The Supply and Demand for Data Privacy: Evidence from Mobile Apps

Bo Bian[*]   Xinchen Ma[†]   Huan Tang[‡]

_____

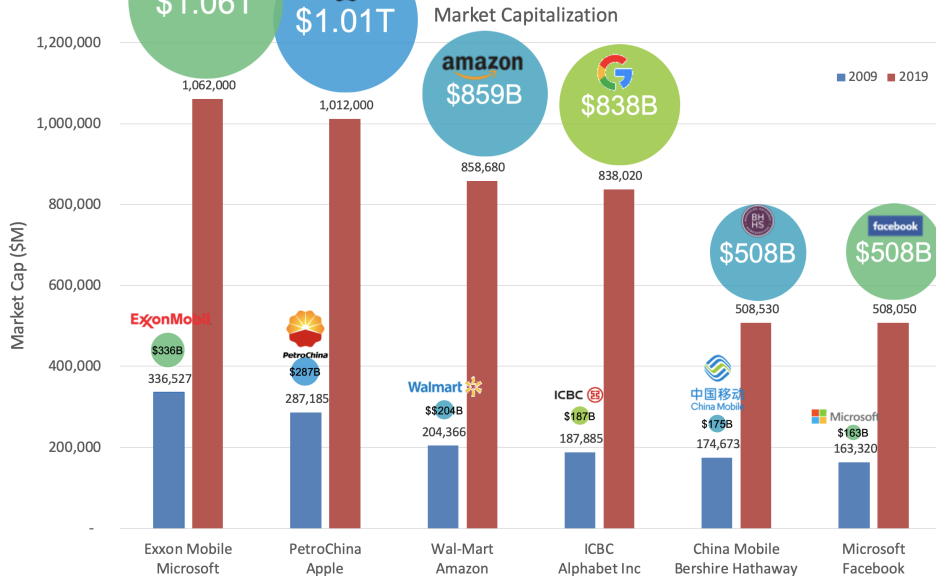[*]University of British Columbia
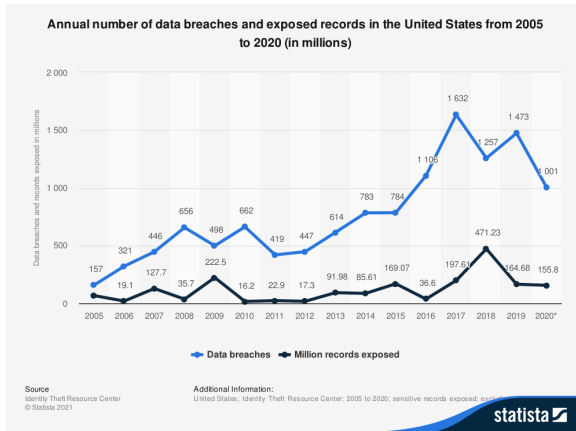[†]London School of Economics
[‡]London School of Economics; CEPR

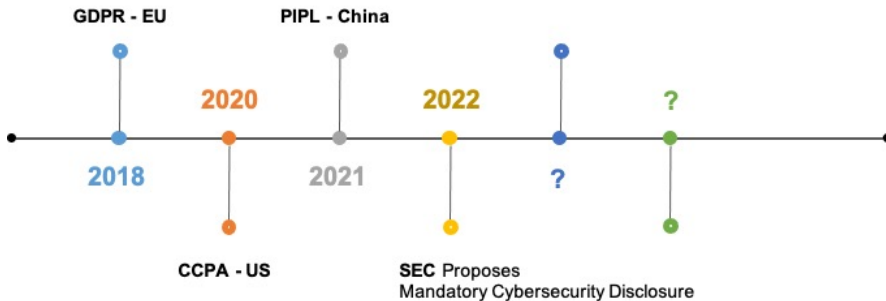Market capitalization of the largest companies in 2009 and 2019

Market Capitalization

Market Cap ($M)

■ 2009 ■ 2019

Microsoft $1.06T — 1,062,000
Apple $1.01T — 1,012,000
amazon $859B — 858,680
G $838B — 838,020
BH HS $508B — 508,530
facebook $508B — 508,050

ExxonMobil $336B — 336,527
PetroChina $287B — 287,185
Walmart $$204B — 204,366
ICBC 工 $187B — 187,885
中国移动 China Mobile $175B — 174,673
Microsoft $163B — 163,320

| Exxon Mobile Microsoft | PetroChina Apple | Wal-Mart Amazon | ICBC Alphabet Inc | China Mobile Bershire Hathaway | Microsoft Facebook |

Motivation

▸ **Data is not costless – growing public concerns & tightening regulation**

Annual number of data breaches and exposed records in the United States from 2005 to 2020 (in millions)

Motivation

‣ **Data is not costless – growing public concerns & tightening regulation**

Motivation

‣ **Data is not costless – growing public concerns & tightening regulation**

‣ **Scarce empirical evidence on the supply and demand for privacy**

‣ **Supply: how much data (privacy) do firms collect (offer)?**
  ‣ Lack of consistent measures of firms' data collection
  ‣ What data are collected, and for what purpose?

‣ **Demand: is there a demand for privacy & implication for firm valuation?**
  ‣ Privacy paradox: what users say $\neq$ what users do
  ‣ Impacts valuations of firms in the digital economy

This Paper: Mobile App Market

- **Mobile apps become increasingly important for consumers and firms**
  - Smartphones account for 70% of the total digital media time (Comscore)
  - > 600 billion app revenues and mobile advertising

Introduction
○○●○

supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○○○

This Paper: Mobile App Market

- **Mobile apps become increasingly important for consumers and firms**
  - Smartphones account for 70% of the total digital media time (Comscore)
  - > 600 billion app revenues and mobile advertising

- **Introduction of Apple Store's privacy labels** – Dec 14, 2020
  - Mandates the disclosures of data collection
  - Visible and digestible, $\sim$ food nutrition labels

Introduction
○○●○

supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○○○
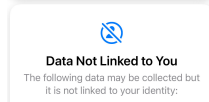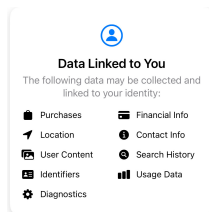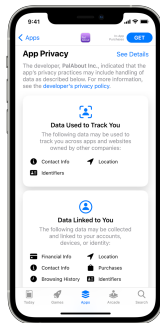
This Paper: Mobile App Market

- **Supply: scrape privacy labels of most popular apps in US and worldwide**
  - Amount of data collected, intrusiveness of data uses
  - Consistent formatting allows for cross-country/app/category/firm comparisons

- **Demand: download and revenue responses to Apple's policy**
  - Causal evidence from Diff-in-Diff: Android apps as control
  - Event study: stock market reactions and firm earnings

Preview of Findings

‣ **Supply of privacy based on privacy labels**
  - ‣ 80% of the data collected for purposes unrelated to app functionality
  - ‣ Advertising and marketing: Games, news, shopping, entertainment apps
  - ‣ Top data collectors: public firms, larger market shares, better ratings

Introduction
○○○●

supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○○○

Preview of Findings

- **Supply of privacy based on privacy labels**
  - 80% of the data collected for purposes unrelated to app functionality
  - Advertising and marketing: Games, news, shopping, entertainment apps
  - Top data collectors: public firms, larger market shares, better ratings
- **Consumers react negatively to disclosure of data collection practices**
  - iOS apps experience a 14-15% drop in weekly downloads/revenues, relative to Android
  - Stronger effect for more privacy-invasive and substitutable apps
  - Cross-country heterogeneity: data privacy laws/enforcement, consumer attitudes, trust

Preview of Findings

- **Supply of privacy based on privacy labels**
  - 80% of the data collected for purposes unrelated to app functionality
  - Advertising and marketing: Games, news, shopping, entertainment apps
  - Top data collectors: public firms, larger market shares, better ratings
- **Consumers react negatively to disclosure of data collection practices**
  - iOS apps experience a 14-15% drop in weekly downloads/revenues, relative to Android
  - Stronger effect for more privacy-invasive and substitutable apps
  - Cross-country heterogeneity: data privacy laws/enforcement, consumer attitudes, trust
- **Adverse stock market reactions**
  - -5% to -10% cumulative abnormal returns in 6 months
  - Driven by firms that harvest more data - supported by decline in earnings

Introduction
oooo

supply of privacy
●oooooooooo

demand for privacy
oooooooooooooooooooo

Laboratory, Sample, and Measurement

Introduction
OOOO

supply of privacy
O●OOOOOOOO

demand for privacy
OOOOOOOOOOOOOOOOOOOO

App Privacy: Facebook as an example

- ▸ Data Used to Track You (or your device) and shared across different apps, ad networks, and data brokers

- ▸ Data Linked to You (and your real identity) that is collected by the app and company but not shared

- ▸ Data not Linked to You that the company generally aggregates into larger statistics

Introduction
○○○○

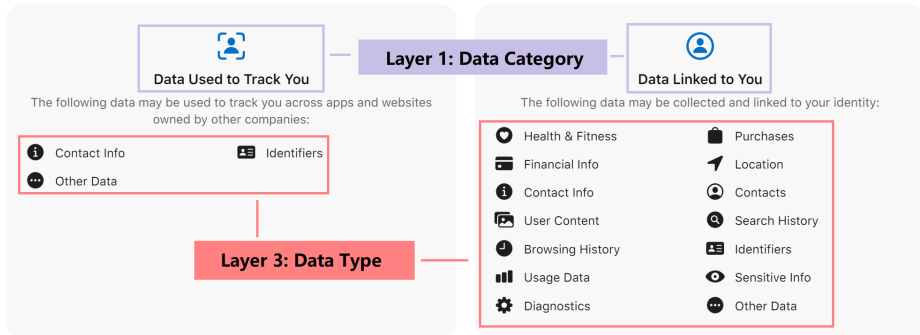supply of privacy
○○●○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○

App Privacy: Facebook as an example
Front page

App Privacy

See Details

The developer, **Facebook, Inc.**, indicated that the app's privacy practices may include handling of data as described below. For more information, see the developer's privacy policy.

Layer 1: Data Category

Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

Contact Info   Identifiers
Other Data

Data Linked to You

The following data may be collected and linked to your identity:

Health & Fitness	Purchases
Financial Info	Location
Contact Info	Contacts
User Content	Search History
Browsing History	Identifiers
Usage Data	Sensitive Info
Diagnostics	Other Data

Layer 3: Data Type

Privacy practices may vary, for example, based on the features you use or your age. Learn More

Introduction
○○○○

supply of privacy
○○○●○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○

App Privacy: Facebook as an example
Click "See Details": full information

Introduction
oooo

supply of privacy
ooooo●ooooo

demand for privacy
oooooooooooooooooooo

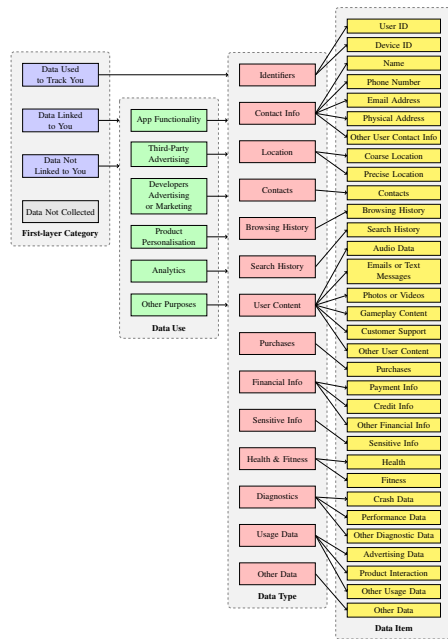# Structure of Privacy Labels

1. Data Category: 3
   - Data used to track you
   - Data linked to you
   - Data not linked to you
   - Data not collected

2. Purpose: 6
   - App functionality
   - 3-party advertising
   - Developers' ad and mkt
   - Product personalisation
   - Analytics
   - Other purposes

3. Data Type: 14

4. Data Item: 34

Introduction
○○○○

supply of privacy
○○○○○●○○○○

demand for privacy
○○○○○○○○○○○○○○○○○○

Summary Statistics: Data Collection

	min	mean	p50	sd	max	count
Data Collection Intensity						
1(Data used to track you)	0	0.6	1	0	1	9,728
# Data Types Collected	0	14.2	12	11	80	9,728
# Data Items Collected	0	21.5	17	18	167	9,728
By Purpose: # Data Items Collected						
Third-party Advertising	0	1.4	0	3	24	9,728
Product Personalization	0	1.7	0	3	25	9,728
Developer's Advertising or Marketing	0	1.7	0	3	24	9,728
Analytics	0	3.2	1	4	30	9,728
Other Purposes	0	0.5	0	2	28	9,728
App Functionality	0	4.0	2	5	32	9,728

‣ Sample: Top 10,000 apps that generated > 80%/90% of store-wide downloads/revenue
‣ 60% of apps collect *data used to track you*
‣ Large variations across apps (even within each purpose)

Heterogeneity across app category

Layer 3: Data Types Collected
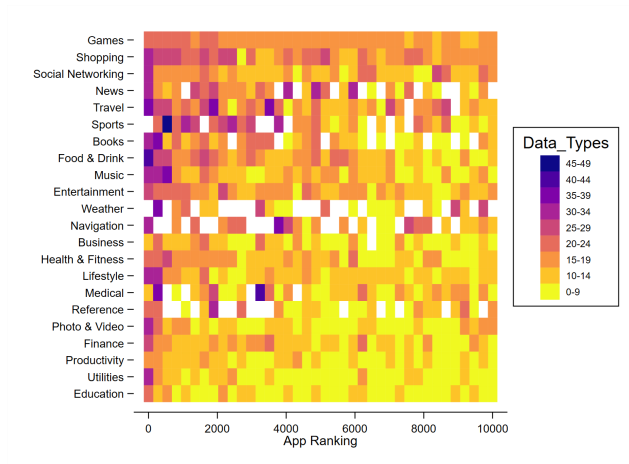
- ‣ Highly ranked apps collect more information
- ‣ Large within-category variation

Heterogeneity across app category
Layer 3: Data Types Collected
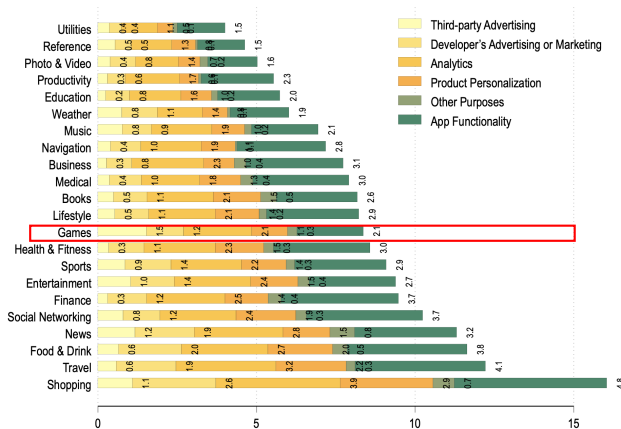
- ▸ 3rd party ads: games, news, entertainment, and shopping

Introduction
oooo

supply of privacy
ooooo○○○●○

demand for privacy
oooooooooooooooooo

Little time-series changes (1/2)
Data Used to Track Users
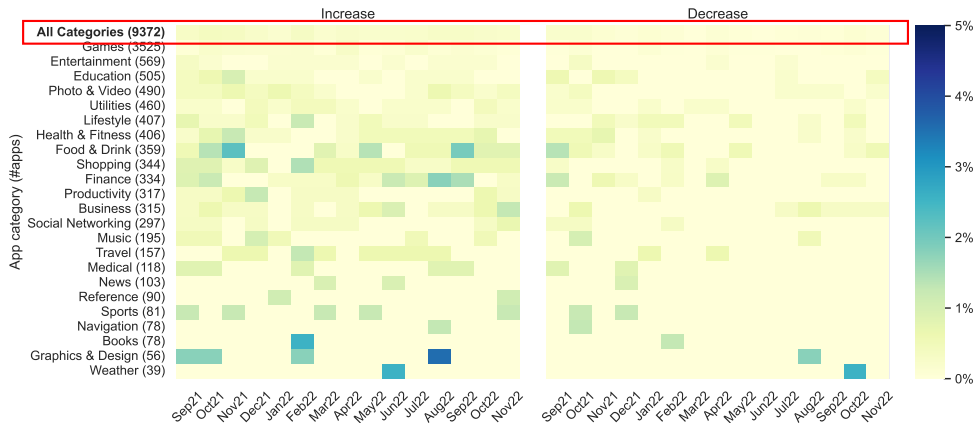
▸ 0.34% (0.15%) of apps turn on (off) tracking each month

Little time-series changes (2/2)

Total Data Items
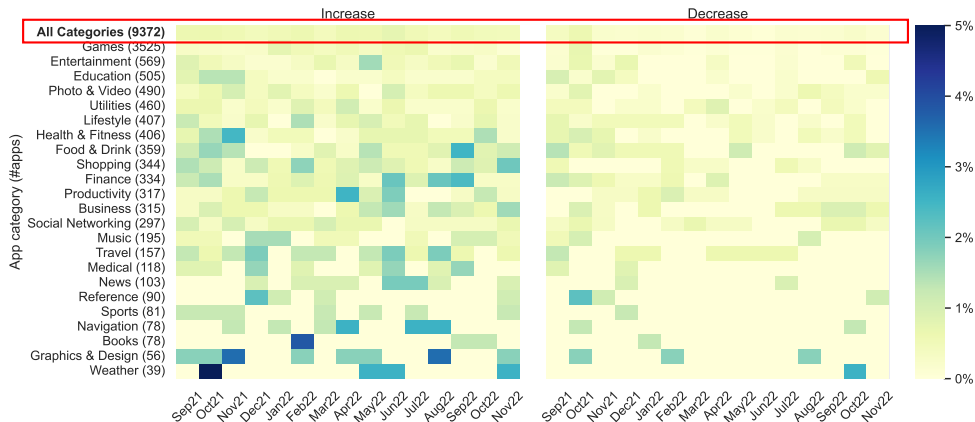
▸ 0.57% (0.21%) of apps increase (decrease) #total data items collected each month

▸ back

Introduction
○○○○

supply of privacy
○○○○○○○○○○

demand for privacy
●○○○○○○○○○○○○○○○○○○○

Consumers' Reaction to Privacy Label Release

Release Dates
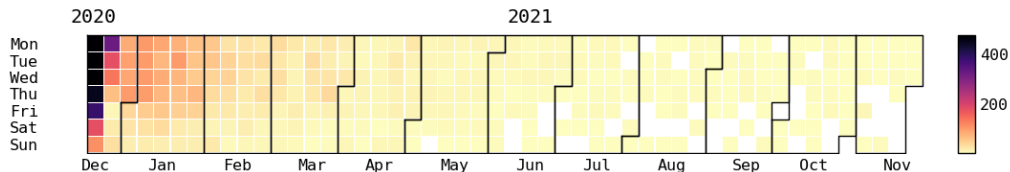
- ▸ Many apps release their privacy labels in a few weeks but there are variations in timing
  ⇒ Controls for shocks to digital consumption (Covid-19) and supply of privacy (CCPA)

- ▸ Privacy label policy only applies to iOS apps and affects iOS users
  ⇒ DiD: pre vs. post; iOS (treatment) vs. Android (control)

Introduction
0000

supply of privacy
0000000000

demand for privacy
00●0000000000000000

iOS vs. Android Parallel Trends
Download and revenue by platform
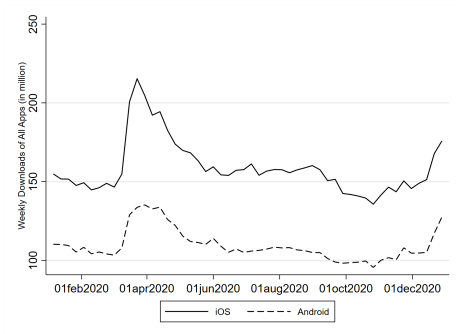
(a) weekly downloads
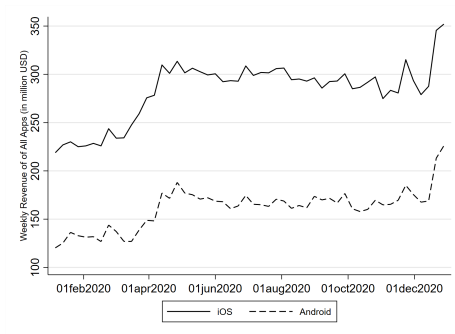
(b) weekly revenue

‣ Android tracks closely with iOS
‣ Downloads: new users; Revenue: usage of both existing and new users

Introduction
o o o o

supply of privacy
o o o o o o o o o o

demand for privacy
o o o ● o o o o o o o o o o o o o o o

Share of iOS downloads: pre vs. post

Top Quartile: # Data Items Collected
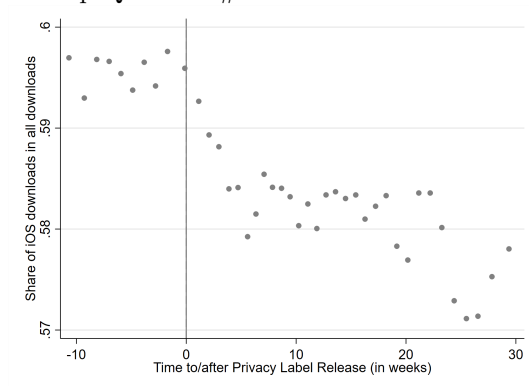
- Share of iOS downloads: sharp drop around privacy label release
- 60-61% pre-event ⇒ 57-58% post-event

Baseline DiD Estimation

	Downloads			Revenue		
	(1) Main version	(2) All versions	(3) All versions	(4) Main version	(5) All versions	(6) All versions
Post	0.240***	0.238***	0.203***	0.255***	0.284***	0.207***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
iOS × Post	-0.117**	-0.117**	-0.138***	-0.135***	-0.195***	-0.151***
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Linear Trend	Y	Y	Y	Y	Y	Y
Year-week FE	Y	Y	Y	Y	Y	Y
App FE	Y	Y	Y	Y	Y	Y
Platform-Age FE	Y	Y	Y	Y	Y	Y
Sample	Full	Full	Updated apps	Full	Full	Updated apps
Observations	966,216	966,216	871,470	966,216	966,216	871,470
R-sq	0.557	0.558	0.569	0.917	0.917	0.920

‣ Weekly downloads: $\sim$ 11-14% drop

‣ Weekly revenue: $\sim$ 14-20% drop

Introduction
0000

supply of privacy
0000000000

demand for privacy
00000●000000000000000

DiD dynamics - by data collection intensity
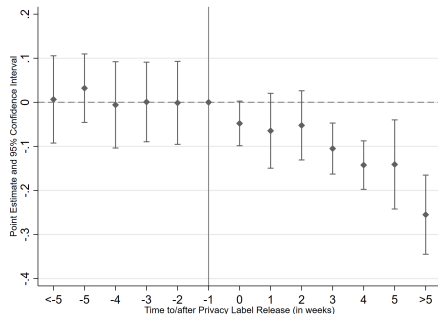
(a) # Data Types Collected - high
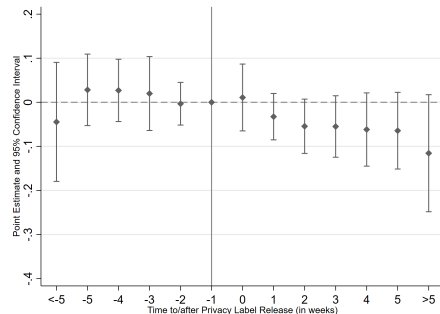
(b) # Data Types Collected - low

‣ A larger decline for more privacy-invasive apps
‣ No positive spillover to privacy-centric apps → consumer underestimate data collection

Introduction
○○○○

supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○●○○○○○○○○○○○○

Heterogeneity
Data Use

	Downloads (all versions)					
Post	0.281***	0.231***	0.180***	0.215***	0.213***	0.143**
	(0.05)	(0.04)	(0.06)	(0.05)	(0.04)	(0.06)
iOS × Post	-0.096**	-0.107**	-0.015	-0.110**	-0.131***	-0.091
	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.08)
iOS × Post × Third-party Ad	-0.049**					
	(0.02)					
iOS × Post × Developer Ad or Mkt		-0.044*				
		(0.02)				
iOS × Post × Analytics			-0.085***			
			(0.03)			
iOS × Post × Product Personalization				-0.040*		
				(0.02)		
iOS × Post × Other Purposes					-0.033	
					(0.03)	
iOS × Post × App Functionality						-0.039
						(0.03)
Linear Trend	Y	Y	Y	Y	Y	Y
Year-week FE	Y	Y	Y	Y	Y	Y
App FE	Y	Y	Y	Y	Y	Y
Platform-Age FE	Y	Y	Y	Y	Y	Y
Observations	871,470	871,470	871,470	871,470	871,470	871,470
R-sq	0.571	0.570	0.569	0.569	0.569	0.570

▸ A larger drop in downloads for apps that collect data for more intrusive purposes

Introduction
oooo

supply of privacy
oooooooooo

demand for privacy
ooooooo○●oooooooooooo

Heterogeneity
Substitutability

	Downloads (all versions)			Revenues (all versions)		
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.078*	0.200***	0.166***	0.193***	0.191***	0.248***
	(0.04)	(0.04)	(0.06)	(0.04)	(0.03)	(0.06)
iOS × Post	-0.046	-0.143***	-0.225***	-0.160***	-0.153***	-0.195***
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)	(0.07)
iOS × Post × Platform-wide ranking decile	-0.016**			-0.001		
	(0.01)			(0.01)		
iOS × Post × $\mathbb{1}$(Market share above 90th pct.)		0.059*			0.013	
		(0.03)			(0.06)	
iOS × Post × Age			0.015**			0.008
			(0.01)			(0.01)
Linear Trend	Y	Y	Y	Y	Y	Y
Year-week FE	Y	Y	Y	Y	Y	Y
App FE	Y	Y	Y	Y	Y	Y
Platform-Age FE	Y	Y	Y	Y	Y	Y
Observations	871,470	871,470	871,470	871,470	871,470	871,470
R-sq	0.571	0.569	0.569	0.920	0.920	0.920

‣ A large drop when the app is more substitutable

Introduction
supply of privacy
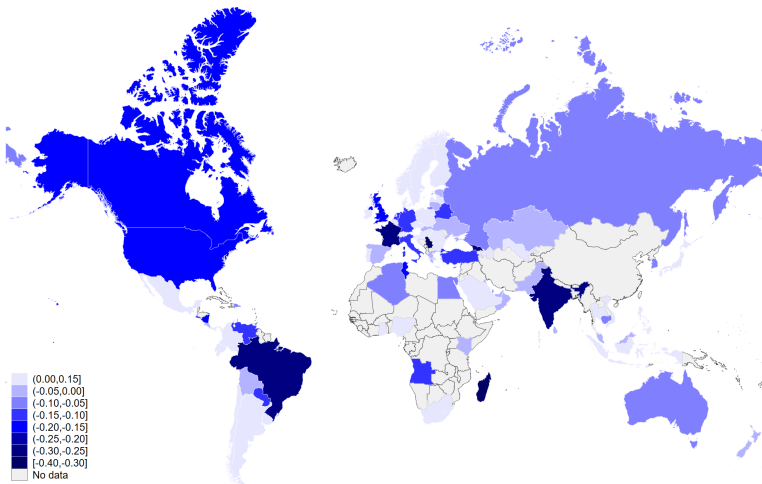demand for privacy
○○○○
○○○○○○○○○○
○○○○○○○●○○○○○○○○

Robustness and Placebo
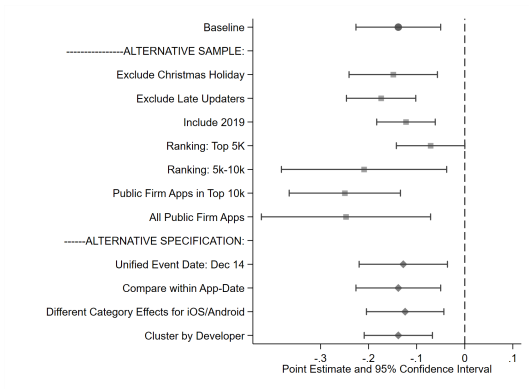
Robustness checks
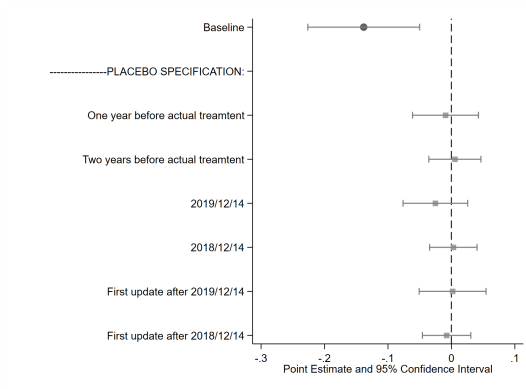
Placebo treatment date

Cross-country comparison - Public Firm Apps
A world map

(0.00,0.15]
(-0.05,0.00]
(-0.10,-0.05]
(-0.15,-0.10]
(-0.20,-0.15]
(-0.25,-0.20]
(-0.30,-0.25]
(-0.40,-0.30]
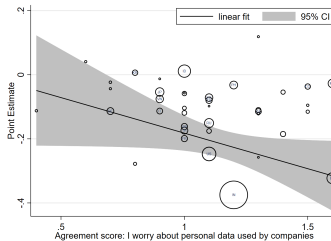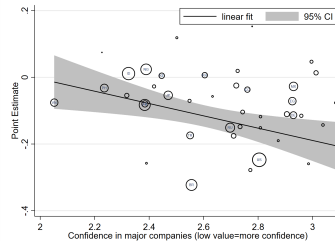No data

What explains the heterogeneity
Public Firm Apps
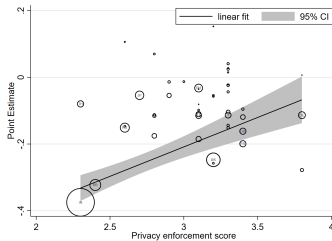
legal protection   data privacy confidence   trust in private sector

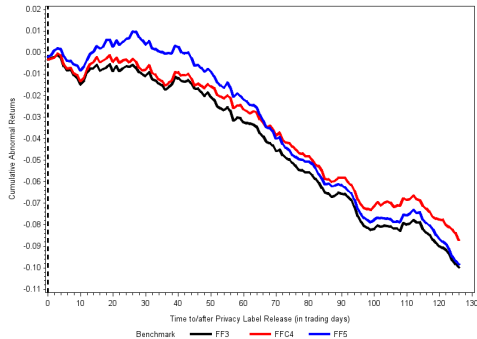▸ poorer legal protection, lower data privacy confidence, lower trust in private sector
  ⇒ stronger reactions

Introduction
OOOO

supply of privacy
OOOOOOOOOO

demand for privacy
OOOOOOOOOOOOO●OOOOOO

Stock Market Reactions

Introduction
○○○○
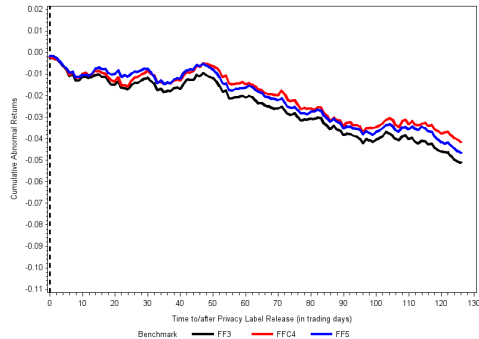
supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○●○○○○○○

Stock Market Reaction

Data Types Collected - High

Data Types Collected - Low

▸ More negative stock market reaction when the firm collects more data: -10% vs. -3%

Introduction
oooo

supply of privacy
oooooooooo

demand for privacy
ooooooooooooooo●oooo

Evidence on Investor Underreaction

‣ **No immediate reaction after the privacy label policy, consistent with...**
- "privacy paradox": will consumers react at all?
- investor inattention

‣ **Supporting evidence: investors react swiftly when...**
- the first post-policy earnings report is released → consumers do react
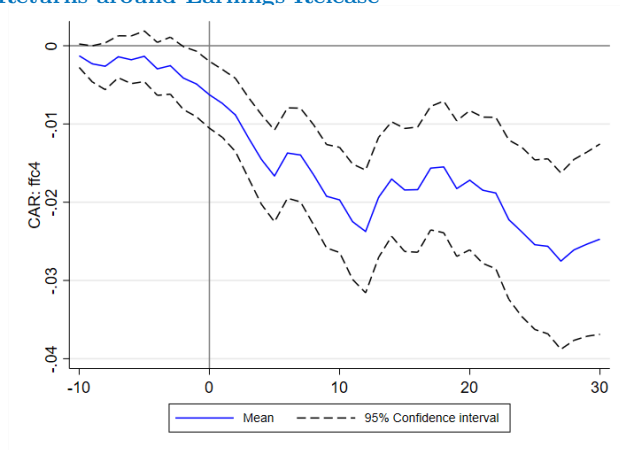- the policy is more salient

Privacy Paradox
Impact of Privacy Labels on Earnings

Sample	All	All	Retail & Service	All	All	Retail & Service
	(1)	(2)	(3)	(4)	(5)	(6)
Post × %(Data Used to Track You)	-0.641**	-0.833**	-1.300**	-1.407***	-1.568**	-2.252**
	(0.28)	(0.40)	(0.50)	(0.37)	(0.60)	(0.86)
$Size_{q-4}$		-0.607	-1.198		-1.556	-1.083
		(0.63)	(0.73)		(0.94)	(0.81)
$Cash_{q-4}$		1.883	2.354		1.098	-1.919
		(2.68)	(3.53)		(2.18)	(2.13)
$Tangible_{q-4}$		-1.696	-4.844		-5.462	-18.908
		(5.40)	(6.85)		(7.03)	(12.76)
$Leverage_{q-4}$		-1.216	-1.851		-3.173**	-3.641**
		(1.54)	(1.81)		(1.45)	(1.72)
$EBITDA_{q-4}$		0.915	0.287		2.033	4.117
		(1.94)	(2.31)		(6.03)	(8.75)
Weighted by	Downloads	Downloads	Downloads	Revenue	Revenue	Revenue
Firm	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y
Observation	3260	3095	1364	521	505	277
R-sq	0.72	0.73	0.73	0.86	0.87	0.88

▸ Large drop in EPS for firms with a higher fraction of apps tracking users

Introduction
OOOO

supply of privacy
OOOOOOOOOO

demand for privacy
OOOOOOOOOOOOOOOO●OO

Privacy Paradox
Cumulative Abnormal Returns around Earnings Release

‣ CAR drops around the first earnings announcement: consistent with investor underreaction

Introduction
oooo

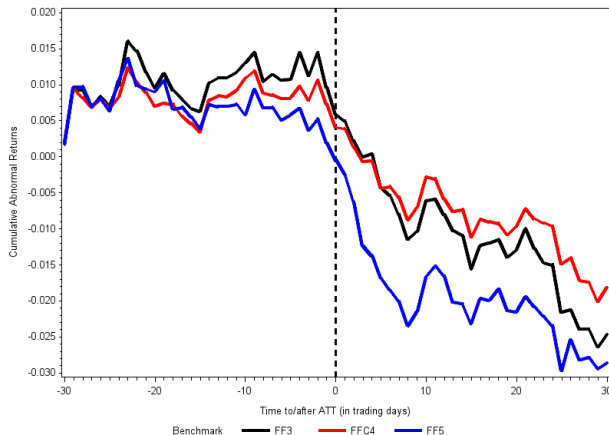supply of privacy
oooooooooo

demand for privacy
oooooooooooooooooooo

Investor inattention
Stock Market Reaction around *App Tracking Transparency*
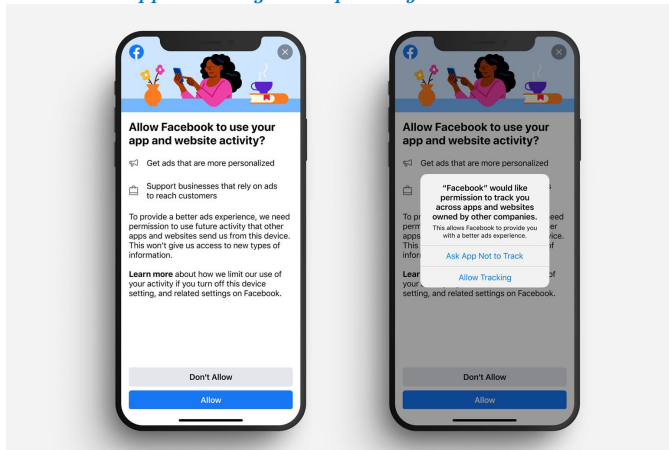
▸ Pop-up notification displayed to all *existing* users of Apps

Introduction
○○○○

supply of privacy
○○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○●○

Investor inattention
Stock Market Reaction around *App Tracking Transparency*

▸ Sharp decline in CAR following the enforcement of ATT on April 26, 2021

Conclusion

‣ **First comprehensive analysis of the supply and demand for data privacy**
  ‣ Setting: Apple's privacy labels + weekly app downloads/revenue

‣ **Lack of consumer awareness of firms' data collection practices**
  ‣ One explanation for privacy paradox; A key barrier to privacy protection

‣ **Ability to collect personal data matters for firms in the digital era**
  ‣ Race to the top: competition on privacy protection/product quality?

‣ **Policy implications**
  ‣ Transparency and disclosure quality are important
  ‣ Collaboration between regulators and platforms (Apple and Google)
    e.g., Google: privacy safety form, phase out third-party cookies in 2023