NBER WORKING PAPER SERIES

CONSUMER SURVEILLANCE AND FINANCIAL FRAUD

Bo Bian
Michaela Pagel
Huan Tang
Devesh Raval

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2023, Revised August 2025

We thank Simona Abis, Pat Akey (discussant), Hunt Allcott, Tania Babina, Nathan Blascak, Mark Eichorn, Emma Fletcher, Lewis Kirvan (discussant), Samuel Kruger (discussant), Kai Li, Markus Mobius, Jordan Nickerson, David Rothschild, Dave Schmidt, James Thomas (discussant), Andreanne Tremblay-Simard (discussant), Guillaume Vuillemey (discussant), Jonathan Wallen, Brett Wendling, Paul Witt, Con-stantine Yannelis, Xiaoyun Yu, 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, CFPB Research Conference, Future of Financial Information Conference, Georgia Tech - Atlanta Fed House-hold Finance Conference, Goethe University, Household Finance Brownbag Series, IMF, Microsoft Research, NBER SI (Household Finance, Digital Economics and Artificial Intelligence), NFA, Peking University, Remin University, UBC Winter Finance Conference, University College London, UW-Madison Junior Finance Con-ference, and University of Innsbruck. Yile Chen, Rohan Joseph, Yiming Ma, Niels Wagner, and Andy Xin 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, any of its Commissioners, or the National Bureau of Economic Research. Bo Bian acknowledges financial support from the Social Sciences and Humanities Research Council of Canada (Grant Number: 430-2023-00553). We have nothing else to disclose. First Version: November 1, 2022.

Consumer Surveillance and Financial Fraud
Bo Bian, Michaela Pagel, Huan Tang, and Devesh Raval
NBER Working Paper No. 31692
September 2023, Revised August 2025
JEL No. G5

ABSTRACT

In today's digital economy, firms continuously collect, store, share, and sell personal data, exposing customers to risks of financial fraud. Leveraging Apple's App Tracking Transparency policy as a natural experiment, we show that restricting data tracking and sharing significantly reduces consumer fraud complaints, particularly those involv-ing personal information misuse. Effects are stronger in areas dominated by firms with risky data practices and coincide with a decline in dark web discussions and higher prices for sensitive data. By tracing effects along the fraud supply chain, our find-ings suggest that data regulations can benefit consumers by constraining the flow of exploitable information.

Bo Bian
University of British Columbia
bo.bian@sauder.ubc.ca

Michaela Pagel
Washington University in St Louis
Olin Business School
and NBER
mpagel@wustl.edu

Huan Tang
University of Pennsylvania
huan.ht.tang@gmail.com

Devesh Raval
Federal Trade Commission
devesh.raval@gmail.com

1 Introduction

We live in a commercial surveillance economy. Mobile phones are essential to daily life and enable companies to monitor consumers' private lives almost continuously.[1] The collection, tracking, sharing, and selling of personal data expose individuals to identity theft, deception, manipulation, and other abuses by fraudsters. The rise of AI is amplifying these risks, enabling increasingly sophisticated schemes—from deepfake calls to fabricated documents. Lax privacy practices, security breaches, and a thriving data market have made it easier for fraudsters to target consumers, while limited public awareness and costly consent offer little protection. Each year, millions of fraud complaints are filed, with the Federal Trade Commission (FTC) estimating that over 10% of US adults fall victim annually and reported losses reached nearly $12.5 billion in 2024.[2] In response, policymakers have tightened data protection standards, with the European Union (EU) strengthening rules in 2018, the state of California (CA) following in 2020, and the FTC, Consumer Financial Protection Bureau (CFPB), and the US Congress considering new federal laws and regulations.[3]

Against this backdrop, we ask whether an industry-led initiative that substantially limited the tracking and sharing of personal information across mobile applications and websites can reduce financial fraud. In doing so, we offer novel evidence linking privacy protections to fraud outcomes across multiple layers of the data ecosystem, from consumer victimization to firms' data practices to illicit data markets. Our findings inform the broader policy debate by quantifying one benefit of stronger privacy protections that restrict how companies collect, use, and share consumer data. By showing that such restrictions can generate downstream spillover benefits from reduced fraud, our analysis highlights the broader societal value of regulating data practices, particularly for firms engaging in extensive consumer surveillance.

[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]See FTC press release: "New FTC Data Show a Big Jump in Reported Losses to Fraud to $12.5 Billion in 2024". Other estimates suggest even higher prevalence, with surveys indicating up to half of Americans have experienced fraud (Anderson, 2019; Huff et al., 2010; DeLiema et al., 2017).

[3]For the FTC, see press release "FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices". For the CFPB, see "Required Rulemaking on Personal Financial Data Rights". For the US Congress, see "Senate's summary of the American Privacy Rights Act of 2024".

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. Introduced by Apple on April 26, 2021, ATT requires all mobile applications (apps) to obtain explicit user permissions before accessing users' mobile identifiers. With such identifiers, one can track consumers across apps or websites owned by other companies and share their data with third parties. Crucially, ATT reversed the default from *opt-in* to *opt-out*: users are now automatically excluded from tracking unless they actively grant permission. As of March 2022, only 17% of iOS users had opted in (Kraft et al., 2023). As a result, ATT greatly limited the volume and scope of personal data collected and shared across mobile apps, websites, companies, and platforms. This reduced the availability of high-quality real-time data for fraudsters to exploit.

We exploit the fact that ATT affects only iOS users and not Android users. We first measure to what extent people in any given zip code are affected by ATT using granular foot traffic data from SafeGraph to calculate pre-ATT shares of iOS users out of all smartphone users. In turn, we examine the effects of the ATT policy on fraud activities in zip codes with different shares of iOS versus Android users. In our analysis, we use three different datasets of consumer fraud complaints in the US.[4] We mainly rely on CFPB's public complaint database, which is especially relevant for studying financial fraud because it centers on disputes involving personal information and financial institutions. To broaden the scope, we supplement our analysis with two proprietary datasets from the FTC. First, the FTC's Identity Theft database, which contains a broad set of all identity theft complaints. And second, the Consumer Sentinel Network database, a comprehensive repository aggregating fraud and consumer complaints from federal and state agencies, private firms, and other sources, providing near-universal coverage of all types of fraud in the US.

Our results indicate that limiting the tracking and sharing of personal information sub-

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

stantially reduces consumer complaints. We find that a 10 percentage point ($\sim$ one standard deviation) increase in the share of iOS users in a zip code leads to a 6.1% reduction in the number of CFPB complaints post-ATT. Given that 83% of iOS users opted out of tracking after ATT, a 10% increase in consumers opting out of tracking translates to a 7.3% reduction in CFPB complaints. Our identification assumes that, absent ATT, complaint trends would have evolved similarly across areas with high and low iOS penetration. While areas with more iOS users may differ in baseline complaint rates or consumer behavior, our dynamic Difference-in-Differences (DiD) estimates show no evidence of differential pre-trends and reveal a negative trend emerging shortly after ATT and intensifying over time, consistent with rapid exploitation of personal data and prompt consumer reporting.

To sharpen identification, we study a similar but opposite policy shock—the 2017 repeal of the Federal Communications Commission (FCC)'s broadband privacy rules, which permanently lifted impending restrictions on how Internet Service Providers (ISPs) collect and share consumer data. ISPs occupy a uniquely powerful position in the data ecosystem, with the ability to track virtually all unencrypted online activity across websites, apps, and devices—visibility that is difficult for consumers to avoid. The repealed rules, adopted in 2016 but never implemented, would have required ISPs to obtain explicit opt-in consent before using or sharing sensitive customer information. The repeal eliminated these pending requirements and barred the FCC from issuing similar rules in the future. In contrast to the ATT policy, which restricted mobile data flows, the repeal expanded them, enhancing ISPs' ability to monetize consumer data with minimal consent or oversight.

We measure the invasiveness of ISPs' data collection practices using two metrics: the number of unique data items they collect and the number of third-party platforms they exchange data with. By combining these metrics with geographic variation in ISP presence, we can capture local exposure to high-risk ISPs. We then conduct a DiD analysis around the repeal. After the repeal, we find that areas served by ISPs with poor data protection practices experienced an increase in financial fraud. The dynamic DiD plot confirms that this effect only appears after the repeal, with no significant changes before the repeal. Like our analysis based on ATT, this setting also links changes in data practices to shifts in fraud, lending additional credence to our narrative by showing that it operates in both directions.

We corroborate our main findings using proprietary data from the FTC, which provides uncensored coverage of all CFPB complaints, a complete census of fraud and identity theft cases from the Consumer Sentinel Network and Identity Theft databases, and information on reported dollar losses from fraud. The uncensored data confirms that censoring in the public CFPB database does not drive our results. The identity theft records show significant post-ATT declines consistent with reduced misuse of personal data. Broader complaints to the Consumer Sentinel Network exhibit smaller effects, as they include complaints on non-finance products and services and cases not triggered by data breaches. Importantly, the unique dollar loss data from Consumer Sentinel complaints suggest that ATT curtails the severity as well as frequency of fraud. A simple back-of-the-envelope calculation suggests that ATT reduced consumer fraud losses by roughly $274 million annually.

The second part of our analysis explores the mechanisms behind how greater privacy protections reduce fraud. Providing direct evidence on how ATT leads to reduced fraud is exceptionally challenging. The fraud economy operates through opaque, unregulated channels where neither the flow of illicit data nor the actors involved are directly observable. Some personal and financial information is collected covertly, trafficked through hidden online markets, and monetized in ways that rarely leave transparent traces. Even when consumers or firms are victimized, the pathway from compromised data to fraudulent transactions is typically invisible in public or administrative records. We introduce several new databases to piece together complementary evidence that trace the effects of ATT from downstream victims, through data-collecting firms targeted by ATT, to the illicit markets supplying stolen or leaked data.

First, we dig deeper into the most visible point in the chain of events: downstream consumers and their complaints. Using keyword searches and machine-learning classifications of complaint narratives, we separate fraud types that depend on compromised personal or financial data—such as account takeovers and identity theft—from those that do not. We find that ATT-induced declines are concentrated in data-dependent categories while unrelated categories like mortgage or student loan show little change.

We next examine financial institutions' role as intermediaries in the illicit data supply chain, where the sensitive consumer information they collect and store can be exploited by upstream fraudsters to fuel downstream fraud victimization and drive up complaints. Financial institutions with consumer-facing, popular mobile apps are particularly exposed, as these apps facilitate extensive data collection and sharing. We find that, after ATT, these firms see significant declines in cyber incidents, especially those caused by data breaches and leading to violations of consumer laws. This is consistent with Ramadorai et al. (2025), who find that data extraction activities are linked to increased cybersecurity risks. We also observe corresponding reductions in data-driven consumer complaints, with no comparable changes in unrelated incidents.

To connect our baseline results more directly to information intermediaries in the fraud supply chain, we study whether ATT's effects are largest in places where local midstream entities—financial institutions and ISPs—collect and transmit more consumer data. We measure exposure to poor data practices by combining local bank deposit and ISP market shares with the intensity of their data collection or sharing practices. The latter is captured by the number of unique data items they collect and the presence of third-party data-sharing platforms connected through their mobile apps. A triple difference design, interacting our DiD term with these exposures, reveals that the drop in fraud complaints is significantly larger in markets where high-risk ISPs and banks are more prevalent. This analysis ties midstream data collection and sharing practices to the size of the downstream fraud reductions, consistent with ATT disrupting channels through which data flows to fraudsters.

Finally, we move upstream and examine the illicit markets that trade compromised personal and financial data, focusing on both dark web forum discussions (the conversation layer) and marketplace listings and prices (the transaction layer). Using novel data on 38 dark web forums and tens of millions of posts, we find sharp post-ATT declines in iOS/Apple-related discussions and smaller but still significant declines in broader topics that depend on personal or financial data. These effects are estimated relative to a control group of illicit trade categories (e.g., drugs, human trafficking, weapons) that operate in segmented markets with minimal overlap in actors or infrastructure with the data-driven fraud economy. Complementary analysis of dark web marketplace listings shows that prices for mobile-sourced and financial data rise after ATT, consistent with reduced supply. While these upstream markets are inherently opaque and our evidence is not definitive, the results align with ATT constraining the availability of compromised data, thereby limiting the supply chain that feeds downstream fraud.

Taken together, our analysis provides a rare account of how privacy regulation reshapes the entire fraud supply chain. ATT reduced downstream consumer fraud in categories dependent on compromised data; these reductions were strongest for firms and localities with greater pre-ATT exposure to risky data practices; and upstream, ATT disrupted the illicit markets that supply this data. By linking visible consumer outcomes to the hidden supply chain of illegally obtained information, we provide the first evidence tracing the full chain through which data privacy initiatives may curb fraud.

Literature Review. We first contribute to the literature on data privacy concerns. Surveys show that identity theft is the most cited reason for privacy worries among US households (Armantier et al., 2021).[5] We complement this perception-based evidence by showing that weak data collection and sharing safeguards can cause higher rates of identity theft and financial fraud. This shifts the discussion from stated concerns to measurable harm, and relates to work on the consequences of excess data collection by financial institutions, including discrimination (Fuster et al., 2022), behavioral manipulation (Acemoglu et al., 2025), and privacy intrusion (Tang, 2019).

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

Relatedly, our paper contributes to the literature on the economic implications of data privacy regulations. Prior work has largely examined the European General Data Protection Regulation (GDPR), connecting it to changes in web traffic (Goldberg et al., 2024), app entry and exit (Janssen et al., 2021), VC financing (Jia et al., 2021), firms' abilities to collect, monetize, store, and use consumer data (Aridor et al., 2023; Bessen et al., 2020; Demirer et al., 2024; Peukert et al., 2022), as well as the visibility and quality of firms' data collection disclosures (Ramadorai et al., 2025).[6] Other studies have looked at open banking (Babina et al., 2025) and the effect of the California Consumer Privacy Act on mortgage lending (Doerr et al., 2023) and firm risk (Wu, 2023). A small but growing set of papers focuses on Apple's privacy initiatives, such as the privacy label policy (Bian et al., 2021) and the ATT framework (Kesler, 2022; Cheyre et al., 2023; Bian et al., 2024; Abis et al., 2025). We further extend the literature by examining whether stronger privacy protections—in our case, a standardized and uniform opt-out consent framework that led most users to decline tracking—can curb financial fraud.

Our research leverages consumer complaints data in a new way. We link patterns of consumer victimization to variation in data collection and sharing practices, both cross-sectionally and over time. In contrast, existing work uses CFPB data to study the service quality and discrimination of financial institutions, especially mortgage providers (Begley and Purnanandam, 2021; Huang et al., 2024; Dou and Roh, 2024; Li, 2023; Mazur, 2024; Jou et al., 2024). Haendler and Heimer (2025) employ confidential CFPB data to investigate racial disparities in restitution for disputed financial services. Our use of the data is different from prior work in two ways. First, we dig into the complaint narratives and product categories to identify fraud that may originate from digital surveillance. By complementing it with confidential data obtained from the FTC covering a broader range of consumer problems, our paper draws on one of the most comprehensive publicly visible records of consumer fraud activity. Second, we extract the identities of financial institutions from the

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

complaints and construct novel measures of their data-collection and data-sharing practices, which can be used to study a broader set of questions around data handling and outcomes of financial institutions or their consumers.

Our study also relates to the forensic finance literature that Griffin and Kruger (2024) surveys comprehensively. An important line of this work investigates how traditional information intermediaries—such as banks (e.g., Wang et al., 2010; Gao et al., 2020) and financial advisors (e.g., Egan et al., 2019, 2025)—mitigate or amplify misconduct that victimizes consumers, investors, or shareholders.

We contribute to this literature in two key ways. First, alongside banks, we identify ISPs as emerging and influential information intermediaries that, through their role in the digital economy and persistent consumer surveillance, shape the flow and use of personal data. Second, while prior studies focus on intermediaries that shape what downstream victims know, we highlight those that expand what upstream fraudsters know, making it easier for them to target and deceive consumers. This distinction yields clear, actionable policy implications: regulating the data practices of these intermediaries may produce spillover effects that reduce fraud. On the data and methods side, we make progress in addressing long-standing empirical challenges in this literature, where fraud measures based on regulatory actions, litigation, or restatements capture only a subset of actual misconduct, often with delays, and suffer from false positives, omissions, and inconsistent definitions (e.g., Yu, 2013; Amiram et al., 2018; Griffin and Kruger, 2024). In contrast, we leverage granular, individual-level complaints from the CFPB and FTC, supplemented with scraped dark net activity, to observe fraud as it emerges in near real-time.[7]

2 Institutional Background

In this section, we first provide background information on how the collection and sharing of personal data can lead to financial fraud. We then describe an industry-led privacy initiative

[7] There is growing use of even richer transaction-level data (e.g., Griffin et al., 2023, 2024; Griffin and Kruger, 2024).

implemented by Apple and its relevance for data-driven fraud.

2.1 Characteristics of Data-Driven Fraud and Enforcement Activities

Modern financial fraud increasingly relies on personal data that is both real-time and linkable across sources—two characteristics that have triggered regulatory enforcement and are directly disrupted by our policy shock.

First, fraud schemes depend heavily on access to real-time personal data. Unlike static identifiers such as Social Security numbers, dynamic information—current addresses, recent transactions, active credentials, and live location—enables fraudsters to act quickly and exploit vulnerabilities before detection. Stolen credit card numbers can be used immediately for unauthorized purchases; newly breached login credentials can be tested in credential-stuffing attacks before passwords are reset; and geolocation data can help time account takeovers when victims are away from home. In contrast, delayed or outdated data rapidly loses value: consumer behavior changes, digital traces expire, and institutions implement protective countermeasures. The time-sensitive nature of financial fraud makes real-time data especially valuable for carrying out successful attacks.

Second, fraud also depends on the ability to link and merge data across different sources and time periods. Persistent identifiers, especially device-level ones like Apple's Identifier for Advertisers (IDFA), allow data brokers, advertisers, and malicious actors to combine disparate records and construct detailed behavioral profiles. Even minimal sets of linked data, such as names, dates of birth, addresses, and bank account numbers, can enable impersonation, account takeovers, or targeted social engineering attacks. Linkability across data sources allows fragmented information to be consolidated into a unified profile, thereby improving the targeting and persuasiveness of fraudulent schemes.

These two characteristics are central to recent enforcement actions by US regulators, as illustrated by the two examples below. In the case against Ideal Financial Solutions Inc., the FTC charged the company with purchasing Social Security numbers and bank account information from payday-loan applicants and using the merged data to debit consumer accounts without consent. The FTC also settled with the data brokers who had sold the information.[8] In another case, the Department of Justice settled criminal charges against Epsilon Data Management LLC, a major data broker that sold consolidated lists of financially vulnerable consumers to fraud rings promoting sweepstakes, astrology, and psychic scams. These schemes depended on timely and detailed data to identify and repeatedly target susceptible individuals.[9]

More importantly, the FTC has recently turned its attention to mobile location data, which can be used to profile consumers and pursue time-sensitive fraud opportunities. The agency settled with data brokers X-Mode and Outlogic and is currently litigating against Kochava for selling precise geolocation information.[10] When combined with mobile device IDs and home addresses, this data can reveal where people are in real time, how they move, and what behaviors they engage in, and so has the potential to enable highly personalized and well-timed fraud. Kochava's widely promoted "Household Mapping" is one example of such data.

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, which reduced the availability of high-quality data for fraudsters.

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. These mobile identifiers were the primary way that allowed apps and data brokers to combine user data from different sources. This feature, dubbed "App Tracking Transparency (ATT)", grants users both greater and easier control over their data. An

[8] 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".

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

[10] 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".

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 the user's mobile identifier. Importantly, the opt-in design and the uniform consent prompt apply to all firms that serve iOS users, unlike firm-specific cookie banners. 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).[11] Bian et al. (2021) document a sharp and negative stock market reaction 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).

ATT breaks down a core mechanism through which personal data is collected and circulated in the mobile ecosystem. Before ATT, tracking was enabled by default, allowing data collected by an app to flow far beyond the developer to a dense network of third-party firms—including ad networks, data brokers, and analytics providers—many of which had no direct relationship with the user and operated with limited oversight. This structure created multiple downstream copies of sensitive data, increasing the risk of resale, leakage, or misuse. ATT interrupts this chain by requiring user permission for cross-app tracking, thereby reducing both the volume of personal data collected and its broad distribution across the digital supply chain.

This disruption also directly targets the two characteristics of personal data that make it valuable for financial fraud: its real-time nature and its linkability. ATT limits the availability of up-to-date behavioral, transactional, and location data by curbing passive background tracking, reducing the chances that fraudsters can act before detection. At the same time, by blocking access to Apple's device-level identifier, ATT undermines the ability to stitch together data across apps and time, breaking the coherence needed to build detailed user

[11]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/`.

profiles. In doing so, the policy weakens both the speed and precision of fraud operations built on mobile-sourced data.

3 Data

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 zip-code level. We construct this variable using data from SafeGraph, a company that tracks foot traffic using GPS location data from mobile devices. This data has information on daily visits to 6 million points of interest across the US. For each point of interest, SafeGraph reports a rich set of information, including time-invariant information such as the POI's operating company (e.g., a certain bank or retail store), NAICS code, postal code, and time-varying information, such as monthly visit or visitor counts. Crucial for our study, SafeGraph also reports whether visitors use Android vs. iOS devices. SafeGraph aims to provide a representative sample of US consumers. Li et al. (2024) document a near-perfect correlation (>0.97) between the number of sampled devices and census population, for both urban and rural areas, and minor sampling biases among a number of demographic categories such as age, gender, and moderate income, with less than 5% under- and over-representation.

For the purpose of our analysis, we aggregate all visits 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. The data 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 our measurement.

3.2 Financial Fraud Data: Consumer Complaints

Consumers who suspect they have been subjected to financial fraud can complain to a range of authorities, both to obtain redress and to alert other consumers to potential risks. Consumers can direct such complaints to state and federal regulatory bodies, as well as to private entities such as the Better Business Bureau (BBB). Our principal measure of financial fraud is derived from publicly available data on complaint submissions to the Consumer Financial Protection Bureau (CFPB), the federal agency with a mandate to address consumer grievances involving financial products and services.

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.

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. The major product categories include "credit reporting," "debt collection," and "mortgage." They can also provide more specific information by selecting a "subproduct" or "subissue" if applicable.[12] Each reported issue has the potential to be relevant for 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

[12] 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".

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 removing all personal information. We analyze the product, issue, subproduct, subissue, and narrative statements to gain insights on how the specific details and circumstances of each complaint relate to financial fraud originating from tracking and sharing personal data.

The CFPB also reports the company being complained about. However, sometimes consumers and even firms themselves may be unable to trace the source in a data-related case. For example, a consumer with accounts at multiple banks may have personal information stolen from one, leading to a fraudulent credit line being opened in their name. The consumer may first learn of this when reviewing their credit bureau record, prompting them to file a complaint against the bureau over inaccurate credit report information rather than the unidentified bank where the breach occurred. As a result, roughly half of the complaints are directed toward the three credit bureaus over issues like inaccurate information on credit reports or improper use of credit reports. We thus primarily rely on aggregate complaint counts at a granular locality level rather than company-specific ones as the outcome variables.

We supplement complaints from the CFPB with the Consumer Sentinel database, which combines complaints from several sources, including the CFPB, and the FTC's Identity Theft database. We discuss these datasets further in Subsection 4.4.

3.3 Other Data Sources and Measures

Data Collection and Sharing Practices. We create novel measures of the data collection and sharing practices of financial institutions and ISPs. Prior work has employed conceptually similar measures of data extraction activities, drawing on privacy policies and website cookies of public firms (Ramadorai et al., 2025). Our approach parallels these efforts but adapts the metrics to the mobile setting and to specific data intermediaries. Information on ISP data collection practices comes from Apple's App Store privacy labels, which provide standardized disclosures of firms' data collection practices. We scraped and parsed these labels to count the number of unique data items collected (details in Appendix Section A.2). We measure data sharing activities using data on the third-party Software Development Kits (SDKs) usage from Apptopia. SDKs are pre-built software components that enable specific functionalities, including data analytics and targeted advertising, within an app (details in Appendix Section A.3). We identify all SDKs that facilitate data sharing across firms following the approach in Bian et al. (2024). We obtain ISP market-share data from the FCC's Broadband Data Collection, which reports provider coverage and market shares at the local level (details in Appendix E).

We construct equivalent measures for financial institutions appearing in the CFPB complaint database, using the same scraped privacy label and SDK installation data, and obtain their local market shares from the FDIC's Summary of Deposits (SOD). Appendix Section A.2 provides details on how we identify the mobile presence of banks.[13]

Cyber Events and Dark Net Activities. We turn to Advisen to identify data breaches and other cyber security events at the firm level (details in Appendix G). Finally, we use novel data on Dark Web forum discussions and listings from the Cambridge Cybercrime Centre (CCC, details in Appendix H) and Top10VPN (details in Appendix I).

3.4 Summary Statistics

The main regression sample consists of a balanced panel of the public CFPB complaints at the zip code level, spanning from January 2019 to June 2022. We impute zeros for zip codes with no reported complaints and exclude outlier zip codes using a multi-criterion procedure detailed in Appendix D, which flags persistent irregularities in complaint patterns. As shown in Appendix D, these zip codes exhibit abnormal and recurring volatility, such as frequent

[13] Recent studies have explored the effects of mobile banking penetration on local competition (Haendler, 2022; Koont, 2023; Jiang et al., 2025), financial inclusion (Jiang et al., 2025), bank franchise value (Koont et al., 2024), as well as local small business lending and economic growth (Haendler, 2022). We contribute by showing that, conditional on mobile penetration, lax data practices by banks increase financial fraud.

spikes far beyond typical levels, that standard winsorization techniques will not remove.[14]

Table 1 presents summary statistics. For the balanced sample around ATT, approximately 26% 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 0.05, indicating that around 5 residents out of every 100,000 file a complaint in a given 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 rich spatial variation in complaint rates. For the balanced sample around the FCC Broadband rule (see discussions on the institutional background in Section 4.3 and Appendix E), both the probability and total number of complaints are lower. Examining local ISPs' data collection and sharing intensities, we also observe substantial variation (see Appendix Figure A.3 for the distributions at ISP level).

Using foot traffic data, we find that the average iOS share across US zip codes is 46%, close to Statista's estimate that the national iOS share fluctuated around 50% during 2019–2022.[15] This share varies substantially across zip codes, with a share of 39% for the 25th percentile zip code and 52% for the 75th percentile. Figure 2b further illustrates the geographical variation across the US. Consistent with DeviceAtlas estimates, iOS adoption is higher in the Southern states and along the Northeast Corridor.[16]

4 Effect of Privacy Rules on Financial Fraud Complaints

4.1 Regression Specification

We begin by examining the extensive margin using a linear probability model:

$$\mathbb{1}\{Complaints\}_{z,t} = \alpha_z + \alpha_{state/county(z),t} + \beta\, iOS\ Share_{z,pre-ATT} \times Post_t + \varepsilon_{z,t} \qquad (1)$$

where the dependent variable $\mathbb{1}\{Complaints\}_{z,t}$ equals one if there is at least one complaint in zip code z during month t, and zero otherwise.

[14] For example, outlier zip codes in North Carolina exhibit multiple spikes ranging from 200 to 800 complaints per 1,000 residents.

[15] See https://www.statista.com/statistics/266572, the 4% gap may reflect that our measure better captures usage frequency, and iOS and Android users may spend different amounts of time on their devices, making our measure a usage-weighted iOS share.

[16] See https://deviceatlas.com/blog/mobile-os-popularity-by-us-state.

To move beyond the extensive margin, we next estimate the average treatment effect on the total number of complaints using a Poisson model. Specifically, the regression specification is:

$$Complaints_{z,t} = \exp\left(\alpha_z + \alpha_{state/county(z),t} + \beta \, iOS\ Share_{z,\ pre-ATT} \times Post_t\right) \varepsilon_{z,t} \tag{2}$$

where $Complaints_{z,t}$ is the number of complaints in zip code z and month t. We scale it by population of the zip code, using the 2020 Census data, to obtain complaints per 1,000 residents, and winsorize the measure at the top and bottom 1%. We estimate Equation (2) using Poisson pseudo-maximum likelihood (PPML) via `ppmlhdfe` in Stata. This approach is well-suited for proportional count data, as it naturally accommodates the non-negative and skewed distribution of the outcome while allowing for high-dimensional fixed effects.

In both regression equations, we use the variable $iOS\ Share_{z,pre-ATT}$ to capture the variation in exposure to ATT. The $iOS\ Share_{z,pre-ATT}$ represents the average *pre*-ATT iOS share of users at the zip code level, calculated over the nine quarters preceding the policy's introduction.[17] 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 on April 26, 2021. The coefficient β on the DiD term, $iOS\ Share_{z,pre-ATT} \times Post_t$, captures the differential change in consumer complaints between zip codes with higher versus lower iOS device shares.

To account for time-invariant characteristics that contribute to fraud, we include zip code fixed effects. Additionally, we incorporate state-by-year-month or county-by-year-month fixed effects, denoted as $\alpha_{state(z),t}$ or $\alpha_{county(z),t}$, to control for time-varying confounders at the state or 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 checks, we aggregate the data to larger units, such as the quarter-month level or the county-month level, and conduct similar DiD analysis.

[17] We use the *pre*-ATT measure because SafeGraph's foot traffic data rely on location tracking from mobile devices. Since ATT may reduce the precision of these measurements, using the pre-policy period avoids measurement error caused by the policy itself.

4.2 Baseline Results

Table 2 presents the regression results using the CFPB complaints database. At the extensive margin, Columns 1 and 2 show a significant, negative coefficient on the interaction term $iOS\ Share_{z,pre-ATT} \times Post_t$. This estimate implies that, within a given county and month, zip codes with a higher proportion of iOS users experienced a larger decline in the probability of consumer complaints following the implementation of the ATT policy, relative to zip codes with lower iOS user shares. In Column 2, the coefficient is approximately -0.059 when controlling for county-specific shocks, indicating that a zip code with a 10% ($\approx$ one standard deviation) higher iOS share experienced a 0.59 percentage point reduction in the probability of a complaint, equivalent to about 2.4% of the mean.

The extensive margin analysis, however, does not capture changes in complaint intensity. Moving to Columns 3 and 4, we examine the total number of complaints per 1,000 residents. We again find negative and statistically significant coefficients at the 1% level. The same one standard deviation increase in iOS share translates into a 5.4%–6.1% decline in complaints per 1,000 residents.[18]

The magnitude of the effect is larger using the total number of complaints because it captures both whether any complaints occur and the number of complaints filed, conditional on any complaints filed. Given the observed opt-out rate of 83%, our estimates imply that complaints to the CFPB could decline by roughly 6.5%-7.3% if 10% of mobile app users were to disallow data tracking (calculated as 5.4%/0.83 or 6.1%/0.83). While the implied magnitude appears large, it reflects a relative drop within an overall upward trend in CFPB complaints. Broader forces, such as growth in digital transactions, expanding attack surfaces, and increasingly sophisticated criminal tools, continue to push fraud upward, so even sizable relative effects may not be reflected in the aggregate trends.

[18]In Poisson models, coefficients can be interpreted as semi-elasticities. The percentage change in the dependent variable for a change Δx in the regressor is given by $\left(e^{\beta \cdot \Delta x} - 1\right) \times 100$. In our case, the coefficient in Column 3 is 0.548 and we evaluate the effect of a one standard deviation increase in iOS share, yielding $\left(e^{-0.548 \times 0.108} - 1\right) \approx -5.4\%$.

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 trends of consumer complaints. For example, ownership of Apple products predicts higher income and better education (Bertrand and Kamenica, 2023), 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 3, 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.[19] 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 probability of complaints in Panel a of Figure 3, we observe that, prior to the introduction of ATT, the coefficients for all quarters are not statistically significantly different from zero. However, there is a clear negative post-trend, suggesting a decline following the policy's implementation in zip codes with larger shares of iOS users. Examining complaints per 1,000 capita in Panel b of Figure 3 reveals a similar pattern. We also estimate the monthly dynamic treatment effects and report the results in Appendix Figure J.1. While the estimates are noisier at the monthly frequency, the overall pattern is consistent with the quarterly figures.

These dynamic figures show that the effects of ATT manifest within two months after ATT's implementation. One might expect a longer lag for increased data protection to lead to less fraud and then less reports by consumers about fraud. However, in today's surveillance economy, stolen credentials and personal information can be monetized almost immediately

[19] As discussed in Subsection 4.5, we also extend the sample through the end of 2023 as well as 2024 and find larger, statistically significant DiD coefficients (Appendix Table J.6 Columns 5 and 6). We use an earlier cutoff in our main analysis because our foot-traffic data ends then, and the pre-ATT iOS share may no longer accurately capture treatment intensity as the post-ATT period extends further out.

through well-organized fraud-as-a-service markets and automated attack tools.[20] On the victim side, expanded use of real-time transaction monitoring, mobile banking alerts, and identity-theft protection services means that consumers detect suspicious activity and file complaints more timely than in the past.[21]

To validate the short timeline from a data breach or leakage to a reported complaint, we examine CFPB consumer complaints following the 2017 Equifax breach and detect a surge in Equifax-related complaints within three months (Panel a of Appendix Figure J.2). Previous reports by the FTC (Synovate, 2003) show that even twenty years ago, more than half (one-third) of victims discovered the misuse of their personal information within a month (week). The 2021 National Crime Victimization Survey (NCVS) reports that 44% of identity theft cases were discovered by victims within a day while another 33% were discovered within a month (Panel b of Appendix Figure J.2).[22] These statistics are consistent with fraudsters quickly acting upon obtaining leaked or hacked data to maximize returns as well as consumers detecting and reporting fraud quickly.

4.3 A Reverse Shock: Repeal of the FCC's Broadband Privacy Rules

To further address identification concerns, we examine a second policy shock that also targets the collection and sharing of consumer data but reduces data protections. Specifically, we study the 2017 repeal of the FCC's broadband privacy rules, which granted ISPs greater latitude in consumer surveillance.

ISPs occupy a uniquely central position in the data economy due to their ability to observe user activity across websites, apps, and devices, unlike platforms constrained to logged-in or embedded environments. Their data collection is persistently tied to households and is difficult for consumers to avoid, given limited broadband competition. These characteristics make ISP data especially valuable for profiling and monetization. Bian et al. (2024) indeed

[20]Dark web marketplaces facilitate near real-time resale of compromised data, while credential-stuffing and account-takeover attacks can be launched within hours of a breach. Emerging AI-driven tools further accelerate these processes. See the discussion on fraud-as-a-service here and AI tools here.

[21]See Mckinsey's discussion here and Thomson Reuters' report here.

[22]The full NCVS-Identity Theft Supplement can be found here.

documents that ISPs are among the most connected firms in the data-sharing ecosystem.

In 2016, the Federal Communications Commission (FCC) adopted the Broadband Privacy Rules, which were set to require Internet Service Providers (ISPs) to obtain explicit customer consent before using or sharing sensitive information such as web browsing history, app usage data, and precise geolocation for advertising or other non-service purposes. The rulemaking process started in 2015, when the FCC reclassified broadband as a telecommunications service under Title II, shifting privacy oversight from the FTC to the FCC and signaling that stronger privacy restrictions were forthcoming. From 2015 until the repeal in 2017, ISPs operated under the shadow of these pending rules, even though the rules themselves never fully took effect, which constrained their ability to expand data collection practices.

In March 2017, Congress repealed the FCC's privacy rules under the Congressional Review Act (henceforth "the Repeal"), and President Trump signed the repeal into law. This action prevented the FCC from ever enacting "substantially similar" rules, and so eliminated potential regulatory overhang and uncertainty about future privacy regulation of broadband. It thus granted ISPs greater flexibility to collect, share, and monetize customer data under the FTC's more general consumer protection framework, without the strict consent requirements envisioned by the FCC.[23] A detailed description of the background and timeline of the Repeal is provided in Appendix E.

This reverse shock complements the main ATT setting by allowing us to causally examine how an expansion in consumer surveillance, rather than a contraction, affects financial fraud. To measure a locality's exposure to the Repeal, we construct two metrics that combine the local market shares of ISPs with the invasiveness of their data practices. Specifically, we consider (1) the number of unique data items collected (# Data Types Collected) and (2) the

[23]The Congressional Review Act includes a "substantially similar" clause (5 U.S.C. §801(b)(2)) that prohibits regulatory agencies from reissuing any future rules that closely resemble the repealed regulation, unless explicitly authorized by subsequent legislation. As a result, the Repeal not only removed the immediate threat of stringent privacy requirements but also credibly signaled to ISPs a significantly lower likelihood of facing similar restrictions in the future.

number of third-party platforms with which ISPs share or transmit data (# Data SDKs).[24] To capture exposure to ISPs with invasive data practices, we aggregate the two metrics across ISPs within each zip code. Specifically, we first weight ISPs by their local market shares, measured at the census block level and available at semi-annual frequency. We then aggregate from census blocks to zip-code level using the population shares of blocks within each zip code, as defined in Equation (3) and Equation (4).[25]

$$\text{ISP Exposure}_z^{(m)} = \sum_{b \in \mathcal{B}(z)} \underbrace{\frac{p_b}{P_z}}_{\text{block population weight}} \left(\sum_i s_{ib} \, I_i^{(m)} \right), \tag{3}$$

where i indexes ISPs, b indexes census blocks, and $\mathcal{B}(z)$ denotes the set of census blocks in zip code z. Data collection and sharing intensities, $I_i^{(m)}$, are defined as

$$\begin{aligned} I_i^{(\text{Collection})} &= \#\,\text{Data Types Collected}_i, \\ I_i^{(\text{Sharing})} &= \#\,\text{Data SDKs}_i. \end{aligned} \tag{4}$$

We then compute the average of this exposure measure over the pre-Repeal period (January 2015 to October 2016).

We employ a DiD design around the Repeal, between January 2015 and December 2018, and report the results in Table 3.[26] Panel a uses the exposure measure based on ISP data collection intensity, while Panel b focuses on data sharing intensity. We find that, following the Repeal, areas served by ISPs with weaker data protection practices experienced an increase in complaints. For example, in Panel a, a one-standard-deviation increase in local exposure based on data collection practices is associated with a 2.4%-2.8% increase in the probability of complaints and a 3.9%-5.7% increase in the number of complaints per 1,000

[24] Since ISPs often operate multiple apps, we compute the average of relevant metrics across all apps owned by a given ISP. Our results remain robust when using the aggregate or unique counts of data types collected and SDKs installed across all apps owned by the same ISP. These robustness checks are provided in Appendix Table J.1 and Table J.2.

[25] Given the high concentration in ISP markets, we manually identify the 12 app-owning ISPs, listed in Appendix Section A.3, which together represent 85% (92%) of the market as of June 2016 (June 2021). Two factors may lead to an underestimation of local exposure. First, we assume that the data collection and sharing intensity of the substantially smaller remaining ISPs is zero. This assumption is consistent with evidence that larger companies engage in heavier data extraction activities, as documented in Bian et al. (2021). Second, our measures of ISP data practices are based on their mobile apps, which may understate the true extent of surveillance, as ISPs can also track consumers through web-based activity.

[26] We restrict the sample to observations before 2019 to avoid overlap with the ATT analysis.

residents. Results based on data sharing practices in Panel b are quantitatively similar. The dynamic DiD estimates in Appendix Figure J.3 show that the effects arise abruptly following the Repeal, with no evidence of pre-trends in the preceding period.

Like our analysis based on ATT, this setting links changes in data practices to shifts in fraud, strengthening our identification by showing that the narrative operates in both directions.

4.4 Analysis Using Consumer Sentinel and Identity Theft Complaints

We complement our main analysis using proprietary complaint-level data from the FTC, which offers three key advantages over the public CFPB database. First, the Consumer Sentinel data is uncensored: it includes all complaints made public by the CFPB as well as those withheld from the public release for privacy reasons (e.g., when disclosure could identify a consumer in a small zip code) and those excluded when firms fail to acknowledge a consumer relationship within 15 days.[27] Second, it includes two complementary non-public datasets: the Consumer Sentinel Network (CSN), which aggregates fraud and consumer complaints from a wide range of federal, state, and local agencies as well as private-sector partners, and the Identity Theft (IDT) database, which focuses specifically on identity theft cases.[28] Together these datasets provide a complete census of complaints. Third, the availability of reported dollar losses in the Consumer Sentinel data allows us to assess the effect on the costs of fraud in addition to its incidence.

Table 4 presents the results. Column 1 shows that using uncensored CFPB complaints from Consumer Sentinel yields an effect size of 5.6%, very close to the baseline 6.1% reduction in public CFPB complaints for a one-standard-deviation increase in iOS share. Thus, public-data censoring is unlikely to explain our results, and rules out the alternative explanation that firms might have been less able to acknowledge such relationships if ATT impaired firms'

[27] The Consumer Sentinel database has about 34% more CFPB complaints per month than the public CPFB database used in our previous analyses. See https://www.consumerfinance.gov/data-research/consumer-complaints/ for more details on which CFPB complaints are made public.

[28] We provide a description of the Consumer Sentinel and Identity Theft databases in Appendix C.

ability to identify consumers, leading to filed complaints being omitted from the public CFPB database. Columns 2 and 3 turn to the Consumer Sentinel and Identity Theft datasets. For the Identity Theft database, we find a statistically significant decline in complaints in high-iOS-share areas after ATT, indicating that identity theft, as one of the most common forms of fraud leveraging illicit personal data, is affected. The broader CSN shows a smaller decline of 1.1%. This decline in magnitude is not surprising, as both databases cover a much broader range of complaints that concern non-finance products and services and cases not triggered by data breaches.

Finally, Columns 4 and 5 exploit the FTC's monetary loss data. The similar magnitudes in Columns 3 and 4 of a 1.1% decline suggest that both fraud cases involving a reported loss and those without a reported loss decline at similar rates after ATT. In contrast, the larger magnitude in Column 5 (1.8%) relative to Column 4 suggests that ATT may disproportionately reducs high-loss cases, implying that the policy curtails not only the frequency of fraud but also its more severe financial impacts.

These estimates also enable a simple back-of-the-envelope calculation of ATT's potential monetary benefit to consumers. The Consumer Sentinel Network received reports of 6$ billion in consumer fraud losses in 2021. With iOS accounting for roughly 50% of the US smartphone market, an 83% opt-out rate from tracking under ATT, and the estimated 1.1% reduction in CSN fraud complaints with dollar losses per 10% increase in iOS device share, the implied reduction in fraud losses associated with ATT is approximately $0.011 \times (50\%/10\%) \times 83\% \times 6$ billion $\approx$ \$274 million annually. These estimates may be conservative given that reported fraud losses have doubled since 2021 (at \$12.5 billion in 2024), and fraud losses reported to Consumer Sentinel are known to substantially understate total fraud losses.[29]

[29] See the discussion on underreporting in the FTC's 2024 Older Adults Report https://www.ftc.gov/reports/protecting-older-consumers-2023-2024-report-federal-trade-commission.

4.5 Robustness

COVID-19. Because our sample period partly overlaps with the COVID-19 pandemic, one concern is that fraud complaints related to the pandemic increased for Android users relative to iOS users at the same time as the implementation of ATT. The pandemic led to many changes that could foster fraud, such as increases in online shopping and a surge in demand plus shortages for COVID-19 related items. We address this concern in the following ways.

We focus first on the Treasury's Economic Impact Payments (EIP). To examine whether these payments affect our results, we directly control for the interaction between three variables related to EIP payments and the post-ATT indicator. The EIP variables are constructed using IRS zip code-level tax return data and include the total amount of EIPs received, the average household income, and the average number of children and other dependents in a zip code. The amount of EIPs received by a zip code captures the actual payments, while household income and the number of children determine eligibility for EIPs. If our baseline results are driven by rising EIP-related fraud in low-iOS-share zip codes, we should see a substantial reduction in the point estimates after controlling for these factors. However, Table 5 Panel a shows that our DiD estimates remain similar in both economic and statistical significance, regardless of whether the interaction terms are based on actual EIP payments (Columns 1 and 3) or on eligibility (Columns 2 and 4).

Beyond EIP-related controls, we further address potential confounding factors by interacting zip code-level socioeconomic characteristics with the post-ATT indicator. The rationale is that such factors jointly influence COVID-19 and EIP exposure as well as the propensity to use iOS devices. Haendler and Heimer (2025) show that the propensity to claim fraud and request refunds in the complaint process correlates with education and income. We consider a broad set of variables, including age, gender, education, income, and unemployment. Unemployment is included to capture the potential effects of pandemic-related job support, such as generous unemployment insurance, as well as the increased rates of unemployment during COVID-19. Adding these interactions leaves the direction of our

DiD estimates unchanged (Appendix Table J.7), though the magnitude is smaller (at most halved) as some of the identifying variation is absorbed by these socioeconomic factors.

To further address concerns that our ATT estimates could be confounded by COVID-19 relief fraud disproportionately affecting low-iOS-share areas, we conduct placebo tests using two alternative event dates: April 2020 and December 2020. These dates correspond to the disbursement of the first and second major waves of pandemic-related financial support under the CARES Act and subsequent legislation.[30] By shifting the event time to these earlier points, we can test whether our baseline results also appear around the timing of major COVID-19 relief disbursements. As shown in Panel b of Table 5, after the COVID-19 relief disbursements, high-iOS-share areas do not experience declines in complaints compared to low-iOS-share areas, providing further assurance that our main results are not driven by fraud related to COVID-19 relief programs.

Propensity to File Complaints. The propensity to complain after victimization can vary across different communities, so declines in consumer complaints may not immediately translate to declines in fraud victimization. Raval (2020b) examines how several zip code-level demographic variables affect the likelihood of complaining by comparing complaints and victims for several consumer protection cases and found much lower complaint rates, conditional on victimization, in heavily Black and Hispanic areas. Using these estimates, Raval (2020b) develops zip code-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 weight for the median zip code was normalized to 1, with Black zip codes averaging about 2 because residents there were roughly half as likely to file complaints.

In Column 1 of Appendix Table J.3, we multiply our complaint counts by these weights to examine changes in fraud victimization. This exercise relies on the assumption that

[30]The first round of stimulus payments and expanded unemployment benefits began reaching individuals in late April and early May 2020, while the second round was distributed in late December 2020. Both periods saw sharp increases in fraud complaints, particularly related to identity theft and benefit scams.

the differences between the propensity to complain across locations found in Raval (2020b) extrapolate to CFPB complaints and that the adoption of ATT did not change the propensity to complain.[31] We find that the estimated post-ATT decline in fraud victimization is 6.9% per one standard deviation increase in iOS share, which is larger than the unadjusted 6.1% from Column 4 of Table 2. This pattern suggests that individuals most vulnerable to data-driven fraud are, on average, less likely to file a complaint, implying that the unadjusted estimates modestly understate ATT's true effects.

Measurement Error in iOS Shares. We address potential measurement error in our foot-traffic-based iOS share measure by restricting the sample in several ways. First, we exclude zip codes in the bottom quartile or bottom half of foot traffic. Second, we drop zip codes with a small number of grocery stores (bottom quartile or bottom half), as residents in such areas may shop in other zip codes, making our measure less representative of local residents. Finally, we weight regressions by population, as larger populations are more likely to support multiple grocery and retail stores, ensuring that the measured iOS share better reflects local residents rather than visitors. These alternative sampling criteria improve the precision of our estimates by reducing potential noise in low-traffic areas. Columns 2 to 6 in Appendix Table J.3 show that the estimated ATT effect ranges from -6.0% to -6.8% across specifications and remains negative and statistically significant with a slightly larger magnitude than our baseline estimates.

Alternative Aggregation. Our baseline analysis is conducted at the zip code-month level. While this fine-grained panel makes full use of the identifying variation in our data, it also raises concerns about sparsity and noise in the data, as the number of complaints in many zip code-month cells is small. To address this concern, we aggregate the data to both the quarterly level and to the county-month level to develop datasets at a coarser level of aggregation. Estimates using the zip code-by-quarter panel are both qualitatively and quantitatively similar to our baseline estimates (Column 1 of Appendix Table J.4).

[31] ATT could raise awareness of data-driven fraud among iOS users and induces them to complain more, but that would lead to an increase in fraud complaints after ATT rather than a decrease.

When aggregating to the county level, the estimated effect is -3.7% (Column 2 of Appendix Table J.4), about one-third smaller than our baseline estimates (Column 3 of Table 2), but remains highly significant. This difference could reflect weighting effects from aggregating to larger geographic units, as the treatment effect likely differs across zip codes.

Excluding Complaints about Credit Bureaus. Another potential concern is that reports concerning the three major US credit reporting agencies make up about half of the complaints in the CFPB complaints database. If our estimated treatment effects were driven primarily by complaints about these firms, the results could reflect idiosyncratic dynamics in that segment rather than patterns of consumer fraud driven by ATT. To address this, we re-estimate our main specifications after excluding all complaints related to these three agencies. The results are similar in magnitude and significance (Column 1 of Appendix Table J.5), so our findings are not driven solely by complaints about the credit bureaus.

Monotonic Effects and Spillovers. Another concern is that fraudsters could shift their focus from iOS users to Android users after privacy is strengthened on the iOS platform. Such spillovers would still imply that ATT increased the costs to defraud iOS users, which should reduce overall fraud. However, if this spillover effect is strong, we may overestimate the effect of ATT on consumer complaints. While fully accounting for spillover effects is difficult, we provide evidence inconsistent with within-zip-code spillovers and within-firm spillovers.

First, our alternative identification strategy in Subsection 4.3 is based on changes to ISP data collection practices, which should not generate spillovers between iOS and Android users. Second, if spillovers within zip codes drove ATT's effects, we would expect the largest effects in areas with more Android users as there would be more consumers to whom fraudulent activity could be redirected. However, as shown in Appendix Figure J.4, we find that the effect of ATT is strongest in regions predominantly populated by iOS users. Third, in Subsection 5.2, we examine firm-level cyber incidents using variation in firms' heterogeneous presences in the mobile space for identification. We find reductions in complaints for firms with a mobile presence, providing evidence against within-firm spillovers such as from the firm's iOS-using consumers to its Android-using consumers.

Other Robustness. In Appendix Table J.6, we augment our regression specification with several additional robustness checks. First, we replace our baseline 1% two-sided winsorization with a more conservative threshold (0.5% on each side), and find almost no change in the DiD coefficient (Column 1). Second, we re-estimate the Poisson model without scaling the outcome variable by population, which yields results that are both economically and statistically similar to the baseline (Column 2). Third, we relax the criteria for identifying outlier zip codes and continue to find robust effects when we do not remove any outlier zip codes (Column 3). Next, we double-cluster standard errors by state and year-month. Column 4 of Appendix Table J.6 shows that the results remain highly significant with a t-statistic almost identical to the baseline (6.21 vs. 6.60). Lastly, as mentioned previously, we extend the sample through the end of 2023 as well as 2024 and find larger, statistically significant DiD coefficients (Columns 5 and 6). This pattern indicates that the effect of ATT is not short-lived but rather strengthens over time, consistent with the dynamics shown in Panel b of Figure 3.

5 Mechanism

5.1 Downstream: Data-Related Fraud Complaints

The downstream analysis follows our earlier analysis but zooms in on consumer fraud complaints as the most visible consequences of compromised data. While many forms of fraud exist, only some rely directly on access to compromised personal or financial information. By singling out the more relevant categories, we can test whether ATT primarily affects fraud mechanisms that plausibly depend on compromised upstream data suppliers.

5.1.1 Identifying Relevant Complaints

Not all consumer complaints are directly linked to the collection and misuse of personal information by thieves or hackers. The "issues" or "sub-issues" fields in the CFPB database do not explicitly distinguish between more or less relevant categories for fraud arising from lax data privacy regulations. This is even more true in the Consumer Sentinel data, as it amalgamates complaints from many different sources.[32] 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".[33] 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 might be affected by changes in data privacy standards.

Second, we utilize a machine learning method called zero-shot learning (ZSL) to assess the likelihood that a narrative is 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 F. 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 F.1 presents the mean and standard deviation of complaint-level scores by product category.

[32] 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.

[33] 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 complaints. 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 data security regulations.

5.1.2 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—Student Loan and Mortgage—and report the results in Panel a of Table 6.

Consistent with the hypothesis that ATT reduces financial fraud enhanced by lax data privacy, 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. Following the implementation of ATT, a 10% increase in the share of iOS users in a zip code is associated with a 5.5% decrease in the number of complaints related to credit reporting per 1,000 residents (Column 1) and a 5.7% decline in the number of complaints related to debt collection (Column 2). In contrast, ATT has an insignificant effect on complaints related to student loans or mortgages (Columns 3 and 4). Applying the same method to the Consumer Sentinel complaint database, we find similar results in Appendix Table J.8.

We additionally leverage the full coverage of narratives in the Identity Theft and Consumer Sentinel complaint 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 6. 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 7.2% decline for complaints about identity theft (Column 1) and a 2.7% decline in complaints to Consumer Sentinel (Column 2). We find null effects (0.0%, -0.4%) for complaints from both sources that do not include one of the relevant keywords. The lack of an effect 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 to be driven by concurrent shocks or differential time trends for iOS and Android users.

5.2 Midstream: Financial Institutions and ISPs as Data Gateways

Midstream actors—firms, banks, and ISPs—form the key link between fraudsters and consumers. They routinely collect large volumes of individual-level behavioral and financial data, which, if stolen or misused, can fuel downstream fraud. Indeed, Ramadorai et al. (2025) document a positive association between data extraction activities and cyber incidents. These institutions are thus both gateways through which data can reach malicious actors and targets whose data repositories attract cybercriminals. Understanding how their data-collection and sharing practices shape fraud risk is essential to connecting consumer fraud outcomes to upstream data markets. We study this link using variation in the presence in the mobile market and data practices of these institutions.

5.2.1 Firm-Level Evidence Exploiting Mobile Market Presence

Institutions with consumer-facing iOS apps are more likely to collect and share rich consumer data, making them especially exposed to ATT's privacy restrictions. If ATT curtails fraud by limiting access to exploitable personal data, its effects should be visible in both reduced exposure to data breaches and related cyber incidents as well as in downstream consumer fraud complaints.

We examine this question using Advisen's cyber-incident database, which covers over 90,000 events from publicly verifiable sources between 2000 and 2023.[34] For each firm-month in our sample, we identify whether the firm experienced any cyber incident, whether it stemmed from malicious breaches or privacy violations, whether it resulted in violations of the Fair Debt Collection Practices or Fair Credit Reporting Acts—the two regulations most commonly linked to fraud in CFPB complaint narratives, or whether it was unrelated to lax data standards (e.g., lost devices, accidental disclosure).

The results in Panel a of Table 7 show that ATT significantly reduces cyber-incident exposure for app-owning firms. The likelihood of any incident falls by 4.9 percentage points relative to firms without an app, which is half of the baseline. Effects are strongest for privacy-related incidents and for those involving violations of the two key fraud-related regulations (Columns 1 and 3), with no significant effect for unrelated causes (Column 4).

We next examine whether the reduction in firm exposure to data breaches and privacy violations translates into fewer consumer complaints about these firms. In Panel b of Table 7, we find that app-owning firms are significantly less likely to receive fraud-related complaints post-ATT than non-app-owning firms. Column 1 shows that the monthly probability of any complaint falls by 2.1 percentage points. The effect is concentrated in the most fraud-relevant categories—credit reporting, credit repair services, and debt collection—with no detectable change in unrelated categories such as student loans or mortgages.

5.2.2 Heterogeneity by Local Exposure to High-Risk Data Collectors

While the above firm-level analysis is informative, it deviates from our main identification strategy, which exploits geographic variation in iOS share. To connect the two, we extend the zip code-level design into a triple difference one that incorporates local variation in the prevalence and data practices of two key data collectors: ISPs and financial institutions. ISPs sit at the center of online activity, capturing broad swaths of consumer behavior, while banks handle high-value identity and transaction data directly relevant to financial fraud. Both

[34] More information on Advisen's data sources can be found here.

institutions can act as gateways, either directly or indirectly, supplying data to upstream fraudsters.

We measure these institutions' data-handling practices by analyzing their mobile apps, parsing Apple's iOS privacy-label disclosures, and identifying data-sharing SDKs installed in those apps. We then aggregate these measures to the local market level using ISP market shares and bank deposit shares to capture the intensity of local exposure to aggressive data collectors. Interacting these measures with zip code-level iOS share in a triple difference regression allows us to test whether ATT's effect on fraud is concentrated in areas where high-risk banks and ISPs are more prevalent.

Panel a of Table 8 reports results exploiting local exposure to problematic data practices in the banking sector. We focus on institutions active in the mobile market and construct two exposure measures, each interacted with Post × iOS Share. The first measure captures local data-collection exposure. For each bank, we multiply the number of unique data items collected by its app with the app's popularity, proxied by download counts, and then sum across all banks.[35] The second measure, local data-sharing exposure, is constructed analogously, replacing the number of data items collected with the number of third-party data-sharing SDKs integrated into the app. Both measures are constructed at the county-quarter level and then averaged across all quarters in the pre-ATT period. To facilitate comparison of coefficients across specifications, we further standardize the measures using z-scores. To approximate the share of downloads originating from local users, we weight exposure by the within-bank, across-county deposit share, as formally defined in Equation (5).

$$\text{Bank Exposure}_c^{(m)} = \sum_{j \in \mathcal{B}(c)} \left(I_j^{(m)} \times \text{Downloads}_j \times \%\text{Deposit}_{jc} \right), \tag{5}$$

where c indexes counties, j indexes banks, and $\mathcal{B}(c)$ denotes the set of banks with branches in county c. The data collection and sharing intensities, $I_j^{(m)}$, are defined analogously to

[35]This product reflects both how many users are using the app and how much data is collected per user, providing an estimate of the total volume of data extracted in a given area. If a bank owns multiple mobile apps, we consider its flagship app to be the one with the highest number of downloads.

Equation (4). Downloads_j is the total number of app downloads for bank j, and $\%\text{Deposit}_{jc}$ is the within-bank share of deposits held by bank j in county c.

Across all specifications, the coefficients on the triple interaction terms Post × iOS × Exposure are negative and statistically significant, indicating that ATT's fraud-reducing effect is strongest in markets where mobile banking is more prevalent and local banks are more aggressive in collecting or sharing consumer data. In terms of magnitude, based on Column 4 in Panel a, a one-standard-deviation increase in the interaction between iOS share and exposure based on banks' data-sharing intensity is associated with a 1.9% additional decrease in the number of complaints per 1,000 residents. In this way, banks act as information intermediaries, often inadvertently playing a key role in collecting and transmitting personal data that is later exploited by fraudsters.

Panel b of Table 8 repeats the analysis for ISPs, using county-level exposure to their data-collection and data-sharing intensities, defined analogously to Equation (3). These terms help us test whether ATT's fraud-reducing effect is stronger in markets where ISPs are more active in gathering and transmitting consumer data. The coefficients are negative and statistically significant across specifications, indicating that ISPs amplify ATT's impact.

Overall, these results on intermediaries show that ATT's fraud-reducing effects are most pronounced in markets where key local entities, including financial institutions, collect and share large volumes of consumer data. This pattern is consistent with ATT disrupting the pathways through which personal information reaches fraudsters. By limiting these midstream data supplies, ATT likely reduces the volume and richness of data available for trafficking and resale in criminal marketplaces, which we examine next.

5.3 Upstream: Illicit Data Markets

Our upstream analysis digs into the supply side of the fraud economy—the illicit markets where compromised personal and financial data are traded. As the FTC notes, the dark web functions as a critical conduit for this trade, enabling cybercriminals to buy and sell compromised information that fuels a variety of fraudulent activities.[36] These markets operate in hidden, unregulated environments and are only partially observable through dark web forum discussions and marketplace listings, which together capture both the conversation layer (coordination and exchange of know-how) and the transaction layer (actual sales and pricing). We track these two layers and investigate discussions, availability, and prices of compromised data on the dark web. Importantly, we compare trends in categories directly exposed to ATT with those in unaffected control groups to assess whether privacy restrictions disrupted the supply chain feeding downstream fraud.

In contrast to downstream complaints and the data-collection practices of midstream entities, which can be readily measured in structured administrative or standardized datasets, upstream activity is inherently opaque. Dark web market prices and volumes can be noisy, influenced by strategic behavior, enforcement actions, or migration to less visible channels. As such, our evidence based on the dark web should be viewed as indicative rather than conclusive. Even so, these data offer a rare opportunity for forensic finance—direct observations of fraudsters' actions rather than inferences pieced together from more distant outcomes. To our knowledge, they have not been used in finance research, and we hope our work encourages further studies that leverage such rare opportunities to shed light on the economy of fraudsters.

5.3.1 Posts on Dark Web Forums

We begin with dark web forum discussions—the conversation layer of the illicit data economy—where actors share techniques, coordinate activities, and exchange market intelligence. Changes in discussion patterns can signal shifts in supply conditions or attacker behavior.

Our data comes from CrimeBB, a structured collection of textual data scraped and maintained by the Cambridge Cybercrime Centre (CCC). CrimeBB contains approximately 124 million posts from 7 million members across 38 cybercrime and extremist forums, regularly updated and stored on CCC servers. We use posts from all forums from 2015 onward.

[36]See FTC discussion's discussion on dark web here.

We apply a textual classification procedure to label each post according to the labels and keywords listed in Appendix Table H.2 to Table H.6. We then aggregate monthly post counts by label and forum to construct our regression panel.[37] The most exposed category includes posts explicitly related to Apple devices or the iOS ecosystem. While direct exposure to ATT is difficult to measure, the policy applies only to iOS, making Apple related discussions most likely to be affected. The partially treated category covers broader illicit activities that plausibly depend on personal or financial data—such as tracking, data breaches, identity theft, and financial data theft. Many of these activities are known to rely heavily on data originating from mobile devices, particularly in financial fraud schemes, so changes in these discussion volumes can provide an early signal of shifts in the supply of mobile-sourced and financially sensitive data.

Our control group consists of illicit-trade categories that operate in a fundamentally different, segmented market. These markets, such as drugs, human trafficking, and weapons, are characterized by distinct participants, supply chains, and transaction methods, with minimal overlap in actors or infrastructure with the data-driven fraud economy. As such, they provide a benchmark plausibly insulated from shifts in digital data availability and attacker strategies that could result from ATT.

This setup allows us to estimate DiD regressions comparing changes in post volume for the exposed and partially treated groups against the control group around ATT's implementation. The results are displayed in Panel a of Table 9. We find that post volumes in the most exposed group fell by over 30% following ATT (Column 1). The decline is smaller but still statistically significant for the partially treated group (Column 2), consistent with these illicit activities being affected through reduced access to mobile-sourced and financially sensitive data. Although some actors discuss potential workarounds immediately after ATT, overall activity declines, likely due to a combination of reduced supply and higher barriers to circumvention, which limit participation in these markets.

[37] Our classification of labels follows Avarikioti et al. (2018), which use supervised machine learning to develop a comprehensive categorization of the dark web content.

We show in Appendix Table J.9 that the results are robust to using alternative, more restrictive keyword lists and to alternative counting methods (fractional assignment versus double counting ambiguous posts). The richness of forum topics also enables a placebo test: we examine unrelated digital technology categories such as zero-days, malware, and botnets, which do not rely on personal data for their operation. We find little or no change in these categories, helping to rule out the possibility that our results are driven by broader trends in digital-related discussions rather than ATT (Column 3 of Table 9 Panel a).[38]

5.3.2 Dark Web Listings and Prices

To connect conversations on dark web forums to actual transactions, we analyze the transaction layer through dark web marketplace listings and prices. Using data on products listed on the dark web, we provide suggestive evidence that ATT has driven up the price of data relevant for financial fraud and identity theft, potentially by reducing the supply of stolen or hacked data. The listing data, assembled by a research organization focused on fraud, is described in detail in Appendix I.

We use two snapshots of dark web listings from 2020 and 2023 to compare the prices of data generated from mobile app user activities with those of other listings, and of financial information with other categories, before and after ATT.[39] Applying a DiD design, we find that ATT is associated with higher prices for data likely generated from consumers' mobile activities and for financial information, relative to other products (Panel b of Table 9). This pattern is consistent with ATT reducing the availability of such data on the dark web.

While our listings and forum data provide an important glimpse into upstream markets, we caution that they likely represent only the tip of the iceberg. More sophisticated actors operate in closed or invitation-only markets, conduct transactions via encrypted communi-

[38]As with other work using dark web posts, some categories inevitably fall into a gray area. Even in the computer science literature, where more sophisticated classification algorithms are employed, researchers continue to face challenges in drawing sharp boundaries. In our case, the results are not sensitive to reassignments of these fuzzy categories (Column 5 of Table J.9).

[39]Because only two cross-sectional snapshots are available, we cannot analyze changes in the number of listings over time and instead focus solely on the price. The detailed regression specification and results are provided in Appendix I.

cation channels, or route sales through intermediaries to avoid detection.[40] There are also gray areas where fraudsters engage in transactions of compromised data through seemingly legitimate channels, creating the appearance of lawful business activity while still trafficking in compromised information.[41]

6 Conclusion

In this paper, we show that Apple's App Tracking Transparency (ATT) policy, that greatly limited the tracking of personal data across apps and websites, reduced financial fraud in areas where more iOS users are present. Our findings shed light on the full fraud supply chain and highlight how data privacy regulation benefit customers by constraining illicit data flows. We interpret our estimates as capturing the effects of tighter constraints on data collection and sharing practices. The observed reductions in fraud do not arise from directly suppressing online engagement—an outcome that is neither practical nor desirable—but from altering the ways in which data is collected, processed, shared, and sold. These findings suggest that improving data practices and their regulation, rather than restricting digital consumption, have the potential to reduce fraud and generate spillovers by disrupting the broader illicit data economy that sustains financial fraud. We caveat that, although the policy is not designed to curb digital activity, such activity could nonetheless be affected—for instance, if restrictions on data use reduce app monetization or, conversely, if greater consumer trust encourages higher engagement (Armantier et al., 2024). Assessing the broader equilibrium consequences of improving privacy standards for online activity and welfare remains an important task for future research.

[40] For example, on many dark web platforms, user-to-user transactions reveal only minimal metadata, such as the timestamps and user-submitted ratings, while concealing the actual content or price of the transaction.

[41] The FTC charged data brokers (Gen X) for selling sensitive financial information to scammers who used it to drain consumer accounts, under the cover of routine data market activity. The scammers operated through a seemingly legitimate business, Ideal Financial Solutions Inc. See the FTC newsletter here.

References

Abis, S., H. Tang, and B. Bian (2025). Breaking the Data Chain: The Ripple Effect of Data Sharing Restrictions on Financial Markets. *Available at SSRN 5334566*.

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2025). When Big Data Enables Behavioral Manipulation. *American Economic Review: Insights 7*(1), 19–38.

Amiram, D., Z. Bozanic, J. D. Cox, Q. Dupont, J. M. Karpoff, and R. Sloan (2018). Financial Reporting Fraud and Other Forms of Misconduct: a Multidisciplinary Review of the Literature. *Review of Accounting Studies 23*(2), 732–783.

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 (2023). The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR. *RAND Journal of Economics 54*(4).

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.

Armantier, O., S. Doerr, J. Frost, A. Fuster, and K. Shue (2024). Nothing to hide? gender and age differences in willingness to share data. *Gender and age differences in willingness to share data (April 26, 2024). Swiss Finance Institute Research Paper* (24-99).

Avarikioti, G., R. Brunner, A. Kiayias, R. Wattenhofer, and D. Zindros (2018). Structure and Content of the Visible Darknet. *arXiv preprint arXiv:1811.01348*.

Babina, T., S. Bahaj, G. Buchak, F. De Marco, A. Foulis, W. Gornall, F. Mazzola, and T. Yu (2025). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. *Journal of Financial Economics 169*, 103950.

Begley, T. A. and A. Purnanandam (2021). Color and Credit: Race, Regulation, and the Quality of Financial Services. *Journal of Financial Economics 141*(1), 48–65.

Bertrand, M. and E. Kamenica (2023). Coming Apart? Cultural Distances in the United States over Time. *American Economic Journal: Applied Economics 15*(4), 100–141.

Bessen, J. E., S. M. Impink, L. Reichensperger, and R. Seamans (2020). GDPR and the Importance of Data to AI Startups. Working paper, New York University, Boston University.

Bian, B., Q. Huang, Y. Li, and H. Tang (2024). Data as a Networked Asset. *Available at SSRN 5183829*.

Bian, B., X. Ma, and H. Tang (2021). The Supply and Demand for Data Privacy: Evidence from Mobile Apps. *Available at SSRN 3987541*.

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*.

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*.

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*.

Dou, Y. and Y. Roh (2024). Public Disclosure and Consumer Financial Protection. *Journal of Financial and Quantitative Analysis 59*(5), 2164–2198.

Egan, M., G. Matvos, and A. Seru (2019). The Market for Financial Adviser Misconduct. *Journal of Political Economy 127*(1), 233–295.

Egan, M., G. Matvos, and A. Seru (2025). Arbitration with Uninformed Consumers. *Review of Economic Studies*, Forthcoming.

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.

Gao, J., J. Pacelli, J. Schneemeier, and Y. Wu (2020). Dirty Money: How Banks Influence Financial Crime. *Available at SSRN 3722342*.

Goldberg, S. G., G. A. Johnson, and S. K. Shriver (2024). Regulating Privacy Online: An Economic Evaluation of the GDPR. *American Economic Journal: Economic Policy 16*(1), 325–358.

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

Griffin, J. M., S. Kruger, and P. Mahajan (2023). Did FinTech Lenders Facilitate PPP Fraud? *The Journal of Finance 78*(3), 1777–1827.

Griffin, J. M., S. Kruger, and P. Mahajan (2024). Did Pandemic Relief Fraud Inflate House Prices? *Available at SSRN 4487877*.

Haendler, C. (2022). Keeping Up in the Digital Era: How Mobile Technology is Reshaping the Banking Sector. *Available at SSRN 4287985*.

Haendler, C. and R. Heimer (2025). The Hidden Costs of Financial Services: Consumer Complaints and Financial Restitution. *Available at SSRN*.

Huang, R., J. S. Linck, E. J. Mayer, and C. Parsons (2024). Can Human Capital Explain Income-Based Disparities in Financial Services. *Review of Financial Studies*, 22–16.

Huff, R., C. Desilets, and J. Kane (2010). The 2010 National Public Survey on White Collar Crime. *National White Collar Crime Center 44*.

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 40*(4), 661–684.

Jiang, E. X., Y. G. Yu, and J. Zhang (2025). Bank Competition amid Digital Disruption: Implications for Financial Inclusion. *Journal of Finance*, Forthcoming.

Johnson, G. (2022). Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond. Technical report, National Bureau of Economic Research.

Jou, J., A. Kleymenova, A. Passalacqua, L. Sándor, and R. Vijayaraghavan (2024). Disciplining Banks through Disclosure: Evidence from CFPB Consumer Complaints. *Available at SSRN*.

Kesler, R. (2022). The Impact of Apple's App Tracking Transparency on App Monetization. *Available at SSRN 4090786*.

Koont, N. (2023). The Digital Banking Revolution: Effects on Competition and Stability. *Available at SSRN 4624751*.

Koont, N., T. Santos, and L. Zingales (2024). Destabilizing Digital "Bank Walks". Technical report, National Bureau of Economic Research.

Kraft, L., B. Skiera, and T. Koschella (2023). Economic Impact of Opt-In versus Opt-Out Requirements for Personal Data Usage: The Case of Apple's App Tracking Transparency (ATT). *Available at SSRN 4598472*.

Li, X. (2023). Does the Disclosure of Consumer Complaints Reduce Racial Disparities in the Mortgage Lending Market. *Available at SSRN 4741819*.

Li, Z., H. Ning, F. Jing, and M. N. Lessani (2024). Understanding the Bias of Mobile Location Data across Spatial Scales and over Time: a Comprehensive Analysis of SafeGraph Data in the United States. *Plos one 19*(1).

Mazur, L. (2024). The Impact of the CFPB Complaint Database Disclosure on Banks' Provision of Mortgage Credit to Low-and Moderate-Income Communities.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science 41*(4), 746–768.

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

Synovate (2003). Federal Trade Commission - Identity Theft Survey Report. Technical report, Federal Trade Commission.

Tang, H. (2019). The Value of Privacy: Evidence from Online Borrowers. *Available at SSRN*.

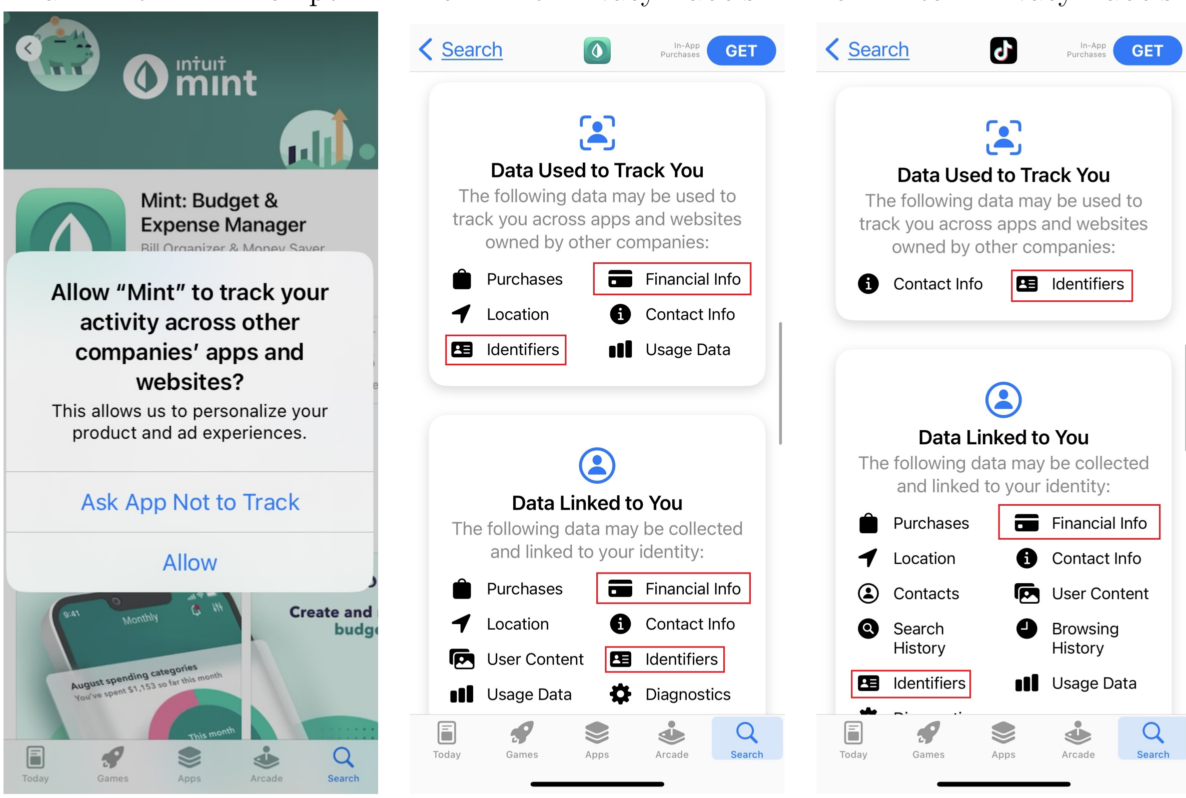
Wang, T. Y., A. Winton, and X. Yu (2010). Corporate Fraud and Business Conditions: Evidence from IPOs. *The Journal of Finance 65*(6), 2255–2292.

Wu, X. (2023). Mobile App, Firm Risk, and Growth. *Available at SSRN 4519061*.

Yu, X. (2013). Securities Fraud and Corporate Finance: Recent Developments. *Managerial and Decision Economics 34*(7-8), 439–450.

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

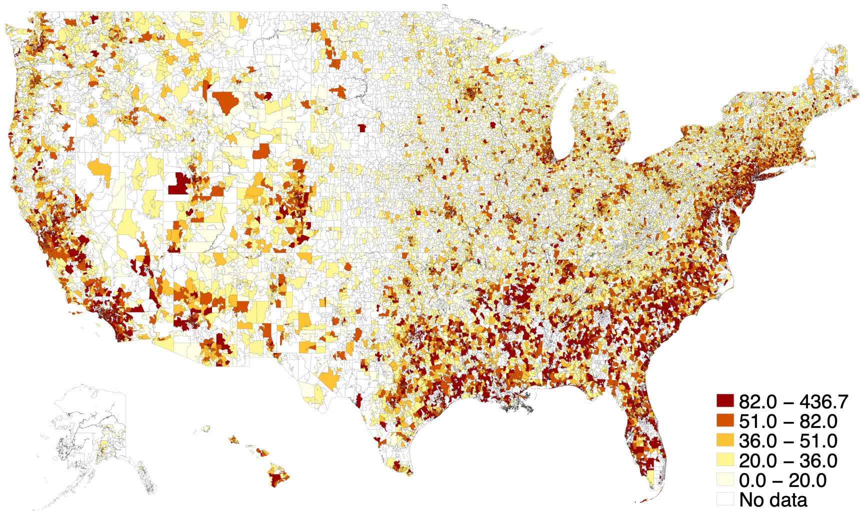
a. Mint ATT Prompt b. Mint Privacy Labels c. Tiktok Privacy Labels



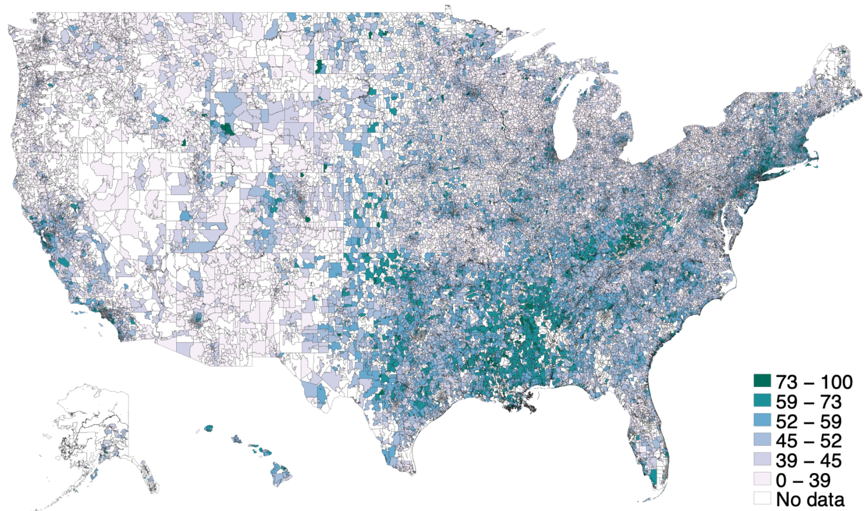
Apple's App Tracking Transparency (ATT) policy was introduced on April 26, 2021. Panel a shows the ATT prompt through which apps (Intuit's Mint in this example) could get user permission to obtain mobile identifiers that allow them to track, share, or sell consumer data with or to other apps and websites. 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, which describes that TikTok uses mobile identifiers to obtain, e.g., financial data of consumers.

Figure 2: Number of Complaints Scaled by Population and iOS Share by Zip Code

a. # Complaints per Million Residents



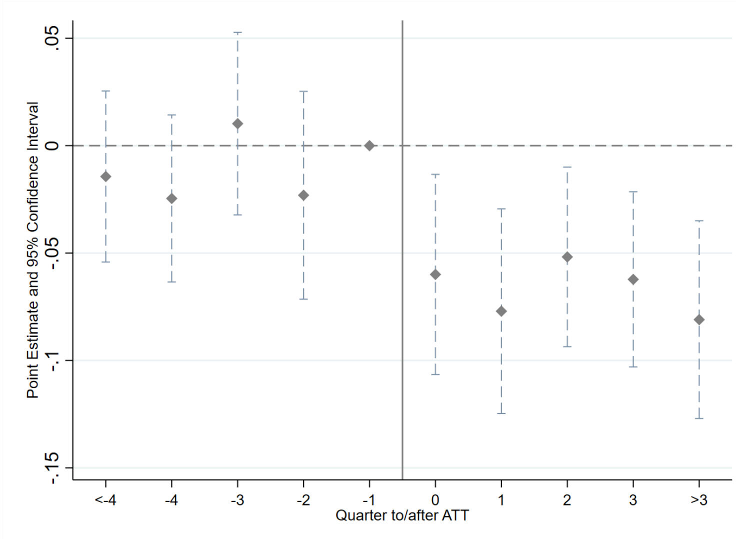
b. iOS Share



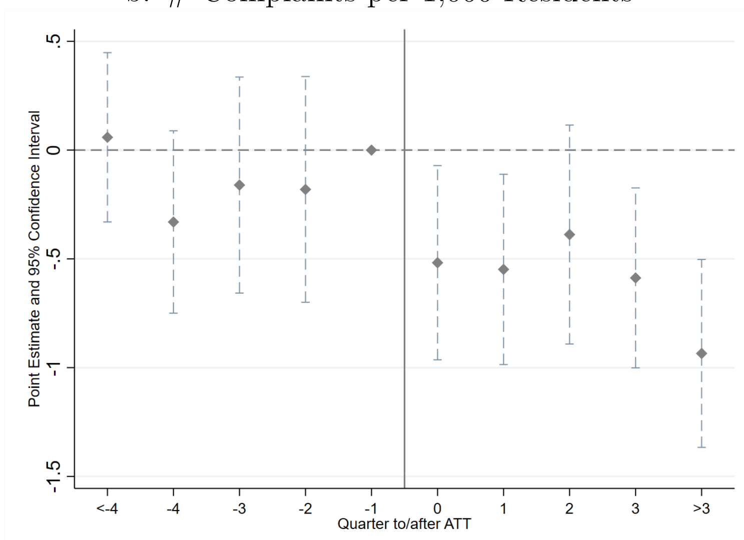
This figure presents the number of CFPB complaints per million residents (Panel a) and the average iOS share prior to the ATT rollout (Panel b), across zip codes over the full sample period (January 2019 to June 2022). For readability, Panel a displays CFPB complaints per million residents; however, our empirical analyses use CFPB complaints per 1,000 residents.

Figure 3: Dynamic Effects of ATT on CFPB Complaints

a. Any Complaints



b. # Complaints per 1,000 Residents



This figure illustrates the dynamic effect of ATT on CFPB complaints around the implementation of ATT (April 26, 2021). 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 an indicator for whether a zip code has at least one complaint and estimates are from a linear probability model. In Panel b, the outcome variable is the number of complaints per 1,000 residents and estimates are from a Poisson model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Coefficients on the interaction terms between indicators for the relative timing to ATT and the pre-ATT iOS device share are plotted. The sample period is January 2019 to June 2022. Zipcode fixed effects and county × year-month fixed effects are included. Standard errors are clustered at the state level.

Table 1: Summary Statistics: Zip code-Month Level

variable	mean	sd	p25	p50	p75	count
			ATT sample (2019m1–2022m7)			
Any complaints (0/1)	0.26	0.44	0.00	0.00	1.00	927,780
# Complaints per 1,000 residents	0.05	0.15	0.00	0.00	0.03	927,780
iOS share	0.46	0.11	0.39	0.45	0.52	927,780
			FCC Broadband Rule sample (2015m1–2018m12)			
Any complaints (0/1)	0.21	0.40	0.00	0.00	0.00	1,166,496
# Complaints per 1,000 residents	0.03	0.09	0.00	0.00	0.00	1,166,496
Exposure (data collection)	0.93	0.78	0.04	0.85	1.51	1,166,496
Exposure (data sharing)	0.15	0.14	0.00	0.13	0.26	1,166,496

This table presents summary statistics for our key explanatory and outcome variables at the zip code-month levels. The top part shows the summary statistics for the main sample that spans 2019m1 and 2022m6, which we use to examine the effect of the ATT. The bottom part shows the summary statistics for the sample that spans 2015m1 and 2018m12, which we use to examine the effect of the 2017 Repeal of the FCC Broadband Rule. We primarily examine two outcomes variables constructed from the CFPB complaints: an indicator for whether a zip code has at least one complaint (*Any complaints (0/1)*) and the total number of complaints per 1,000 residents (*# Complaints per 1,000 residents*). *iOS share* denotes the share of iOS devices in a given zip code. *Exposure (data sharing)* refers to the market-share-weighted number of data SDKs used by ISPs in their mobile applications in a given zip code. *Exposure (data collection)* represents the market-share-weighted number of data types collected by the Top 12 fixed broadband ISPs, as disclosed in their privacy labels, in a given zip code.

Table 2: Effect of ATT on Consumer Complaints - CFPB

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post × iOS share	−0.036*	−0.059***	−0.548***	−0.621***
	(0.018)	(0.015)	(0.095)	(0.094)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.264	0.264	0.053	0.053
Magnitude (↑1 SD iOS share)	-1.4%	-2.4%	-5.4%	-6.1%
Observations	926,772	926,772	926,772	926,772
R^2 / Pseudo R^2	0.441	0.432	0.094	0.134

This table presents the estimated effects of ATT on CFPB complaints. The unit of observation is at the zip-code-month level. Columns 1 and 2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3 and 4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022. All columns include zip code fixed effects. The odd-numbered columns include state × year-month fixed effects, while the even-numbered columns include county × year-month fixed effects. 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: Effect of 2017 Repeal of FCC Broadband Privacy Rule on CFPB Complaints

Panel a. Local Exposure Based on ISP Data Collection Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data collection)	0.008***	0.007***	0.074***	0.058***
	(0.002)	(0.002)	(0.014)	(0.022)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	2.8%	2.4%	5.7%	3.9%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

Panel b. Local Exposure Based on ISP Data Sharing Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data sharing)	0.044***	0.038***	0.414***	0.311***
	(0.008)	(0.011)	(0.059)	(0.102)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	2.9%	2.5%	6.1%	3.9%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

This table presents the estimated effects of the 2017 FCC Broadband Privacy Rule repeal on CFPB complaints. The unit of observation is at the zip-code-month level. Columns 1 and 2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3 and 4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Panel a uses the market-share-weighted number of data types collected by ISPs, as disclosed in their privacy labels, as the measure of exposure to the shock. Panel b uses the the market-share-weighted number of data SDKs used by ISPs in their mobile applications. The sample period is January 2015 to December 2018. All columns include zip code fixed effects. The odd-numbered columns include state × year-month fixed effects, while the even-numbered columns include county × year-month fixed effects. 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: Effect of ATT on Consumer Complaints -
Consumer Sentinel Network and FTC Identity Theft Databases

	# Complaints per 1,000 residents				
	(1) Uncensored CFPB	(2) Identity Theft	(3) CSN	(4) Any loss	(5) Loss ≥ $1k
Post × iOS share	−0.575***	−0.255**	−0.114***	−0.108*	−0.183**
	(0.085)	(0.124)	(0.044)	(0.064)	(0.093)
Zip code FE	✓	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML	PPML
Mean outcome var.	0.125	0.223	0.733	0.165	0.062
Magnitude (↑1 SD iOS share)	-5.6%	-2.5%	-1.1%	-1.1%	-1.8%
Observations	757,635	955,332	1,131,115	1,024,758	825,936
Pseudo R^2	0.209	0.240	0.143	0.099	0.096

This table presents the estimated effects of ATT on consumer complaints using proprietary internal data from the Consumer Sentinel Network (CSN) and FTC Identity Theft databases. The unit of observation is at the zip-code-month level. In all specifications, we estimate a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Column 1 uses all uncensored CFPB complaints from the Consumer Sentinel Network. Column 2 uses all Identity Theft complaints. Column 3 uses all Consumer Sentinel Network complaints. Column 4 uses Consumer Sentinel Network complaints that report a positive dollar loss. Column 5 uses Consumer Sentinel Network complaints that report a loss greater than $1,000. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. All columns include zip code and county × year-month fixed effects. 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: Effect of ATT on Consumer Complaints - CFPB
Addressing COVID-19 Related Concerns

Panel a. Controlling for Exposure to EIP

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post × iOS share	−0.073***	−0.056***	−0.684***	−0.483***
	(0.017)	(0.016)	(0.098)	(0.090)
Post × EIP amount	−0.012***		−0.015	
	(0.003)		(0.011)	
Post× Total income		−0.006***		−0.019**
		(0.002)		(0.008)
Post × Child care credit		0.001		−0.022*
		(0.002)		(0.013)
Constant	0.291***	0.288***	−2.084***	−2.120***
	(0.003)	(0.002)	(0.019)	(0.018)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.280	0.280	0.070	0.070
Magnitude (↑1 SD iOS share)	-2.7%	-2.1%	-6.6%	-4.7%
Observations	876,414	876,414	876,288	876,288
R^2 / Pseudo R^2	0.512	0.512	0.133	0.133

Panel b. Placebo Tests

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post-EIP 1st × iOS share	0.005		−0.170	
	(0.014)		(0.146)	
Post-EIP 2st × iOS share		0.007		0.059
		(0.019)		(0.182)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.250	0.250	0.065	0.065
Magnitude (↑1 SD iOS share)	0.2%	0.3%	-1.7%	0.6%
Observations	617,848	617,848	617,848	617,848
R^2 / Pseudo R^2	0.504	0.504	0.131	0.131

This table presents the estimated effects of ATT on CFPB complaints by controlling for COVID-19 stimulus payments (Panel a) and reports results from placebo tests (Panel b). The unit of observation is at the zip-code-month level. Columns 1 and 2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3 and 4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Panel a controls for exposure to economic impact payments (EIPs) by adding the interaction terms between the post-ATT indicator and variables related to EIPs in the regression. The variables include total amount of EIPs received (Columns 1 and 3), the average household income, and the amount of childcare credit (Columns 2 and 4), all constructed using IRS data. Panel b considers placebo tests using two alternative event dates: April 2020 and December 2020. These dates correspond to the disbursement of the first and second major waves of pandemic-related financial support under the CARES Act and subsequent legislation. The sample period is January 2019 to June 2022 in Panel a. In Panel b, the sample is restricted to the pre-ATT period (January 2019-April 2021) to avoid contamination from ATT's actual treatment effect. All columns include zip code fixed effects and county × year-month fixed effects. 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: Effect of ATT on Consumer Complaints

More vs. Less Relevant Categories of Complaints—Classified by Narratives

Panel a. CFPB

	# Complaints per 1,000 residents			
	Top 2 Fraud Categories		Bottom 2 Fraud Categories	
	(1) Credit Reporting and Repair	(2) Debt Collection	(3) Student Loan	(4) Mortgage
Post × iOS Share	−0.574***	−0.600***	0.019	−0.285
	(0.163)	(0.178)	(0.597)	(0.209)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML
Mean outcome var.	0.022	0.008	0.002	0.005
Magnitude (↑1 SD iOS share)	-5.5%	-5.7%	0.2%	-2.8%
Observations	782,292	782,292	782,292	782,292
Pseudo R^2	0.138	0.088	0.059	0.072

Panel b. Identity Theft and Consumer Sentinel Network

	# Complaints per 1,000 residents			
	Narratives w/ Keywords		Narratives w/o Keywords	
	(1) Identity Theft	(2) CSN	(3) Identity Theft	(4) CSN
Post × iOS Share	−0.745***	−0.274***	−0.001	−0.042
	(0.154)	(0.060)	(0.142)	(0.043)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML
Mean outcome var.	0.106	0.276	0.139	0.472
Magnitude (↑1 SD iOS share)	-7.2%	-2.7%	0.0%	-0.4%
Observations	800,386	1,034,586	880,678	1,089,791
Pseudo R^2	0.199	0.137	0.214	0.117

This table presents the estimated effects of ATT on consumer complaints after classifying complaints by their likelihood to be affected by ATT. The unit of observation is at the zip-code-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 Identity Theft and Consumer Sentinel Network complaints to classify complaints into those with and without any relevant words in the narrative. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Effect of ATT on Firm-Level Cyber Incidents and Consumer Complaints

Panel a. Cyber Incidents

	Any cyber incidents (0/1)			
	(1) All	(2) Breach/Data Misuse	(3) Debt Collection/Credit Reporting Regulation	(4) Other Causes
Has an app × Post	-0.049***	-0.046***	-0.009***	-0.003
	(0.010)	(0.009)	(0.003)	(0.004)
Firm FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Mean outcome var.	0.097	0.072	0.015	0.021
Observations	306,933	306,933	306,933	306,933
R^2	0.183	0.185	0.198	0.123

Panel b. Consumer Complaints

	Any complaints (0/1)				
	(1) All	(2) Credit Reporting and Repair	(3) Debt Collection	(4) Student Loan	(5) Mortgage
Has an app × Post	-0.021***	-0.031***	-0.024***	-0.001	-0.001
	(0.006)	(0.005)	(0.005)	(0.001)	(0.003)
Firm FE	✓	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓	✓
App-size-specific linear trend	✓	✓	✓	✓	✓
Mean outcome var.	0.037	0.037	0.079	0.006	0.030
Observations	394,243	394,243	394,243	394,243	394,243
R-square	0.396	0.316	0.352	0.423	0.475

This table presents the estimated effects of ATT on cyber incidents (Panel a) and firm-level consumer complaints received via CFPB (Panel b). The unit of observation is at the firm-month level. In Panel a, 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 violated the Fair Debt Collection Practices Act or the Fair Credit Reporting Act, and (4) cyber incidents that were caused by other reasons unrelated to data breach. All columns report linear probability model estimates for the interaction term between the post-ATT indicator and an indicator for whether the firm owns an app. In Panel b Column 1, the outcome variable is an indicator variable for whether the firm was exposed to any cyber incident. In Panel b Column 2-5, the outcome variable is an indicator for whether the firm was exposed to any cyber incident in the two most relevant and two least 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). All columns include firm fixed effects and county × 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.

Table 8: Effect of ATT on Consumer Complaints

Heterogeneity by Local Exposure to Data Practices of Banks and ISPs

Panel a. Data Practices of Banks

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post × iOS share	−0.059***	−0.060***	−0.597***	−0.603***
	(0.015)	(0.015)	(0.091)	(0.094)
Post × iOS share × Exposure (data collection)	−0.042***		−0.167*	
	(0.012)		(0.087)	
Post × iOS share × Exposure (data sharing)		−0.040***		−0.189**
		(0.014)		(0.086)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.267	0.267	0.068	0.068
Magnitude (↑1 SD iOS share)	-2.4%	-2.4%	-5.8%	-5.9%
Magnitude (↑1 SD iOS share×Exposure)	-1.7%	-1.6%	-1.7%	-1.9%
Observations	925,428	925,428	925,428	925,428
$R^2/PseudoR^2$	0.514	0.514	0.131	0.131

Panel b. Data Practices of ISPs

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post × iOS share	−0.060***	−0.059***	−0.604***	−0.598***
	(0.015)	(0.015)	(0.095)	(0.092)
Post × iOS share × Exposure (data collection)	−0.040***		−0.197**	
	(0.014)		(0.087)	
Post × iOS share × Exposure (data sharing)		−0.042***		−0.174*
		(0.012)		(0.089)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.267	0.267	0.069	0.069
Magnitude (↑1 SD iOS share)	-2.4%	-2.4%	-5.9%	-5.9%
Magnitude (↑1 SD iOS share×Exposure)	-1.6%	-1.7%	-2.0%	-1.7%
Observations	925,428	925,428	925,428	925,428
R^2 / Pseudo R^2	0.514	0.514	0.134	0.134

This table reports the heterogeneous effects of ATT on consumer complaints by local exposure to the data practices of banks and ISPs. The unit of observation is at the zip-code-month level. Columns 1 and 2 of both panels report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3 and 4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The variables used in the triple interaction terms capture heterogeneity in the data collection practices of local banks and ISPs across counties. In Panel A, *Exposure (data collection/sharing)* represents the download-weighted number of data types collected or data SDKs used by all banks that own a mobile app as of April 2021. The precise definitions are provided in Equation (5). In Panel b, *Exposure (data sharing)* refers to the market-share-weighted number of data SDKs used by ISPs in their mobile applications. *Exposure (data collection)* represents the market-share-weighted number of data types collected by the Top 12 fixed broadband ISPs, as disclosed in their privacy labels. The sample period is January 2019 to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Effect of ATT on Dark Web Activities

Panel a. Posts on Dark Web Forums

	# Posts		
	(1)	(2)	(3)
Post × Treated	−0.413***		
	(0.024)		
Post × Partially treated		−0.241*	
		(0.126)	
Post × Placebo			−0.006
			(0.173)
Label FE	✓	✓	✓
Forum × Year-month FE	✓	✓	✓
Model	PPML	PPML	PPML
Mean outcome var.	46.0	60.4	101.8
Magnitude (Treated vs. Control)	-33.8%	-21.5%	-0.6%
Observations	27,612	43,719	57,525
Pseudo R^2	0.871	0.863	0.879

Panel b. Listing Price of Data

	Listing price per unit (log)			
	Worldwide		US	
	(1)	(2)	(3)	(4)
Post × Mobile footprint data	0.474***		0.520***	
	(0.090)		(0.080)	
Post × Financial info data		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^2	0.862	0.863	0.636	0.636

This table presents the estimated effects of ATT on upstream dark web activities. Panel a focuses on posts on dark web forums and the unit of observation is at forum-label-month level. The full list of labels and their exposure to ATT can be found in Appendix H. The outcome variable is the number of posts. Column 1 examines the explicitly treated group (Apple/iOS-related posts), Column 2 the partially treated group (categories reliant on personal or financial data such as tracking, breaches, identity theft, financial data theft), and Column 3 a placebo group (e.g., zero-days, malware, botnets). All groups are compared to a control group of illicit trade categories (e.g., drugs, human trafficking, weapons) operating in separate markets. Standard errors are double clustered at the forum and label level and reported in parentheses. The sample period is January 2015 to December 2022. Panel b 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, with details described in Appendix I. 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. One snapshot of listings is in 2020 and the other one is in 2023. Columns 1-2 use all the listings, while Columns 3 and 4 use only listings in which the price is quoted in USD. Company and year fixed effects are included in all columns. We additionally include currency fixed effects in Columns 1 and 2. Standard errors clustered at the brand level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Internet Appendix to "Consumer Surveillance and Financial Fraud"

This Appendix has 10 sections. Section A contains information on policy shocks in the mobile app economy and the construction of our data collection/sharing intensity variables. Section B presents an example form for consumer complaints. Section C adds details on the Consumer Sentinel Network and Identity Theft supplementary databases from the FTC. Section D discusses irregularities in the CFPB complaints data and methods to deal with outliers. Section E covers background information on the 2017 repeal of the FCC broadband privacy rule. Section F presents the classification methods of fraud-related complaints and example narratives on data-driven fraud incidents. Section G contains information on data on cyber incidents from Advisen. Section H covers information on the posts from dark web forums, including classification methods, summary statistics, and examples. Section I describes the data on dark net listings related to data and the analysis. Section J covers additional analysis, including tables and figures.

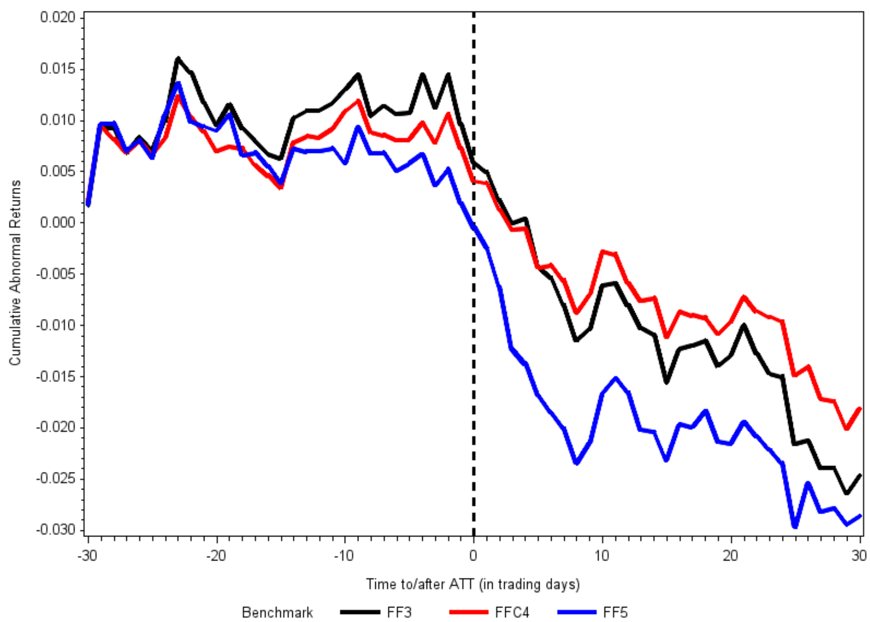
List of Appendix Figures

List of Appendix Tables

A Background: Data Collection and Sharing in the Mobile Economy

A.1 The App Tracking Transparency and Privacy Nutrition Label Policies

Figure A.1: Stock Market Reactions around ATT



This figure is from Bian et al. (2021). This figure plots the average cumulative abnormal returns (CARs) around the implementation of the App Tracking Transparency 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.

A.2 Measuring Data Collection Intensity via Privacy Nutrition Labels

Following Bian et al. (2021), we use iOS app privacy nutrition labels to quantify the data collection intensity of firms that own these apps. Since December 14, 2020, Apple has required all app developers to disclose their data collection practices in a standardized and accessible format. An example of such a privacy label can be found at: `https://apps.apple.com/kg/app/facebook/id284882215`.

The standardized structure of these labels enables consistent comparison across firms. Specifically, we measure data collection intensity by counting the number of unique data types an app collects across all declared purposes (i.e., the second layer of the label). These data types correspond to the third-layer items illustrated in Figure A.2, including categories such as identifiers, location, and financial information.

Identifying Mobile Presence of ISPs and Banks. For banks, we identify mobile app presence using the following procedure:

1. We obtain an exhaustive list of iOS apps in the *Finance* category from SensorTower and identify the app with the highest number of downloads as the flagship app for each financial institution.

2. We extract the URL of the official website of each bank from SensorTower. This information is typically found under the `"Developer Website ↗"`, `"App Support ↗"`, and `"Privacy Policy ↗"` links at the bottom of the "Information" section of the app's page. See `https://apps.apple.com/us/app/citi-mobile/id301724680` for an example.

3. We also collect the official website URLs of banks (field: from the FDIC database `WEBADDR`) .

4. We apply a standard string-cleaning procedure to both sets of URLs, such as removing protocol prefixes like "https://" and "http://".

5. To supplement URL-based matching, we perform name matching using developer names from SensorTower and registered bank names from the FDIC. We standardize names by removing common suffixes such as "THE", "NATIONAL ASSOCIATION", and "LTD".

In total, we identify 2,618 banks that own a mobile app that remained active as of the data collection date.

For ISPs, given the high concentration in ISP markets, we manually identify the 12 largest app-owning ISPs, listed in Panel B of Figure A.3, which together accounted for 85% of the market as of June 2016. Since ISPs often operate multiple apps, we compute the average of relevant metrics across all apps owned by a given ISP. For example, Comcast owns 248 apps and Verizon 124, corroborating their central position in the digital surveillance economy.

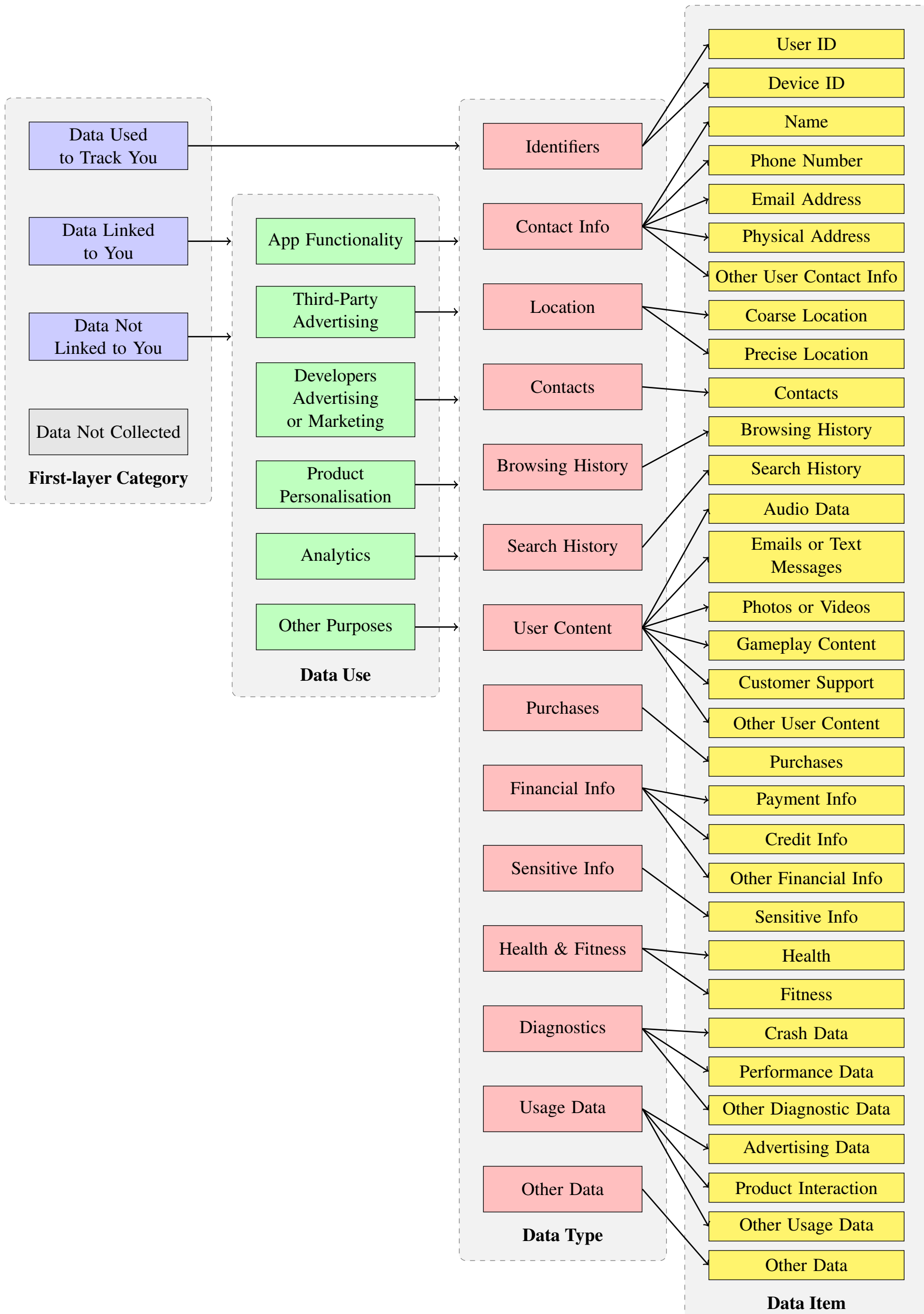
A.3 Measuring Data Sharing Intensity via Third-Party Data Sharing SDKs

Following Bian et al. (2024), we construct a measure of data sharing intensity based on the use of third-party Software Development Kits (SDKs), using data from Apptopia. SDKs are pre-built software components that provide functionalities such as analytics and targeted advertising. We identify SDKs that facilitate data sharing across firms by linking user activity through unique identifiers, thereby enabling the creation of detailed consumer profiles. Firms integrate these SDKs into their apps to gain access to such profiles, which can inform marketing strategies, product development, and other business decisions.

We measure data sharing intensity as the number of data-sharing SDKs used in a firm's iOS apps. When a firm owns multiple apps, we take the average number of such SDKs across its app portfolio.

The distributions of data collection and sharing intensity across the top 12 ISPs and banks are provided in Figure A.3.

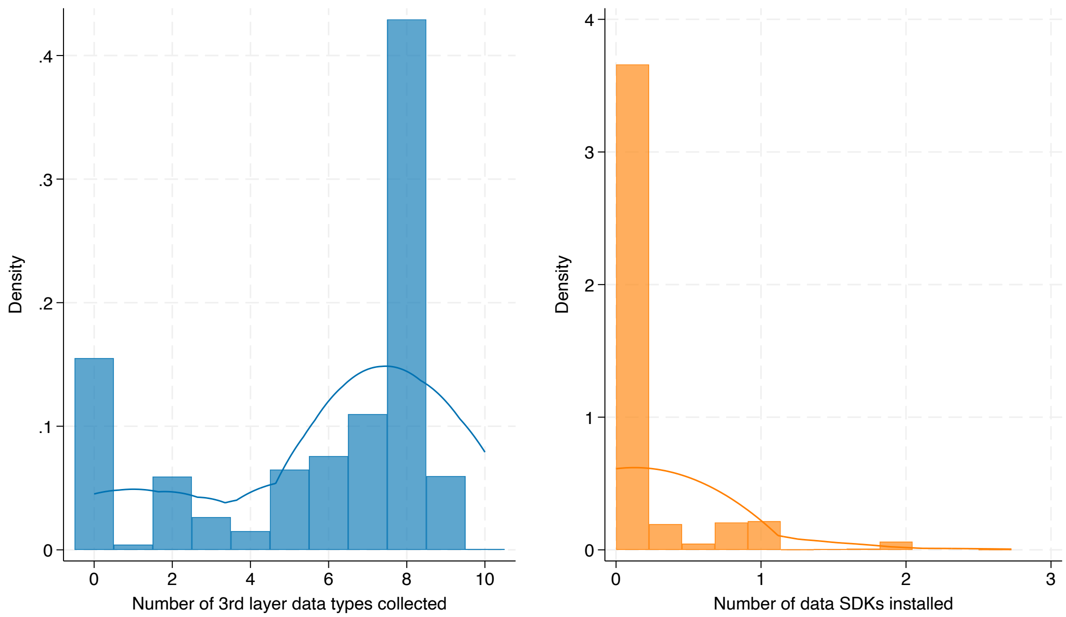
Figure A.2: Apple Privacy Nutrition Label



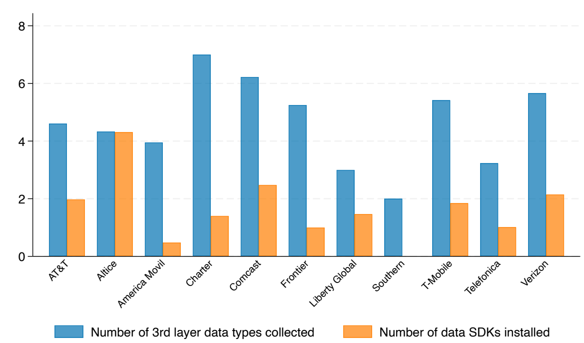
This figure shows the structure of the mandatory privacy nutrition label from Apple for iOS apps. Apple privacy label has four layers. The first layer consists of three categories: *Data Used to Track You*, *Data Linked to You*, and *Data Not Linked to You*. If an app doesn't collect any data, it will 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.

Figure A.3: Distributions of Data Collection & Sharing Intensity

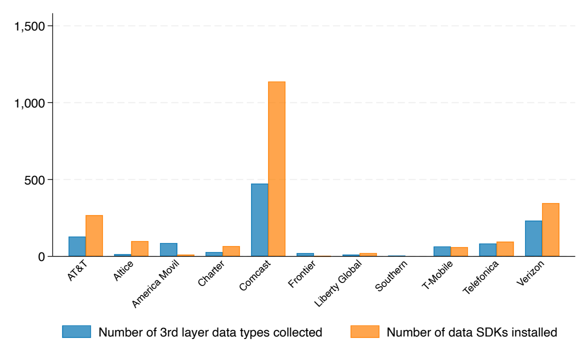
a. Banks



b. ISPs - Average Across Apps



c. ISPs - Total Across Apps



This figure presents the distribution of data collection intensity, measured by the number of data types collected (blue bars), and data sharing intensity, measured by the number of data-related SDKs used (orange bars). Panel (a) shows the distributions for banks, with the horizontal axis showing the number of data types collected or SDKs installed. Panels (b) and (c) display the average and total data collection and sharing intensity across app portfolios of the 12 ISPs with the largest market shares.

B Consumer Sentinel Complaint Data Fields

Untitled Page Page 1 of 2

CONSUMER SENTINEL NETWORK
Law enforcement's source for consumer complaints

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:**	

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

Untitled Page Page 2 of 2

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:**	

Provided by the Federal Trade Commission

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

C Supplementary Databases Obtained from the FTC

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.[1] 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 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 unlikely to be affected by ATT.

For analysis in Table J.8, 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,

[1] 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. 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.

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. 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.

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

D Dealing with Irregularities in CFPB Complaints Data

This section documents our approach to identifying irregularities in complaint activity. We remove zip codes that exhibit persistently extreme levels or volatility in consumer complaints, as these units can exert disproportionate influence on estimates and obscure broader patterns. Our filtering strategy is based on multiple criteria, such as persistently high complaint levels, frequent large monthly changes, and high long-run volatility, all measured per 1,000 residents to account for differences in population size. Zip codes are classified as outliers if they satisfy at least two of the four conditions. We show that these excluded units differ markedly in both temporal behavior and socioeconomic composition, often concentrated in specific geographic clusters with elevated complaint activity.

While one alternative would be to winsorize the outcome variable, we find that aggressive right-tail winsorization (e.g., at the 5% or 10% level) yields results similar to our main estimates. This reinforces the pattern documented below, suggesting that the distortion is not driven by a few isolated outliers, but rather by a small subset of structurally atypical zip codes. Removing these units, therefore, offers a cleaner and more interpretable estimation sample without materially affecting our substantive conclusions.

Procedure to Identify Outliers. To mitigate the influence of highly skewed complaint volumes and extreme outliers, we identify and exclude a subset of zip codes based on four criteria. All criteria are based on monthly complaint volumes scaled by population and are designed to capture persistent or structurally atypical patterns rather than transient noise. A zip code is flagged as an outlier if it satisfies at least two of the following conditions:

1. **Frequent large changes in monthly volume:** The absolute month-over-month change exceeds the 90^{th} percentile in more than 10 months.

2. **Frequent extreme complaint levels:** Monthly complaint volume exceeds the 90^{th} percentile in more than 10 months.

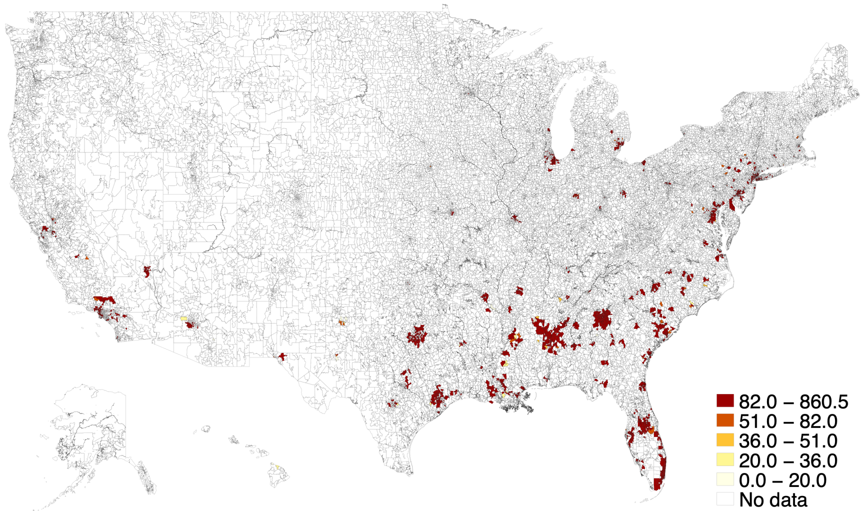
3. **High volatility:** Standard deviation of monthly complaints (over the full sample) exceeds the 90^{th} percentile.

4. **High average level:** Average monthly complaint volume (over the full sample) exceeds the 90^{th} percentile.

Our filtering strategy aims to remove zip codes that show persistently high levels or excessive volatility, which can disproportionately influence regression results and mask broader trends. This approach is more robust than standard winsorization, as it targets structural outliers rather than individual extreme values. We then classify as outliers the zip codes that satisfy at least two of the four conditions. This yields a total of 2,147 zip codes, while our main regression sample after excluding the outliers includes 24,237 zip codes. The geographic distribution of outlier zip codes is plotted in Figure D.1. Outliers exhibit excessively high complaint intensity and are concentrated in distinct metropolitan clusters, particularly across the Southeast, mid-Atlantic, and parts of California and Texas. Notable concentrations appear in urban areas such as Los Angeles, Miami, Atlanta, Washington, D.C., and throughout central and northern Florida.

Table D.1 summarizes the distribution of consumer complaints per 1,000 residents across

Figure D.1: Number of complaints per Million Residents
Outlier Sample



This figure presents the number of CFPB complaints per million residents for the outlier zip codes.

the regression sample and the excluded outlier sample. Zip codes in the outlier group are substantially more likely to report complaints in a given month (mean of 0.83 vs. 0.26) and exhibit significantly higher complaint volumes, with a mean of 0.35 complaints per 1,000 residents compared to 0.07 in the main sample. The outlier sample also shows markedly greater dispersion, with the 99$^{\text{th}}$ percentile complaint rate reaching 1.95, nearly triple that of the main sample (0.74).

Table D.1: The Number of Complaints per 1,000 Residents
Regression Sample vs. Outlier Sample

	mean	sd	p25	p50	p75	p90	p99	count
Regression Sample								
Any complaints (0/1)	0.26	0.44	0.00	0.00	1.00	1.00	1.00	927,780
# Complaints per 1,000 residents (raw)	0.07	4.65	0.00	0.00	0.03	0.14	0.74	927,780
Outlier Sample								
Any complaints (0/1)	0.83	0.38	1.00	1.00	1.00	1.00	1.00	90,174
# Complaints per 1,000 residents (raw)	0.35	5.99	0.05	0.17	0.35	0.63	1.95	90,174

This table reports summary statistics for our two main outcome variables: whether a zip code has at least one complaint and the total number of complaints per 1,000 residents, separately for the regression sample and the outlier sample.

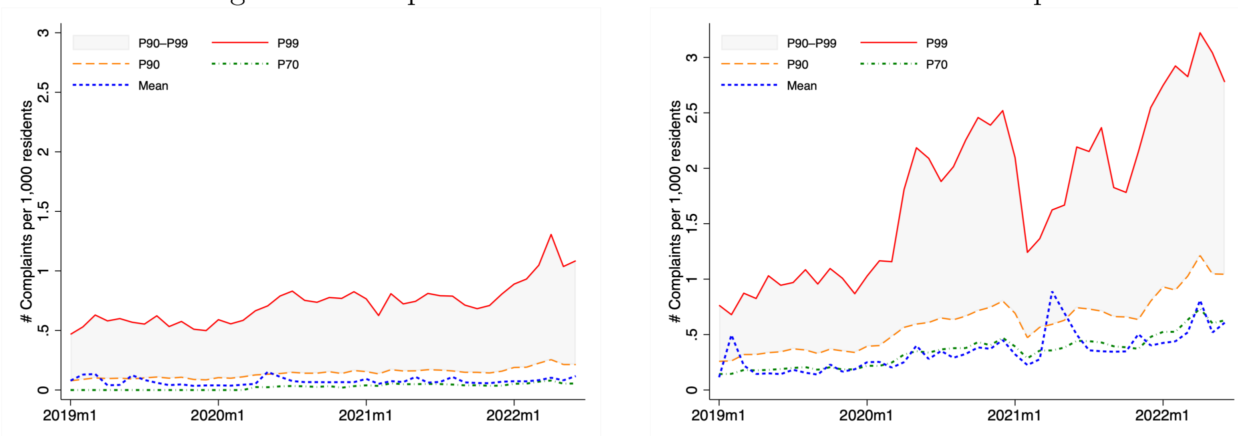
Time-Series of Complaints for Outlier Zip Codes.

To further understand the source of the dispersion, we examine the time series of complaints. Figure D.2 compares the time series of population-scaled complaint volumes between the regression sample (Panel a) and the outlier sample (Panel b). While the central tendency (mean, 70^{th} , and 90^{th} percentiles) remains relatively stable in both groups, the outlier sample exhibits substantially greater upper-tail variation, as reflected in the rising and more volatile 99^{th} percentile. In particular, the P99 trajectory in the outlier sample diverges sharply after mid-2020, indicating persistent and growing concentration of complaints in a subset of zip codes.

Motivated by the geographical clustering of outlier zip code, we zoom into twelve states that are most representative of the irregular complaint patterns observed in the data. Figure D.3 displays time series of population-scaled complaint volumes for these twelve states, focusing on zip codes identified as outliers. The y-axis ranges vary substantially across states, with several panels, such as Indiana and North Carolina, showing complaint spikes at the 99^{th} percentile exceeding 800 complaints per 1,000 residents. Even states with relatively low averages, such as Arizona and Nevada, exhibit repeated short-term surges, highlighting the volatility that characterizes these outlier zip codes. For comparison, in the regression sample (Figure D.2), the 99^{th} percentile rarely exceeds 1, underscoring the disproportionate scale of complaint activity in the excluded zip codes. These patterns suggest that outlier zip codes are not only defined by persistently elevated levels, but also by abrupt, episodic deviations. Their exclusion helps prevent such extreme behavior from distorting estimates in the main analysis.

Figure D.2: Time Series of the Number of Complaints per 1,000 Residents

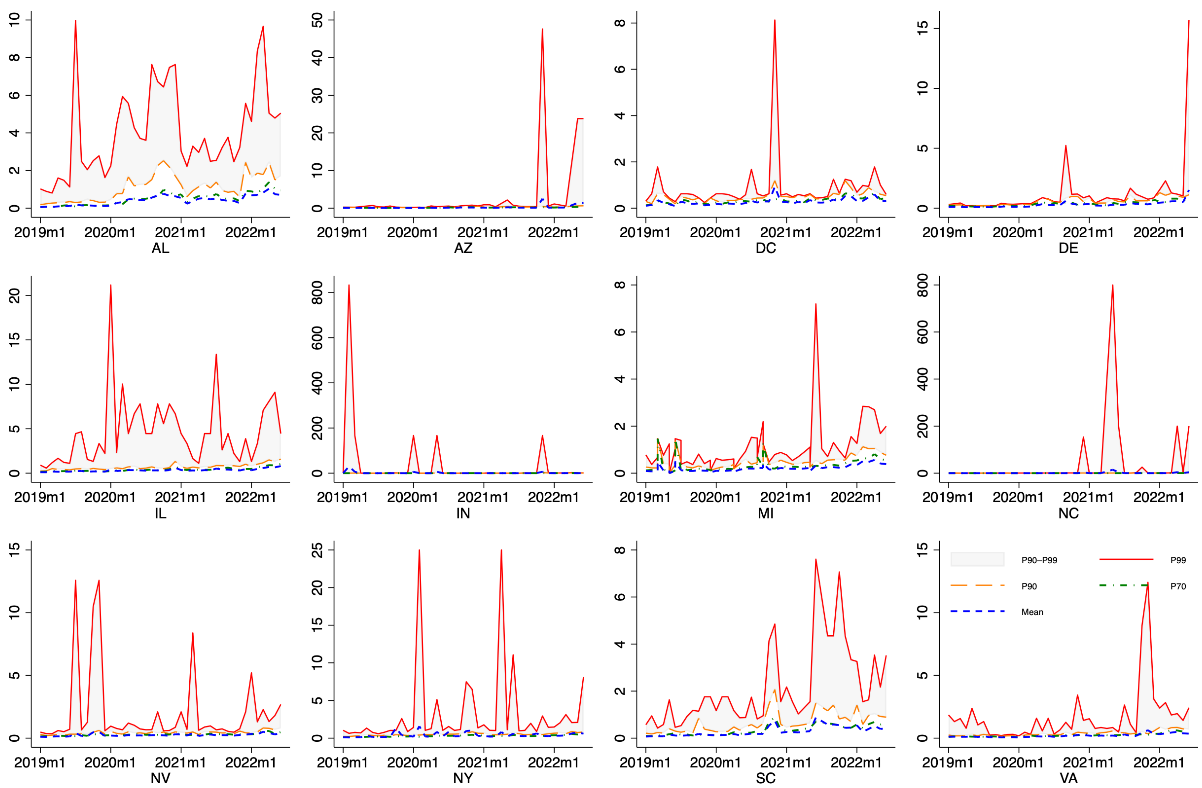
a. Regression Sample b. Outlier Sample



This figure compares various statistics of the number of complaints (scaled by population) between the regression sample and the outlier zip code sample. For each month in the sample period (January 2019 to June 2022), we compute the mean, 75th percentile, 90th percentile, and 99th percentile across zip codes.

Comparison of Zip Codes in Outlier Sample vs. Regression Sample. Finally, Table D.2 compares zip code-level socioeconomic characteristics between the regression and outlier samples. Outlier zip codes differ systematically from those in the main sample along several dimensions. They tend to have slightly lower iOS usage, higher shares of young adults (age 20-39), and a greater proportion of female residents. Outlier zip codes also have significantly higher unemployment rates and lower median income levels. These areas are less likely to have older populations (age 50+) and show a modestly higher share of residents with a bachelor's degree.

Figure D.3: Time Series of the Number of Complaints per 1,000 Residents
Outlier Sample



This figure presents various statistics for outlier zip codes by state. We display twelve states that are most representative of the irregular complaint patterns observed in the data. For each month in the sample period (January 2019 to June 2022), we compute the mean, 75th percentile, 90th percentile, and 99th percentile across zip codes.

Table D.2: Regression Sample vs. Outlier Sample

	(1) Regression sample		(2) Outlier sample		(3) Outlier − Main	
	mean	sd	mean	sd	b	t
iOS share	0.46	0.11	0.45	0.10	-0.01***	(-6.18)
Female	0.50	0.04	0.52	0.03	0.02***	(23.11)
Age 10-19	0.13	0.04	0.13	0.03	-0.00	(-0.13)
Age 20-29	0.11	0.06	0.14	0.05	0.03***	(25.80)
Age 30-39	0.12	0.03	0.15	0.04	0.03***	(30.43)
Age 40-49	0.12	0.02	0.12	0.02	0.01***	(20.15)
Age 50+	0.41	0.10	0.34	0.08	-0.07***	(-39.33)
Unemployment rate	0.05	0.04	0.07	0.04	0.02***	(20.52)
Bachelor	0.27	0.17	0.31	0.17	0.04***	(9.86)
Median income	65,962	27,5176	63,043	25,342	-2,919***	(-5.02)
Observations (# ZIP codes)	22,090		2,147		24,237	

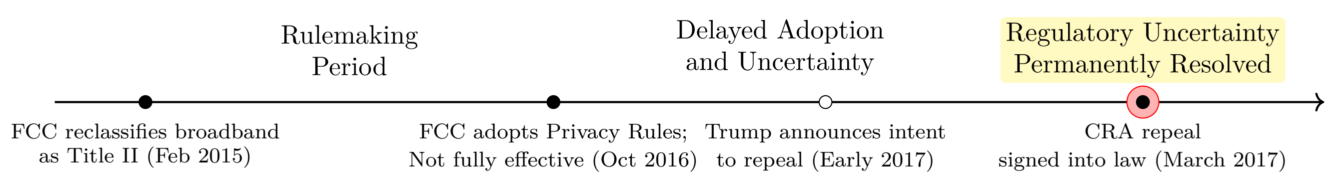
This table reports two-sample *t*-tests comparing socioeconomic factors between the regression sample and the outlier sample of zip codes.

Extreme complaint intensity in certain zip codes may arise from small population denominators, concentrated legal aid activity, or localized service disruptions. As such patterns often reflect structural or institutional dynamics rather than random variation, their exclusion provides a cleaner estimation sample and strengthens the interpretability of aggregate results.

E 2017 FCC Privacy Rule Repeal—Timeline and Nature of the Shock

This Appendix Section provides additional detail on the timeline and interpretation of the FCC privacy rule repeal as a regulatory shock to consumer surveillance capacity. We also elaborate on the distinctive role of ISPs in the modern data economy, motivating their centrality in our empirical design.

Timeline of Events.



In 2015, the FCC reclassified broadband internet as a telecommunications service under Title II of the Communications Act, thereby transferring privacy oversight authority from the Federal Trade Commission (FTC) to the FCC. In October 2016, the FCC formally adopted the Broadband Privacy Rules, which would have required ISPs to obtain opt-in consent before using or sharing sensitive customer data. These rules were scheduled to take effect in phases starting in 2017, but were repealed before full implementation. The repeal was enacted in April 2017 through the Congressional Review Act (CRA), which also barred the FCC from issuing substantially similar rules in the future absent new congressional authorization.

Nature of the Shock: A Positive Shift in Surveillance Capacity. Although the FCC's rules never fully took effect, ISPs faced regulatory overhang during 2015-2017, which limited their ability to expand surveillance practices. The 2017 repeal abruptly removed these constraints, restoring FTC jurisdiction and granting ISPs greater discretion under a less prescriptive enforcement regime. In contrast to the FCC's ex-ante consent requirements, the FTC framework relies primarily on ex-post enforcement against unfair or deceptive practices.

The CRA's prohibition on reissuing substantially similar regulations further signaled that stringent privacy rules were unlikely to return in the near term. This combination of deregulatory policy and legal durability constitutes a clear and plausibly exogenous positive shock to the permissible scope of ISP surveillance.

The Central Role of ISPs in the Data Economy. ISPs occupy a structurally privileged position in the digital ecosystem due to their role as the conduit through which consumers access the internet. Unlike individual web platforms or applications, ISPs have the technical capability to observe nearly all unencrypted internet traffic originating from a subscriber, such as which websites a user visits, regardless of the destination or service used. This includes domain-level browsing activity, app usage patterns, device identifiers, and potentially fine-grained geolocation, especially for mobile broadband.

Several features distinguish ISP-collected data in the broader data economy. First, ISPs can track user behavior across websites, apps, and devices, in contrast to platforms like Google or Facebook whose visibility is limited to logged-in users or embedded services. Second, ISP data is typically tied to long-lived customer accounts or IP address allocations, enabling persistent tracking over time and facilitating linkage to household characteristics. Third, broadband markets in the U.S. are often characterized by limited competition, which reduces consumers' ability to avoid surveillance by switching providers.

These properties make ISP data particularly valuable for targeted advertising, profiling, and data brokerage. As such, ISPs are central—not peripheral actors in the contemporary surveillance economy, especially following the 2017 repeal, which relaxed the regulatory restrictions on how such data could be monetized. Bian et al. (2024) shows that ISPs exhibit the highest level of data connectedness with other firms in the data economy, as measured by the overlap in data-related SDKs.

While the FCC privacy rules applied to both fixed and mobile broadband ISPs, we focus on fixed ISPs for two reasons. First, fixed ISPs are the dominant providers of household internet access and account for the majority of internet traffic volume, particularly in the study period (2015–2017). Second, market-level variation in surveillance practices is more pronounced and measurable among fixed ISPs, enabling us to construct a granular exposure measure. To the extent that mobile ISPs also expanded surveillance post-repeal, our estimates may understate the full effect of the policy.

F Classification of Fraud-related Complaints

Table F.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 cer-

Table F.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

This table reports the likelihood of relevant fraud cases by product category based on 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, sub-issue, 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 is the number of observations in each product category. The sample only includes complaints with narratives, and about 40% of complaints have narratives.

tain keywords in the issue, sub-issue, 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, sub-issue, 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, which 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 manually annotated dummy score. The estimated model has a non-zero coefficient for 17 out of the possible 51 features (3×17), with 10 entailments, 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 complaints, we obtain the following out-of-sample statistics: an 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 and identity theft and that the unverified inquiries/accounts/debts 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 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 have a less than 25% 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.

G 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 system 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 are more likely to be affected by ATT (as highlighted in bold). The share of each case type is reported in 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 laws and regulations. 5,566 or 10.89% of incidents are recorded as leading 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 America National Association (44 incidents), and Portfolio Recovery Associates LLC (35 incidents). These observations suggest that incidents that resulted in violations of FDCPA are more likely to result in fraud that involves financial companies.

H Dark Web Forum Posts: Classification, Statistics, and Examples

We use data from CrimeBB, a structured collection of posts scraped from 37 dark web forums by the Cambridge Cybercrime Centre (CCC). CCC started their work on 1 October 2015. Access to their dataset is restricted to accredited academic researchers under the CCC's legal and ethical framework, requiring formal applications and compliance with their strict data-handling agreements. We report the volume of posts by each forum in Table H.1.

Classification. We first present the classification of posts into different labels (Avarikioti et al., 2018)) and groups, including treated (Table H.2), partially treated (Table H.3), control (Table H.4), placebo (Table H.5), and excluded (Table H.6). Both the full keyword list and the narrower, more precise keyword lists are presented in these tables. The treated group consists of posts explicitly related to Apple devices or the iOS ecosystem, which are directly exposed to ATT's impact. The partially treated group includes posts covering illicit activities that plausibly depend on access to personal or financial data—such as tracking, data breaches, identity theft, and financial data theft—and could therefore be indirectly affected by ATT through reduced availability of such data.

Table H.1: Total Posts and Average Monthly Posts by Forum

Forum	Total Posts	Avg. Monthly Posts
Antichat	44,133	1,583
Blackhatworld	2,486,478	38,915
Breached	136,991	24,560
Cracked	174,050	10,168
Crackedto	50,149	6,310
Deutschland-Im-Deep-Web	13,568	1,621
Dread	148,875	48,224
Elhacker	63,044	1,508
Envoy-Forum	1,412	779
Forum-Team	12,301	447
Freehacks	2,398	54
Garage-For-Hackers	327	22
Greysec	5,599	403
Hack-Forums	2,998,744	109,157
Hackers-Armies	1,725	74
Ifud	5,090	170
Indetectables	8,279	441
Kernelmode	3,614	186
Lolzteam	4,270,894	441,309
Mmo4Me	733,294	13,831
Multiplayer-Game-Hacking	982,005	24,998
Nulled	818,866	28,121
Offensive-Community	71,281	32,319
Ogusers	1,322,813	123,882
Piratebay-Forum	13,566	835
Probiv	396,350	12,855
Raidforums	117,860	6,972
Runion	43,996	1,698
Safe-Sky-Hacks	15,893	5,113
Stresser-Forums	2,912	366
The-Hub	26,410	1,598
Torum	17,313	3,846
Underc0De	28,976	736
Unknowncheats	991,618	21,791
V3Rmillion	613,149	28,009
Xss-Forum	33,966	1,738
Zismo	509,137	20,238

The table presents total posts and average monthly posts by forum over the period 2015–2022. A total of 37 forums are scraped and covered by CCC during this period.

The control group contains illicit-trade categories, such as drugs, human trafficking, and weapons, that operate in segmented markets with minimal overlap in actors or infrastructure with the data-

Table H.2: Classification of Dark Web Forum Posts - Treated

Label	Full Keyword List	Precise Keyword List
iOS/Apple	ios, iphone, apple, apple id, icloud, idfa, app store, apple pay, jailbreak, mdm bypass, apple watch	ios, iphone, apple

Table H.3: Classification of Dark Web Forum Posts - Partial Treated

Label	Full Keyword List	Precise Keyword List
Tracking & AdTech & ATT	att, app tracking, tracking transparency, app tracking transparency, cross-app tracking, sdk, mobile attribution, spyware, device fingerprinting, tracking app, tracking id, advertising id, attribution	att, app tracking, tracking transparency
Data Breaches & Dumps	leak, data dump, database leak, db dump, combo list, combolist, logs, fullz, dox, cracked db, stealer logs, public dump, private dump	leak, data dump, database leak
PII & Identity Theft	ssn, dob, driver's license, passport, email+pass, identity, kyc, selfie with id, gov id, biometric, facial recognition, personal data	ssn, dob, driver's license
Phishing & Exploits	phishing, spoofing, keylogger, rat, stealer, scraper, clone app, malware, fake update, exploit, 0day, drive-by download	phishing, spoofing, keylogger
Financial Data Theft	bank login, paypal, credit card, cvv, bin, apple pay, crypto wallet, account takeover, 2fa, sms bypass, otp	bank login, paypal, credit card
Scam & Fraud	scam, fraud, rip, ripoff, fake, counterfeit, trick, deception, fake review, fake escrow	scam, fraud, rip
Personal Security & Privacy	opsec, operational security, privacy, anonymity, encryption, vpn, tor, secure messaging	opsec, operational security, privacy
Stolen Accounts & Credentials	account, login, credential, password, username, email+pass, netflix, spotify, social media account	account, login, credential

driven fraud economy, providing a benchmark insulated from ATT-related shocks. The placebo group includes unrelated digital technology categories such as zero-days, malware, and botnets, which do not rely on personal data collection for their operation and thus serve to detect whether observed effects reflect broader shifts in cyber-related discussion rather than ATT. All remaining categories, including those with ambiguous or mixed exposure, are assigned to the excluded group. We provide summary statistics for the number of posts for each group in each forum in Table H.7.

Examples. We next provide a few examples, showcasing the discussions in both the (partially) treated and control groups.

A user named "sludgepuppy" posted on Dread in a thread about phishing Apple logs (post label: iOS/Apple, treated):

> Ok so ive successfully phished some apple logs, they're full info including ssn, mmn, dob, live cvv and email access. through email access i gained license front and back as-well as license plate number and other tax info. The original goal was to use it with apple pay but

Table H.4: Classification of Dark Web Forum Posts - Control

Label	Full Keyword List	Precise Keyword List
Hardware & Electronics	hardware, electronics, device, component, circuit, mining rig, usb, phone, computer	hardware, electronics, device
Drugs	cannabis, weed, cocaine, heroin, lsd, mdma, ecstasy, fentanyl, oxy, xanax, stimulants, opiates, psychedelics, darknet drugs	cannabis, weed, cocaine
Weapons & Explosives	gun, firearm, pistol, rifle, shotgun, ammo, ammunition, explosive, bomb, grenade, c4, tnt, ar-15, ak-47	gun, firearm, pistol
Counterfeit & Forged Items	fake id, counterfeit, forged, replica, passport, driver's license, currency, credit card, branded goods	fake id, counterfeit, forged
Child Exploitation Material (CSAM)	child porn, csam, child sexual abuse material, cp, illegal images, underage	child porn, csam, child sexual abuse material
Guides & Tutorials	guide, tutorial, how to, guide for, handbook, manual, wiki, method, tips, tricks	guide, tutorial, how to
Human Trafficking	human trafficking, slavery, forced labor, organ trafficking, sex trafficking, exploitation	human trafficking, slavery, forced labor
Leaks & Whistleblowing	wikileaks, leak, whistleblower, dump, disclosure, anonymous tip	wikileaks, leak, whistleblower
Vendors & Markets	vendor, market, shop, store, seller, buyer, product list, listing, escrow, review	vendor, market, shop
Tools & Utilities	tool, utility, software, script, generator, checker, builder, scanner, tester	tool, utility, software
Legal & Law Enforcement	law enforcement, fbi, police, customs, court, lawyer, seized, arrest, bust, investigation	law enforcement, fbi, police

> the card unfortunately isnt linked to the same email, regardless i know theres much more i can do with this info. Ive worked with pros for a while so im not new but im wondering what you would do with all of this information. Hearing peoples ideas might spark something up.

A user named "ilovetobun69" posted on Multiplayer-Game-Hacking in a thread about fake IDs and PayPal, sharing advice on account setup and identity details (post label: Financial Data Theft, partially treated):

> Thanks for understanding, i did some testing and that seems to be the problem (made a fake account under my real info and a different ISP.). It has the same problem. It could be the SSN but I am not sure about the SSN. It most likely was caused by the address, maybe a mix.
>
> Think about it like this, you are PayPal and you see two accounts and one has a different ISP and is newer. The bot would flag it and won't let anyone send money to the account until they call and send in their ID (that is what they would have said if I called).
>
> Also I really do not think he should sell accounts with real SSNs; fake ones work just as well and are not identity theft. That was misleading - I thought I was buying an account

Table H.5: Classification of Dark Web Forum Posts - Placebo

Label	Full Keyword List	Precise Keyword List
Android OS & Device Terms	android, apk, apk mod, apk crack, android device, android id, device fingerprint, imei, rooted, root access, custom rom, magisk, bootloader	android, apk, apk mod
Market Listings & Services	[sell], [leak], autoshop, market, marketplace, vendor, escrow, buyer, seller	[sell], [leak], autoshop
Hacking & Cracking Tools	exploit kit, bypass, crack, patch, keygen, license key, loader, ddos, botnet, vpn service, rdp, ssh, proxy, shell, web shell, admin panel, cms exploit, remote access tool, trojan, rootkit, ransomware	exploit kit, bypass, crack
Malware Viruses	malware, virus, worm, ransomware, trojan, rootkit, spyware, adware, backdoor, dropper, exploit, keylogger	malware, virus, worm
Cryptocurrencies & Finance	bitcoin, btc, monero, xmr, zcash, zec, ethereum, eth, crypto, blockchain, wallet, exchange, mixer, tumbler, fiat, usd, eur, gbp	bitcoin, btc, monero
Hosting & Servers	hosting, vps, server, dedicated server, dns, domain, ssl, ip address, bulletproof hosting	hosting, vps, server
VPN & Anonymity Tools	vpn, tor, socks, proxy, anonymity, privacy, secure communication, ip address mask	vpn, tor, socks
Botnets & DDoS	botnet, ddos, distributed denial of service, stresser, booter, attack service	botnet, ddos, distributed denial of service
Exploits & Zero-days	exploit, zero-day, 0day, vulnerability, patch, cve	exploit, zero-day, 0day
Hacking Services	hack for hire, ddos service, website deface, database hack, social media hack, email hack, phone hack	hack for hire, ddos service, website deface
Cryptography/Encryption	encryption, decryption, pgp, gpg, cipher, hash, private key, public key	encryption, decryption, pgp
Software Development & Programming	code, script, api, programming, dev, development, open source, repository, github, framework	code, script, api
Security & Anonymity Guides	guide, tutorial, opsec, vpn, tor, secure, anonymity, privacy, encryption, safety	guide, tutorial, opsec
Counter-Exploitation & Defense	honeypot, anti-malware, firewall, intrusion detection, security research, vulnerability assessment	honeypot, anti-malware, firewall

> with a fake SSN attached. Most sellers sell fake SSNs, if they attach one at all.

A user named "vcpu" posted on Breached in a thread about credential theft and account compromise (post label: Phishing & Exploits, partially treated):

> Essentially, most threat actors will try to find as much stuff on you - your email address, any compromised passwords, your phone number, and other PII. I'd say most people reuse passwords or at least variations of the same password. Most hackers just try to use "public" knowledge to gain access. 2FA would require a SIM swap, or access to their email depending on verification method. Sometimes, hackers will run phishing campaigns or compromise websites and use harvested credentials from other sites.

A user named "afin" posted on Runion in a thread offering firearms for sale (post label: Weapons & Explosives, control):

> You will NEVER have to pay for postage on delivery, and we make sure all customs forms

Table H.6: Classification of Dark Web Forum Posts - Excluded

Label	Full Keyword List	Precise Keyword List
Illegal Services	money laundering, carding, hacking for hire, assassination, hitman, fake documents, drug trafficking, weapon sales, cyber attack for hire	money laundering, carding, hacking for hire
Forums & Communities	forum, board, community, discussion, thread, post, member, user	forum, board, community
Darknet Operations & Community	darknet, onion, tor network, hidden service, deep web, community, forum, market	darknet, onion, tor network
General Discussion & News	news, discussion, current events, forum, thread, post, general chat, updates	news, discussion, current events

Table H.7: Summary Statistics by Group

Group	Mean	SD	p25	p50	p75	N
Number of Posts (Treated)	23	66	0.25	3.58	17.90	2,441
Number of Posts (Partially Treated)	84.6	469	0.111	1.75	15.00	19,528
Number of Posts (Placebo)	238	2,143	0.45	3.82	27.10	34,174
Number of Posts (Control)	45	192	0	0.75	9.27	26,851
Number of Posts (Excluded)	626	4,269	0.167	5.21	113.00	9,764

The table presents summary statistics for the number of posts for each group. The unit of observation is at the forum-label-month level. The sample includes all the posts over the period 2015–2022.

> are filled out correctly. Sending the parts at once or in separate packages is your choice. You do have to pay upfront for the item so most people choose to buy the first part, then the second, then the third. However, we are easily able to ship the firearm in multiple parts or even all at once. Note that the firearm will NOT be assembled in any event. This means we are going to have to disassemble the firearm (strip, not completely disassemble) and put it in the shielding (decoy packaging). We do ask that you pay for at least 40% of the firearm upfront......

A user named "Jubs" posted on Multiplayer-Game-Hacking in a thread about drug consumption, sharing advice on how to pass a drug test (post label: Drugs, control):

> There are many ways to pass a drug test.
>
> The first and probably hardest one is to chug 2 liters of water every 8 hours. It dilutes your pee. Make sure to pee frequently.
>
> Second would be taking supplements with garnicia or matcha extract and working out a lot.
>
> Third would be borrowing urine from your friend. Buy a heat pack at Walmart or whatever, put the pee in a condom, tuck it in your waistband or pelvis area, and apply the heat pack

> 5 minutes before your test.

A user named "Diplopia" posted on Breached in a thread about his choice of VPN (post label: VPN & Anonymity Tools, placebo):

> I recommend Mullvad vpn its a solid privacy respecting vpn and it also provides decent speeds for a good price its in my top 3 favourite vpns for sure.

A user named "Moody Exchanger" posted on Multiplayer-Game-Hacking in a thread advertising an online currency exchange and cash out service for various payment methods and digital assets (post label: Cryptography/Encryption, placebo):

> We are happy to bring biggest online exchange service to this forum! Probably all of you people heard for us... You can google about Moody Exchange. We are biggest online vendors.
>
> Buying/Selling:
>
> BTC
>
> WMZ
>
> PP
>
> PM
>
> WU
>
> MG
>
> PSC
>
> GIFTCARDS
>
> PREPAID CARDS
>
> Game cards
>
> Many more!
>
> Please contact us on ICQ only: 679751619
>
> If you don't have ICQ and you use email or Skype please PM ME your contact details and someone from our support team will contact you!
>
> We can cashout your BTC with best rates!!!

Robustness. Table J.9 shows that our main results are robust to a range of alternative samples and measurement approaches. Column 1 restricts the analysis to forums active for at least five years, ensuring results are not driven by transient or short-lived communities. Column 2 changes the counting rule so that posts matching multiple labels are counted in each relevant category rather than split across them, capturing the full presence of multi-topic discussions. Column 3 instead allocates posts proportionally to each label based on keyword frequency, addressing concerns that some labels may dominate a post's content. Column 4 redefines the control group by excluding "Tools & Utilities" and "Leaks & Whistleblowing", which could plausibly interact with digital products and be weakly affected by ATT. Column 5 applies a narrower, more restrictive keyword list, following the definitions in Table H.2-Table H.5, to test whether results are robust to a more conservative labeling approach. Across all these specifications, the estimated ATT effect on Apple/iOS-related posts remains negative, statistically significant, and similar in magnitude to the baseline estimate in Column 1 of Table 9 Panel a, suggesting that the decline in related dark web discussions is not sensitive to these alternative definitions or sample restrictions.

I 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 that capture listings in 2020 (July-August) and 2023 (February-March), respectively.[2] Each listing contains the following information: market, listing name, company, category, listing price, listing currency, units, unit price, and listing URL. Examples of listing names 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 companies 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

[2]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/.

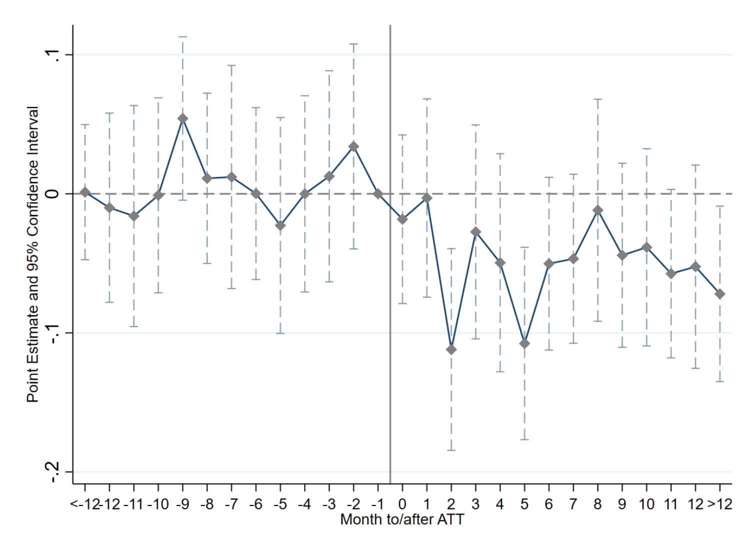
these listings.

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 are likely to take place via laptops or desktops 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.

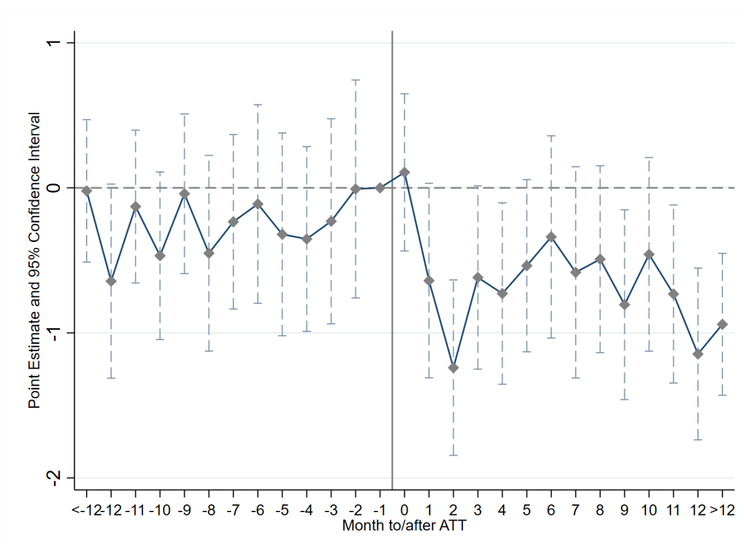
We run a DiD regression using listing-level data to investigate the effect of privacy regulation on the price of data sold on the dark web. The variable of interest is the interaction term between the post-ATT indicator and the two indicator variables constructed above. We add company and year fixed 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 Panel b of Table 9. 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.

J Additional Figures and Tables

Figure J.1: Dynamic Effects of ATT on CFPB Complaints - Monthly Frequency



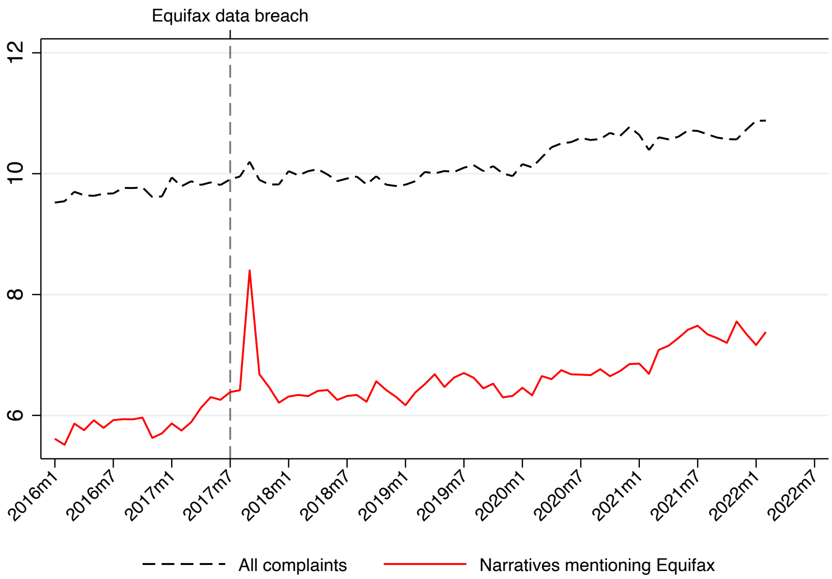
a. Any Complaints



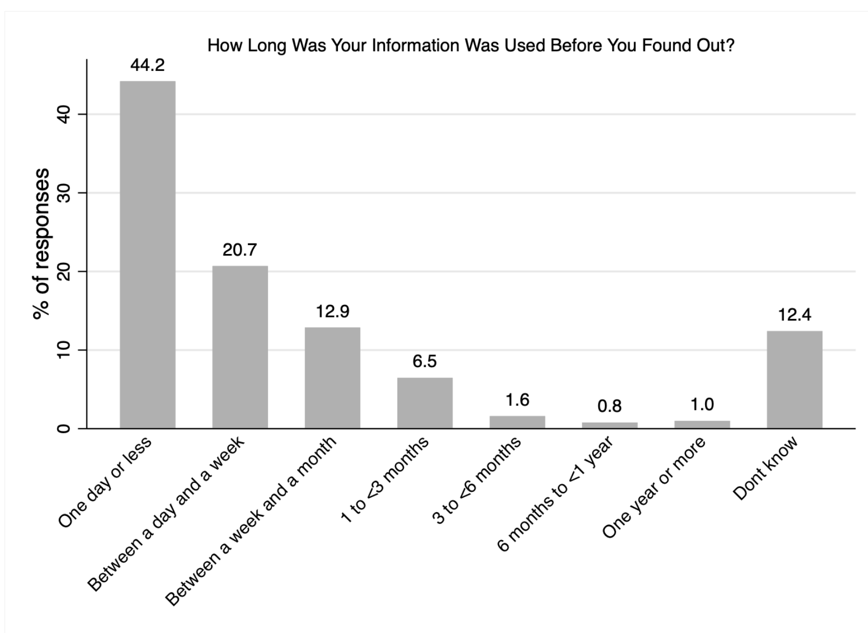
b. # Complaints per 1,000 Residents

This figure illustrates the dynamic effect of ATT on CFPB complaints around the implementation of ATT at the monthly frequency. Month -1 is the month before the implementation (March 2021) and is the omitted category. In Panel a, the outcome variable is an indicator for whether a zip code has at least one complaint and estimates are from a linear probability model. In Panel b, the outcome variable is the number of complaints per 1,000 residents and estimates are from a Poisson model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Coefficients on the interaction terms between indicators for the relative timing to ATT and the pre-ATT iOS device share are plotted. The sample period is January 2019 to June 2022. Zip code fixed effects and county × year-month fixed effects are included. Standard errors are clustered at the state level.

Figure J.2: Event Timeline from Data Breach to Reported Complaint



a. Changes in CFPB Complaints around Equifax Data Breach in 2017

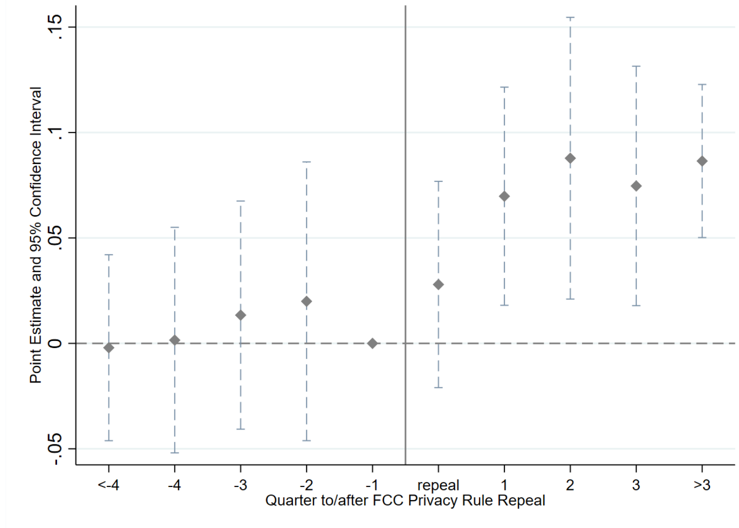


b. Time to Discover Data Misuse from National Crime Victimization Survey (NCVS)

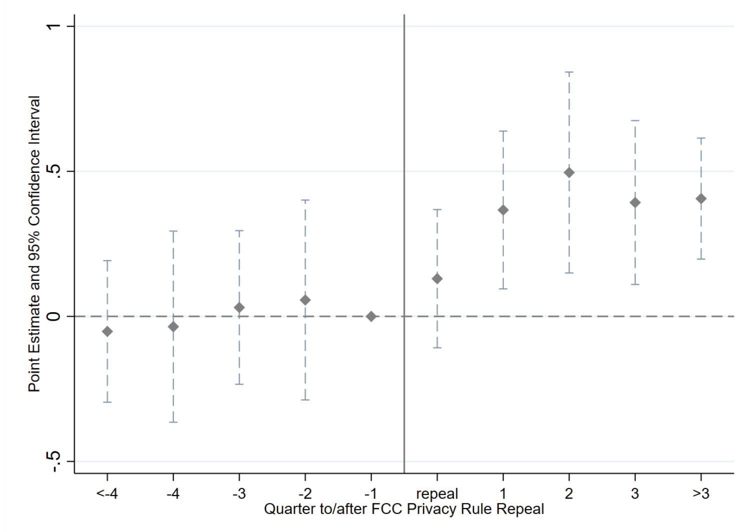
This figure illustrates the timeline from data breach or leakage to reported complaint. Panel a shows changes in CFPB consumer complaints related to Equifax relative to all other complaints around the July 2017 Equifax breach. Panel b presents the fraction of identity theft cases discovered within different timeframes, based on the most recent National Crime Victimization Survey (NCVS) in 2021. Survey respondents were asked, *"How long was the information used before you found out?"*

Figure J.3: Dynamic Effects of 2017 FCC Privacy Rule Repeal on CFPB Complaints

a. Local Exposure Based on ISP Data Collection Intensity

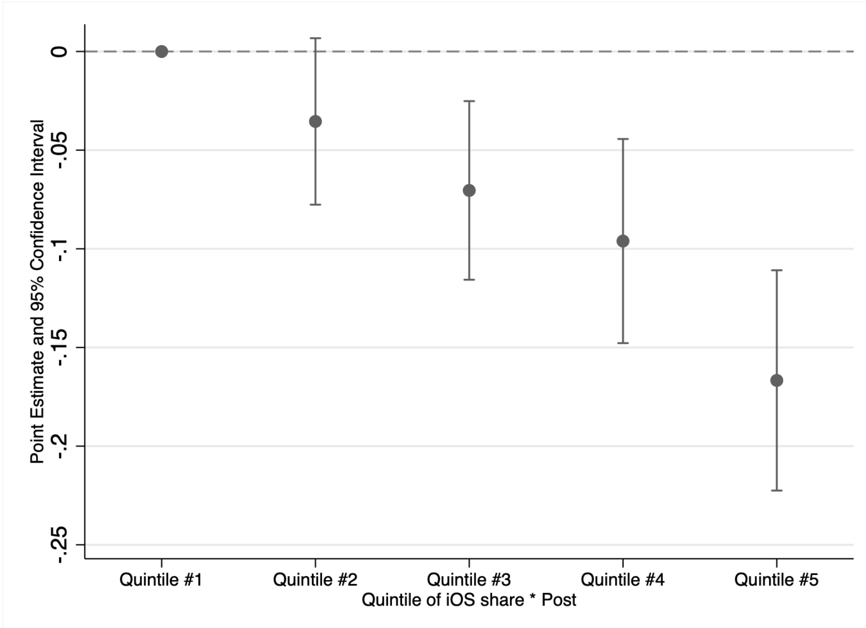


b. Local Exposure Based on ISP Data Sharing Intensity



This figure shows the dynamic effects of the 2017 FCC privacy rule repeal on CFPB complaints, estimated at a quarterly frequency using a Poisson model. Quarter $t = -1$ (2016Q4), the quarter immediately preceding the repeal, serves as the omitted reference category. In both panels, the outcome variable is the number of complaints per 1,000 residents. Panel a uses the market-share-weighted number of data types collected by ISPs, as disclosed in their privacy labels, as the measure of exposure to the shock. Panel b uses the the market-share-weighted number of data SDKs used by ISPs in their mobile applications. Plotted coefficients correspond to the interaction terms between relative-quarter indicators and pre-Repeal exposure levels. Zip code fixed effects and state × year-month fixed effects are included. Standard errors are clustered at the state level.

Figure J.4: Monotonicity—Categorizing iOS Share into Quintiles



This table illustrates the effect of ATT on CFPB complaints, in which we interact the post-ATT indicator with indicators for each of the quintiles of the pre-ATT iOS share. The outcome variable is the number of complaints per 1,000 residents and estimates are from a Poisson model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022. Zip code fixed effects and county × year-month fixed effects are included. Standard errors are clustered at the state level.

Table J.1: Effect of 2017 Repeal of FCC Broadband Privacy Rule on CFPB Complaints
Robustness: Unique # Data Types Collected and SDKs Installed Across ISP-Owned Apps

Panel a. Local Exposure based on ISP Data Collection Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data collection)	0.004***	0.004***	0.039***	0.025*
	(0.001)	(0.001)	(0.007)	(0.013)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	3.0%	2.7%	6.4%	4.0%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

Panel b. Local Exposure based on ISP Data Sharing Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data sharing)	0.002***	0.002***	0.018***	0.010*
	(0.000)	(0.001)	(0.003)	(0.006)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	3.2%	2.9%	6.5%	3.6%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

This table presents the estimated effects of the 2017 FCC Broadband Privacy Rule repeal on CFPB complaints. The unit of observation is at the zip-code-month level. Columns 1–2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3–4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Panel a (b) uses the count of unique data types collected (the count of unique data SDKs installed) across all apps owned by the same ISP. The sample period is January 2015 to December 2018. All columns include zip code fixed effects. The odd-numbered columns include state × year-month fixed effects, while the even-numbered columns include county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.2: Effect of 2017 Repeal of FCC Broadband Privacy Rule on CFPB Complaints
Robustness: Total # Data Types Collected and SDKs Installed Across ISP-Owned Apps

Panel a. Local Exposure Based on ISP Data Collection Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data collection)	0.012***	0.008**	0.109***	0.050
	(0.003)	(0.004)	(0.021)	(0.034)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	2.5%	1.7%	5.1%	2.4%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

Panel b. Local Exposure Based on ISP Data Sharing Intensity

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post Repeal × Exposure (data sharing)	0.005***	0.003**	0.044***	0.020
	(0.001)	(0.002)	(0.010)	(0.014)
Zip code FE	✓	✓	✓	✓
State × Year-month FE	✓		✓	
County × Year-month FE		✓		✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.207	0.207	0.034	0.034
Magnitude (1SD increase in iOS share)	2.3%	1.6%	4.6%	2.2%
Observations	1,166,496	1,166,496	1,166,496	1,166,496
R^2 / Pseudo R^2	0.406	0.460	0.061	0.109

This table presents the estimated effects of the 2017 FCC Broadband Privacy Rule repeal on CFPB complaints. The unit of observation is at the zip-code-month level. Columns 1–2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3–4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Panel a (b) uses the aggregate number of data types collected (the aggregate number of data SDKs installed) across all apps owned by the same ISP, allowing for double counting. The sample period is January 2015 to December 2018. All columns include zip code fixed effects. The odd-numbered columns include state × year-month fixed effects, while the even-numbered columns include county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.3: Effect of ATT on Consumer Complaints - CFPB
Robustness: Propensity to Complain and Measurement Error

	# Complaints per 1,000 residents					
	(1) victimization	(2) foot traffic>25^{pct}	(3) foot traffic>50^{pct}	(4) # grocery store>25^{pct}	(5) # grocery store>50^{pct}	(6) pop. weight
Post × iOS share	−0.719***	−0.742***	−0.657***	−0.658***	−0.701***	−0.723***
	(0.111)	(0.103)	(0.099)	(0.093)	(0.087)	(0.094)
Zip code FE	✓	✓	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome var.	0.070	0.073	0.077	0.072	0.076	0.069
Magnitude (↑1 SD iOS share)	-6.9%	-6.8%	-6.0%	-6.1%	-6.5%	-7.0%
Observations	878,346	698,601	469,885	701,673	472,122	926,772
Pseudo R^2	0.131	0.125	0.110	0.124	0.107	0.086

This table presents the estimated effects of ATT on CFPB complaints using alternative regression specifications. The unit of observation is at the zip-code-month level. Column 1 reports estimates adjusted for victimization, using outcome variables weighted following Raval (2020b) to account for differences in the propensity to complain across zip code level demographics. Column 2 reports estimates for observations with pre-ATT foot traffic above the 25th percentile and Column 3 restricts to those above the 50th percentile. Column 4 reports estimates for observations with a pre-ATT number of grocery stores above the 25th percentile, and Column 5 restricts to those above the 50th percentile. Column 6 reports estimates from a population weighted regression. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.4: Effect of ATT on Consumer Complaints - CFPB

Robustness: Alternative Aggregation by Zipcode-quarter and County-month

	# Complaints per 1,000 residents	
	(1)	(2)
	Zip-code-quarter	County-month
Post × iOS share	-0.510***	-0.444***
	(0.116)	(0.162)
Zip code FE	✓	
County FE		✓
County × Year-quarter FE	✓	
State × Year-month FE		✓
Model	PPML	PPML
Mean outcome var.	0.193	0.042
Magnitude (↑1 SD iOS share)	-5.0%	-3.7%
Observations	309,161	134,177
Pseudo R^2	0.175	0.105

This table presents the estimated effects of ATT on CFPB complaints using alternative aggregation by zip code-quarter (Column 1) and county-month (Column 2). The unit of observation in Column 1 is at the zip code-quarter level. The unit of observation in Column 2 is at the county-month level. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.5: Effect of ATT on Consumer Complaints - CFPB

Robustness: Excluding and Including Complaints to the Three Credit Bureaus

	(1)	(2)
	# Complaints per 1,000 residents	
	Excl. 3 credit bureaus	3 credit bureau only
Post × iOS share	−0.420***	−0.670***
	(0.077)	(0.150)
Zip code FE	✓	✓
County × Year-month FE	✓	✓
Model	PPML	PPML
Mean outcome var.	0.069	0.087
Magnitude (↑1 SD iOS share)	-4.1%	-6.3%
Observations	908,292	908,292
Pseudo R^2	0.091	0.164

This table presents the estimated effects of ATT on CFPB complaints by separating complaints related to the three major credit bureaus from all others. The unit of observation is at the zip code-month level. Column 1 reports estimates using only complaints not concerning the three credit bureaus (Equifax, Experian, and TransUnion), while Column 2 reports estimates using only complaints concerning these three credit bureaus. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.6: Effect of ATT on Consumer Complaints - CFPB

Robustness: Alternative Regression Specifications

	# Complaints per 1,000 residents or # Complaints (unscaled)					
	(1) winsorize 0.5%	(2) unscaled count	(3) incl. outlier zip codes	(4) double cluster	(5) sample to 2023Q4	(6) sample to 2024Q4
Post × iOS share	−0.618***	−0.686***	−0.602***	−0.621***	−0.824***	−1.034***
	(0.101)	(0.081)	(0.090)	(0.100)	(0.152)	(0.157)
Zip code FE	✓	✓	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome var.	0.069	0.069	0.093	0.069	0.073	0.100
Magnitude (↑1 SD iOS share)	-6.0%	-6.8%	-5.9%	-6.1%	-7.9%	-10.0%
Observations	926,772	943,488	1,025,472	926,772	1,325,640	1,590,768
Pseudo R^2	0.149	0.445	0.172	0.134	0.187	0.241

This table presents the estimated effects of ATT on CFPB complaints using alternative regression specifications. The unit of observation is at the zip code-month level. Column 1 reports estimates with less aggressive winsorizaton at 0.5% on each side. Column 2 reports estimates using unscaled number of complaints as the outcome variable. Column 3 reports estimates including outlier zip codes in the regression. Column 4 reports estimates with standard errors double clustered at state and year-month level. Column 5 extends the sample to include CFPB complaints from January 2019 through the end of 2023. Column 6 extends the sample to include CFPB complaints from January 2019 through the end of 2024. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 to June 2022 in Columns 1 to 4. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level (except Column 4) are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.7: Effect of ATT on Consumer Complaints - CFPB

Robustness: Controlling for Socialeconomic Factors

	Any complaints (0/1)		# Complaints per 1,000 residents	
	(1)	(2)	(3)	(4)
Post × iOS share	−0.038***	−0.035***	−0.446***	−0.345**
	(0.014)	(0.013)	(0.099)	(0.139)
Post × %Female	0.010***		0.013	
	(0.002)		(0.019)	
Post × %Age 10-19	−0.010***		−0.001	
	(0.003)		(0.033)	
Post × %Age 20-29	−0.004		−0.051**	
	(0.003)		(0.024)	
Post × %Age 30-39	0.002		0.005	
	(0.002)		(0.024)	
Post × %Age 40-49	−0.002		−0.061***	
	(0.002)		(0.021)	
Post × %Age 50+	−0.025***		−0.113***	
	(0.005)		(0.038)	
Post × Median income		−0.019***		−0.051***
		(0.004)		(0.017)
Post × Unemployment rate		0.004***		0.041***
		(0.001)		(0.014)
Post × %Bachelor		0.011***		0.021
		(0.004)		(0.019)
Zip code FE	✓	✓	✓	✓
County × Year-month FE	✓	✓	✓	✓
Model	Linear	Linear	PPML	PPML
Mean outcome var.	0.267	0.273	0.069	0.070
Magnitude (↑1 SD iOS share)	-1.5%	-1.3%	-4.4%	-3.4%
Observations	926,772	900,942	926,772	900,648
R^2 / Pseudo R^2	0.514	0.513	0.134	0.133

This table presents the estimated effects of ATT on CFPB complaints by adding interactions between the Post-ATT indicator and various socioeconomic factors. The unit of observation is at the zip code-month level. Columns 1–2 report estimates from a linear probability model, where the outcome is a binary indicator equal to one if at least one complaint is filed in a zip code during a given month. Columns 3–4 report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. Columns 1 and 3 control for demographic factors (gender and age), whereas Columns 2 and 4 control for economic factors (income, unemployment, and education). The sample period is January 2019 to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.8: Effect of ATT on Consumer Complaints - Consumer Sentinel Network
Robustness: Top vs. Bottom Fraud Categories by Relevance

	# Complaints per 1,000 residents	
	Top Fraud Category (1)	Bottom Fraud Category (2)
Post × iOS share	−0.223***	−0.024
	(0.079)	(0.040)
Zip code FE	✓	✓
County × Year-month FE	✓	✓
Model	PPML	PPML
Mean outcome var.	0.215	0.556
Magnitude (↑1 SD iOS share)	-2.2%	-0.2%
Observations	1,029,942	1,122,071
Pseudo R^2	0.161	0.126

This table presents the estimated effects of ATT after classifying complaints from Consumer Sentinel Network (CSN) by their relevance to ATT. The unit of observation is at the zip code-month level. The outcome variable is the number of complaints per 1,000 residents. Column 1 includes complaints from products for which at least 50% of complaints include at least one of the relevant words in the narrative. Column 2 includes complaints from products for which less than 25% of complaints include one of the relevant words in the narrative. All columns report estimates from a Poisson model where the dependent variable is the number of complaints per 1,000 residents, using the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The sample period is January 2019 (February 2019 for Identity Theft) to June 2022. All columns include zip code fixed effects and county × year-month fixed effects. Standard errors clustered at the state level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table J.9: Effect of ATT on Dark Web Activities—Posts on Dark Web Forums
Robustness: Alternative Samples and Measures

	# Posts				
	(1) forum age>60m	(2) double count	(3) keyword-freq. weighted	(4) alt. control	(5) alt. keywords
Post × Treated	−0.433***	−0.359***	−0.420***	−0.362***	−0.390***
	(0.017)	(0.057)	(0.021)	(0.074)	(0.070)
Label FE	✓	✓	✓	✓	✓
Forum × Year-month FE	✓	✓	✓	✓	✓
Model	PPML	PPML	PPML	PPML	PPML
Mean outcome var.	44.5	46.0	46.0	42.8	20.8
Magnitude (Treated vs. Control)	-35.1%	-30.2%	-34.3%	-30.4%	-32.3%
Observations	27,600	27,060	27,612	27,612	23,010
Pseudo R^2	0.872	0.877	0.869	0.898	0.806

This table presents the estimated effects of ATT on number of posts on dark web forums using alternative samples and measures. The unit of observation is a forum-label-month. The full list of labels and their exposure to ATT can be found in Internet Appendix Section H. All columns compare the explicitly treated group (Apple/iOS-related posts) with a control group of illicit-trade categories (e.g., drugs, human trafficking, weapons) that operate in separate markets. Column 1 restricts the sample to forums active for at least five years. Column 2 counts a post in every applicable label rather than splitting it evenly across labels when it meets multiple label criteria. Column 3 allocates posts proportionally based on the frequency of matching keywords within each label. Column 4 uses an alternative control group that excludes "Tools & Utilities" and "Leaks & Whistleblowing", which may interact with digital products and thus be tangentially treated. Column 5 applies a narrowly defined keyword list, as in Table H.2-Table H.5. Standard errors double clustered at the forum and label level are reported in parentheses. The sample period is January 2015 to December 2022. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.