# Consumer Surveillance and Financial Fraud*

Bo Bian†, Michaela Pagel‡, Devesh Raval§, and Huan Tang¶

April 1, 2024

**Abstract**

In today's digital economy, firms near constantly collect, analyze, and profit from consumers' personal information, which might expose consumers to financial fraud. We examine the effects of Apple's App Tracking Transparency (ATT) policy, which significantly curtailed data collection and sharing on the iOS platform. Using zip code level variation in iOS user shares, we show that ATT substantially reduced fraud complaints. The effects are concentrated in complaints that have more relevant narratives and in complaints about companies engaging in intensive consumer surveillance and lacking data safeguards. Our evidence quantifies one of the main harms of lax privacy standards.

**Keywords:** Financial fraud, data security, privacy regulation, data sharing and tracking, App Tracking Transparency
**JEL Codes:** G5, G28, L86

*We thank Simona Abis, Hunt Allcott, Tania Babina, Nathan Blascak, Mark Eichorn, Emma Fletcher, Samuel Kruger, Kai Li, Markus Mobius, Jordan Nickerson, David Rothschild, Dave Schmidt, James Thomas, Jonathan Wallen, Brett Wendling, Paul Witt, Constantine Yannelis, and Anthony Lee Zhang for valuable comments as well as seminar and conference participants at the Chicago Conference on Empirical Finance, Chinese University of Hong Kong, Future of Financial Information Conference, Household Finance Brownbag Series, IMF, Microsoft Research, NBER SI (Household Finance), UBC Winter Finance Conference, University College London, UW-Madison Junior Finance Conference, and University of Innsbruck. Niels Wagner provided outstanding research assistance. The views expressed in this article are those of the author. They do not necessarily represent those of the Federal Trade Commission or any of its Commissioners. We acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. We have nothing else to disclose.

†University of British Columbia, Sauder School of Business. E-mail: bo.bian@sauder.ubc.ca
‡Washington University Olin Business School, NBER, and CEPR. E-mail: mpagel@wustl.edu
§Federal Trade Commission. E-mail: draval@ftc.gov
¶University of Pennsylvania, The Wharton School and CEPR. E-mail: huan.ht.tang@gmail.com

## 1 Introduction

We live in a commercial surveillance economy. Mobile phones are essential to everyday life but enable firms to near constantly surveil consumers' private lives.[1] A prevailing concern among regulators is that the collecting, tracking, sharing, and selling of private data 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. Because consumers are largely unaware of the different forms of commercial surveillance practices, consumer consent and public scrutiny may not alleviate these problems.

In response to these risks, the European Union (EU) strengthened its data protection standards in 2018, and the state of California 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.[2]

Lax data privacy measures, security breaches, and the market for personal data enable financial fraudsters to target and harm consumers. Every year, law enforcement and federal agencies receive millions of fraud complaints. The FTC estimates that more than 10% of U.S. adults fall victim to fraud each year (Anderson, 2019), with more than 5 million consumers reporting having lost almost $9 billion in 2022 (Federal Trade Commission, 2023). 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 substantially limited the tracking and sharing of personal information across mobile applications (apps) and websites can reduce financial fraud. Our findings are useful for quantifying the benefits of new reg-

[1] According to Comscore (2019), smartphones account for 70% of all digital media time in the US; 88% of mobile phone time is spent on apps (eMarketer, 2020).

[2] For the FTC, see press release "FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices". For the CFPB, see press release "Required Rulemaking on Personal Financial Data Rights" .

ulations or other regulatory alternatives concerning the ways in which companies collect, aggregate, protect, use, analyze, and retain the financial information of consumers, as well as transfer, share, sell, or otherwise monetize that data.

We use Apple's App Tracking Transparency (ATT) policy as a source of variation in consumer surveillance practices and investigate its effect on financial fraud. The ATT policy, introduced by Apple on April 26, 2021, requires all apps to obtain explicit user permission before tracking users across apps or websites owned by other companies and sharing user data. By default, users are opted out of tracking, and as of March 2022, only 17% of iOS users opted in for data sharing (Kraft et al., 2023). As a result, ATT greatly limited the volume and scope of personal data collected and shared across companies, reducing the availability of high-quality, real-time data for fraudsters to exploit.

We exploit the fact that ATT only affects iOS users, and not Android users, to examine the effects of the ATT policy on fraud activities in zip codes with different exposure to the policy. To measure exposure to ATT, we use granular foot traffic data from Safegraph to calculate pre-ATT zip-code-level shares of iOS users out of all smartphone users.

In our analysis, we use three different datasets of consumer complaints, which cover the near-universe of fraud complaints in the US.[3] The CFPB's public complaint database and the FTC's Identity Theft database focus on complaints concerning financial institutions and identity theft, respectively. In addition, we use data from the Consumer Sentinel Network, a consortium run by the FTC that collects complaints from many sources, including the FTC, CFPB, Better Business Bureau (BBB), state law enforcement agencies, and private actors like major money transfer firms. The Consumer Sentinel Network data give us a broad perspective of complaints about many different kinds of fraud that may lead consumers to suffer from financial losses.

[3]We refer to consumers' voluntary submission of information about fraud and other scams as "complaints" throughout this paper. Although the FTC and other institutions long described this information as "complaints," the FTC now describes this information as "reports" in order to emphasize the problems that consumers may observe as opposed to whether the consumers were directly affected or lost money as a result.

Our results indicate that limiting the tracking and sharing of personal information substantially reduces consumer complaints in all three datasets. We find that a 10 percentage point ($\sim$ 1 standard deviation) increase in the share of iOS users in a zip code leads to a 13.3% reduction in the number of CFPB complaints post-ATT. Given that 83% of iOS users opted out of tracking after ATT, a 10 percentage point increase in consumers opting out of tracking translates to a 16% reduction in CFPB complaints. We also find reductions in the number of Consumer Sentinel and Identity Theft complaints. Our identification relies on parallel pre-trends in complaints between areas with high and low iOS shares, which we confirm using various checks.

Not all consumer complaints relate to financial fraud originating from lax data security standards. We thus go on to classify the complaints into more or less relevant cases using two methods. First, we conduct simple keyword searches in the issue, sub-issue, and consumer narrative fields for relevant words such as "scam," "imposter," "unauthorized," or "data breach." 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 first show that these two measures of complaint relevance are highly correlated. We then document large estimates for the more relevant complaint categories (e.g., credit reporting and debt collection for CFPB complaints) and close-to-zero effects for the less relevant complaints across all three complaint databases.

We link our evidence on consumer complaints to actual fraud victimization in two ways. First, many of the data contributors of the Consumer Sentinel Network record whether a consumer suffered a financial loss. We find that the number of complaints reporting financial losses also declines after ATT. Second, Raval (2020b) develops a set of weights that translate consumer complaints into variation in fraud victimization by accounting for differences in the propensity to complain across zip codes with the same degree of victimization.[4] After

[4]Using data on consumers affected by nine consumer protection law enforcement actions, Raval (2020b) find that residents of heavily Black and Latino areas who lost money in the cases were about half as likely to complain as residents of heavily White areas. The paper develops weights that account for these differences.

applying these weights to better align our complaint data with fraud victimization, the estimated effects of ATT increase substantially.

To further strengthen the link between consumer surveillance and fraud, we examine how the effect of ATT varies depending on firms' surveillance practices and data security measures. 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.[5] Both platforms require firms to disclose 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.

All CFPB complaints report the specific financial companies in question. We identify 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, via mobile device identifiers. 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 an 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 after ATT. 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 after ATT. These effects account for 25% to 50% of the average number of download-scaled complaints. Moreover, we show that the effects on firm-level complaints are particularly pronounced in categories that are more

[5]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.

likely related to data issues (as identified using our two approaches described above), while there are no discernible effects in less-relevant categories.

Finally, we supplement the firm-level analysis with comprehensive data on cyber incidents from Advisen, a data provider for insurance companies to assess cyber security risks. We find that companies with an app are 33% less likely to experience cyber incidents compared to firms without an app after ATT. This effect is only present among cyber events that were caused by malicious data breaches or identity theft (as opposed to other causes such as network disruption) and is stronger when the incidents 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 categories most susceptible to fraud, as shown above.

These findings lend support to the notion that ATT has indeed mitigated the adverse effects of firms' data sharing, vulnerabilities, and breach risks on consumers. To the best of our knowledge, our study is the first to provide empirical evidence on the link between lax data practices and financial fraud. Regulators are growing more worried about the potential harm to consumers from the sales and sharing of mobile data; our research highlights the risks of such consumer tracking.

**Literature review**

Our paper first contributes to the burgeoning literature on consumer data privacy concerns. Armantier et al. (2021) report that identity theft is the most cited reason for consumers' privacy concerns about data sharing among US households.[6] Complementing this survey evidence, our findings provide justification for consumers' specific privacy concerns and quantify one of the costs of firms' lax data security standards – identity theft and financial fraud. In the same vein, our paper also relates to prior work on the consequence of big data analytics and excess data collection by financial institutions, such as racial discrimination and privacy

[6] Around 90% of respondents in each demographic group report identity theft as an important concern, followed by abuse of data, personal safety, and reputation.

intrusion (Fuster et al., 2022; Tang, 2019).

Moreover, our paper adds to recent work that examines data privacy regulations. Most of this work has focused on the European General Data Protection Regulation (GDPR), linking GDPR to changes in 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, monetize, store, and use consumer data (Aridor et al., 2020; Bessen et al., 2020; Demirer et al., 2024; Peukert et al., 2021).[7] Babina et al. (2022) show that open banking policies spur investments into FinTech startups. Doerr et al. (2023) examines the impact of the California Consumer Privacy Act on fintech lending in the mortgage markets.

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. Cheyre et al. (2023) find that the adoption of ATT led to a temporary decline in app entry and a reduction in app updates and that apps adapted to ATT by changing their model of monetization.

Compared to the existing literature, we focus on the effects of data privacy regulations on curbing financial fraud. The ATT policy, compared to other data privacy regulations, offers distinct advantages in answering this question. First, it establishes a standardized and uniform consent framework, unlike, for instance, cookie consent banners. Second, companies are forbidden from displaying the consent prompt to users who have already declined the request. These unique policy characteristics suggest a significant decrease in firms' capacity

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

for data collection and sharing, which may reduce opportunities for financial fraud.[8]

In addition to the literature we have discussed, there are three related areas of research that focus on specific subsets of financial fraud. The first literature examines cases of financial fraud involving elderly victims (e.g., Alves and Wilson, 2008; DeLiema, 2018; DeLiema et al., 2012, 2020; James et al., 2014; Lichtenberg et al., 2013). Notably, Carlin et al. (2020) analyze the impact of new regulations aimed at combating elder financial abuse by deputizing financial professionals across different US states. The second literature focuses on the misconduct of financial advisers or investment managers. Griffin and Kruger (2024) provide a comprehensive survey of this line of work. For example, Deliema et al. (2020) examine investment fraud. The third examines consumer complaints and victim data from consumer protection cases to inform our knowledge about fraud victimization. This literature has found that fraud victimization varies across demographic groups; for example, consumers in Black communities are more likely to suffer from fraud, especially related to specific financial issues such as payday or student loans, and are less likely to complain when victimized (Raval, 2021, 2020b; Sweeting et al., 2020). Consumer complaints can also help assess which countries disproportionately account for cross-border fraud (Raval and Grosz, 2022) and examine specific frauds such as the Social Security imposter scam (DeLiema and Witt, 2021). We add to the fraud literature by identifying digital surveillance as a novel factor driving financial fraud.

## 2 Institutional Background

In this section, we first provide background information on how the sharing of personal data can lead to financial fraud through the lens of enforcement efforts. We then describe an industry-led privacy initiative implemented by Apple and its relevance for data-driven fraud.

[8] Kraft et al. (2023) provide a comprehensive discussion of the policy's unique features and its impact on firms' ability to collect and share data.

### 2.1 Enforcement Activities against Data-driven Fraud

A major concern with the collection and sharing of personal data is that such data could be used to commit fraud against consumers. Indeed, enforcement agencies in the US have brought several actions against fraudsters exploiting such data and the brokers who sold them such data. We give two illustrative examples. First, the FTC won in court against Ideal Financial for purchasing bank account and social security numbers from consumers applying for payday loans online and then charging them without their consent; the FTC also settled charges against the data brokers involved for selling such data to Ideal Financial and others.[9] Second, the Department of Justice (DOJ) settled criminal charges against Epsilon, a major data broker, for selling lists of vulnerable consumers to fraudsters that used the lists to deceptively market sweepstakes and astrology and psychic services; Epsilon was required to pay $150 million as part of the settlement.[10]

More recently, regulators have become increasingly concerned with how sales of mobile data could be used to harm consumers. For example, the FTC just settled allegations against multiple mobile data brokers (X-Mode and Outlogic), and is currently in court against another (Kochava), for selling consumers' location data.[11] Such data can be used to link consumers to their offline activities, such as visits to reproductive health clinics, domestic violence centers, or addiction recovery centers. In addition, as the FTC complaint in the Kochava case describes, such location data can link a consumers' mobile application ID, or MAID, to their home address and thus to more traditional information sources such as other websites, offline stores, advertising networks, and data brokers. Kochava in fact advertised "Household Mapping" as one use case for its service.

[9] See FTC press releases: "FTC Action Leads Court Orders Against Scheme Charged Millions of Dollars to Consumers' Bank and Credit Card Accounts" and "Data Broker Defendants Settle FTC Charges They Sold Sensitive Personal Information to Scammers".

[10] See FTC press release: "Marketing Company Agrees to Pay $150 Million for Facilitating Elder Fraud Schemes".

[11] See FTC press releases: "FTC Order Prohibits Data Broker X-Mode Social and Outlogic from Selling Sensitive Location Data" and "FTC Sues Kochava for Selling Data that Tracks People".

### 2.2 Apple's App Tracking Transparency Policy

In this paper, we study an industry-led privacy initiative. We use the implementation of the ATT policy as an exogenous shock to the gathering, sharing, selling, or leaking of detailed data of iOS users, and argue that ATT reduces the availability of high-quality data for fraudsters.

In June 2020, Apple announced its plans to move to a new version of its iOS operating system, iOS 14. Starting from the release of iOS 14.5 on April 26, 2021, Apple introduced a new privacy feature that required all apps to ask for explicit user permission before obtaining users' mobile identifiers that allow them to track users across apps or websites. This feature, dubbed "App Tracking Transparency (ATT)", grants users both greater and easier control over their data. An example of the prompt notification is provided in Panel a of Appendix Figure 1. By default, a user is opted-out of tracking. That is, Apple would no longer provide apps and websites with an Apple-assigned user identifier. This Apple-assigned device-level identifier is the most universally applied in the mobile world for data aggregation and user profiling, compared to other types of user-supplied ID (e.g., email address or name) that could differ across apps or government-issued ID (e.g., Social Security Number, or SSN) that is less commonly collected. Importantly, the opt-in design and the uniform consent prompt (in contrast to firm-specific cookie banners) apply to all firms that serve iOS users. In addition, companies are forbidden from displaying the consent prompt to users who have already declined the request.

Industry reports suggest that the vast majority of users did not opt-in for tracking upon seeing the notification (Kraft et al., 2023).[12] Bian et al. (2021) document a sharp and negative stock market reactions for firms owning an active iOS app around the implementation of ATT, corroborating its substantial impact on the data economy (see Appendix Figure A.1).

[12]For example, Flurry, a mobile app analytics platform, shows that only 18% of iOS users allowed tracking among those who were asked for permissions. For details, see source: https://www.flurry.com/blog/att-opt-in-rate-monthly-updates/.

**Why is ATT relevant for financial fraud?** ATT's unique policy features greatly limit the amount and scope of personal data collected and shared across firms. When data sharing is restricted, fewer entities possess copies of personal data and so there is less data available for fraudsters to exploit. Moreover, with the mobile identifier, fraudsters gain the ability to link data on the same users from different periods or different sources and so build comprehensive profiles of individuals for fraudulent activities. For example, the aggregation of a minimum set of personal information, such as names, birth dates, and up-to-date addresses, phone numbers, and bank account details makes it much easier to successfully impersonate individuals or commit identity theft.

While ATT reduces the accessibility of high-quality, real-time data to fraudsters, one may argue that fraudsters can still exploit the historical data accumulated before ATT. However, real-time data often contains more accurate and up-to-date information about potential targets, such as their current address and recently used financial products, compared to historical or stale data. In addition, real-time data allows fraudsters to exploit vulnerabilities or weaknesses before they are identified and addressed. For example, fraudsters may use real-time data to quickly make unauthorized purchases before the cardholder or fraud detection systems can detect fraudulent activities.

## 3 Data

In this section, we describe our main variable on exposure to ATT – the share of iPhone users in a zipcode – as well as our measures of financial fraud from several consumer complaint databases.

### 3.1 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 zipcode 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 (>0.97) between the number of sampled devices and census population at the county level, for both urban and rural areas, and minor sampling biases among various demographic categories such as age, gender, and moderate income, with less than 5% under- and overrepresentation.[13]

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-digit NAICS code 52) based on the device operating system (iOS or Android) and zip codes. 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 little effect on the measurement.

In our analysis, we primarily use 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

[13] Li et al. (2023) also reports a lower correlation at the census-track level. Therefore, it is likely that the zip-code-level correlation is lower too. In our robustness checks, we use the subsample of zip codes with large foot traffic and population to alleviate concerns about the measurement error in the iOS share due to the sampling of devices.

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 affect the ability to precisely capture foot traffic information.

### 3.2 Financial Fraud Data: Consumer Complaints

When individuals believe they have been victims of financial fraud, they can submit a complaint to various state and federal government agencies as well as to private actors like the Better Business Bureau (BBB). We construct our outcome variables using consumer complaints reported from three sources: the publicly available data on CFPB complaint filings, the Consumer Sentinel database, which combines complaints from several sources including the CFPB, and the FTC's Identity Theft database.[14] We provide a description of each dataset below. Our empirical analyses mainly rely on the public CFPB data because it concerns financial products, but whenever possible, we use the Consumer Sentinel and Identity Theft database to show that the broad effects of ATT on fraud.

**CFPB complaints** The CFPB provides a portal for consumers to submit complaints on its website, as well as a phone number that consumers can call to complain. In addition, the FTC's Report Fraud website forwards consumers complaining about debt collection, credit cards, credit reporting, or banking to the CFPB's complaint portal. 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

[14] Consumer Sentinel has about 34% more CFPB complaints per month than used in our previous analyses of public CPFB complaints, both because some small zipcodes are suppressed in the public data and because some CFPB complaints are not reported publicly but are forwarded to Consumer Sentinel. The CFPB states that "We do not publish complaints referred to other regulators, such as complaints about depository institutions with less than $10 billion in assets." and that 5 digit zipcodes with a population of less than 20,000 are suppressed. See https://www.consumerfinance.gov/data-research/consumer-complaints/ for more details.

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", and "Fraud or scam". 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.

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, credit and debt card fraud, and debt collection scams.

**Consumer Sentinel complaints** In addition to the public CFPB database, we use data from the non-public Consumer Sentinel Network. The Consumer Sentinel Network is a consortium run by the FTC that collects complaints from federal government agencies such as

the FTC and CFPB; private sector organizations such as the BBB; state and local government agencies such as state attorneys general and police departments; and private companies such as Western Union and MoneyGram.[15] The complaints in the CFPB database thus represent a subset of the reports in Consumer Sentinel.

Like the public CFPB database, Consumer Sentinel complaints include information on the incident that the consumer is complaining about, including a narrative text field and the company involved. In addition, the non-public data contains identifying information on the complaining consumer, such as their name and address, and many data contributors provide information on the dollar losses of the consumer.

The Consumer Sentinel database provides a nice contrast to the CFPB complaints. The CFPB complaints are focused on financial fraud and have received a lot of attention from academic researchers because they are public. On the other hand, the Consumer Sentinel database covers a much broader range of products and services – including imposter scams, online shopping, internet services, data security, and cyber fraud – which are directly affected by ATT but may not be forwarded to the CFPB. It also receives complaints from a very broad range of sources. This diversity allows us to gain a more complete view of the effects of ATT on fraud incidents, but also means that more complaints may be about issues unlikely to be affected by ATT.

**Identity Theft complaints** The FTC also maintains a separate database of Identity Theft complaints that consumers file using different channels from the complaints in Consumer Sentinel (for example, by visiting `identitytheft.gov` instead of `reportfraud.ftc.gov`). The Identity Theft database contains complaints with broadly similar information to those in Consumer Sentinel. Like the Consumer Sentinel complaints, the Identity Theft complaints are non-public. Like the CFPB complaints, they are more focused on a specific

[15] Raval (2020a) finds that the three largest contributors are the FTC, BBB, and CFPB and provides more details on how consumer demographics vary across organizations and consumer protection issues. See https://www.ftc.gov/enforcement/consumer-sentinel-network/reports for the Consumer Sentinel Data Book, which contains further detail on the Consumer Sentinel and statistics on the complaints included in it.

issue (here, identity theft rather than financial problems) compared to Consumer Sentinel.

### 3.3 Summary Statistics

Table 1 presents summary statistics for the variables used in our regression analysis for all three datasets. The main regression sample consists of a balanced panel at the zipcode level, spanning from January 2019 to June 2022.[16] Zeros were filled in for zip codes without any reported complaints.

Panel a of Table 1 presents summary statistics for all three complaint databases. For the public CFPB complaints, approximately 31% of zip codes have at least one complaint in any given month. The mean number of complaints per 1,000 residents 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.[17]

Both the Consumer Sentinel and Identity Theft datasets are significantly larger than the public CFPB complaint database, with an average of 0.72 complaints per 1,000 people in a zipcode-month reported to Consumer Sentinel and 0.19 complaints per 1,000 people to the Identity Theft database. Both of these databases also include more zip codes (since the public CFPB database suppresses small zip codes), so the Consumer Sentinel panel has about 7.2 times more complaints per month in total than the public CFPB panel, and the Identity Theft panel about 2.5 times more complaints per month.

Figure 2a 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. Figure 2b illustrates the total number of complaints per month over the entire sample period, showing temporal variations in complaint volume.

The bottom part of Panel a also displays summary statistics for the iOS device share.

[16] For the Identity Theft data, we do not have complaints for a few days in January 2019, so the panel begins in February 2019.

[17] The annual number of public CFPB complaints in 2019-2021 are 277,325, 444,347, and 496,018, respectively.

Using foot traffic data, we find that the average iOS share across US zip codes is 46%. This estimate is similar to evidence from Statista that the iOS share aggregated over the entire US fluctuates around 50% during the period of 2019-2022.[18] However, this share varies substantially across zip codes, with a share of 39% for the 25th percentile zip code and 52% for the 75th percentile.

## 4 The Impact of the Privacy Regulation on Financial Fraud

### 4.1 Regression Specification

Our main regression specification is:

$$Complaints_{z,t} = \alpha_z + \alpha_{county(z),t} + \beta iOS\ Share_z \times Post_t + \varepsilon_{z,t} \tag{1}$$

The outcome variable, $Complaints_{z,t}$, is constructed by aggregating complaints to each zip-month. Our main outcome measure is the winsorized number of complaints per 1,000 residents, although we examine alternative measures and Poisson regression models in robustness tests.[19]

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, $Post_t$, takes a value of one starting from May 2021, the first month after the ATT policy took effect.

To account for time-invariant characteristics that contribute to fraud, we include zip code fixed effects. Additionally, we incorporate county-by-year-month fixed effects, denoted as $\alpha_{county(z),t}$, to control for time-varying confounders at the county level. These confounders may include region-specific data regulations, local fraud news, or local economic developments. To be conservative, we cluster the standard errors by state. In our robustness

[18] 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.

[19] We winsorize the top 0.5% of complaints and top 0.5% of complaints per 1,000 residents.

checks, we explore alternative sets of fixed effects and clustering choices.

### 4.2 Baseline Results

Table 2 presents the regression results based on each of our complaint databases. In all three datasets, we observe a significant and negative coefficient on the interaction term $iOS\ Share_z \times Post_t$. That is, within a given county and month, zip codes with a higher proportion of iOS users experience a decline in consumer complaints following the implementation of the ATT policy, compared to zip codes with a lower iOS user share.

The effects of ATT are economically meaningful. For the CFPB complaint database, a zip code with a 10% ($\sim$ 1 standard deviation) higher iOS user share sees a decrease of 0.0093 (calculated as $10\% \times 0.093$) reduction in the number of complaints per 1,000 residents, the latter representing 13.3% of the sample mean. Given the observed tracking opt-out rate of 83%, our results suggest that if 10% of mobile app users were to disallow data tracking, complaints to the CFPB could be reduced by approximately 16.2% (calculated as 13.3% divided by 0.83).

We continue to find reductions in complaints for zip codes with more iOS users after the adoption of ATT using the other complaint databases. For the Consumer Sentinel and Identity Theft dataset, a zip code with a 10% higher iOS user share sees a 0.01 and 0.0073 reduction in the number of complaints per 1,000 residents, respectively. While the absolute effects are comparable to those obtained from the CFPB complaints, the effects relative to their respective sample averages are only 1.4% and 3.8%. This is not surprising given that both databases cover a much broader range of complaints that concern non-finance products and services and cases not triggered by data breaches.

To validate that the iOS share is a good proxy for the exposure to ATT, in Table D.1, we group the iOS share at the zip code level into deciles and report all the interaction terms for the CFPB complaints. Using zip codes in the bottom decile of the iOS share distribution as the base group, we find that the magnitude of the treatment effect is monotonically increasing in the treatment intensity. This is in line with our expectations. Focusing on Column 1,

where the outcome variable is scaled by the local population, in contrast to the lack of results for the bottom five deciles, zip codes in the top five deciles experience treatment effects significant at the 5% level. The magnitude of the effects is -0.016 for the 6th decile and gradually increases to -0.043 for the top decile. Similar patterns hold for other outcome variables, as reported in Columns 3-4.

### 4.3 Dynamics

While we compare zip codes 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 consumer complaints. For example, ownership of Apple products predicts higher income and better education (Bertrand and Kamenica, 2018), which can lead to changes in consumer complaints among iOS users over time if higher-income or better-educated consumers are hit with different shocks than other consumers. To rule out alternative explanations, we examine pre-trends in consumer complaints and plot dynamic DiD coefficients.

In Figure 3a, we present the results. To reduce estimation error, we group 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.[20] We define quarter $-1$ as the quarter immediately preceding the implementation (2021Q1), which serves as the benchmark quarter. Quarters prior to $-4$ are combined into a single period.

The dynamic DiD coefficients confirm that the reduction in complaints manifests after the ATT policy's implementation. Examining the population-scaled number of complaints, we observe that 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.[21] Because the CFPB database contains less complaints relative to the

[20] 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.

[21] We also estimate the monthly dynamic treatment effects and report the results in Appendix Figure C.1. While the estimates are noisier at the monthly frequency, almost all the point estimates for the post-event months are negative and highly significant.

other two data sources, we also plot the dynamic DiD coefficients for the extensive margin of fraud complaints, measured by whether there is any complaint in a given zip code month in Figure 3b and observe a similar pattern.

The dynamic figures show that the effects of ATT manifest quickly, within two months after ATT's implementation, and amplify over time. These findings are consistent with fraudsters quickly acting upon obtaining leaked or hacked data to maximize returns and evade detection and consumers detecting and reporting fraud quickly as well.

For the other two complaint databases, we report the dynamic DiD coefficients in Appendix Figure E.1. For Consumer Sentinel complaints, we find insignificant coefficients in the year before the implementation of ATT and negative, significant coefficients after implementation. However, the coefficient for the period more than a year before implementation is negative and significant. For Identity Theft complaints, there is some evidence of increasing (albeit insignificant) pre-trends in the year before implementation of ATT, followed by negative and significant coefficients after implementation. If the increasing pre-trends would continue after implementation, we would understate the full effects of the ATT policy on complaints.

## 5 Threats to Identification

In this section, we examine several potential threats to our identification strategy. First, we conduct placebo tests that show that our results are not driven by factors unrelated to ATT. Second, we discuss an alternative mechanism that could explain our results – changes in fraud due to the COVID-19 pandemic. Third, we examine potential spillovers from the ATT policy to Android users. Finally, we examine several alternative specifications, including alternative measures of fraud.

### 5.1 Placebo Tests

We conduct two placebo tests – in both the time series and cross-section – to demonstrate that the reduced complaints 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, 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. Unless other events perfectly coincide with ATT and generate a similar effect, we can confidently attribute the observed decline in fraud complaints to the implementation of ATT. For the complaints per 1,000 residents, the placebo DiD coefficients were slightly negative in April-May 2020, which coincided with the first wave of economic impact payments issuance ("EIP") in the US. These payments may be associated with an uptick in fraud that 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. We address the effects of EIP separately in the next subsection.

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. Based on Figure 5, our main estimate is unlikely to be driven by peculiarities or unaccounted correlation structure.

### 5.2 COVID-19 Concurrent Shocks

Because our sample period covers the COVID-19 pandemic, one concern is that many fraud complaints related to the COVID-19 pandemic also increased for Android users relative to

iOS users at the same time as the adoption of ATT. The pandemic led to many changes that could foster fraud, such as increases in online shopping and a surge in demand for COVID-19 cures and face masks. We focus here, however, on the largest change – the large government stimulus and transfer payments, and in particular on the Treasury's Economic Impact Payments (EIP).

To examine whether these payments affect our results, we directly control for the interaction between variables related to the EIP and the post-ATT indicator. These variables, constructed using zipcode-level tax return data from the IRS, include the total amount of EIP received, the average household income, and the average number of children and other dependents in a zip code. While the amount of EIP received by a zip code captures actual payments, the other two variables, household income and number of children, determine the eligibility of EIPs. The motivation for the latter two variables is that households that are eligible for EIPs are the potential targets for EIP-related fraud. If our baseline results are driven by rising EIP-related fraud in low-iOS-share zip codes, after controlling for these factors, we should see a substantial reduction in the point estimates.

The results are reported in Table 4. Our coefficients of interest remain similar for the CFPB complaints and increase in magnitude for Consumer Sentinel complaints. The coefficient for the CFPB is -0.072 (compared to -0.093 at baseline), and the coefficient for Consumer Sentinel is -0.111 (compared to -0.103 at baseline). However, the estimates for the Identity Theft database fall by about half from -0.073 at baseline to -0.039 after EIP controls.

We also leverage the product categories and identify categories that may be directly related to stimulus payments; we discuss these categories in more detail in the next section. First, for CFPB complaints, 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 (Consumer Financial Protection Bureau, 2021).[22] Second,

[22]See section 3 of Consumer Financial Protection Bureau (2021) for a detailed discussion.

for Consumer Sentinel complaints, we isolate the product code of "Unemployment Insurance Fraud," which should pick up effects from more generous unemployment insurance, as well as more unemployment in general, during the pandemic. Finally, for Identity Theft, we examine the product code "Government Benefits or Documents Fraud", which should encompass both fraud related to EIP payments and unemployment insurance fraud.

To assess the differential impact of EIP on complaints by iOS and Android users, we re-run the regression separately for these products for each dataset. The results, reported in Table G.3, show economically small coefficients on the interaction term between iOS shares and the post-ATT indicator for CFPB complaints. The magnitude of the effect on these categories is about 1/50 of the baseline effects (-0.002 compared to -0.093). This suggests that the differential responses to ATT in high versus low iOS share zip codes for CFPB complaints are largely not driven by EIP-related fraud incidents.

For Consumer Sentinel complaints, in panel b, complaints related to unemployment insurance fraud are *increasing* after ATT in high iOS share zip codes, which would lead us to understate the benefits of ATT in reducing complaints. However, these estimates are also quantitatively small at about 1/10 of the baseline effects (a coefficient of 0.009, compared to the baseline estimate of -0.103).

Finally, for Identity Theft complaints, we find substantial reductions in complaints about Government Benefits and Documents fraud in high iOS share areas after the adoption of ATT. For example, the coefficient on complaints per 1,000 people for this topic is -0.047, compared to the baseline effect of -0.073 for all Identity Theft complaints, or about 65% of the overall effect. This suggests that fraud linked to EIP has the potential to explain some, but not all, of the changes in Identity Theft complaints associated with ATT; however, it does not appear to explain the decline in CFPB and Consumer Sentinel complaints.

### 5.3 Spillovers

Another concern is that fraudsters could shift their focus from iOS users to Android users as privacy regulations tighten on the iOS platform. Such spillovers still mean that privacy

regulations increased the costs of fraud for iOS users, and so should reduce overall fraud. However, if this spillover effect is strong, we may overestimate the effect of ATT on aggregate fraudulent activity by using Specification 1. While fully accounting for spillover effects is difficult, we provide evidence against within location and within firm spillovers.

First, if the ATT effects were driven by spillovers within locations, we should find the largest effects in areas with more Android users where there are more consumers to redirect fraudulent activity to. However, we find in Table D.1 that the effect of ATT is strongest in regions predominantly populated by iOS users. Second, in Section 7, we examine firm-level cyber incidents using variation in firms' heterogeneous presence in the mobile space, surveillance practices, and data security measures for identification. We find reductions in complaints for firms with a mobile presence, worse surveillance practices, or worse data security practices, which is evidence against within firm spillovers (for example, from consumers of the firm using iOS to those using Android).

In addition, the estimated effect from Specification 1 captures the average treatment effect for the treated iOS users rather than the average treatment effect for the entire population. Android users might be more susceptible to data collection and security breaches due to a lower privacy awareness and more vulnerability in the Android ecosystem (Garg and Baliyan, 2021), and potentially benefit more from privacy regulations like ATT. In this case, the average treatment effect could be even larger than our estimation.

### 5.4 Additional Robustness Checks

Finally, we examine a battery of robustness checks using the CFPB complaint data.

First, we consider multiple additional complaint measures, including an indicator for any complaint in a zip-month, a winsorized count of complaints, and the logarithm of one plus the number of complaints. We report these results in Table 3. The effects of ATT are economically meaningful for all measures for the CFPB complaints. 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, a zip code with a 10% higher iOS user share sees a decrease of 0.17 complaints. 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.[23]

In addition, we 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. Figure 6 presents results from these alternative regression specifications and samples. We display the baseline estimate at the top for ease of comparison. Following the baseline estimate is the sample excluding 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.

We also 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. Additionally, we only include zip codes with more than 10,000 people. These alternative sampling criteria help improve the precision of regression estimates by mitigating potential measurement error in areas with low foot traffic. For example, because the iOS device share is based on visits to all grocery and retail stores in a zip code, one may argue that it does not reflect the iOS share of people residing in a zip code that does not host grocery stores. Given that zip codes with 10,000 population are likely to have multiple grocery and retail stores, we alleviate this concern. The point estimate is still negative and statistically significant, with a somewhat larger magnitude.

We next consider alternative regression specifications. First, we replace county-by-month

[23]For the Consumer Sentinel and Identity Theft complaints, we find significant effects for all measures except whether a zip code had any complaint in a given month. The difference may be that the baseline complaint rate is much higher for these other datasets, so the zip codes on the margin of having a single complaint or not are different.

fixed effects with state-by-month fixed effects. The results remain largely similar. We also cluster standard errors at the zip code level. Again, the results are consistent with the baseline. Double clustering at zip code and year-month level or state and year-month level has little impact on the significance of the point estimate.

Last, we apply the Poisson model to the total number of complaints at the zip-code level and examine the same set of specifications. These results are presented in Appendix D. Despite all these variations, our main findings remain consistent and robust, suggesting that the observed effect of ATT on reducing financial fraud complaints is not specific to a particular choice of the empirical model.

## 6 Cause of Complaints and Fraud Victimization

While we have shown above that aggregate complaints fall in zip codes with more iOS users after the adoption of ATT, complaints are only a signal of our ultimate object of interest – fraud victimization related to data privacy issues. We take several approaches to show that our results reflect reduced fraud victimization from data-privacy-related issues. First, we use the narrative fields to focus on relevant complaints. Second, we examine Consumer Sentinel complaints reporting financial losses from fraud. Third, we adjust our complaint measures using weights that account for differences in the propensity to complain for fraud victims in different demographic communities and examine differences across demographic groups potentially more vulnerable to fraud.

### 6.1 Identifying Relevant Complaints

Not all consumer complaints are directly linked to the collection and misuse of personal information by thieves or hackers. For example, the "issues" or "sub-issues" field in the CFPB database does not explicitly distinguish between more or less relevant categories specifically related to fraud arising from lax data privacy regulations. These issues are compounded in

the Consumer Sentinel data, as it amalgamates complaints from many different sources.[24]

To determine the relevance of complaints for data privacy issues, we employ two approaches using the consumer narrative field. We have narratives for 40% of the complaints in the public CFPB dataset, as well as for all of the complaints in the Consumer Sentinel and Identity Theft databases.

First, we conduct keyword searches based on the sub-product, issue, sub-issue, and narrative complaint fields for the CFPB database and based on the narrative field for the other databases. We compile a list of keywords related to fraud and data privacy, such as "incorrect", "fraud", "theft", "identity", and "data breach".[25] 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 the intersection of data privacy concerns and fraud.

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 privacy 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 G. 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. Given the computational burden of this technique, we only apply it to the public CFPB complaint database.

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 in the CFPB data to classify categories with higher average scores as more relevant. Appendix Table G.1 presents the mean and standard deviation of complaint-level scores by product category.

[24] A consumer with the same underlying issue complaining to a specific data contributor like the CFPB could potentially classify the same complaint into different categories, and each data contributor also has its own way of classifying complaints into different categories, which then have to be translated into the Consumer Sentinel categorization.

[25] The full list of keywords used is "incorrect", "improper", "false", "wrong", "missing", "fraud", "scam", "theft", "embezzlement", "imposter", "unauthorized", "unsolicited", "identity", "sharing", "advertising", "marketing", "security", "data breach", "not owed".

Two patterns are worth noting. First, both the keyword search method and the machine learning approach generate meaningful variation in the average scores at the product level, allowing us to distinguish between more and less relevant cases. For example, the highest and lowest product-level 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 loans" and "Mortgages" receive the lowest scores, suggesting lower relevance for our purposes.

### 6.2 Heterogeneity by Complaint Relevance

We first estimate our main regression specification, separately for the two CFPB categories with the highest relevance – Credit Reporting and Credit Repair and Third Party Debt Collection – and the category with the least relevance – Mortgages – and report the results in Panel a of Table 5.

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 0.0081 (12.5%) decrease in the number of complaints related to credit reporting per 1,000 residents (Column 1). For complaints related to debt collection, there is a statistically significant 1.2% decline in the number of complaints (Column 2).

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 student loans or mortgages (Columns 3 and 4).

Applying the same method to the Consumer Sentinel complaint database, we find sim-

ilar results. We classify product codes with more than 50% relevant word shares as high relevance and product codes with less than 25% shares as low relevance. Issues related to Debt Collection, Credit Bureaus, Tech Support Scams, Government Imposters, and Unemployment Insurance Fraud (among others) are rated as highly relevant, whereas issues about Sweepstakes, New Auto Sales, Funeral Services, Unsolicited Text Messages, and Diet Plans are rated as less relevant. As reported in Table G.2, for complaints in highly relevant product categories, following the implementation of ATT, a 10% increase in the share of iOS users in a zip code is associated with a decline of 0.01 complaints per 1,000 residents or a 5% percentage decline (Column 1). In less relevant categories, we continue to find an insignificant effect on complaints (Column 2).[26]

We additionally leverage the full coverage of narratives in the Consumer Sentinel and Identity Theft complaints databases to directly split complaints into those including one of the keywords described above and those not including one of those keywords. We estimate the main regression specification separately for these two sets of complaints and report these estimates in Panel b of Table 5.

Once again, we find negative and statistically significant results for complaints that are more related to data privacy issues. Following the implementation of ATT, a 10% increase in the share of iOS users in a zip code is associated with a decline of 0.0083 complaints per 1,000 residents or a 3.2% percentage decline for the Consumer Sentinel database and a decline of 0.007 complaints per 1,000 residents or a 9.3% percentage decline for the Identity Theft database. Again, we find insignificant effects for complaints from sources that do not include one of the relevant keywords.

The lack of effects of ATT on complaints that are not related to the misuse of personal information serves as an additional placebo test, suggesting that our results are unlikely driven by concurrent events or differential time trends in the complaint activities among iOS and Android users.

[26]For the Identity Theft database, the categories did not vary substantially in the relevant word share, as the complaints are all of similar type. We therefore did not conduct a sub-group analysis.

### 6.3 Fraud Losses

While consumer complaints provide a measure of fraud victimization, not all complaining consumers suffer from financial losses, which are reported in the Consumer Sentinel data. This information comes with some caveats. First, not all data contributors record losses; for example, no public CFPB complaints report losses. Second, some consumers may mention having lost money in the narrative fields but not separately record the dollar loss in the loss field.

We examine the effect of ATT on complaints in the Consumer Sentinel data reporting a positive dollar loss in Table 6 and find similar effects as for the full sample. A zip code with a 10% higher iOS user share sees a 0.002 reduction in the number of complaints reporting a financial loss per 1,000 capita, which gives us confidence that consumer losses from fraud also fall after the introduction of ATT. After categorizing dollar losses into distinct bins – ranging from $1 to $99, $100 to $999, and $1,000 and above – we observe negative effects across all loss ranges, with the most pronounced statistical significance for complaints reporting losses between $1 and $99.

### 6.4 Demographics and the Propensity to Complain

In addition, the propensity to complain after victimization can vary across different communities, so declines in consumer complaints do not immediately translate to declines in fraud victimization. Raval (2020b) examines how several zipcode-level demographic variables affect the likelihood of complaining by comparing complaints and victims for the same consumer protection case across several cases and found much lower complaint rates, conditional on victimization, in heavily Black and Hispanic areas. Using these estimates, Raval (2020b) develops zipcode-level weights designed to be the inverse of the predicted complaint-to-victim ratio based on those demographics in order to "correct" complaint data for differences in the likelihood of complaining across demographic groups. The median zip code is set to 1, with the majority of Black zip codes having an average weight of 2, as residents of those zip codes

were about half as likely to complain as the median zip code.

In Table 7, we multiply our complaint counts by these weights to examine changes in fraud victimization. This exercise relies on the assumption that the differences between the propensity to complain across locations found in Raval (2020b) extrapolate to all three of the datasets examined in this paper and that the adoption of ATT did not change the propensity to complain.[27] We find larger decreases in fraud victimization than in aggregate complaints; for example, fraud victimization using the weighted CFPB, Consumer Sentinel, and Identity Theft complaints fall by 0.0139, 0.0356, 0.0109 per 1,000 residents in zip codes with a 10% larger share of iOS users after ATT, compared to 0.0093, 0.0103, 0.0073 examining the unweighted complaints.

We then directly examine whether the adoption of ATT affected demographic groups differentially in two ways. First, we construct demographic variables using zipcode-level data from the 2020 census release. We interact these demographic variables (constructed as an indicator for above or below the US median zip code) with the iOS share and ATT adoption. Table F.1 to Table F.3 examine these heterogeneous effects for each of the demographic variables separately across all three main datasets.

The results reveal that the effects of ATT are stronger among communities with a higher share of black or Asian people, as well as a higher share of teenagers or women. However, these differences are the strongest in the CFPB data and are insignificant statistically in the Identity Theft data. Thus, they provide only suggestive evidence that minorities, females, and relatively young or old populations might be more susceptible to fraud triggered by personal data sharing.

Second, for the Consumer Sentinel complaints, we construct race and ethnicity probabilities for each complaint using consumer first and last names and zip codes using Bayesian Improved Surname Geocoding (BISG) (Consumer Financial Protection Bureau, 2014; Zhang,

[27] It is possible that ATT raises awareness of data-driven fraud among iOS users and induces them to complain more, but this would lead to a positive instead of a negative effect of ATT on fraud complaints.

2018), and then aggregate these probabilities to zipcode-month level.[28] Table F.4 examines our main specifications for each race/ethnicity group. We find large and significant negative estimates for Black and Hispanic consumers, insignificant positive estimates for Non-Hispanic White consumers, and positive and significant estimates for Asian consumers. Since Raval (2020b) found that fraud victims in Black and Hispanic communities were less likely to complain than victims in White communities, large declines for complaints for Black and Hispanic consumers likely mean even greater reductions in fraud victimization.

There are two potential explanations for these patterns. First, user groups with greater declines in complaints after ATT 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 target these populations as they may be more vulnerable to fraud; previous work has shown higher rates of victimization for black communities (Raval, 2020b, 2021).

## 7 Firm Exposure to Data Breaches and Cyber Risks

The analyses in the previous section exploit variation in a locality's exposure to the ATT policy. In this section, we examine a complementary source of variation through firm-level differences in exposure to the ATT policy. We use 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 sets of outcome variables at the firm level – CFPB fraud complaints and cyber incidents – and find additional evidence supporting the connection between the privacy policy change and the reduction in financial fraud related to lax data security standards. Finally, we provide suggestive evidence that ATT contributes to an increase in the price of data pertinent to financial fraud and identity theft on the Dark Web.

[28] We match surnames to data from the Census on the distribution of race and ethnicity for more than 150,000 surnames. We also match first names to data from the Home Mortgage Disclosure Act (HMDA) on the distribution for more than 4,200 first names (Voicu, 2018). The zipcode-level race probabilities here are based upon the 2010 Census.

### 7.1 Apple's Privacy Nutrition Label and Google's Data Safety Form

Apple's ATT policy constitutes an arguably exogenous shock to consumer surveillance practices, for which firms are affected to a different degree depending on their pre-ATT consumer surveillance intensities. Apple's "nutrition" privacy labels and Google's Safety form allow us to further measure differences in firms' data collection and security practices.

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.2 (Panel a) provides the structure of privacy labels, covering 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 does not 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 Google's Safety form is illustrated in Appendix Figure A.2.[29]

The information from these disclosure policies serves two purposes. First, we confirm that

[29]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.

firms often collect and share sensitive user identifiers and financial information, including those developing finance and non-finance mobile apps. For example, in Panel b of Appendix Figure 1, Mint, a personal finance and budgeting app, collects financial information and identifiers to track users across apps and companies. In Panel c, we show that popular apps like TikTok also gather and share these data fields that are potentially valuable to fraudsters.[30] More importantly, 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 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.[31] To obtain a list of financial companies as app developers, we first pull the universe of finance apps from Sensor Tower, a digital intelligence company, and extract the developers' identifying information (name and website). Using fuzzy name matching and extensive manual checks, we map companies in the CFPB complaint database to app developers and develop several indicators to measure their data practice.

Panel b of Table 1 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 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

[30] Although TikTok is not a finance-related app, the data it collects and shares may nevertheless be misused and lead to financial fraud.

[31] 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/.

them do not encrypt data in transit.[32]

### 7.2 Firm-level CFPB Fraud Complaints

We present evidence that the effectiveness of ATT in reducing consumer complaints varies depending on the company's data vulnerability. The results, shown in Table 8, indicate that ATT has a stronger effect 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 effect of ATT on consumer complaints at the firm-product category level 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 (Panels a-b), as well as debt collection. In contrast, the number of complaints in unrelated categories, such as student loans and mortgages, is not affected by the policy (Panels c-d). These findings further confirm the economic

[32]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.

mechanism: the reduction in consumer complaints is due to ATT limiting the amount and scope of personal data subject to breaches and exploitation.

### 7.3 Cyber Incidents

We next examine whether ATT reduces firms' exposure to cybersecurity risks by leveraging data on cyber incidents from Advisen. This dataset covers more than 90,000 cyber events between 2000 and 2023 collected from publicly verifiable sources, including government websites, keyword-based searches, and official court and litigation sources.[33] We use this data to identify the presence of cyber incidents for each firm in each month during our sample period of January 2019 to June 2022.

The Advisen data provides rich information about each incident including its cause and whether it led to the firm's violations of specific regulations because of data misuse and privacy concerns. This allows us to focus the analysis on 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: an indicator variable for whether the firm was exposed to any cyber incident in a given month; whether the cyber incident was the result of malicious breaches or privacy violations, such as unauthorized data collection and disclosure by the firm; whether the cyber incident was caused by things unrelated to lax data standards, such as devices being physically lost or stolen, network disruption, or unintentional disclosure;[34] and whether the cyber incidents violate the Fair Debt Collection Practices or the Fair Credit Reporting Acts, the two most cited regulations in CFPB complaint narratives. These regulations correspond to the top two fraud product categories and so violations of these regulations directly link to fraud victimization.

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

[33] More information on Advisen's data sources can be found on their website.
[34] We include the full list of ATT-relevant causes in the Appendix.

at the 1% level. The economic magnitudes are substantial: following ATT, companies with an app experience a 3.7-percentage-point reduction in the likelihood of experiencing cyber incidents compared to firms without the app. This represents 33% of the baseline likelihood, reported at the bottom part of the table. Figure 7 plots the dynamic DiD coefficients. There is no evidence of a pre-trend, while the point estimates are all negative and become statistically significant three quarters after ATT. Therefore, other factors or events are unlikely to explain the declining cyber incidents among firms that gather mobile footprints from users.

Moreover, the effects are concentrated among cyber incidents that are more likely to be influenced by ATT: a reduction of 39% for incidents resulting from privacy risks and an insignificant effect for incidents with unrelated causes. Last, we observe a substantial reduction in firms' likelihood of incidents resulting in violations of regulations related to debt collection and credit reporting.

The firm-level evidence is inconsistent with within-firm spillovers from iOS users to Android users at the aggregate level. If such spillovers are large, we should not observe an overall decline in cyber incidents for app-owning firms that collect data from both consumer groups. In addition, this firm-level evidence is inconsistent with pandemic-related explanations, as online activities increased during the pandemic and so would lead to an increase rather than a decrease in cyber incidents for app-owning firms.

### 7.4 Price of Data on the Dark Web

Last, using data on products listed on the dark web, we provide suggestive evidence that ATT drives up the price of data relevant for financial fraud and identity theft, potentially by reducing the supply of stolen/hacked data on the Dark Web. The data is assembled by a website that conducts research on fraud. We describe the sample and our methodology in detail in Appendix I.

Using two snapshots of dark web listings in 2020 and 2023, we compare the price tags of data generated from user activities through mobile apps to other listings, and that of financial information to other listings, before and after the ATT, using a difference in difference design.

While our estimates are based on two snapshots, we find that ATT increases the listing price of data that is likely generated from consumers' mobile activities and financial information compared to other products. This finding is consistent with the notion that ATT has lowered the risk of data leakage or breach, leading to a reduced supply of shared/stolen/hacked data on the Dark Web.

## 8 Conclusion

In this paper, we took the first step in analyzing and quantifying the effects of lax data privacy regulations on financial fraud. We examined the implementation of Apple's App Tracking Transparency (ATT) policy, which restricts the tracking and sharing of personal information on the iOS platform. Using variation in iOS shares across different localities in the US, we have demonstrated that ATT reduced consumer complaints about fraud.

Our results provide evidence of benefits to consumers from privacy initiatives or regulations aimed at constraining consumer surveillance practices. Many more such efforts are planned, including Google's plan to phase out third-party cookies in Chrome by 2024 and potential federal regulations, which may generate similarly beneficial outcomes for consumers.

**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.

Anderson, K. B. (2019). Mass-market consumer fraud in the united states: A 2017 update. *Federal Trade Commission. Washington, DC*.

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.

Armantier, O., S. Doerr, J. Frost, A. Fuster, and K. Shue (2021). Whom do consumers trust with their data? us survey evidence. Technical report, Bank for International Settlements.

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*.

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

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

Consumer Financial Protection Bureau (2014). Using publicly available information to proxy for unidentified race and ethnicity. Technical report.

Consumer Financial Protection Bureau (2021, July). Complaint bulletin: Covid-19 issues described in consumer complaints. Technical report.

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.

DeLiema, M. and P. Witt (2021). Mixed methods analysis of consumer fraud reports of the social security administration impostor scam.

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

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

Federal Trade Commission (2023, February). New ftc data show consumers reported losing nearly $8.8 billion to scams in 2022. Technical report, Staff Report.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance 77*(1), 5–47.

Garg, S. and N. Baliyan (2021). Comparative analysis of android and ios from security viewpoint. *Computer Science Review 40*, 100372.

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*.

Griffin, J. M. and S. Kruger (2024). What is forensic finance. *Available at SSRN*.

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.

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

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

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

Li, 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.

Raval, D. (2020a). Which communities complain to policymakers? evidence from consumer sentinel. *Economic Inquiry 58*(4), 1628–1642.

Raval, D. (2020b). Whose voice do we hear in the marketplace? evidence from consumer complaining behavior. *Marketing Science 39*(1), 168–187.

Raval, D. (2021). Who is victimized by fraud? evidence from consumer protection cases. *Journal of Consumer Policy 44*, 43–72.

Raval, D. and M. Grosz (2022). Fraud across borders. *Available at SSRN 4197797*.

Sweeting, A., D. J. Balan, N. Kreisle, M. T. Panhans, and D. Raval (2020). Economics at the ftc: fertilizer, consumer complaints, and private label cereal. *Review of Industrial Organization 57*, 751–781.

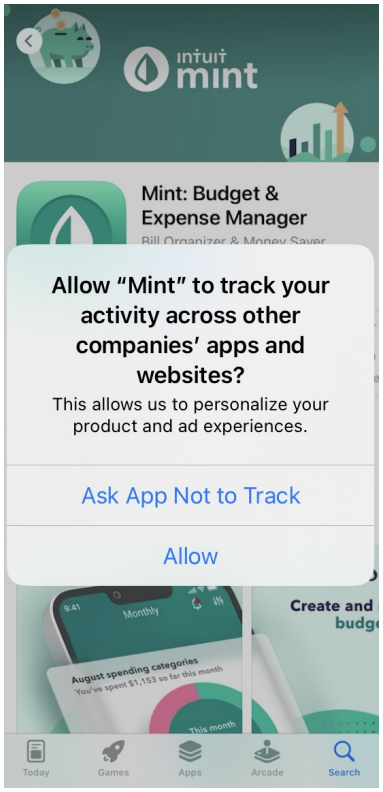
Tang, H. (2019). The value of privacy: Evidence from online borrowers. *Available at SSRN*.

Voicu, I. (2018). Using first name information to improve race and ethnicity classification. *Statistics and Public Policy 5*(1), 1–13.

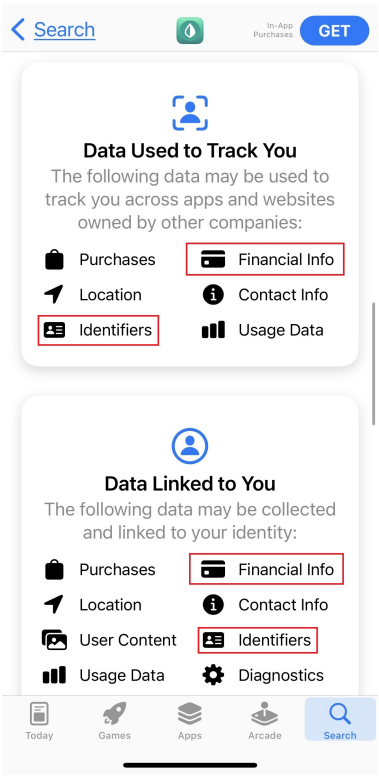
Zhang, Y. (2018). Assessing fair lending risks using race/ethnicity proxies. *Management Science 64*(1), 178–197.

**Figure 1:** Examples of App ATT Prompt and Privacy Nutrition Labels

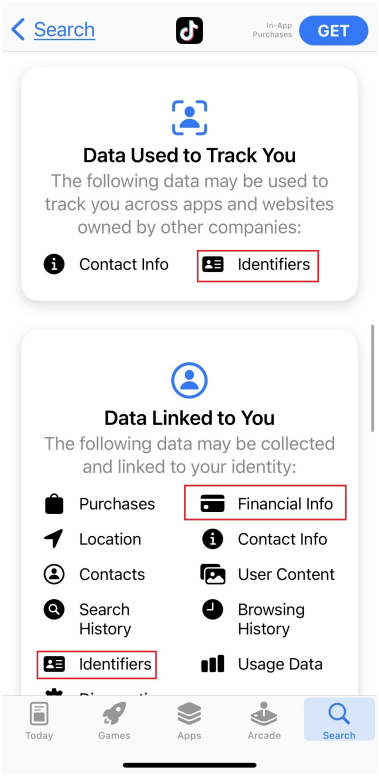
a. Mint ATT Prompt




b. Mint Privacy Labels



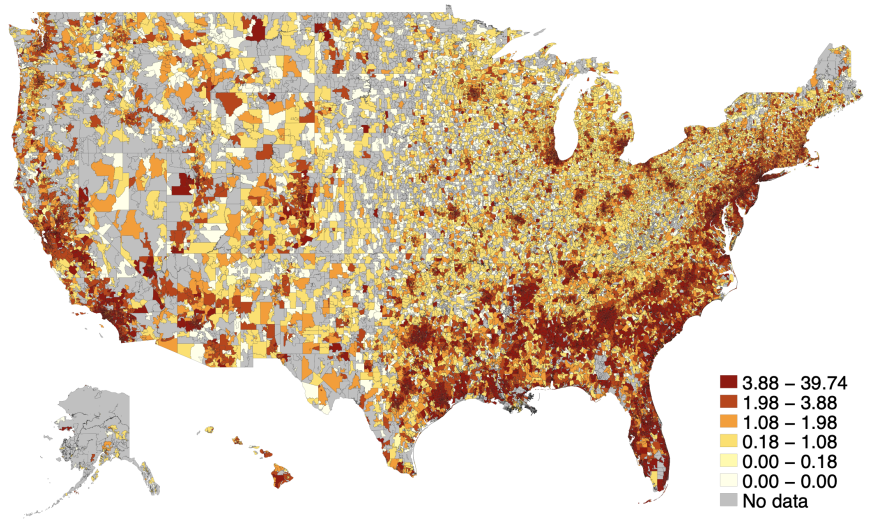

c. Tiktok Privacy Labels



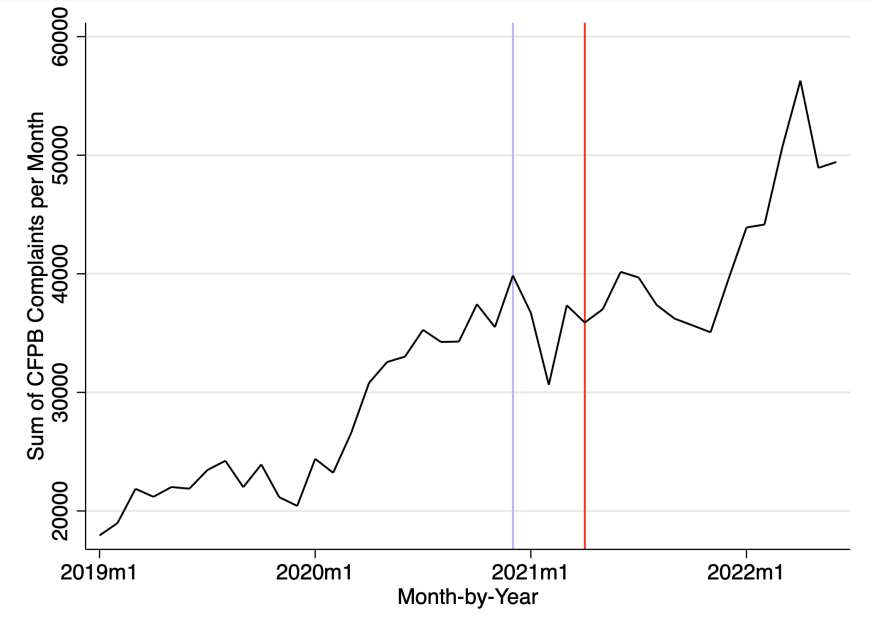

NOTE.— Apple's App Tracking Transparency (ATT) policy was introduced on April 26, 2021. In Panel a this figure shows the ATT prompt through which apps (Intuit's Mint in this example) could get user permission to obtain mobile identifiers that allow them to track and share consumers across other apps and websites. Panel b. shows Mint's privacy label which describes how mobile identifiers are used to track consumers and what data is collected and linked through data sharing. Panel c. shows Tiktok's privacy label showing that Tiktok uses mobile identifiers to obtain, e.g., financial data of consumers.

**Figure 2:** Summary Statistics for CFPB Complaints
*January 2019 to June 2022*

a. Number of complaints per 1,000 residents by zipcode



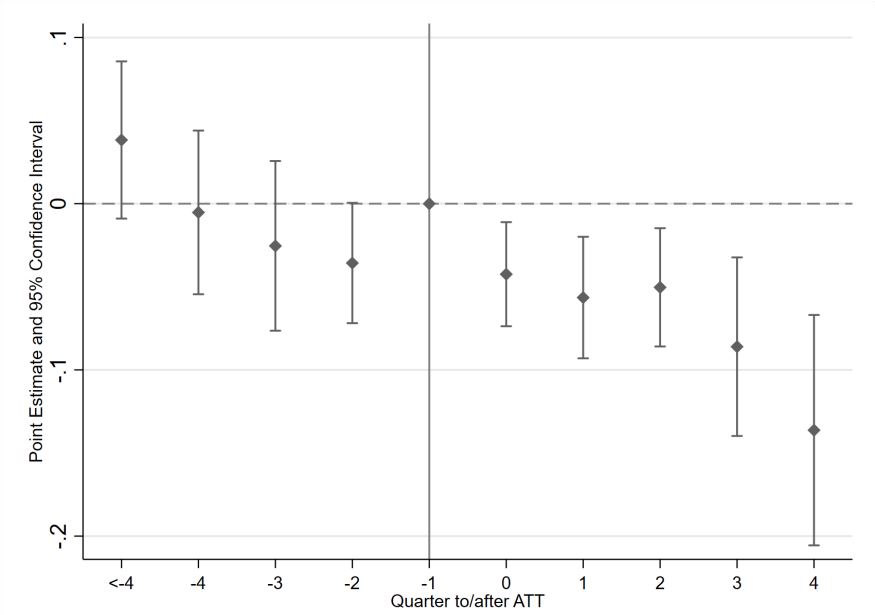

b. Aggregate number of complaints per month



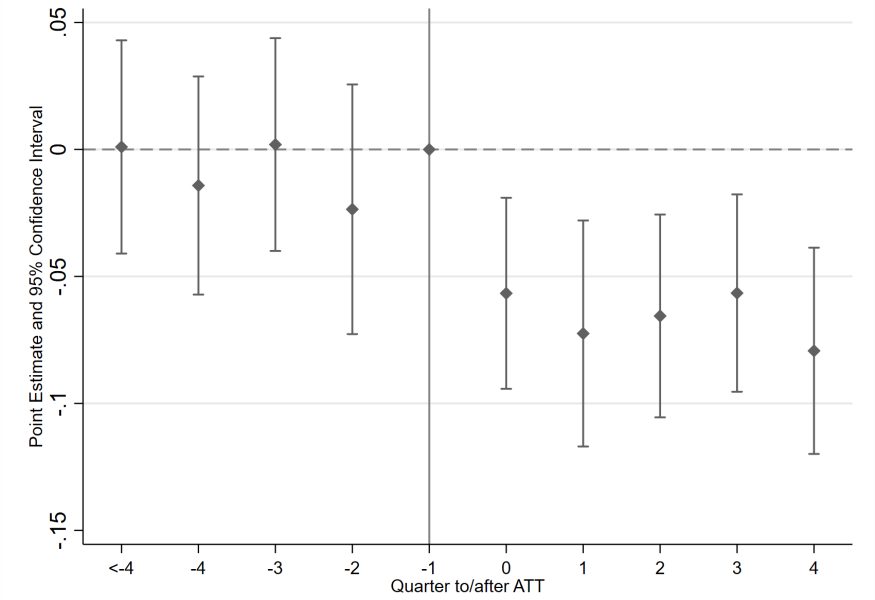

NOTE.— Panel a illustrates the number of CFPB complaints per 1,000 residents for the entire sample period in each zip code. Panel b illustrates the total number of complaints each month, aggregated across all zip codes. The red line indicates the implementation date of Apple's App Tracking Transparency Policy (April 2021). The light blue line indicates the introduction of Apple's privacy label policy (December 2020).

**Figure 3:** The Dynamic Effects of Privacy Regulation on CFPB Complaints

a. Complaints per 1,000 residents



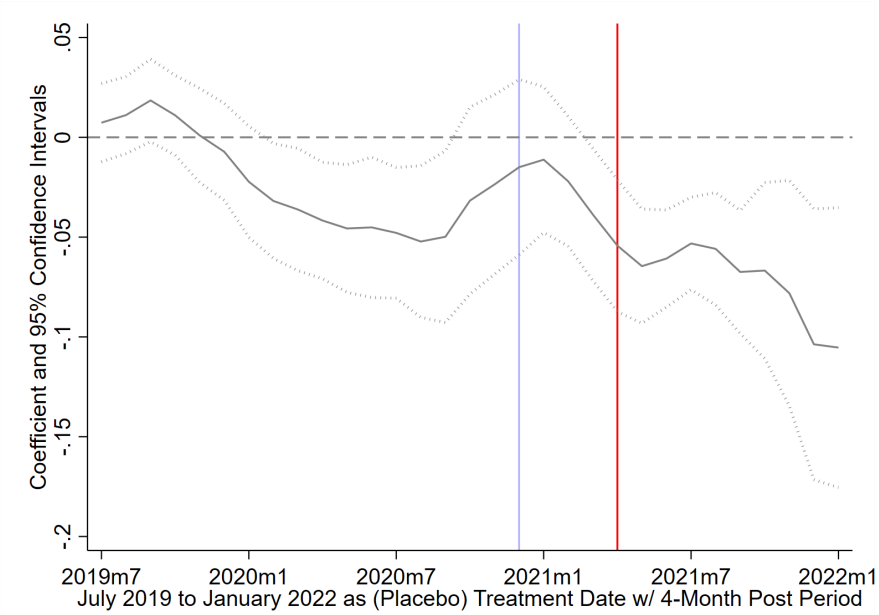

b. Any complaints



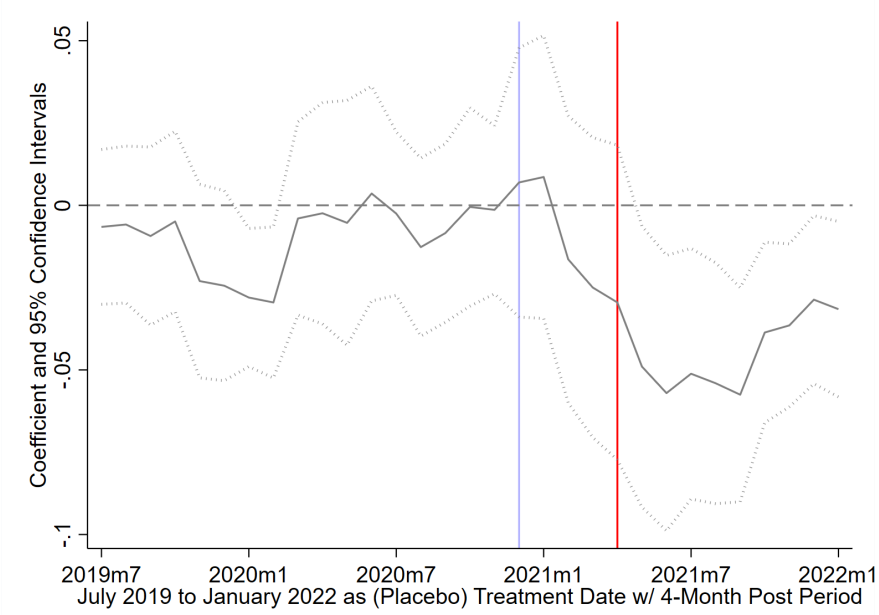

NOTE.— This figure illustrates the dynamic effect of ATT on CFPB complaints around the implementation of ATT (April 2021 or 2021Q2). Quarter -1 is the quarter before the implementation (2021Q1) and is the omitted category. All three months in a corresponding quarter are grouped to reduce estimation error. For example, Quarter 0 corresponds to April, May, and June of 2021. In Panel a, the outcome variable is the number of complaints per 1,000 residents. In Panel b, the outcome variable is an indicator for whether a zip code has at least one complaint. Coefficients on the interaction terms between indicators for the relative timing to ATT and the pre-ATT iOS device share are plotted.

**Figure 4:** Placebo Treatment Date: CFPB Complaints
*4-Month post-treatment window, continuous iOS share*

b. Complaints per 1,000 residents



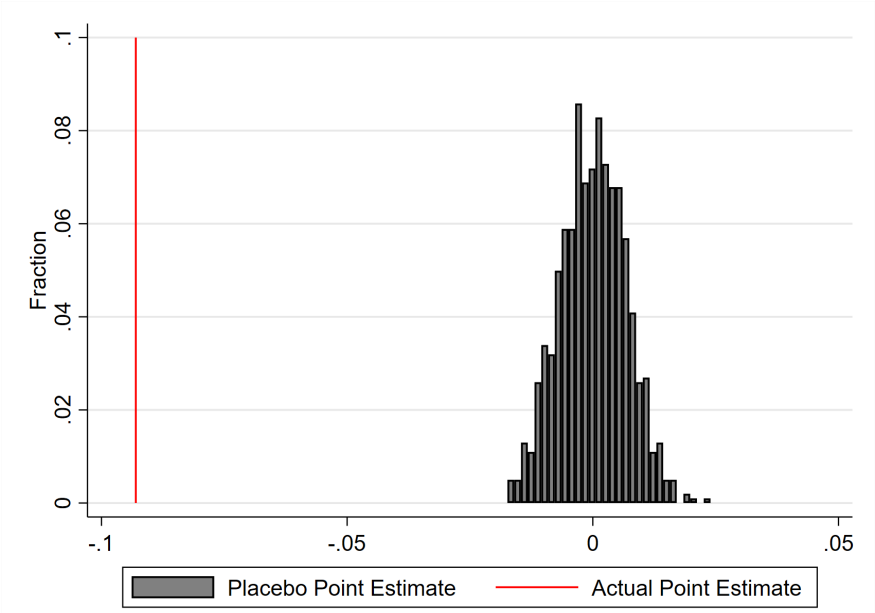

a. Any complaints



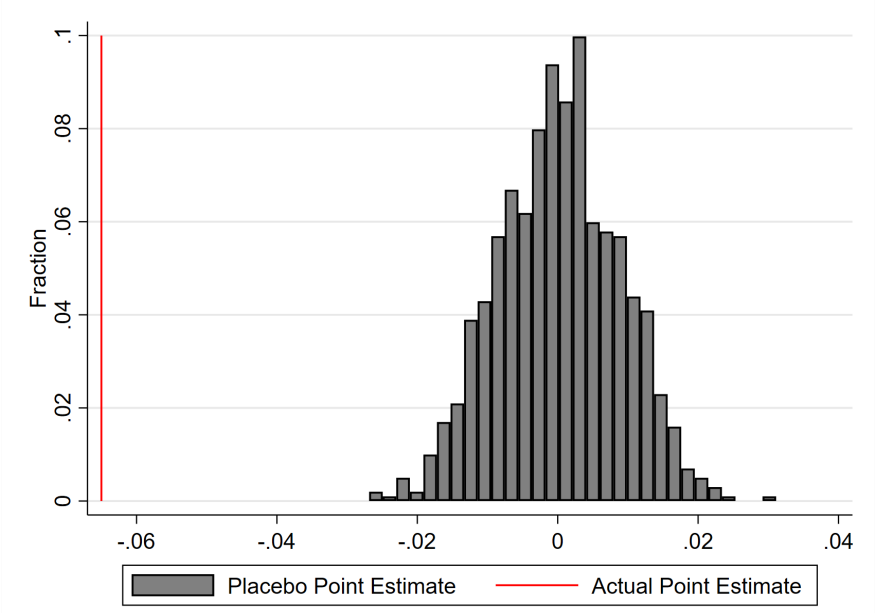

NOTE.— This figure considers any month from July 2019 to January 2022 as a placebo treatment date and plots the effect on CFPB complaints. For any placebo treatment date, a 4-month post-event window (in addition to a pre-event window starting from January 2019) is used for estimation. In Panel a, the outcome variable is the number of complaints per 1,000 residents. In Panel b, the outcome variable is an indicator for whether a zip code has at least one complaint. Coefficients on the interaction term between the indicator for the 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: CFPB Complaints
*Random iOS share*

a. Complaints per 1,000 residents



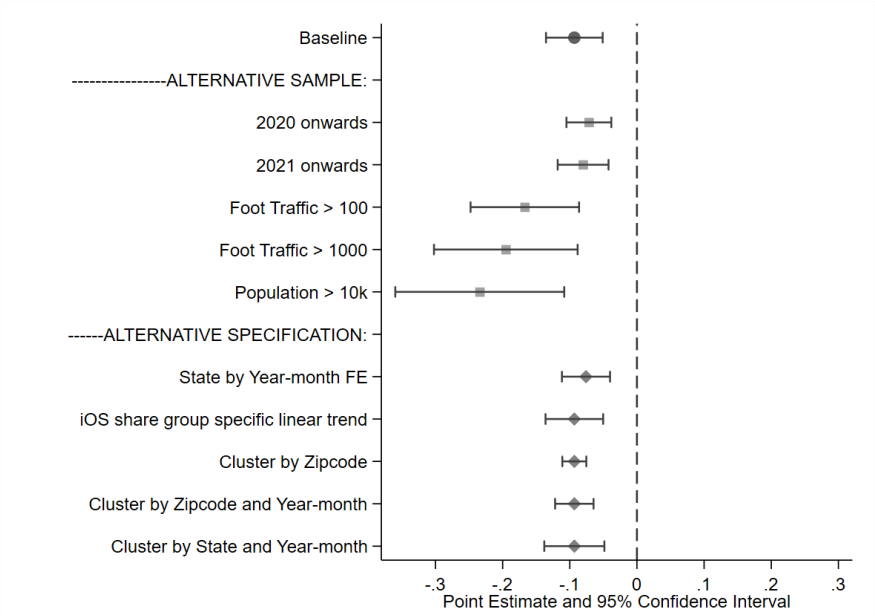

b. Any complaints



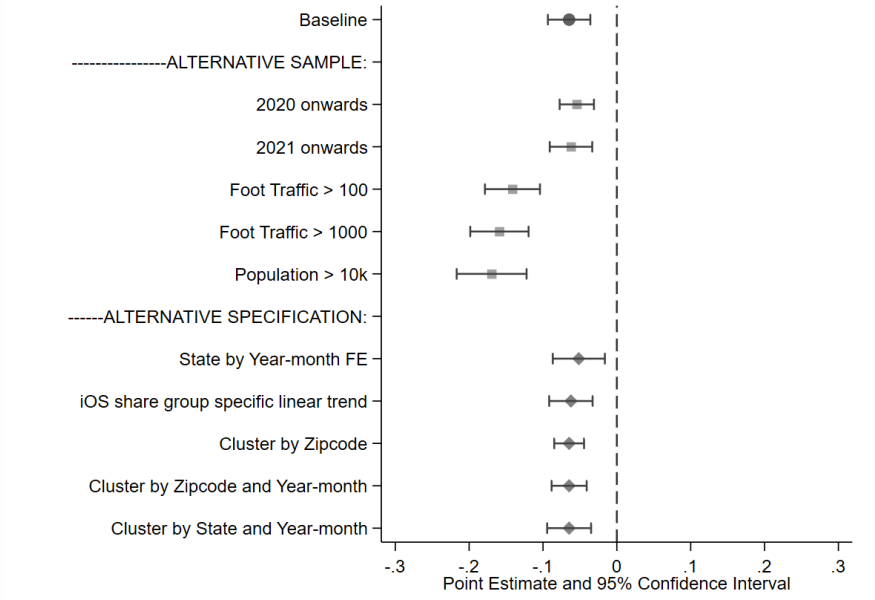

NOTE.— This figure plots the histogram of the estimated effect of ATT on CFPB complaints using 1,000 placebo tests. Each placebo test randomly reshuffles treatment intensities (pre-event iOS device share) and assigns them to zip codes. The sample and regression specifications are the same as those in Table 3. In Panel a, the outcome variable is the number of complaints per 1,000 residents. In Panel b, the outcome variable is an indicator for whether a zip code has at least one complaint.

**Figure 6:** Robustness: The Effects of Privacy Regulation on CPFB Complaints

a. Complaints per 1,000 residents

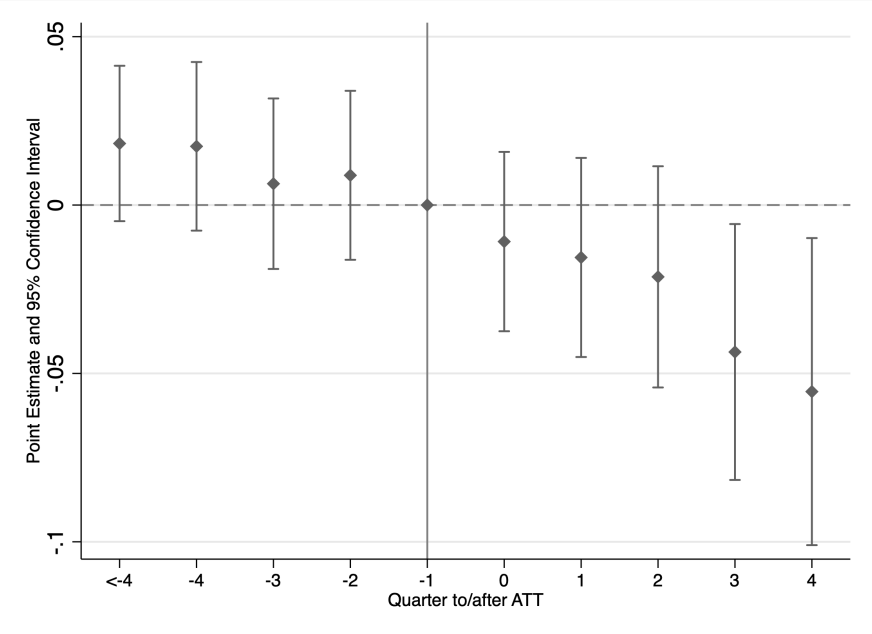



b. Any complaints




NOTE.— This figure plots the effect of ATT on CFPB complaints using alternative regression samples and specifications. The top row of each panel shows the baseline estimate in Columns 1 and 3 of Table 3. In Panel a, the outcome variable is the number of complaints per 1,000 residents. In Panel b, the outcome variable is an indicator for whether a zip code has at least one complaint.

**Figure 7:** The Dynamic Effects of Privacy Regulation on Firm-level Cyber Incidents




NOTE.— This figure illustrates the dynamic effect of ATT on the number of firm-level cyber incidents around the implementation of ATT (April 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 variable is an indicator for whether the firm has experienced a cyber incident in the corresponding quarter. Coefficients on the interaction terms between indicators for the relative timing to ATT and the indicator for whether the firm is exposed to ATT by owning at least one app are plotted.

**Table 1:** Summary Statistics

**Panel a. Zipcode-month level**

| | mean | sd | p25 | p50 | p75 | count |
|---|---|---|---|---|---|---|
| *Complaints per 1,000 residents (zipcode-month panel)* | | | | | | |
| CFPB | 0.07 | 0.20 | 0.00 | 0.00 | 0.06 | 1,003,758 |
| Consumer Sentinel | 0.72 | 0.95 | 0.00 | 0.55 | 0.96 | 1,161,552 |
| Identity Theft | 0.19 | 0.40 | 0.00 | 0.00 | 0.22 | 1,133,896 |
| *Alternative measures for CFPB complaints (zipcode-month panel)* | | | | | | |
| Any complaints (0/1) | 0.31 | 0.46 | 0.00 | 0.00 | 1.00 | 1,026,942 |
| Complaints | 1.38 | 3.55 | 0.00 | 0.00 | 1.00 | 1,026,942 |
| log(1+Complaints) | 0.44 | 0.77 | 0.00 | 0.00 | 0.69 | 1,026,942 |
| *Monthly aggregate number of complaints* | | | | | | |
| CFPB | 36,936 | 12,947 | 24,399 | 38,754 | 43,043 | 42 |
| Consumer Sentinel | 265,834 | 65,128 | 203,352 | 279,620 | 299,504 | 42 |
| Identity Theft | 93,485 | 38,369 | 61,426 | 87,735 | 108,746 | 41 |
| iOS share | 0.46 | 0.11 | 0.39 | 0.45 | 0.52 | 1,026,942 |

**Panel b. Firm-month level**

| | mean | sd | p25 | p50 | p75 | count |
|---|---|---|---|---|---|---|
| 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 |
| Shares data with third party (0/1) | 0.11 | 0.31 | 0.00 | 0.00 | 0.00 | 12,235 |
| Data not 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 |

NOTE.— This table presents summary statistics for our key explanatory and outcome variables at zipcode-month and firm-month levels. Panel a first shows the summary statistics for the number of complaints per 1,000 residents for CFPB, Consumer Sentinel, and Identify Theft complaints, respectively. Panel a then shows additional summary statistics for three more variables using only CFPB complaints: an indicator for whether a zip code has at least one complaint, the total number of complaints, and the logarithm of one plus the total number of complaints. At the bottom of Panel a, we report summary statistics for the monthly nation-wide number of CFPB, Consumer Sentinel, and Identify Theft complaints. At the bottom of Panel a, we also report summary statistics for iOS device share at the zip code level. Panel b reports the summary statistics for the CFPB complaints aggregated to the firm-month level and for firm's app presence and activities. The latter includes an indicator for whether the firm owns an app and conditional on owning an app, whether the app shares data with any third party, whether the app fails to encrypt data in transit, and the number of complaints per 1,000 app downloads. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022.

**Table 2:** Privacy Regulation and Zipcode-level Complaints

| | (1) CFPB | (2) Consumer Sentinel Network | (3) Identity Theft |
|---|---|---|---|
| Post × iOS share | −0.093*** | −0.103** | −0.073*** |
| | (0.021) | (0.041) | (0.022) |
| Zipcode FE | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.074 | 0.720 | 0.190 |
| Observations | 1,003,590 | 1,150,590 | 1,123,195 |
| R-square | 0.397 | 0.317 | 0.553 |

NOTE.— This table displays the results from estimating Specification 1 using all three datasets for consumer complaints. The unit of observation is at the zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Column 1 uses all public CFPB complaints. Column 2 uses all Consumer Sentinel Network complaints. Column 3 uses all Identity Theft complaints. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 3:** Privacy Regulation and Zipcode-level CFPB Complaints - Additional Measures

| | (1) Any complaints (0/1) | (2) Complaints winsorized | (3) log(1+Complaints) |
|---|---|---|---|
| Post × iOS share | −0.065*** | −1.688*** | −0.263*** |
| | (0.014) | (0.327) | (0.045) |
| Zipcode FE | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ |
| Observations | 1,026,942 | 1,026,942 | 1,026,942 |
| R-square | 0.569 | 0.680 | 0.702 |

NOTE.— This table displays the results from estimating Specification 1. The unit of observation is at the zipcode-month level. The outcome variables in Columns 1-3 are an indicator for whether a zip code has at least one complaint, the number of complaints (winsorized at 1%), and the logarithm of one plus the number of complaints. The sample period is January 2019 to June 2022 and all specifications use public CFPB complaints. We include zip code and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 4:** Privacy Regulation and Zipcode-level Complaints With Controls for EIP

| | (1) CFPB | (2) Consumer Sentinel Network | (3) Identity Theft |
|---|---|---|---|
| Post × iOS share | −0.072*** | −0.111*** | −0.039** |
| | (0.017) | (0.043) | (0.016) |
| Post × EIP amount | −0.008*** | −0.016*** | 0.002 |
| | (0.002) | (0.006) | (0.002) |
| Post × Total income | −0.005* | −0.025** | −0.007** |
| | (0.003) | (0.009) | (0.003) |
| Post× Child care credit | −0.004** | 0.009 | −0.009*** |
| | (0.002) | (0.005) | (0.003) |
| Zipcode FE | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ |
| EIP Controls | ✓ | ✓ | ✓ |
| Observations | 957,600 | 1,049,454 | 1,024,467 |
| R-square | 0.415 | 0.361 | 0.610 |

NOTE.— This table displays the results from estimating Specification 1 with controls for economic impact payments (EIP) issuance. We add the interaction term between the post-ATT indicator and variables related to EIP as controls. These variables include the total amount of EIP received, the average household income, and the amount of childcare credit, all constructed using IRS data. The unit of observation is at the zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Column 1 uses all CFPB complaints. Column 2 uses all Consumer Sentinel Network complaints. Column 3 uses all Identity Theft complaints. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. We include zip code and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 5:** Privacy Regulation and Zipcode-level Complaints: Relevance to ATT

**Panel a. CFPB Complaints**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Top 2 Fraud Category | | Bottom 2 Fraud Category | |
| | Credit Reporting and Repair | Debt Collection | Student Loan | Mortgage |
| Post × iOS share | −0.081*** | −0.009*** | −0.0001 | −0.0004 |
| | (0.017) | (0.002) | (0.0003) | (0.0005) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.065 | 0.076 | 0.001 | 0.004 |
| Observations | 1,010,142 | 1,010,142 | 1,010,142 | 1,010,142 |
| R-square | 0.408 | 0.215 | 0.118 | 0.168 |

**Panel b. CSN and Identity Theft Complaints**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Narratives w/ Keywords | | Narratives w/o Keywords | |
| | Consumer Sentinel Network | Identity Theft | Consumer Sentinel Network | Identity Theft |
| Post × iOS share | −0.083*** | −0.071*** | −0.021 | −0.006 |
| | (0.007) | (0.013) | (0.022) | (0.012) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.255 | 0.076 | 0.458 | 0.109 |
| Observations | 1,123,195 | 1,123,195 | 1,123,195 | 1,123,195 |
| R-square | 0.295 | 0.467 | 0.255 | 0.496 |

NOTE.— This table displays the results from estimating Specification 1 after classifying complaints by their relevance to ATT. The unit of observation is at the zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Panel a considers the top 2 relevant and bottom 2 relevant product categories in CFPB complaints: Credit Reporting and Credit Repair Services (Top 1), Debt Collection (Top 2), Student Loans (Bottom 2) and Mortgages (Bottom 1). Panel b leverages the narratives available in the Consumer Sentinel Network and Identity Theft complaints to classify complaints into those with and without any relevant words in the narrative. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 6:** Privacy Regulation and Fraud Victimization - Complaints Reporting Positive Losses

| | (1) Any Positive Loss | (2) Losses between \$1-\$99 | (3) Losses between \$100-\$999 | (4) Losses \$1000 or above |
|---|---|---|---|---|
| Post × iOS share | −0.020** | −0.007** | −0.007 | −0.006* |
| | (0.009) | (0.002) | (0.005) | (0.004) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.147 | 0.035 | 0.058 | 0.045 |
| Observations | 1,150,590 | 1,150,590 | 1,150,590 | 1,150,590 |
| R-square | 0.203 | 0.176 | 0.180 | 0.176 |

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 table only examines Consumer Sentinel complaints reporting any positive losses, positive losses between \$1 and \$99 (first panel), between \$100 and \$999 (second panel), and \$1,000 and above (third panel). The outcome variable is the number of complaints per 1,000 residents at the zip code level. The sample period is January 2019 to June 2022. We 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 7:** Privacy Regulation and Fraud Victimization

| | (1) CFPB | (2) Consumer Sentinel Network | (3) Identity Theft |
|---|---|---|---|
| Post × iOS share | −0.139*** | −0.356*** | −0.109*** |
| | (0.029) | (0.061) | (0.026) |
| Zipcode FE | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ |
| Observations | 959,658 | 1,060,248 | 1,035,004 |
| R-square | 0.446 | 0.500 | 0.688 |

NOTE.— This table displays the results from estimating Specification 1 by accounting for differences in the propensity to complain across zip codes. We multiply the outcome variable by weights developed in Raval (2020b) that adjust aggregate complaints to reflect fraud victimization by accounting for differences in the propensity to complain across zip-level demographics. The unit of observation is at the zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Column 1 uses all CFPB complaints, Column 2 uses all Consumer Sentinel Network complaints and Column 3 uses all Identity Theft complaints. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. We include zip code and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 8:** Privacy Regulation and Firm-level Complaints

| | Any complaints (0/1) | log(1+Complaints) | Complaints per 1,000 downloads | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Post × Has an app (0/1) | −0.011** | −0.039*** | | |
| | (0.005) | (0.011) | | |
| Post × Share data with third party (0/1) | | | −1.693** | |
| | | | (0.695) | |
| Post × Data not encrypted in transit (0/1) | | | | −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 effect of ATT on firm-level complaints received via CFPB. The unit of observation is at the firm-month level. In Columns 1 and 2, we interact the post-ATT indicator with an indicator for whether a firm owns an app. In Column 1, the outcome variable is an indicator of whether the firm has received any complaints in a given month. In Column 2, the outcome variable is the logarithm of one plus the total number of complaints received by a firm in a given month. In Columns 3 and 4, we include only firms with an app and interact the post-ATT indicator with an indicator for firms that share user data with third parties (Column 3) or for firms that do not encrypt data in transit (Column 4). In Columns 3 and 4, the outcome variable is the number of complaints per 1,000 monthly app downloads. We include firm and year-month fixed effects and control for linear time trends specific to app popularity (measured by the log of worldwide all-time downloads). Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 9:** Privacy Regulation and Firm-level Complaints - Relevance to ATT

| | Any complaints (0/1) | log(1+Complaints) | Complaints per 1,000 downloads | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Panel a. Top1 Fraud Category - Credit Reporting and Repair | | | |
| Post × Has an app (0/1) | −0.008** | −0.029*** | | |
| | (0.004) | (0.006) | | |
| Post × Share data with third party (0/1) | | | −0.758* | |
| | | | (0.450) | |
| Post × Data not encrypted in transit (0/1) | | | | −0.064** |
| | | | | (0.032) |
| R-square | 0.534 | 0.821 | 0.544 | 0.677 |
| | Panel b. Top2 Fraud Category - Debt Collection | | | |
| Post × Has an app (0/1) | −0.008** | −0.029*** | | |
| | (0.004) | (0.006) | | |
| Post × Share data with third party (0/1) | | | −0.758* | |
| | | | (0.450) | |
| Post × Data not encrypted in transit (0/1) | | | | −0.064** |
| | | | | (0.032) |
| R-square | 0.534 | 0.821 | 0.544 | 0.677 |
| | Panel c. Bottom2 Fraud Category - Student Loan | | | |
| Post × Has an app (0/1) | −0.007* | −0.018*** | | |
| | (0.004) | (0.006) | | |
| Post × Share data with third party (0/1) | | | −1.060* | |
| | | | (0.616) | |
| Post × Data not encrypted in transit (0/1) | | | | −0.120 |
| | | | | (0.083) |
| R-square | 0.541 | 0.791 | 0.463 | 0.634 |
| | Panel d. Bottom1 Fraud Category - Mortgage | | | |
| Post × Has an app (0/1) | −0.001 | −0.003 | | |
| | (0.003) | (0.004) | | |
| Post × Share data with third party (0/1) | | | −0.140 | |
| | | | (0.145) | |
| Post × Data not encrypted in transit (0/1) | | | | −0.017 |
| | | | | (0.012) |
| R-square | 0.539 | 0.835 | 0.557 | 0.717 |
| 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 |

NOTE.— This table reports the effect of ATT on firm-level complaints received via CFPB after classifying complaints by their relevance to ATT. The unit of observation is at the firm-month level. The four panels consider the top 2 relevant and bottom 2 relevant product categories in CFPB complaints: Credit Reporting and Credit Repair Services (Top 1), Debt Collection (Top 2), Student Loan (Bottom 2), and Mortgage (Bottom 1). In Columns 1 and 2, we interact the post-ATT indicator with an indicator for whether a firm owns an app. In Column 1, the outcome variable is an indicator of whether the firm has received any complaints in a given month. In Column 2, the outcome variable is the logarithm of one plus the total number of complaints received by a firm in a given month. In Columns 3 and 4, we include only firms with an app and interact the post-ATT indicator with an indicator for firms that share user data with third parties (Column 3) or for firms that do not encrypt data in transit (Column 4). In Columns 3 and 4, the outcome variable is the number of complaints per 1,000 monthly app downloads. We include firm and year-month fixed effects and control for linear time downloads). Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 10:** Privacy Regulation and Firm-level Cyber Incidents

| | Cyber incidents (0/1) | | | |
|---|---|---|---|---|
| | (1) All types | (2) Breach/Data misuse | (3) Other causes | (4) Regulation violated |
| Post × Has an app (0/1) | −0.037*** | −0.030*** | −0.007 | −0.007*** |
| | (0.008) | (0.007) | (0.005) | (0.002) |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.111 | 0.077 | 0.077 | 0.011 |
| Observations | 100,197 | 100,197 | 100,197 | 100,197 |
| R-square | 0.198 | 0.183 | 0.173 | 0.185 |

NOTE.— This table reports the effect of ATT on cyber incidents experienced by financial institutions. The unit of observation is at the firm-month level. In all columns, we interact the post-ATT indicator with an indicator for whether a firm owns an app. The outcome variables are an indicator variable for whether the firm was exposed to (1) any cyber incident, (2) cyber incidents that were caused by data breach or data misuse, (3) cyber incidents that were caused by other reasons unrelated to data breach, 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. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

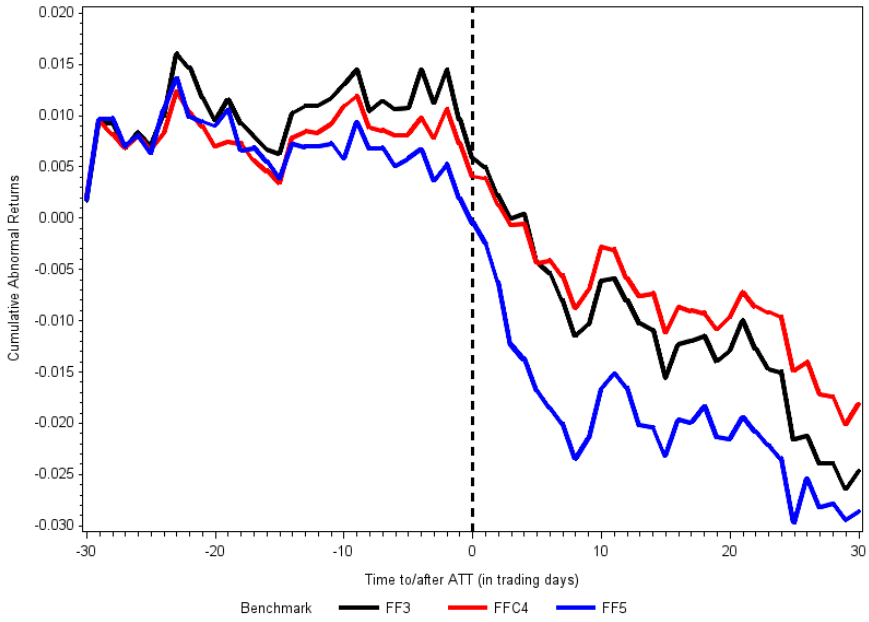
# Consumer Surveillance and Financial Fraud

Online Appendix

**Bo Bian  Michaela Pagel  Devesh Raval  Huan Tang**

# A The App Tracking Transparency and Privacy Nutrition Label Policies

**Figure A.1:** Stock Market Reactions around ATT

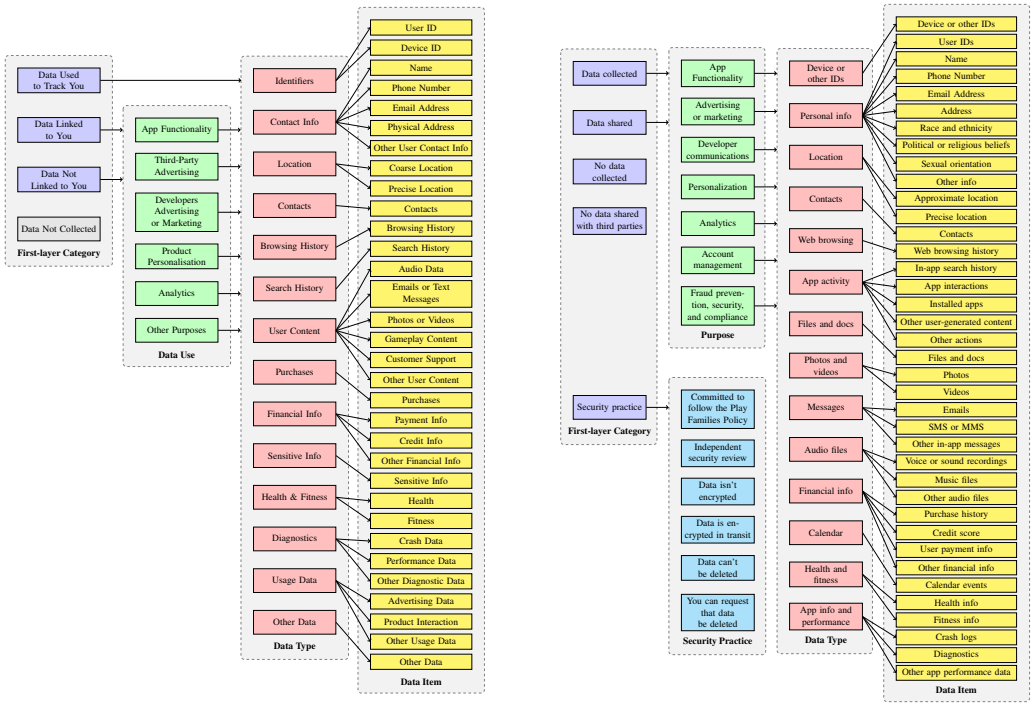



NOTE.— This figure is from Bian et al. (2021). This figure plots the average cumulative abnormal returns (CARs) around the implementation of the App Transparency Tracking Policy on April 26, 2021. The event window includes 30 days before and after the implementation date. CARs are computed using the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

## Figure A.2: Apple Privacy Nutrition Label and Google Safety Form

a. Apple Privacy Nutrition Label

b. Google Safety Form




NOTE.— Panel a shows the structure of the mandatory privacy nutrition label from Apple for iOS apps. Panel b shows the structure of the mandatory safety form for Android apps from Google. Apple privacy label has four layers. 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 types 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 the Google safety form is similar, except that it provides additional information on the firm's data security practice. Firms have to disclose whether data is encrypted in transit, for example.

# B Consumer Sentinel Complaint Data Fields




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: | |



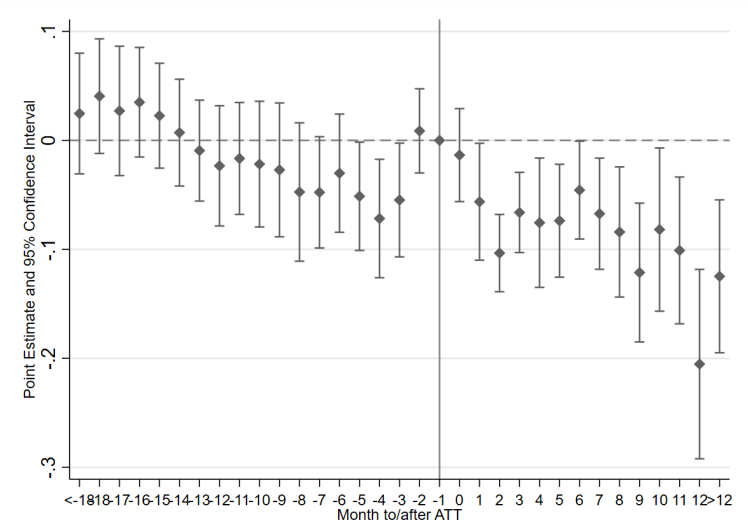


| | | | |
|---|---|---|---|
| ZIP: | | Country: | United States |
| Email: | | URL: | |
| Area Code: | | Phone Number: | |
| Ext: | | Subject ID Type: | |
| Subject ID Issuer State: | | Subject ID Issuer Country: | |
| Representative Name: | | Title: | |

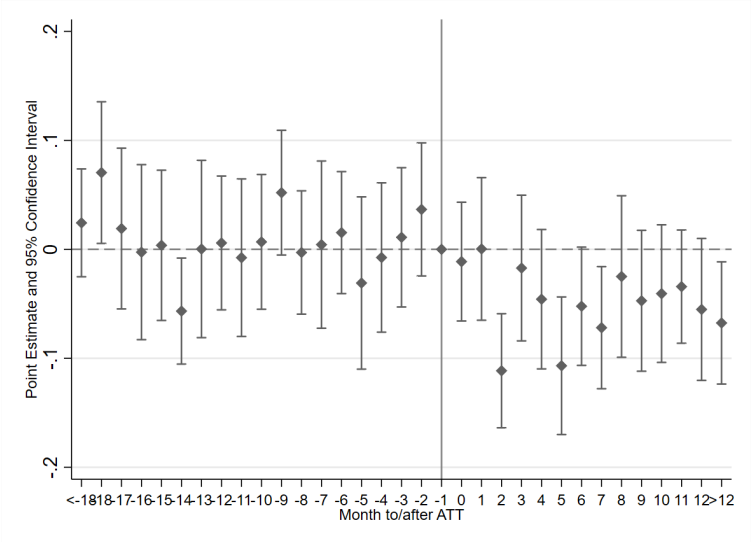

https://www.consumersentinel.gov/_layouts/PrintRecordDetails.aspx?documentNumbers=... 3/20/2013

## C The Dynamic Effect of Privacy Regulation on CFPB Complaints - Monthly Frequency

**Figure C.1:** Dynamic DiD Effects




a. Intensive margin: Complaints per 1,000 residents




b. Extensive margin: Any complaints

NOTE.— This figure illustrates the dynamic effect of ATT on CFPB complaints around the implementation of ATT (April 2021 or 2021Q2) at monthly frequency. Month -1 is the month before the implementation (March, 2021), and is the omitted category. In Panel a, the outcome variable is the number of complaints per 1,000 residents. In Panel b, the outcome variable is an indicator for whether a zip code has at least a complaint. Coefficients on the interaction terms between indicators for the relative timing to ATT and the pre-ATT iOS device share are plotted.

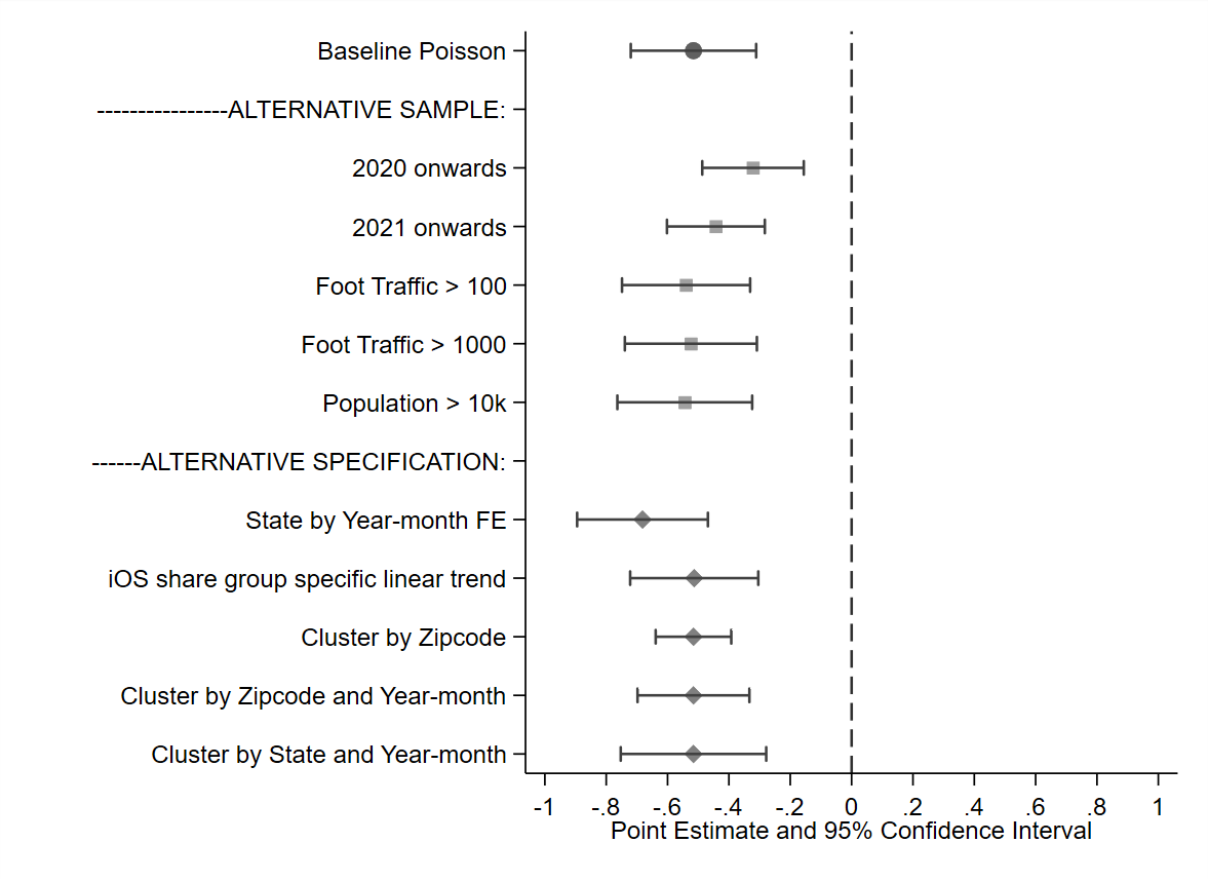
## D Robustness Checks

**Table D.1:** Privacy Regulation and Zipcode-level CFPB Complaints: By iOS Share Decile

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Complaints per 1,000 capita | Any complaints (0/1) | Complaints winsorized | log(1+Complaints) |
| Post × Decile #2 | −0.004 | 0.002 | 0.042 | 0.005 |
| | (0.004) | (0.006) | (0.073) | (0.012) |
| Post × Decile #3 | −0.003 | 0.004 | 0.037 | 0.006 |
| | (0.004) | (0.006) | (0.063) | (0.011) |
| Post × Decile #4 | −0.008* | −0.003 | 0.075 | 0.001 |
| | (0.004) | (0.006) | (0.076) | (0.011) |
| Post × Decile #5 | −0.011* | 0.000 | −0.020 | −0.003 |
| | (0.006) | (0.007) | (0.089) | (0.013) |
| Post × Decile #6 | −0.016** | −0.006 | −0.155 | −0.024* |
| | (0.007) | (0.006) | (0.096) | (0.013) |
| Post × Decile #7 | −0.017** | −0.003 | −0.185* | −0.029* |
| | (0.007) | (0.006) | (0.099) | (0.016) |
| Post × Decile #8 | −0.025*** | −0.013* | −0.326*** | −0.053*** |
| | (0.007) | (0.007) | (0.104) | (0.014) |
| Post × Decile #9 | −0.029*** | −0.010 | −0.399*** | −0.059*** |
| | (0.008) | (0.007) | (0.127) | (0.017) |
| Post × Decile #10 | −0.043*** | −0.031*** | −0.667*** | −0.109*** |
| | (0.008) | (0.006) | (0.135) | (0.017) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Mean outcome var. | 0.318 | 1.380 | 0.069 | 0.437 |
| Observations | 1,010,142 | 1,026,942 | 1,026,942 | 1,026,942 |
| R-square | 0.373 | 0.569 | 0.680 | 0.702 |

NOTE.— This table displays the results from estimating a modified version of Specification 1, in which we interact the post-ATT indicator with indicators for each of the deciles of the pre-ATT iOS share. The omitted base group includes zipcodes that are in the first decile of the pre-ATT iOS share distribution. The unit of observation is at zipcode-month level. The outcome variables from Column 1-4 are the number of (winsorized) complaints per 1,000 residents, an indicator for whether a zip code has at least a complaint, the number of complaints (winsorized at 1%), and the logarithm of one plus the number of complaints. The sample period is January 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

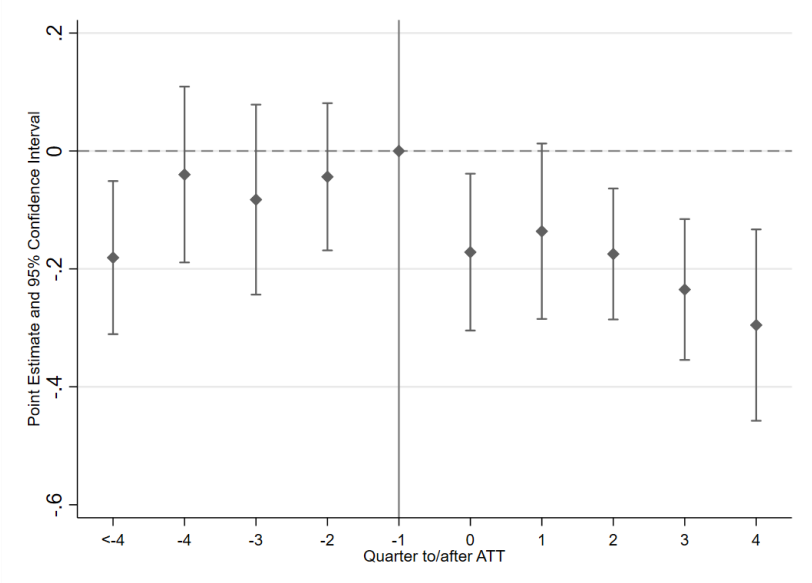
**Figure D.1:** Robustness using Poisson Regressions



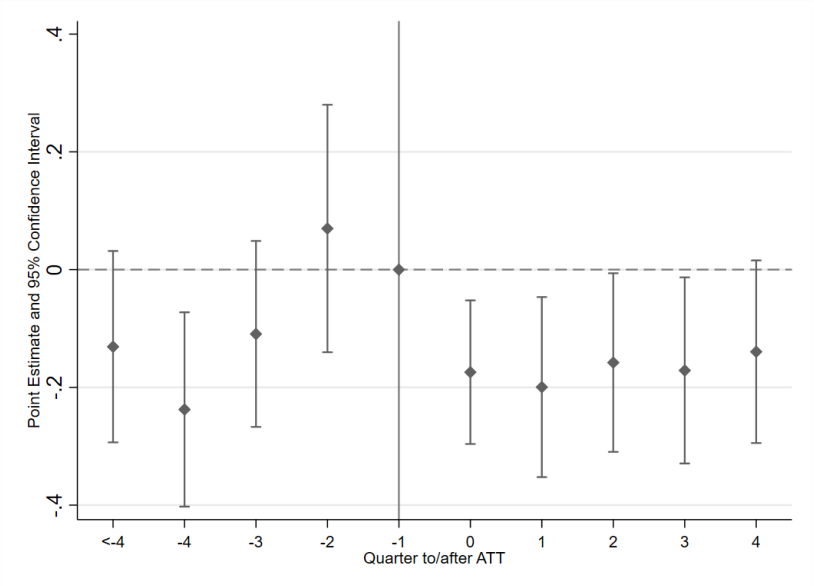

NOTE.— This figure plots the effect of ATT on zipcode-level CFPB complaints using Possion regressions and with different regression samples and specifications. The top row shows the baseline Possion estimate with the same sample and fixed effects as our baseline specification. The outcome variable is the number of complaints at the zip code level (winsorized at 1%).

## E The Dynamic Effect of Privacy Regulation - Additional Complaint Sources

**Figure E.1:** The Dynamic Effect of Privacy Regulation - Additional Complaint Sources




a. Consumer Sentinel Network Complaints




b. Identity Theft Complaints

NOTE.— This figure illustrates the dynamic effect of ATT on complaints from Consumer Sentinel Network (Panel a) and Identity Theft (Panel b) around the implementation of ATT (April 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. Coefficients on the interaction terms between indicators for the relative timing to ATT and the pre-ATT iOS device share are plotted.

## F Heterogeneity Across Different Demographic Groups

**Table F.1:** Privacy Regulation and Zipcode-level CFPB Complaints
*Interaction with Demographics*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Post × iOS share | −0.045*** | −0.073*** | −0.100*** | −0.070*** | −0.078*** |
| | (0.012) | (0.014) | (0.022) | (0.012) | (0.018) |
| Post × Black share above median | 0.070*** | | | | |
| | (0.013) | | | | |
| Post × Black share above median × iOS share | −0.107*** | | | | |
| | (0.025) | | | | |
| Post × Asian share above median | | 0.033** | | | |
| | | (0.014) | | | |
| Post × Asian share above median × iOS share | | −0.062** | | | |
| | | (0.028) | | | |
| Post × Hispanic share above median | | | 0.009 | | |
| | | | (0.013) | | |
| Post × Hispanic share above median × iOS share | | | 0.001 | | |
| | | | (0.027) | | |
| Post × Female share above median | | | | 0.047*** | |
| | | | | (0.012) | |
| Post × Female share above median × iOS share | | | | −0.076*** | |
| | | | | (0.024) | |
| Post × Age 10-19 share above median | | | | | 0.024*** |
| | | | | | (0.007) |
| Post × Age 10-19 share above median × iOS share | | | | | −0.047*** |
| | | | | | (0.015) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | | | | | |
| Observations | 1,010,142 | 1,010,142 | 1,010,142 | 1,010,142 | 1,010,142 |
| R-square | 0.373 | 0.373 | 0.373 | 0.373 | 0.373 |

NOTE.— This table shows the heterogeneous effect of ATT on CFPB complaints across different demographics groups. The unit of observation is at zipcode-month level. We divide zip codes into two groups based on the sample median of the following demographics: share of Black population, share of Hispanic population, share of Asian population, share of female, and share of people aged between 10-19. The outcome variable is the number of complaints per 1,000 residents. The sample period is January 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table F.2:** Privacy Regulation and Zipcode-level Consumer Sentinel Network Complaints

*Interaction with Demographics*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Post × iOS share | −0.013 | −0.044 | −0.140*** | −0.069* | −0.056 |
| | (0.039) | (0.047) | (0.045) | (0.037) | (0.046) |
| Post × Black share above median | 0.133*** | | | | |
| | (0.034) | | | | |
| Post × Black share above median × iOS share | −0.194*** | | | | |
| | (0.067) | | | | |
| Post × Asian share above median | | 0.085*** | | | |
| | | (0.031) | | | |
| Post × Asian share above median × iOS share | | −0.143** | | | |
| | | (0.067) | | | |
| Post × Hispanic share above median | | | −0.029 | | |
| | | | (0.032) | | |
| Post × Hispanic share above median × iOS share | | | 0.090 | | |
| | | | (0.065) | | |
| Post × Female share above median | | | | 0.073*** | |
| | | | | (0.027) | |
| Post × Female share above median × iOS share | | | | −0.096* | |
| | | | | (0.057) | |
| Post × Age 10-19 share above median | | | | | 0.051** |
| | | | | | (0.023) |
| Post × Age 10-19 share above median × iOS share | | | | | −0.093* |
| | | | | | (0.049) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 1,150,590 | 1,150,590 | 1,150,590 | 1,150,590 | 1,150,590 |
| R-square | 0.317 | 0.317 | 0.317 | 0.317 | 0.317 |

NOTE.— This table shows the heterogeneous effect of ATT on Consumer Sentinel Network complaints across different demographics groups. The unit of observation is at zipcode-month level. We divide zip codes into two groups based on the sample median of the following demographics: share of Black population, share of Hispanic population, share of Asian population, share of female, and share of people aged between 10-19. The outcome variable is the number of complaints per 1,000 residents. The sample period is January 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table F.3:** Privacy Regulation and Zipcode-level Identity Theft Complaints
*Interaction with Demographics*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Post × iOS share | −0.056*** | −0.032 | −0.049** | −0.060** | −0.060** |
| | (0.020) | (0.019) | (0.021) | (0.024) | (0.025) |
| Post × Black share above median | 0.032*** | | | | |
| | (0.012) | | | | |
| Post × Black share above median × iOS share | −0.032 | | | | |
| | (0.023) | | | | |
| Post × Asian share above median | | 0.040*** | | | |
| | | (0.011) | | | |
| Post × Asian share above median × iOS share | | −0.087*** | | | |
| | | (0.025) | | | |
| Post × Hispanic share above median | | | 0.030* | | |
| | | | (0.015) | | |
| Post × Hispanic share above median × iOS share | | | −0.054 | | |
| | | | (0.033) | | |
| Post × Female share above median | | | | 0.010 | |
| | | | | (0.008) | |
| Post × Female share above median × iOS share | | | | −0.027 | |
| | | | | (0.017) | |
| Post × Age 10-19 share above median | | | | | 0.017 |
| | | | | | (0.013) |
| Post × Age 10-19 share above median × iOS share | | | | | −0.027 |
| | | | | | (0.033) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 1,123,195 | 1,123,195 | 1,123,195 | 1,123,195 | 1,123,195 |
| R-square | 0.553 | 0.553 | 0.553 | 0.553 | 0.553 |

NOTE.—This table shows the heterogeneous effect of ATT on identity theft complaints across different demographics groups. The unit of observation is at zipcode-month level. We divide zip codes into two groups based on the sample median of the following demographics: share of Black population, share of Hispanic population, share of Asian population, share of female, and share of people aged between 10-19. The outcome variable is the number of complaints per 1,000 residents. The sample period is February 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table F.4:** Privacy Regulation and Zipcode-level Consumer Sentinel Network Complaints
*By Race/Ethnicity*

| | (1) White | (2) Black | (3) Asian | (4) Hispanic |
|---|---|---|---|---|
| Post × iOS share | 0.035 | −0.116*** | 0.009*** | −0.021*** |
| | (0.032) | (0.018) | (0.002) | (0.005) |
| Zipcode FE | ✓ | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ | ✓ |
| Observations | 1,150,590 | 1,150,590 | 1,150,590 | 1,150,590 |
| R-square | 0.292 | 0.691 | 0.650 | 0.575 |

NOTE.— This table shows the heterogeneous effect of ATT on Consumer Sentinel Network across race/ethnicity by aggregating BIFSG race/ethnicity probabilities based upon the consumer's zip code, first name, and last name to zip code-month level. The unit of observation is at zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Columns 1-4 report the result for Non-Hispanic White, Black, Hispanic, and Asian race/ethnicity probabilities, respectively. The sample period is January 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

## G Classification of Fraud-related Complaints

**Table G.1:** Classification of Complaints using Keyword Search and Zero Shot Learning

| | 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 reports the likelihood of relevant fraud cases by product category based two approaches: keyword search and zero shot learning (ZSL). 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 and about 40% of complaints have narratives.

Table G.1 reports the likelihood of fraud cases by product category based on two approaches: keyword search and zero shot learning (ZSL). The two methods deliver similar rankings, with "Credit reporting, credit repair services, or other personal consumer reports" and "Debt collection" being the top two relevant fraud categories, and "Student loan " 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", "data breach", "not owed". 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 ($3 \times 17$), 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.

Using the manually labelled annotated sample, we verify that the ZSL learning has a satisfactory performance. Setting the threshold score at 0.5 for data-driven fraud complaint, we obtain the following out-of-sample statistics: a F1-score of 0.66, an accuracy of 0.81, a precision of 0.61, and a recall of 0.71.

**Example narratives on data-driven fraud incidents** Below we list a few example complaint narratives from the public CFPB complaints 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."

**Consumer Sentinel Product Codes by Relevant Word Share** The following product codes have a greater than 50% of complaints with at least one relevant keyword listed above: Fake Check Scams, Third Party Debt Collection, Government Imposters, Home Protection Devices, Romance Scams, Tech Support Scams, Job Scams & Employment Agencies, Miscellaneous Investments & Investment Advice, Credit Repair, Real Estate, Credit Information Furnishers, Creditor Debt Collection, Credit Bureaus, and Unemployment Insurance Fraud.

The following product codes had a less than 25% share of complaints with at least one relevant keyword listed above: Prizes, Sweepstakes & Lotteries, Unsolicited Text Messages, Diet Products, Plans & Centers, Gasoline, Prepaid Phone Cards, Utilities, Funeral Services, Home Warranties, Home Appliances and Connected Devices, Children's Products, Health Care: Other Products/Supplies, Tobacco Products, Auto Service & Warranties, Cable & Satellite TV, Vacation & Travel, Insurance (excl. Medical), Medical Treatments & Cures, Home Repair, Garments, Wool, Leather Goods & Textiles, New Auto Sales, Health Care: Other Medical Treatments, Telephone: Other, and Home Furnishings.

**Table G.2:** Privacy Regulation and Zipcode-level CSN Complaints

*Product Categories Relevant for Financial Fraud*

| | (1) Top Fraud Category | (2) Bottom Fraud Category |
|---|---|---|
| Post × iOS share | −0.100*** | −0.002 |
| | (0.025) | (0.023) |
| Zipcode FE | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ |
| Mean outcome var. | 0.193 | 0.542 |
| Observations | 1,150,590 | 1,150,590 |
| R-square | 0.320 | 0.277 |

NOTE.— This table displays the results from estimating Specification 1 after classifying complaints from Consumer Sentinel Network (CSN) by their relevance to ATT. The unit of observation is at zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. Column 1 include complaints from products for which at least 50% of complaints include at least one of the relevant words in the narrative. Column 2 include complaints from products for which less than 25% of complaints include one of the relevant words in the narrative. The sample period is January 2019 to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table G.3:** Privacy Regulation and Complaints

*Product Categories Relevant for COVID-related Fraud*

| Relevant Product | (1) CFPB checking and saving | (2) Consumer Sentinel Network unemployment insurance | (3) Identity Theft government benefits |
|---|---|---|---|
| Post × iOS share | −0.002*** | 0.009*** | −0.047*** |
| | (0.001) | (0.003) | (0.016) |
| Zipcode FE | ✓ | ✓ | ✓ |
| County × Year-month FE | ✓ | ✓ | ✓ |
| EIP Controls | ✓ | ✓ | ✓ |
| Observations | 1,003,590 | 1,150,590 | 1,123,195 |
| R-square | 0.200 | 0.458 | 0.659 |

NOTE.— This table displays the results from estimating Specification 1 by including only products categories relevant for COVID-related fraud. The unit of observation is at zipcode-month level. The outcome variable is the number of complaints per 1,000 residents. We only include complaints from checking and saving accounts and credit cards and prepaid cards for CFPB complaints (Column 1), unemployment insurance fraud for Consumer Sentinel complaints (Column 2), and government benefits and documents fraud for Identity Theft complaints (Column 3). The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. We include zipcode and county × year-month fixed effects in all columns. Standard errors clustered at the state level 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, 2019/01-2022/06, there are 51,105 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.

- Data – Unintentional Disclosure (23.56%)
- Data – Physically Lost or Stolen (4.93%)
- **Data – Malicious Breach (44.87%)**
- **Privacy – Unauthorized Data Collection (1.17%)**
- **Privacy – Unauthorized Contact or Disclosure (11.15%)**
- Identity – Fraudulent Use/Account Access (0.69%)
- Industrial Controls & Operations (0.05%)
- Network/Website Disruption (7.11%)
- **Phishing, Spoofing, Social Engineering (4.48%)**
- Skimming, Physical Tampering (0.29%)
- IT – Configuration/Implementation Errors (0.65%)
- IT – Processing Error (0.63%)

**Regulations violated** When specific laws or regulations are violated by the cyber event, Advisen reports the names of the law and regulations. 5,566 or 10.89% of incidents are recorded to lead to violations of laws or regulations. Among those incidents, the three most frequently violated

regulations are Telephone Consumer Protection Act (TCPA) (73%), Fair Debt Collection Practices Act (FDCPA) (29.6%), and General Data Protection Regulation (GDPR) (3.5%). Note that multiple regulations can be violated in a single event.

Conditional on violating the FDCPA, the companies that experienced the most incidents are Synchrony Bank (57 incidents), Midland Credit Management Inc (52 incidents), Capital One Bank (45 incidents), Bank of American National Association (44 incidents), and Portfolio Recovery Associates LLC (35 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.

## I Description of Dark Web Listings for Data

The research team at Top10VPN periodically scrapes fraud-related listings from active darknet markets, including Nemesis, Kingdom, Empire, Bohemia, and Kraken. We take two snapshots which captures listings in 2020 (July-August) and 2023 (February-March), respectively.[1] Each listing contains the following information: market, listing name, brand, category, listing price, listing currency, units, unit price, and listing url. Examples of listing name include "FRESH PAYPAL ACCOUNT WITH KNOWN BALANCE", "Fully Verified USA COINBASE + BANK LINKED + FULL ACCESS", and "PAXFULL DROP VERIFIED ACCOUNT + FULLZ + EMAIL AND MOBILE ACCESS". There are more than 20 categories of products sold on darknet markets, with the most popular categories being streaming, VPN, payment, shopping, entertainment, crypto, and learning. Popular brands in the darknet markets include NordVPN, Netflix, Paypal, Hulu, and Coinbase. Sometimes a certain quantity of accounts are bundled for sales ("PACK OF 5 CVV/CARDS DETAILS OF U.S WITH GOOD VALIDITY "), and a unit price is calculated for these listings.

We append the listings in 2020 and 2023 in one dataset and construct two variables. First, we determine whether the data being sold is likely generated from consumers' mobile activities. The following categories receive a value of zero: Internet Service Providers, Education (e.g. Masterclass Premium Account), Productivity (e.g. Microsoft Office), Reading, and Communication (e.g. phone Verizon PIN). User activities concerning these categories likely take place via laptops or desktops

[1] More details about these two snapshots can be found at https://www.top10vpn.com/research/dark-web-prices/2020/ and https://www.top10vpn.com/research/dark-web-prices/2023/.

as opposed to apps on mobile devices. Second, we determine whether the listing involves financial information. The finance-related categories include payment, crypto, personal finance, trading, and gambling.

We run a DID regression using listing-level data to investigate the effect of privacy regulation on the price of data sold on the Dark Web. The variable of interest is the interaction term between the post-ATT indicator and the two indicator variables constructed above. We add brand and year fixed effects to all regressions and additionally control for currency fixed effects whenever we include non-USD listings. The outcome variable is the logarithm of the unit price in dollar terms. The results are shown in Table I.1. We find that ATT has substantially increased the price tag of data sold on the Dark Web, especially for data generated from users activities through mobile apps, and for financial information. This finding is consistent with the notion that ATT has lowered the risk of data leakage or breach, leading to a reduced supply of shared/stolen/hacked data on the Dark Web.

**Table I.1:** Privacy Regulation and Dark Web Listing Price

| | listing price per unit (log) | | | |
|---|---|---|---|---|
| | Worldwide | | U.S. | |
| | (1) | (2) | (3) | (4) |
| Post × Mobile footprint | 0.474*** | | 0.520*** | |
| | (0.090) | | (0.080) | |
| Post × Financial info. | | 0.470*** | | 0.322* |
| | | (0.145) | | (0.173) |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Currency FE | ✓ | ✓ | | |
| Mean outcome var. | 3.025 | 3.025 | 2.345 | 2.345 |
| Observations | 3,938 | 3,938 | 2,703 | 2,703 |
| R-square | 0.862 | 0.863 | 0.636 | 0.636 |

NOTE.— This table reports the effect of ATT on the price of data sold in the Dark Web. The unit of observation is a listing on the Dark Web. In Columns 1 and 3, we interact the post-ATT indicator with an indicator for whether the listing sells data that is likely generated from consumers' mobile activities. In Columns 2 and 4, we interact the post-ATT indicator with an indicator for whether the listing sells financial information. The outcome variable is the logarithm of the listing price per unit, winsorized at 2.5% at both tails. We have one snapshots of listings in 2020, and another one in 2023. Columns 1-2 use all the listings, while columns 3-4 use only listings in which the price is quoted in USD. We include brand and year fixed effects in all columns. We additionally include currency fixed effects in Columns 1-2. Standard errors clustered at the brand level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.