united states. *Available at SSRN 2914560*.

Deliema, M., D. Shadel, and K. Pak (2020). Profiling victims of investment fraud: Mindsets and risky behaviors. *Journal of Consumer Research 46*(5), 904–914.

Dimmock, S. G. and W. C. Gerken (2012). Predicting fraud by investment managers. *Journal of Financial Economics 105*(1), 153–173.

Egan, M., G. Matvos, and A. Seru (2019). The market for financial adviser misconduct. *Journal of Political Economy 127*(1), 233–295.

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*.

Guabudeanu, L., I. Brici, C. Mare, I. C. Mihai, and M. C. Scheau (2021). Privacy intrusiveness in financial-banking fraud detection. *Risks 9*(6), 104.

Haendler, C. and R. Heimer (2021). The financial restitution gap in consumer finance: insights from complaints filed with the cfpb. *Available at SSRN 3766485*.

Huff, R., C. Desilets, and J. Kane (2010). The 2010 national public survey on white collar crime. *National White Collar Crime Center 44*.

James, B. D., P. A. Boyle, and D. A. Bennett (2014). Correlates of susceptibility to scams in older adults without dementia. *Journal of elder abuse & neglect 26*(2), 107–122.

Janssen, R., R. Kesler, M. Kummer, and J. Waldfogel (2021). GDPR and the lost generation of innovative apps. NBER Working Paper 146409, University of Zurich, University of Minnesota, University of East Anglia, Georgia Institute of Technology.

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*, forthcoming.

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

Khan, J., H. Abbas, and J. Al-Muhtadi (2015). Survey on mobile user's data privacy threats and defense mechanisms. *Procedia Computer Science 56*, 376–383.

Levinson, B. (2008). Unwarranted deputization: Increased delegation of law enforcement duties to financial institutions undermines american competitiveness. *Available at SSRN 2711938*.

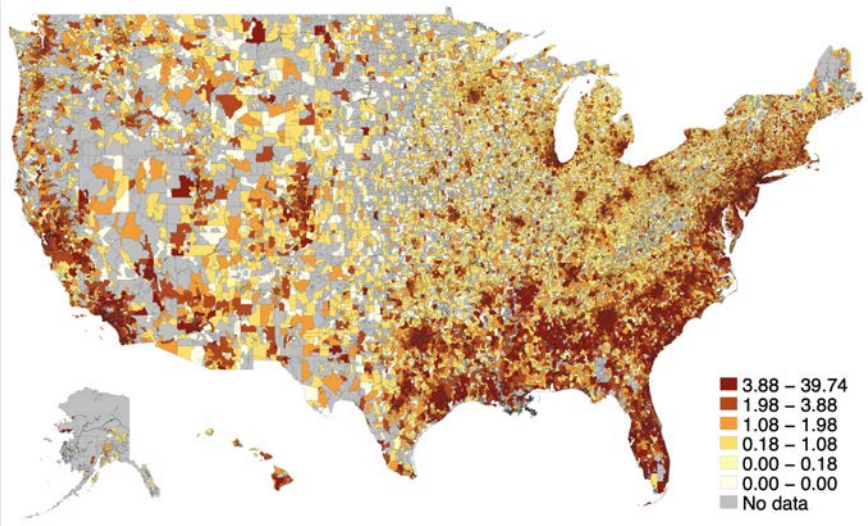
Li, Z., H. Ning, F. Jing, and M. N. Lessani (2023). Understanding the bias of mobile location data across spatial scales and over time: a comprehensive analysis of safegraph data in the united states. *Available at SSRN 4383333*.

Lichtenberg, P. A., L. Stickney, and D. Paulson (2013). Is psychological vulnerability related to the experience of fraud in older adults? *Clinical Gerontologist 36*(2), 132–146.

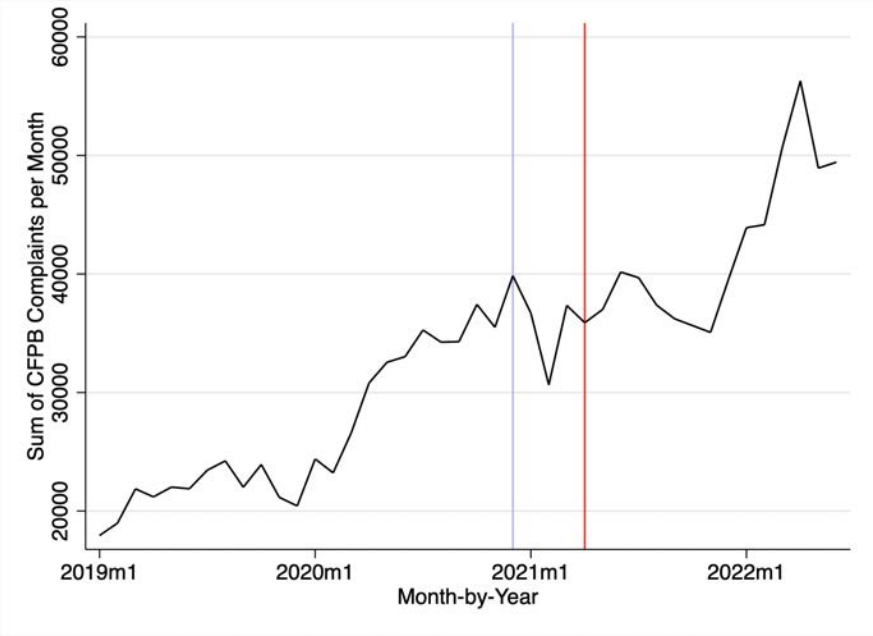
Peukert, C., S. Bechtold, M. Batikas, and K. Tobias (2021). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*, forthcoming.

Figure 1: Summary Statistics
January 2019 to June 2022

a. Number of Complaints per 1,000 People by Zip Code



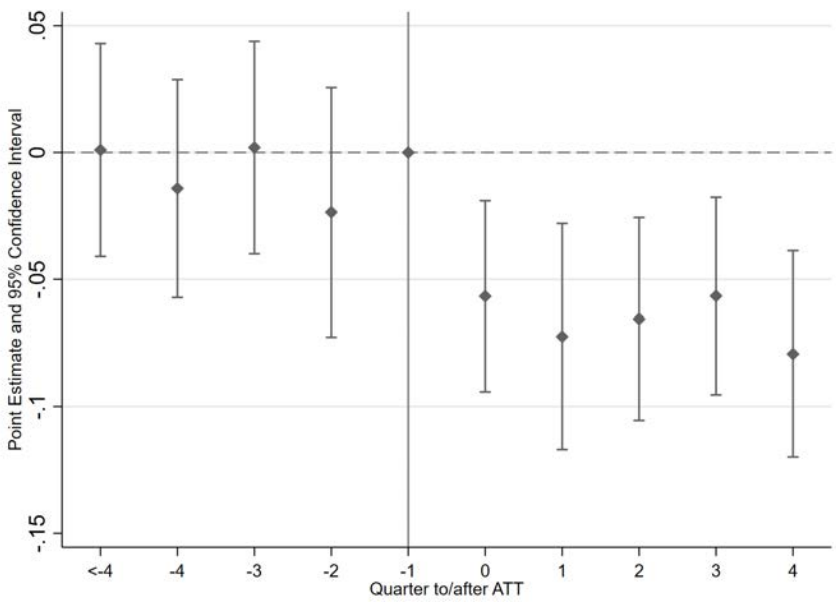
b. Number of Complaints per Month



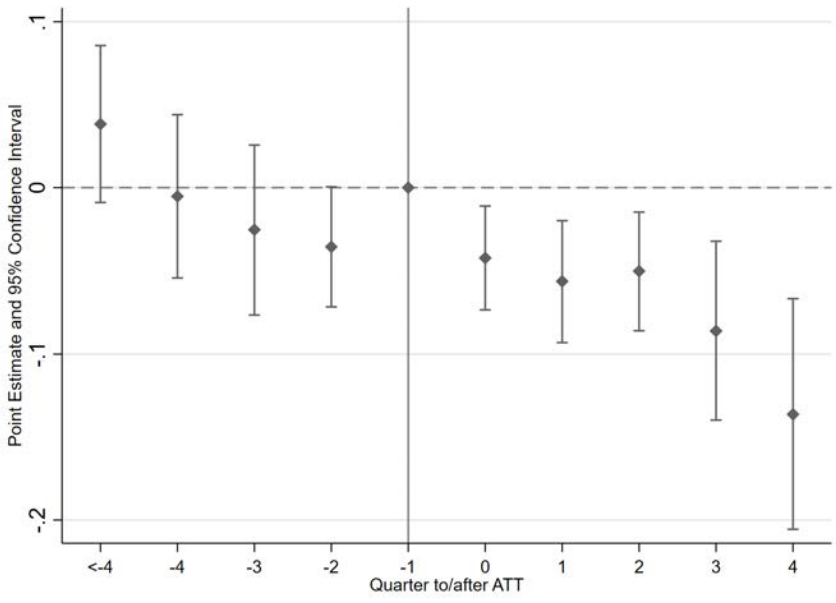
NOTE.—Figure 1a. illustrates the map of complaints over the entire sample period. Figure 1b. illustrates the total number of complaints over the sample period. The red line indicates the implementation date of Apple's App Tracking Transparency Policy (April 2021). The light blue line indicates the introduction of Apple's privacy label policy (December 2020).

Figure 2: Dynamics
Quarterly: continuous iOS share

a. Extensive margin: Any complaints



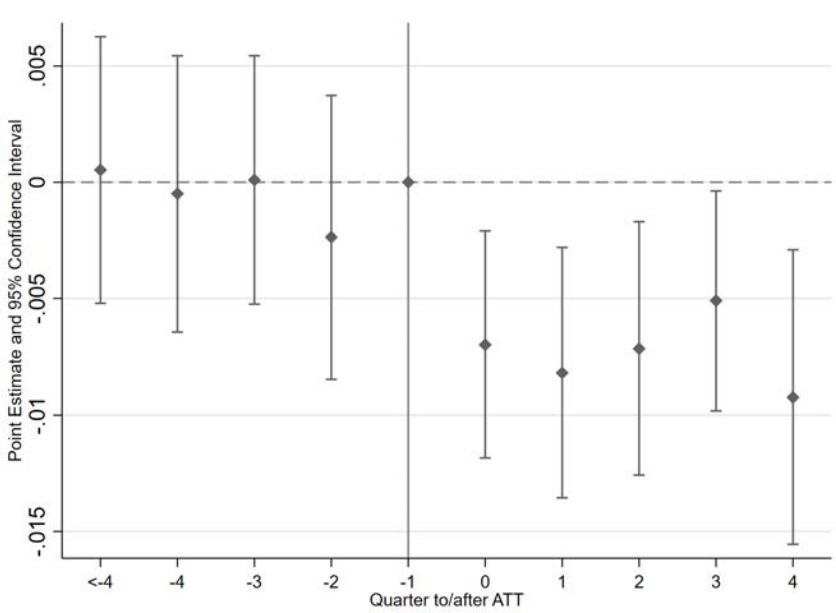
b. Intensive margin: Complaints per 1,000 residents



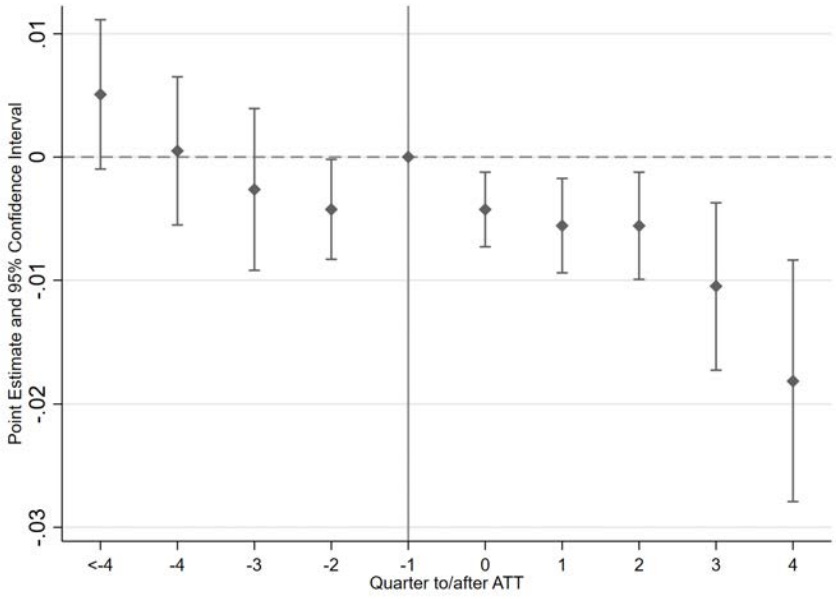
NOTE.—Figure 2 illustrates the dynamic effect of ATT on financial fraud around the implementation of ATT (April 26th, 2021 or 2021Q2). Quarter -1 is the quarter before the implementation (2021Q1), and is the omitted category. All three months in a corresponding quarter are grouped to reduce estimation error. For example, Quarter 0 corresponds to April, May, and June of 2021. The outcome variables in Figure 2a. and Figure 2b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin). Coefficients on the interaction term between the policy indicator and the pre-policy iOS device share are plotted.

Figure 3: Dynamics
Quarterly: iOS share top tercile

a. Extensive margin: Any complaints



b. Intensive margin: Complaints per 1,000 residents

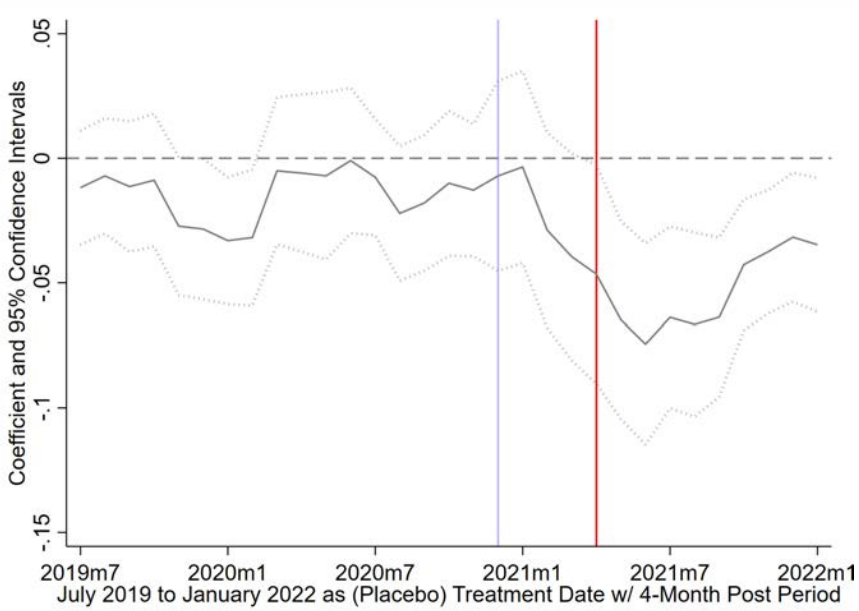


NOTE.—Figure 3 illustrates the dynamic effect of ATT on financial fraud around the implementation of ATT (April 26th, 2021 or 2021Q2). Quarter -1 is the quarter before the implementation (2021Q1), and is the omitted category. All three months in a corresponding quarter are grouped to reduce estimation error. For example, Quarter 0 corresponds to April, May, and June of 2021. The outcome variables in Figure 3a. and Figure 3b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin). Coefficients on the interaction term between the policy indicator and the tercile group for the highest pre-policy iOS device share are plotted.

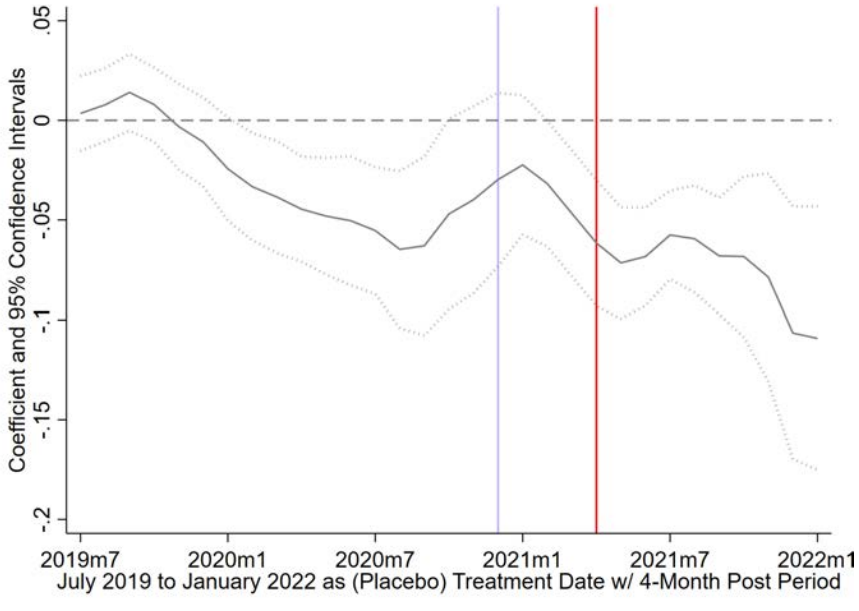
Figure 4: Placebo Treatment Date

4-Month post-treatment window, continuous iOS share

a. Extensive margin: Any complaints



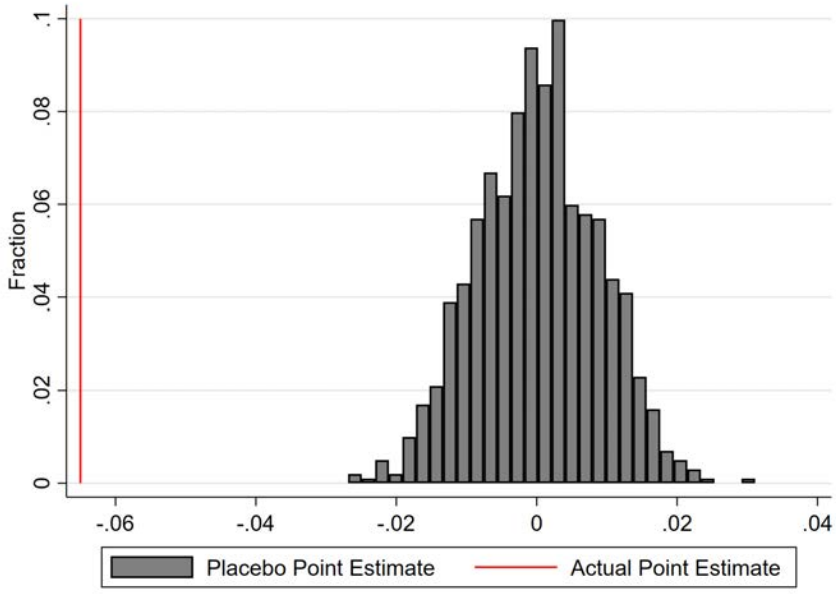
b. Intensive margin: Complaints per 1,000 residents



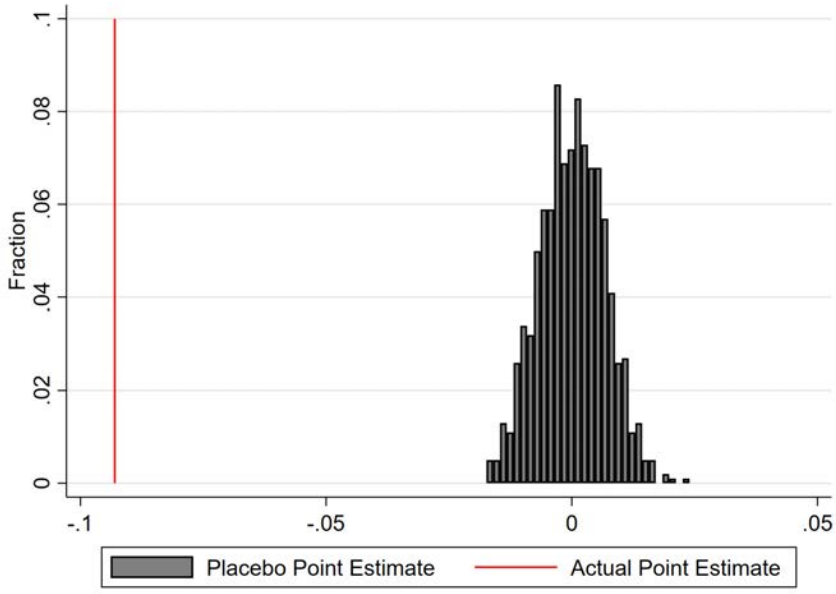
NOTE.—Figure 4 considers any month from July 2019 to January 2022 as a placebo treatment date and plots the effect on financial fraud. For any placebo treatment date, a 4-month post-event window (in addition to a pre-event window starting from January 2019) is used for estimation. The outcome variables in Figure 4a. and Figure 4b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin). Coefficients on the interaction term between an indicator for post-placebo-event period and the pre-event iOS device share are plotted. The red line indicates the implementation date of Apple's App Tracking Transparency Policy (April 2021). The light blue line indicates the introduction of Apple's privacy label policy (December 2020).

Figure 5: Placebo Treatment Intensity
Random iOS share

a. Extensive margin: Any complaints



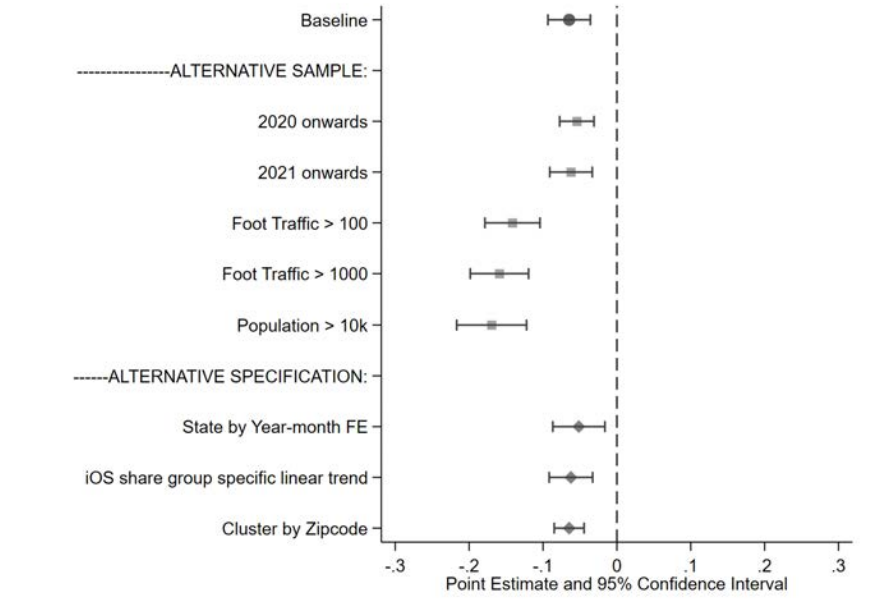
b. Intensive margin: Complaints per 1,000 residents



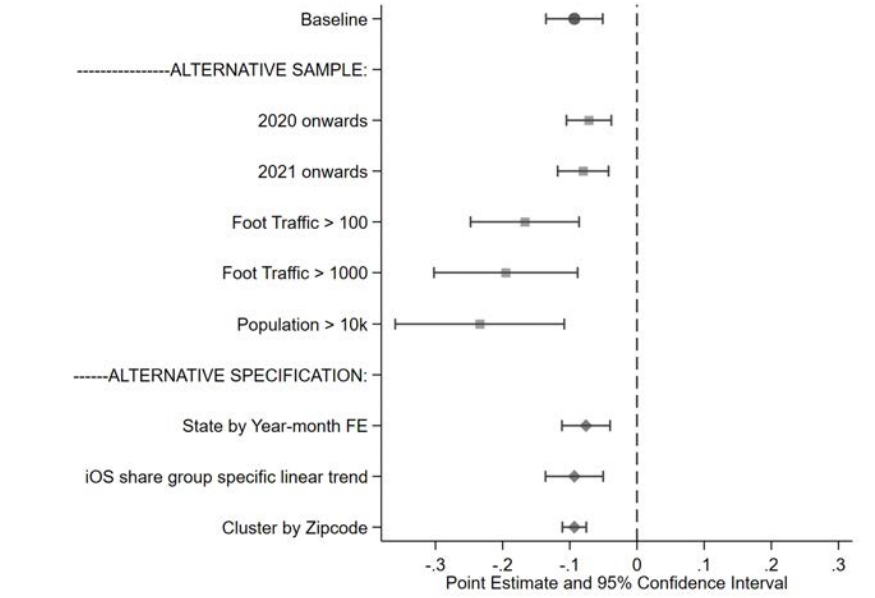
NOTE.—Figure 5 plots the histogram of the estimated coefficients on the privacy regulation from 1,000 placebo tests. Each placebo test randomly reshuffles treatment intensities (iOS share of devices) to zip codes. The sample and regression specifications are the same as those in Table 2. The outcome variables in Figure 5a. and Figure 5b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin).

Figure 6: Robustness

a. Extensive margin: Any complaints



b. Intensive margin: Complaints per 1,000 residents



NOTE.—Figure 6 plots the effect of the privacy regulation for alternative regression samples and specifications. The top row of each panel shows the baseline estimate in Columns 1 and 3 of Table 2. The outcome variables in Figure 6a. and Figure 6b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin).

Table 1: Summary Statistics

Panel a. CFPB sample

	mean	sd	p25	p50	p75	count
Zipcode-month level						
Any complaints (0/1)	0.31	0.46	0.00	0.00	1.00	1,026,942
Complaints winsorized	1.38	3.55	0.00	0.00	1.00	1,026,942
Complaints per 1,000 capita	0.07	0.20	0.00	0.00	0.06	1,003,758
log(1+Complaints)	0.44	0.77	0.00	0.00	0.69	1,026,942
iOS share	0.46	0.11	0.39	0.45	0.52	1,026,942
Month level						
Total complaints per month	36,936	12947	24399	38754	43043	42
Firm-month level						
Any complaints (0/1)	0.20	0.40	0.00	0.00	0.00	271,446
Complaints winsorized	1.61	8.04	0.00	0.00	0.00	271,446
log(1+Complaints)	0.30	0.79	0.00	0.00	0.00	271,446
Has an app (0/1)	0.26	0.44	0.00	0.00	1.00	271,446
Share data with third party (0/1)	0.11	0.31	0.00	0.00	0.00	12,235
Data isn't encrypted in transit (0/1)	0.01	0.10	0.00	0.00	0.00	12,235
Complaints per 1,000 downloads	3.14	8.56	0.00	0.02	1.49	12,235

Panel b. FTC sample

	mean	sd	p25	p50	p75	count
iOS share	0.46	0.11	0.39	0.45	0.52	955,925
Sample Complaints						
Any Complaint (0/1)	0.58	0.49	0.00	1.00	1.00	955,925
log(1+Complaint)	0.95	1.12	0.00	0.69	1.61	955,925
USD loss	8262.29	71527.76	0.00	0.00	217.00	955,925
Internet Services						
Any Complaint (0/1)	0.12	0.32	0.00	0.00	0.00	955,925
log(1+Complaint)	0.12	0.36	0.00	0.00	0.00	955,925
USD loss	173.49	6841.17	0.00	0.00	0.00	955,925
Privacy, Data Security, and Cyber Threats						
Any Complaint (0/1)	0.01	0.09	0.00	0.00	0.00	955,925
log(1+Complaint)	0.01	0.07	0.00	0.00	0.00	955,925
USD loss	18.14	2282.51	0.00	0.00	0.00	955,925

NOTE.—This table presents summary statistics for our key explanatory and outcome variables. Panel a. uses the CFPB dataset and Panel b. the FTC dataset. The CFPB data is aggregated to the zip code by month level. The FTC data is aggregated to the city by month level. The sample period is January 2019 to July 2022.

Table 2: Impact of the Privacy Regulation on Financial Fraud - CFPB Baseline

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.065***	−1.688***	−0.093***	−0.263***
	(0.014)	(0.327)	(0.021)	(0.045)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Observations	1,026,942	1,026,942	1,003,590	1,026,942
R-square	0.569	0.680	0.397	0.702

NOTE.—This table displays the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per zip code. The outcome variables from Column 1 to Column 4 are an indicator variable that equals to one if the zip code has any complaints, the number of complaints per zip code winsorized at 1%, the number of complaints per 1,000 residents at the zip code level, and the logarithm of one plus the number of complaints per zip code. We use all years after 2019 as the sample period, include zip code and county × year-month fixed effects, and cluster standard errors at the state level. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Impact of the Privacy Regulation on Financial Fraud - Demographics

Number of Complaints per 1,000 Residents

	(1)	(2)	(3)	(4)
Post-policy × iOS share	−0.043***	−0.064***	−0.064***	−0.070***
	(0.013)	(0.013)	(0.014)	(0.020)
Post-policy × Black share above median	0.065***			
	(0.013)			
Post-policy × Black share above median × iOS share	−0.092***			
	(0.027)			
Post-policy × Asian share above median		0.036**		
		(0.014)		
Post-policy × Asian share above median × iOS share		−0.063**		
		(0.030)		
Post-policy × Female share above median			0.044***	
			(0.012)	
Post-policy × Female share above median × iOS share			−0.069***	
			(0.024)	
Post-policy × Age 10-19 share above median				0.023***
				(0.008)
Post-policy × Age 10-19 share above median × iOS share				−0.044***
				(0.016)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Observations	1,002,666	1,002,666	1,002,666	1,002,666
R-square	0.398	0.398	0.398	0.398

NOTE.—This table displays the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per zip code, interacted with a number of demographic indicators. The outcome variable in all columns is the number of complaints per 1,000 residents at the zip code level. We use all years after 2019 as the sample period, include zip code and county × year-month fixed effects, and cluster standard errors at the state level. We divide zip codes into two groups based on sample median of the following demographics: share of black population, share of asian population, share of female, and share of people aged between 10-19. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Impact of the Privacy Regulation on Financial Fraud - by Product

Panel a. Top1 Fraud Category - Credit Reporting and Credit Repair Services

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.098***	−1.480***	−0.075***	−0.295***
	(0.023)	(0.288)	(0.019)	(0.053)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.181	0.851	0.042	0.270
Observations	1,026,942	1,026,942	1,003,590	1,026,942
R-square	0.506	0.611	0.420	0.615

Panel b. Top2 Fraud Category - Debt Collection

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.049***	−0.154***	−0.009***	−0.082***
	(0.009)	(0.030)	(0.002)	(0.015)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.100	0.167	0.008	0.096
Observations	1,026,942	1,026,942	1,003,590	1,026,942
R-square	0.333	0.364	0.224	0.375

Panel c. Bottom1 Fraud Category - Mortgage

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.009	−0.011	−0.000	−0.008
	(0.007)	(0.008)	(0.001)	(0.005)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.065	0.076	0.004	0.051
Observations	1,026,942	1,026,942	1,003,590	1,026,942
R-square	0.231	0.239	0.161	0.241

NOTE.—This table displays the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per zip code, by product categories selected by consumers when filing the complaint. The outcome variables from Column 1 to Column 4 are an indicator variable that equals to one if the zip code has any complaints, the number of complaints per zip code winsorized at 1%, the number of complaints per 1,000 residents at the zip code level, and the logarithm of one plus the number of complaints per zip code. We use all years after 2019 as the sample period, include zip code and county × year-month fixed effects, and cluster standard errors at the state level. The three panels respectively report results for top two and bottom one product categories ranked by the likelihood of receiving complaints related to financial fraud: Credit Reporting and Credit Repair Services (Top 1), Debt Collection (Top 2), and Mortgage (Bottom 1). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Impact of the Privacy Regulation on Financial Fraud - by Product and Narratives

Panel a. Top1 Fraud Category - Credit Reporting and Credit Repair Services

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.126***	−0.517***	−0.023***	−0.203***
	(0.024)	(0.123)	(0.006)	(0.043)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.088	0.244	0.011	0.108
Observations	868,266	868,266	853,818	868,266
R-square	0.346	0.391	0.292	0.400

Panel b. Top2 Fraud Category - Debt Collection

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.037***	−0.065***	−0.004***	−0.050***
	(0.011)	(0.020)	(0.001)	(0.017)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.052	0.066	0.003	0.045
Observations	868,266	868,266	853,818	868,266
R-square	0.220	0.229	0.179	0.236

Panel c. Bottom1 Fraud Category - Mortgage

	(1) Any complaints (0/1)	(2) Complaints winsorized	(3) Complaints per 1,000 capita	(4) log(1+Complaints)
Post-policy × iOS share	−0.007	−0.007	−0.000	−0.005
	(0.005)	(0.005)	(0.000)	(0.004)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.035	0.035	0.002	0.026
Observations	868,266	868,266	853,818	868,266
R-square	0.160	0.160	0.140	0.163

NOTE.—This table displays the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per zip code, by product categories selected and narratives provided by consumers when filing the complaint. The outcome variables from Column 1 to Column 4 are an indicator variable that equals to one if the zip code has any complaints, the number of complaints per zip code winsorized at 1%, the number of complaints per 1,000 residents at the zip code level, and the logarithm of one plus the number of complaints per zip code. We use all years after 2019 as the sample period, include zip code and county × year-month fixed effects, and cluster standard errors at the state level. The three panels respectively report results for top two and bottom one product categories ranked by the likelihood of receiving complaints related to financial fraud: Credit Reporting and Credit Repair Services (Top 1), Debt Collection (Top 2), and Mortgage (Bottom 1). Using the consumer narrative, we further refine the sample in each product category using a likelihood of fraud generated by machine learning techniques. For the top two fraud categories, we drop the complaints in the bottom three deciles of the likelihood distribution. For the bottom one fraud category, we drop the complaints in the top two deciles of the likelihood distribution. The choice of the cutoff decile is based the fraction of fraud cases identified using the keyword search. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Impact of the Privacy Regulation on City-level Complaints - FTC Baseline

	Internet Services		**Data Security**	
	(1)	(2)	(3)	(4)
	Any complaints (0/1)	log(1+Complaints)	Any complaints (0/1)	log(1+Complaints)
Post-policy × iOS share	−0.023***	−0.034***	−0.010***	−0.008***
	(0.008)	(0.008)	(0.002)	(0.002)
City FE	✓	✓	✓	✓
State × Year-month FE	✓	✓	✓	✓
Observations	955,925	955,925	955,925	955,925
R-square	0.397	0.607	0.141	0.180

NOTE.—This table examines the impact of the policy using the FTC complaint data. Because the FTC data only contain the city and state of the filers, we aggregate the complaints to the city-month level and and display the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per city. The FTC provides fine product categories, based on which we identify two that are most likely to have fraud-related complaints: Internet Services and Data Security. The outcome variable in the odd column is an indicator variable for whether there is any complaints belonging to that specific product category at the city-month level, and in even columns is the logarithm of one plus the number of complaints at the city-month level. We plan to collect city-level population data to scale the number of complaints as an additional outcome variable. Standard errors are clustered at the state level and are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Impact of the Privacy Regulation on City-level Complaints - FTC Dollar Loss

	(1) Any complaints (0/1) Positive USD loss	(2) log(1+Complaints) Positive USD loss	(3) log(1+USD loss) All Complaints	(4) log(1+USD loss) Internet & Data Security
Post-policy × iOS share	−0.022**	−0.068***	−0.047	−0.401***
	(0.009)	(0.017)	(0.078)	(0.119)
City FE	✓	✓	✓	✓
State × Year-month FE	✓	✓	✓	✓
Observations	955,925	955,925	955,925	955,925
R-square	0.483	0.765	0.574	0.286

NOTE.—This table examines the impact of the privacy regulation using the dollar loss information in the FTC complaint data. Because the FTC data only contain the city and state of the filers, we aggregate the complaints to the city-month level and and display the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per city. In the Columns 1 and 2, we aggregate all complaints with positive reported dollar loss at the city-month level. The outcome variables are an indicator variable for whether there is any such complaint and the logarithm of one plus the number of such complaints. In Columns 3 and 4, we use the logarithm of one plus the total amount of dollar loss as the outcome variable. In Column 3, we include all complaints regardless of the reported dollar loss. In Column 4, we restrict the sample to include only complaints about the most relevant product categories, Internet Service and Data Security products. We plan to collect city-level population data to scale the number of complaints as an additional outcome variable. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Impact of the Privacy Regulation on Complaints - Firm Level

	Any complaints (0/1)	log(1+Complaints)	Complaints per 1,000 downloads	
	(1)	(2)	(3)	(4)
Has an app × Post-policy	−0.011**	−0.039***		
	(0.005)	(0.011)		
Share data with third party (0/1) × Post-policy			−1.693**	
			(0.695)	
Data isn't encrypted in transit (0/1) × Post-policy				−0.814**
				(0.332)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
App-size-specific linear trend	✓	✓	✓	✓
Sample	Full	Full	App sample	App sample
Observations	271,446	271,446	12,235	12,235
R-square	0.534	0.833	0.638	0.638

NOTE.—This table reports the treatment effect of the privacy regulation on financial complaints at the firm-month level. In Columns 1 and 2, the indicator for the post-policy period is interacted with an indicator of whether a firm owns an iOS app. The two outcome variables are an indicator variable for whether the firm received any complaint in a given month and the logarithm of one plus the total number of complaints per firm-month. In Columns 3 and 4, we include only firms with an app and interact the indicator for the post-policy period with an indicator for firms that share user data with third parties (Column 3) or that do not encrypt data in transit (Column 4). The outcome variable in both columns is the number of complaints per 1,000 monthly app downloads. We include firm and year-month fixed effects and control for linear time trend specific to app popularity (measured by the log of world-word all time downloads). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Impact of the Privacy Regulation on Financial Fraud - Firm Level by Product

Panel a. Top Fraud Category - Credit Reporting and Credit Repair Services

	Any complaints (0/1)	log(1+Complaints)	Complaints per 1,000 downloads	
	(1)	(2)	(3)	(4)
Has an app × Post-policy	−0.008**	−0.029***		
	(0.004)	(0.006)		
Share data with third party (0/1) × Post-policy			−0.758*	
			(0.450)	
Data isn't encrypted in transit (0/1) × Post-policy				−0.064**
				(0.032)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
App-size-specific linear trend	✓	✓	✓	✓
Sample	Full	Full	App sample	App sample
Observations	271,446	271,446	12,235	12,235
R-square	0.534	0.821	0.544	0.677

Panel b. Top2 Fraud Category - Debt Collection

	Any complaints (0/1)	log(1+Complaints)	Complaints per 1,000 downloads	
	(1)	(2)	(3)	(4)
Has an app × Post-policy	−0.007*	−0.018***		
	(0.004)	(0.006)		
Share data with third party (0/1) × Post-policy			−1.060*	
			(0.616)	
Data isn't encrypted in transit (0/1) × Post-policy				−0.120
				(0.083)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
App-size-specific linear trend	✓	✓	✓	✓
Sample	Full	Full	App sample	App sample
Observations	271,446	271,446	12,235	12,235
R-square	0.541	0.791	0.463	0.634

Panel c. Bottom Fraud Category - Mortgage

	Any complaints(0/1)	log(1+Complaints)	Complaints per 1,000 downloads	
	(1)	(2)	(3)	(4)
Has an app × Post-policy	−0.001	−0.003		
	(0.003)	(0.004)		
Share data with third party (0/1) × Post-policy			−0.140	
			(0.145)	
Data isn't encrypted in transit (0/1) × Post-policy				−0.017
				(0.012)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
App-size-specific linear trend	✓	✓	✓	✓
Sample	Full	Full	App sample	App sample
Observations	271,446	271,446	12,235	12,235
R-square	0.539	0.835	0.557	0.717

NOTE.—This table reports the treatment effect of the privacy regulation on financial complaints at the firm-month level, by product categories that consumers selected when filing the complaint. In Columns 1 and 2, the indicator for the post-policy period is interacted with an indicator of whether a firm owns an iOS app. The two outcome variables are an indicator variable for whether the firm received any complaint in a given month and the logarithm of one plus the total number of complaints per firm-month. In Columns 3 and 4, we include only firms with an app and interact the indicator for the post-policy period with an indicator for firms that share user data with third parties (Column 3) or that do not encrypt data in transit (Column 4). The outcome variable in both columns is the number of complaints per 1,000 monthly app downloads. We include firm and year-month fixed effects and control for linear time trend specific to app popularity (measured by the log of world-word all time downloads). We further divide the firm-level complaints by category and explore the category-specific treatment effects. The three panels respectively report results for top two and bottom one product categories ranked by the likelihood of receiving complaints related to financial fraud: Credit Reporting and Credit Repair Services (Top 1), Debt Collection (Top 2), and Mortgage (Bottom 1). Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Impact of the Privacy Regulation on Cyber Incidents - Firm Level

	Cyber incidents (0/1)			
	(1)	(2)	(3)	(4)
	All types	Privacy violation	Regulation violated	Regulations on Debt collection credit reporting
Has an app × Post-policy	−0.047***	−0.045***	−0.023***	−0.009***
	(0.011)	(0.009)	(0.005)	(0.003)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.113	0.081	0.029	0.012
Observations	56,224	56,224	56,224	56,224
R-square	0.232	0.221	0.231	0.222

NOTE.—This table reports the treatment effect of the privacy regulation on cyber incidents aggregated at the firm-month level using data from Advisen. In all columns, the indicator for the post-policy period is interacted with an indicator of whether a firm owns an iOS app. The outcome variables are an indicator variable for whether the firm was exposed to (1) any cyber incident, (2) cyber incidents that were result of privacy risks, (3) cyber incidents that violated regulations related to consumer protection, and (4) cyber incidents that violated the Fair Debt Collection Practices Act or the Fair Credit Reporting Act. We include firm and year-month fixed effects throughout. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

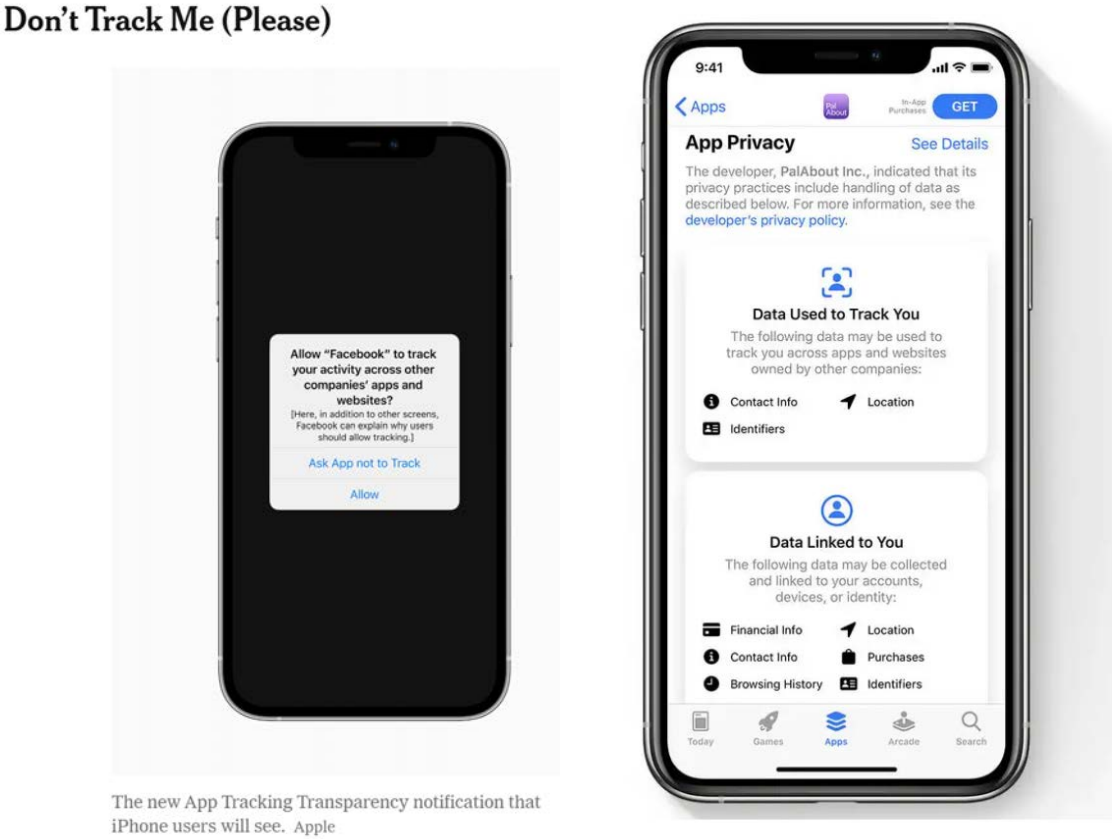
Consumer Surveillance and Financial Fraud

Online Appendix

Bo Bian Michaela Pagel Huan Tang

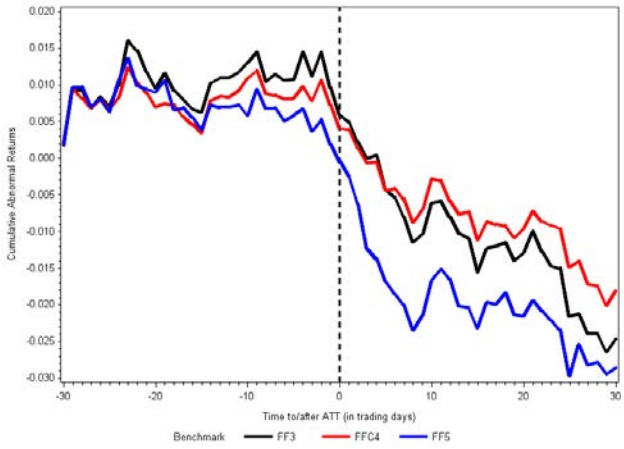
A The App Tracking Transparency and Privacy Nutrition Labels Policies

Figure A.1: Examples of the ATT prompt and the privacy labels



The new App Tracking Transparency notification that iPhone users will see. Apple

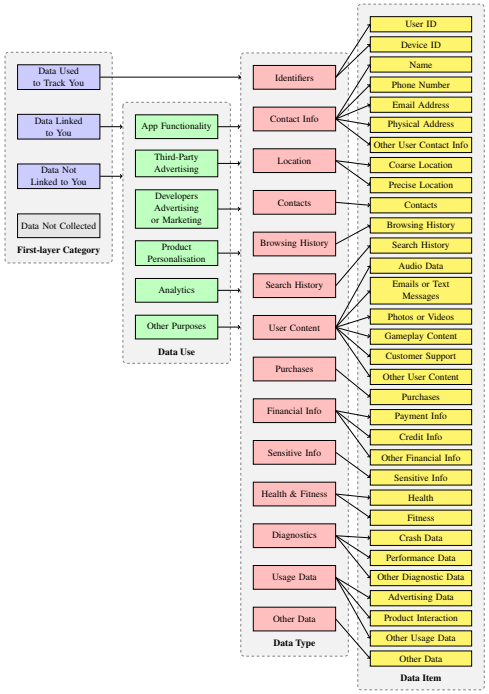
Figure A.2: Stock Market Reactions around ATT



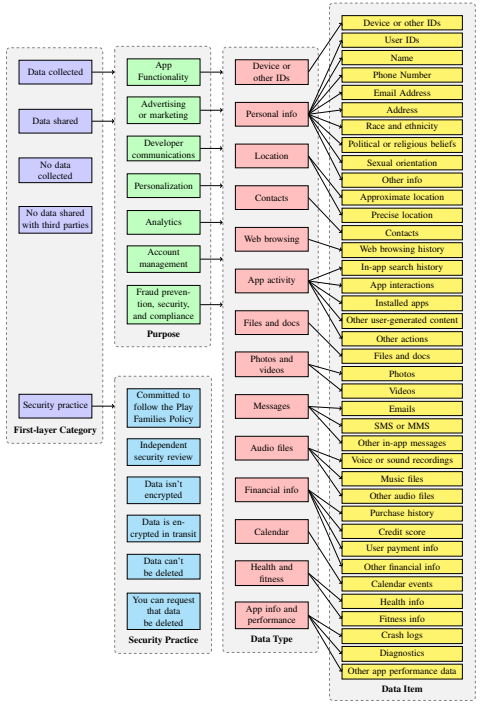
NOTE.— From Bian et al. (2021). This figure plots the average cumulative abnormal returns (CARs) around the implementation of the App Transparency Tracking Policy on April 26, 2021. The event window includes 30 days before and after the implementation date. CARs are computed using the Fama-French factor models

Figure A.3: Apple's App Privacy Label

a. Apple's Privacy Labels



b. Google Safety Form



NOTE.— Panels a. and b. show the structure of the mandatory privacy labels and google safety forms for iOS and Android apps, respectively. There are four layers in the privacy labels. The first layer consists of three categories: *Data Used to Track You*, *Data Linked to You*, and *Data Not Linked to You*. If an app doesn't collect any data, it will have *Data Not Collected* as the only layer in its privacy label. For the second layer, only *Data Linked to You* and *Data Not Linked to You* have this layer which shows 6 different purposes of data use. The third layer includes 14 different data types that the app collects, all data types can appear under each of the 6 purposes of data use in the second layer. The fourth layer reports 32 data items under the corresponding data type in the third layer. The first and the third layers are displayed on the main App Store page while the second and the fourth layers are only displayed in a pop-up window when one clicks on the "See Details" button in the upper right corner of the App Privacy section. The structure of Google safety forms is similar, except that it provides additional information on the data security practice of the firm. Firms have to disclose whether the firm is committed to follow the Play Families Policy, whether the firm received an independent security review, whether data is encrypted in transit, and whether users can request their data to be deleted.

B FTC Complaint Form

PRINT

Consumer Sentinel Network Complaints

Record # 1 / Consumer Sentinel Network Complaints			
Reference Number:	44649725	Originator Reference Number:	
Language:	English	Contact Type:	Complaint
Source:	Consumer	DNC?	N
Comments:			
Was the complaint resolved?:		Complaint Resolution:	
Data Reference:			
Entered By:	FTCCIS-FTCUSER	Entry Date:	3/20/2013
Updated By:		Updated Date:	
Complaint Source:	FTC Online Complaint Assistant (CIS)	Product Service Code:	Other (Note in Comments)
Amount Requested:		Amount Paid:	
Payment Method:		Agency Contact:	Internet
Complaint Date:	3/20/2013	Transaction Date:	
Initial Contact:		Initial Response:	
Statute/Rule:		Law Violation:	Deception/Misrepresentation
Topic:		Dispute with Credit Bureau?:	
Dispute with Credit Bureau - Responded?:		Dispute with Credit Bureau - Resolved to Satisfaction?:	
Member of armed forces or dependent?:	Yes		
Consumer Information			
Consumer			
Complaining Company/Org:			
First Name;		Last Name:	Not Provided
Address 1:		Address 2:	
City:		State:	
Zip:		Country:	UNITED STATES
Home Number:		Work Number:	
Fax Number:		Ext:	
Email:		Age Range:	
Military Service Branch:		Soldier Status:	
Soldier Station:			
Subject			
Subject:	Unknown		
Address:			
City:		State/Prov:	

ZIP:		Country:	United States
Email:		URL:	
Area Code:		Phone Number:	
Ext:		Subject ID Type:	
Subject ID Issuer State:		Subject ID Issuer Country:	
Representative Name:		Title:	

C Classification of Fraud-related Complaints

Table C.1: Classification of Complaints using Two Approaches

	Mean		St. Dev.		
	keyword	ZSL	keyword	ZSL	N
Credit reporting, credit repair services, or other personal consumer reports	0.822	0.526	0.383	0.284	349,106
Debt collection	0.727	0.517	0.445	0.238	99,484
Money transfer, virtual currency, or money service	0.656	0.426	0.475	0.239	18,792
Checking or savings account	0.510	0.436	0.500	0.251	36,263
Credit card or prepaid card	0.494	0.406	0.500	0.250	54,899
Vehicle loan or lease	0.392	0.289	0.488	0.180	13,023
Payday loan, title loan, or personal loan	0.345	0.315	0.475	0.207	8,472
Student loan	0.341	0.248	0.474	0.167	10,490
Mortgage	0.313	0.159	0.464	0.116	42,559

NOTE.—This table report the likelihood of fraud cases by product category based two approaches: keyword search and zero shot learning. The keyword search method returns a binary outcome that is equal to one if any of the keywords is found in the issue, subissue, and consumer narrative fields. The zero-shot-learning method returns a continuous variable that represents the likelihood of a fraud-related complaint. Columns 1 and 2 report the mean of the fraud measure, the next two columns report the standard deviation, and the last column the number of observations in each product category. The sample only includes complaints with narratives.

Table C.1 reports the likelihood of fraud cases by product category based on two approaches: keyword search and zero shot learning. The two methods deliver similar rankings, with "Credit reporting, credit repair services, or other personal consumer reports" and "Debt collection" being the top two fraud categories, and "Mortgage" being the bottom category. Below we describe the details of both approaches.

Keyword search Our goal is to classify complaints into cases that are more versus less likely to be triggered by data privacy related issues. To do this, we search for certain keywords in the issue, subissue, and more importantly, consumer narrative fields. The following keywords are included: "incorrect", "improper", "false", "wrong", "missing", "fraud", "scam", "theft", "embezzlement", "imposter", "unauthorized", "unsolicited", "identity", "sharing", "advertising", "marketing", "security", "breach". The keyword search method returns a binary outcome that is equal to one if any of the keywords was found in the issue, subissue, and consumer narrative fields.

Zero-shot learning We develop an alternative measure to classify complaints using the machine learning approach "zero-shot learning". The method does not require manual annotations and is therefore a more robust approach when few labeled observations are available,

as its understanding of language is rooted in a large diverse sample of text. We use the BART-large-mnli model from Facebook, which uses the pre-trained BART-large language model and adds a task-specific head. Within this model structure, we consider the hypothesis format "I am reporting {label}" with the following 17 labels: "a data breach", "a mistake", "an inaccuracy", "an oversight", "an unauthorized action", "an unrecognized action", "card fraud", "collection scam", "debt collection scam", "embezzlement", "fraud", "harassment", "identity theft", "mistreatment", "mortgage scam", "scam", and "unresponsiveness." Varying the hypothesis from "I am reporting a data breach" to "I am reporting unresponsiveness", for example, while keeping the narrative constant will change the scores generated since the relationship between the narratives and the hypotheses changes.

The relationship between the premise and hypothesis can either be an entailment, neutral, or a contradiction. The model outputs a logit score for each case (e_i, n_i, c_i, respectively). An example of a full query to determine if a specific narrative refers to identity theft includes a narrative (premise) such as *"I am the victim of identity theft. Please remove the fraudulent accounts from my credit report."* and a hypothesis *"I am reporting identity theft."* In this case, a good model outputs a high logit score for entailment and a low score for contradiction.

To combine these multiple logit scores into a single probability, we run a logistic regression with a Lasso penalty on a manually annotated representative sample of 1,400 narratives with the entailment, neutral, and contradiction scores as regressors. By choosing a non-linear combination of the labels' scores, we can slightly tailor the concept of fraud to our context beyond the ZSL's language model's representation.

Specifically, we use as regressor the scores for the whole list of labels and use logistic regression with lasso penalty to calculate a final fraud probability based on the most important entailment, neutral, and contradiction scores. Compared with the ridge penalty that uses the squared magnitude of estimated coefficients, the lasso penalty uses the absolute value of estimated coefficients and sets some coefficients equal to zero. The logistic regression is run on 1400 manually annotated narratives whose product distribution reflects the

total sample's distribution. The outcome variable of the logistic regression is the manual annotated dummy score.The estimated model has a non-zero coefficient for 17 out of the possible 51 features (3×17), with 10 entailment, 3 neutral and 4 contradiction scores. The largest positive coefficient is "identity theft" entailment with 1.05, and the most negative coefficients are "fraud" neutral with -0.57 followed by "mistreatment" entailment with -0.38. Using this estimated model, we then combine the 51 features for all narratives into a final score.

Example narratives on data-driven fraud incidents Below we list a few example complaint narratives that scored highly under both methods. We can see that these narratives clearly reveal that the reporting individuals have been a victim of data breach, identity theft and that the unverified inquiries/account/debt are typical consequences.

> *Complaint ID - 3758105* "I am a victim of identity theft. Due to the Corona Virus Pandemic, we are all facing which has me sitting still at home and I saw the recent news about the multiple XXXX Data breaches. I decided to look at my credit reports from the 3 major credit bureaus and found that someone had used my Identity. I have no idea how the theft took place. I also have no knowledge of any suspects. I did not receive any money, goods, or services as a result of identity theft. I contacted the Credit Bureau and told me to file an Identity Theft Report which I am doing. I appreciate your effort in getting this matter resolved. Thank you. Please let me know if you need any other information from me to block this information from my credit report. Thank you.."

> *Complaint ID - 1904491* "GLOBAL RECEIVABLES SOLUTIONS XXXX have a a unverified account from. I had previously disputed this account. I have never done business with GLOBAL RECEIVABLES SOLUTIONS. Pursuant to the Fair Debt Collection Practices Act (FDCPA) 15 U.S.C.169g, I dispute the validity of the debt GLOBAL RECEIVABLES SOLUTIONS purport I owe. I request that GLOBAL RECEIVABLES SOLUTIONS Provide verification of the following : 1.) The original

Application or contract ; 2.) Any and all statements allegedly related to this debt ; 3.) Any and all signed receipts ; 4.) Any and all canceled checks ; 5.) Original date of default and collection activity begin 6.). Whether you purchased the debt, and if so, the amount paid for the debt 7.) The date (s) the debt allegedly accrued ; 8.) An itemization of the costs, including an accounting, for any additional interest, charges, or other fees placed on this account. I want to request that GLOBAL RECEIVABLES SOLUTIONS Cease and Desist all further communications and collection actives and provide the verification of the purported debt."

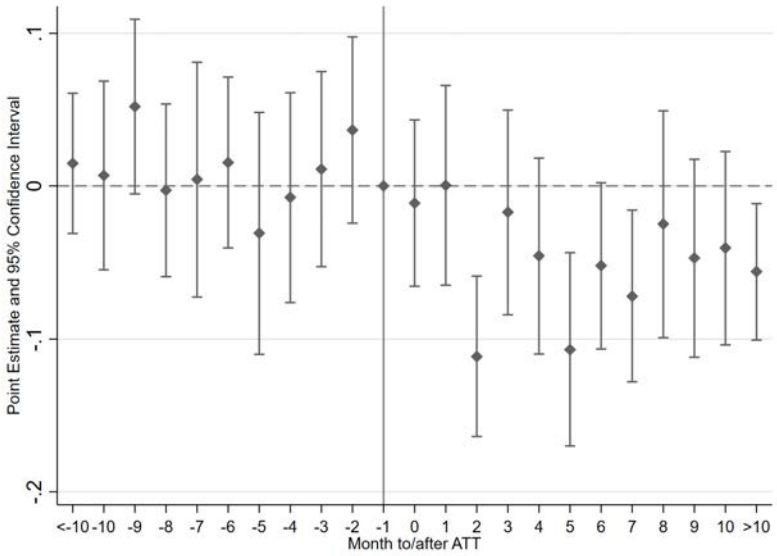
Complaint ID - 1488173 "Today I was Contacted by XXXX from credit control at XXXX on XXXX/XXXX/15 for the purpose of a debt collection. She Previously called on XXXX/XXXX/15 XXXX and was unable to provide information substantiating a debt she was attempting to collect from XXXX XXXX When we first spoke I informed her that There exist the possibility that I may be a victim of identity theft. To day when she called I informed I would not provide her with any verification information and to no longer contact me in regards to the matter or I would be forced to contact your agency and execute my rights under the law. I was very adamant and calm when I informed her of my wishes. XXXX informed me that the calls would continue despite my strict instructions that I do not want her to call my residence any more. To paraphrase her words, "it might not be me who calls but someone will call you".

Complaint ID - 5021069 "On XX/XX/2021 sent a letter regarding inaccurate and unknown things on my credit report, To this day over 60 days later I have not received a response yet. I feel like I'm being taken advantage of and being ignored of my disputes. Section 611 (a), it is plainly stated that a failure to investigate these items within 30days gives a reason to immediately remove those items from my credit report it has been over 60 days so they should be deleted promptly. I demand these accounts be deleted immediately or I will file for litigation due to the stress you caused me. My information was also impacted by the XXXX, Experian and XXXX data breach and may have got into the hands of the wrong person."

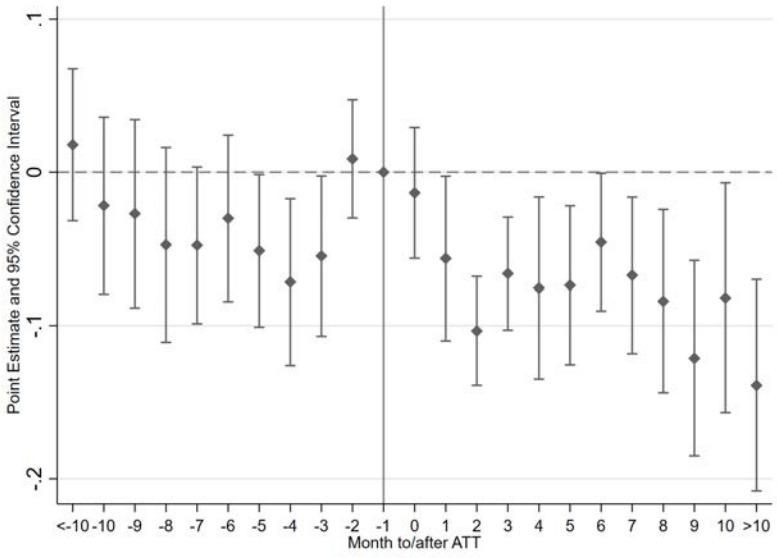
D Dynamic DiD Effects at Monthly Frequency

Figure D.1: Dynamic DiD Effects

Monthly: continuous iOS share



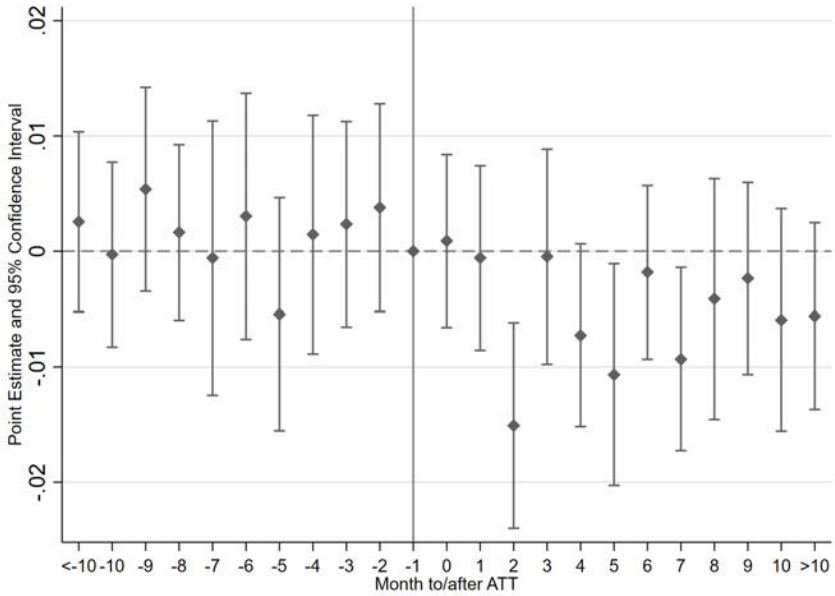
a. Extensive margin: Any complaints



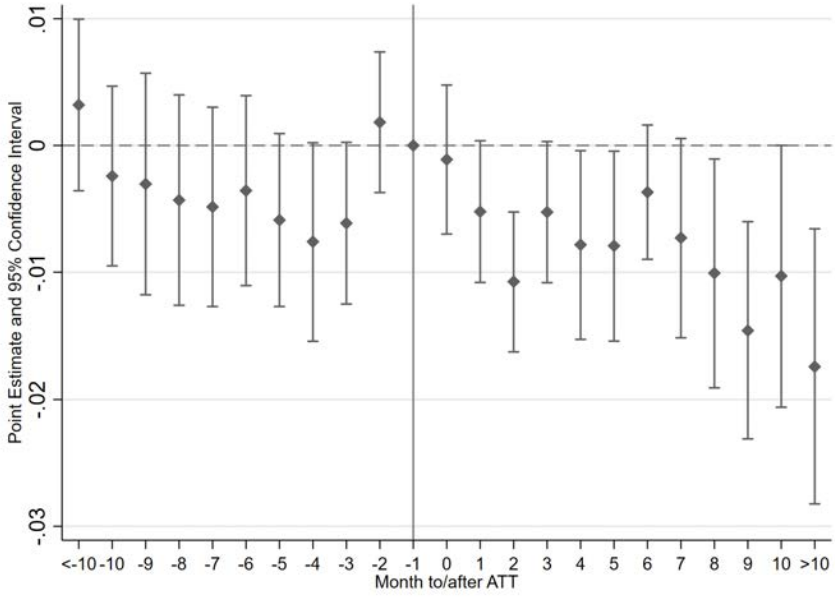
b. Intensive margin: Complaints per 1,000 residents

NOTE.—Figure D.1 illustrates the dynamic effect of ATT on financial fraud around the implementation of ATT (April 26th, 2021 or April, 2021). Month -1 is the month before the implementation (March, 2021), and is the omitted category. The outcome variables in Figure D.1a. and Figure D.1b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin). Coefficients on the interaction term between the policy indicator and the pre-policy iOS device share are plotted.

Figure D.2: Dynamic DiD Effects
Monthly: iOS share top tercile



a. Extensive margin: Any complaints



b. Intensive margin: Complaints per 1,000 residents

NOTE.—Figure D.2 illustrates the dynamic effect of ATT on financial fraud around the implementation of ATT (April 26th, 2021 or April, 2021). Month -1 is the month before the implementation (March, 2021), and is the omitted category. The outcome variables in Figure D.2a. and Figure D.2b. are the indicator for any complaints (extensive margin) and the number of complaints per 1,000 residents (intensive margin). Coefficients on the interaction term between the policy indicator and the tercile group for pre-policy iOS device share are plotted.

E Robustness Checks

Table E.1: CFPB Robustness Checks

Panel a. Extensive Margin: Any Complaints (0/1)

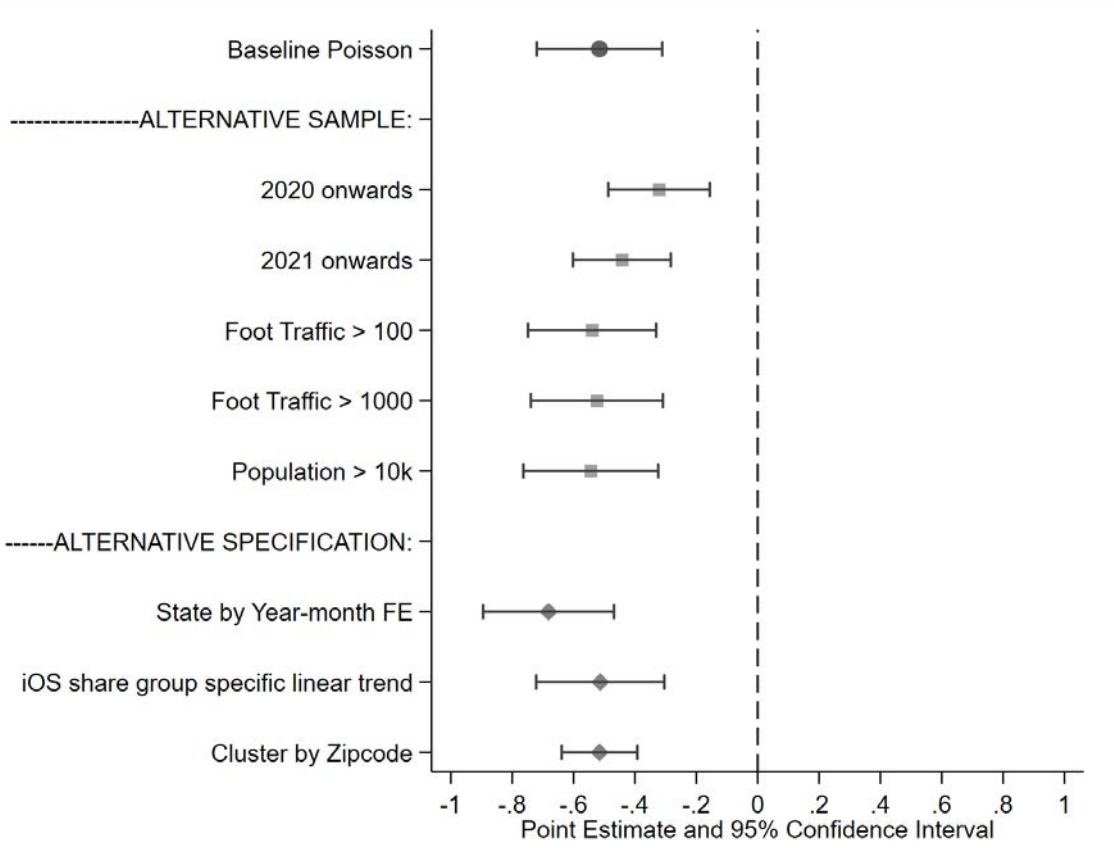
	(1) *baseline*	(2) *fixed effect*	(3) *cluster by zipcode*	(4) *discrete treatment*	(5) *sample period: 2020-*	(6) *foot traffic>100*	(7) *foot traffic>1000*
Post-policy × iOS share	−0.065***	−0.051***	−0.065***		−0.054***	−0.141***	−0.159***
	(0.014)	(0.018)	(0.010)		(0.012)	(0.019)	(0.020)
Post-policy × Top iOS share tercile				−0.019***			
				(0.004)			
Zipcode FE	✓	✓	✓	✓	✓	✓	✓
County × Year-month FE	✓		✓	✓	✓	✓	✓
State × Year-month FE		✓					
Observations	1,026,942	1,043,154	1,026,942	671,580	733,530	847,586	525,445
R-square	0.569	0.522	0.569	0.586	0.593	0.555	0.504

Panel b. Intensive Margin: Number of Complaints per 1,000 Residents

	(1) *baseline*	(2) *fixed effect*	(3) *cluster by zipcode*	(4) *discrete treatment*	(5) *sample period: 2020-*	(6) *foot traffic>100*	(7) *foot traffic>1000*
Post-policy × iOS share	−0.093***	−0.076***	−0.093***		−0.072***	−0.167***	−0.195***
	(0.021)	(0.018)	(0.009)		(0.017)	(0.040)	(0.053)
Post-policy × Top iOS share tercile				−0.025***			
				(0.007)			
Zipcode FE	✓	✓	✓	✓	✓	✓	✓
County × Year-month FE	✓		✓	✓	✓	✓	✓
State × Year-month FE		✓					
Observations	1,003,590	1,019,172	1,003,590	654,192	716,850	836,848	523,253
R-square	0.397	0.294	0.397	0.422	0.438	0.439	0.508

NOTE.—This table shows the baseline results are robust to various alternative specifications and samples. Panels a and b report the results on the extensive margin and intensive margin, measured by an indicator variable that equals to one if the zip code has any complaints and the number of complaints per 1,000 residents at the zip code level, respectively. Column 1 report the baseline result from Table 2 for reference purpose. In Column 2, we replace the county × year-month fixed effects with state × year-month fixed effects. Column 3 reports the result with zip code-level clustering. In Column 4, we use the tercile group for pre-policy iOS device share as the treatment intensity. In Column 5, we only include the observations after 2020. In the last two columns, we restrict the sample to zip codes with more than 100 and 1,000 visits recorded by SafeGraph. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure E.1: Robustness using Poisson Regressions



NOTE.— Figure E.1 plots the effect of the privacy regulation for different regression samples and specifications, using Possion regressions. The top row shows the baseline Possion estimate with the same sample and fixed effects as those in Table 2. The outcome variable in is the number of complaints at the zip code level (winsorized at 1%).

F Products related to Economic Impact Payments

In this table, we explore whether there is a significant reduction in two categories that are most relevant for EIP. These categories are checking and saving accounts and credit cards and prepaid cards.

Table F.1: Product Categories relevant for EIP-related Complaints

	(1)	(2)	(3)	(4)
	Any complaints (0/1)	Complaints winsorized	Complaints per 1,000 capita	log(1+Complaints)
Post-policy × iOS share	−0.022**	−0.046**	−0.002*	−0.024**
	(0.010)	(0.019)	(0.001)	(0.011)
Zipcode FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.127	0.176	0.009	0.108
Observations	1,026,942	1,026,942	1,003,590	1,026,942
R-square	0.347	0.373	0.200	0.380

NOTE.—This table displays the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per zip code, for product categories most relevant for EIP, checking and saving accounts and credit cards and prepaid cards. The outcome variables from Column 1 to Column 4 are an indicator variable that equals to one if the zip code has any complaints, the number of complaints per zip code winsorized at 1%, the number of complaints per 1,000 residents at the zip code level, and the logarithm of one plus the number of complaints per zip code. We use all years after 2019 as the sample period, include zip code and county × year-month fixed effects, and cluster standard errors at the state level. Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

G FTC reports: Other Results

This table examines the impact of the policy on identity theft report using data from FTC. Because the FTC data only contain the city and state of the filers, we aggregate the complaints at city-month level and and display results from our Specification 1. In Columns 1 and 2, the outcome variables are an indicator variable for whether there is any such complaint and the logarithm of one plus the number of such complaints.

Table G.1: The Impact of Privacy Regulation on Identity Theft Reports

	Identity Theft	
	(1)	(2)
	Any complaints (0/1)	log(1+Complaints)
Post-policy × iOS share	0.006	−0.147***
	(0.012)	(0.051)
City FE	✓	✓
State × Year-month FE	✓	✓
Observations	679,049	679,049
R-square	0.554	0.852

NOTE.—This table examines the impact of the policy using the FTC complaint data. Because the FTC data only contain the city and state of the filers, we aggregate the complaints to the city-month level and and display the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per city. Here we focus on Identity Theft as the category of complaints. In Columns 1 and 2, the outcome variables are an indicator variable for whether there is any such complaint and the logarithm of one plus the number of such complaints. Standard errors are clustered at the state level and are reported in parentheses.Standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

The Table below shows that scaling the number of FTC fraud reports by city population does not change the results qualitatively. Note that the sample size is reduced by more than 60%. This is because census do not cover all cities, and additionally, city field is self-reported and noisy in the FTC data. We also drop cities with population more than 1 million and less than 1,000 residents. This is because people living in the largest cities, such as New York, may report boroughs (Manhattan or Brooklyn) as their city.

Table G.2: The Impact of Privacy Regulation on FTC Fraud Reports: Scaled by City Population

	Complaints per 1,000 capita			
	(1)	(2)	(3)	(4)
	Internet	Data Security	Positive USD loss	ID Theft
Post-policy × iOS share	−0.0072	−0.0004***	−0.0290**	−0.1332***
	(0.005)	(0.000)	(0.011)	(0.041)
City FE	✓	✓	✓	✓
State × Year-month FE	✓	✓	✓	✓
Observations	354,944	354,944	354,944	354,944
R-square	0.150	0.061	0.340	0.666

NOTE.—This table examines the impact of the policy using the FTC complaint data. Because the FTC data only contain the city and state of the filers, we aggregate the complaints to the city-month level and and display the results from Specification 1 with the interaction of an indicator for the post-policy period and the treatment intensity, the pre-treatment share of iOS users per city. In all columns, the number of fraud reports are scaled by city population. We report the results for internet service and data security categories in Columns 1-2. In Column 3, we focus on all fraud reports with a positive dollar loss. Column 4 reports the results on identity theft cases. Standard errors are clustered at the state level and are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

H Description of Advisen Data

This section describes the set of information that we use at the incident level. The cases in Advisen's cyber dataset involve billions of unauthorized disclosures, thefts, or serious disruptions of customer and employee identities, corporate assets, and systems capabilities. Over our sample period, 2020/01-2022/06, there are 36,425 incidents.

Causes of incidents We use the subcategory risk to identify cases that more likely to be affected by ATT (as lighted in bold). The share of each case type is reported in the parentheses.

- Cyber Extortion (10.60%)
- Data – Unintentional Disclosure (17.19%)

- Data – Physically Lost or Stolen (3.59%)
- **Data – Malicious Breach (36.57%)**
- **Privacy – Unauthorized Data Collection (1.95%)**
- **Privacy – Unauthorized Contact or Disclosure (18.77%)**
- **Identity – Fraudulent Use/Account Access (0.69%)**
- Industrial Controls & Operations (0.07%)
- Network/Website Disruption (5.06%)
- Phishing, Spoofing, Social Engineering (3.77%)
- Skimming, Physical Tampering (0.21%)
- IT – Configuration/Implementation Errors (0.84%)
- IT – Processing Error (0.49%)

Regulations violated When specific laws or regulations are violated by the cyber event, Advisen reports the names of the law and regulations. 19.79% of incidents lead to violations of laws or regulations. Among those incidents, the three most frequently violated regulations are Telephone Consumer Protection Act (TCPA) (70%), Fair Debt Collection Practices Act (FDCPA) (28.3%), and General Data Protection Regulation (GDPR) (21.1%). Note that multiple regulations can be violated in a single event.

Conditional on violating the FDCPA, the companies that experienced the most incidents are Midland Credit Management Inc (71 incidents), Capital One Bank (66 incidents), Portfolio Recovery Associates LLC (53 incidents), and Bank of American National Association (44 incidents). Conditional on violating the GDPR, the companies that experienced the most incidents are Vodafone Espaia SAU (7 incidents), Veale Wabrough Vizards LLP (4 incidents), and Casa Gracio Operation SL (3 incidents). These observations suggest that incidents that resulted in the violations of FDCPA are more likely to result in financial fraud that involves financial companies.