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

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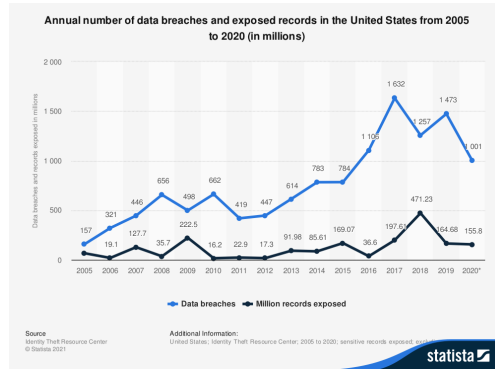
---

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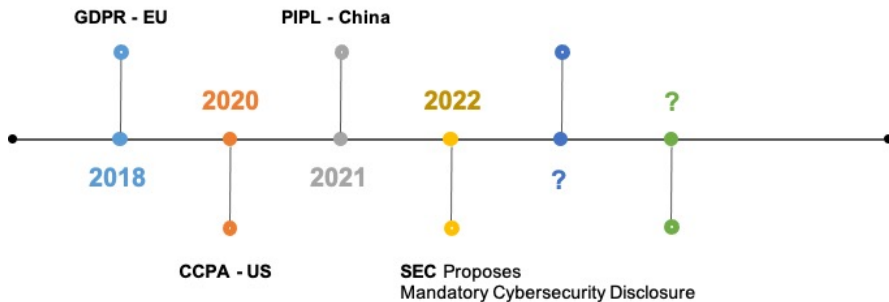[‡]London School of Economics; CEPR

Motivation

‣ **Growing public concerns & regulatory efforts towards privacy issues**

Motivation

‣ **Growing public concerns & regulatory efforts towards privacy issues**

Introduction
●○○

supply of privacy
○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○

Motivation

‣ **Growing public concerns & regulatory efforts towards privacy issues**

‣ **However, 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 and impact on firm valuation?**
  ‣ Privacy paradox: what users say ≠ what users do
  ‣ Help understand the value of data for firms in the digital era

This Paper: Mobile App Markets

- **Mobile apps are taking up an increasing amount of Internet usage time**
    - Smartphones account for 70% of the total digital media time (Comscore)
    - Apps account for 88% of mobile usage time (eMarketer)

Introduction
○●○

supply of privacy
○○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○

This Paper: Mobile App Markets

- **Mobile apps are taking up an increasing amount of Internet usage time**
  - Smartphones account for 70% of the total digital media time (Comscore)
  - Apps account for 88% of mobile usage time (eMarketer)

- **Introduction of Apple Store's privacy labels**
  - Dec 14, 2020: require app developers to disclose data collection
  - Visible and digestible, resembles food nutrition labels

Introduction
supply of privacy
demand for privacy
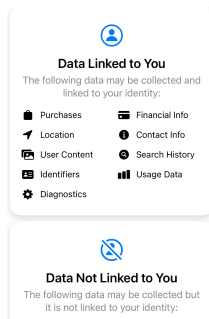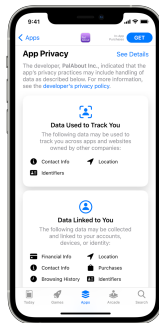
This Paper: Mobile App Markets

- **Supply: scrape privacy labels of most popular apps in 10 countries**
    - 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

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
  - Top data collectors: public firms, larger market shares, better ratings
  - Advertising and marketing: Games, news, shopping, entertainment apps

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
  - Top data collectors: public firms, larger market shares, better ratings
  - Advertising and marketing: Games, news, shopping, entertainment apps

- **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
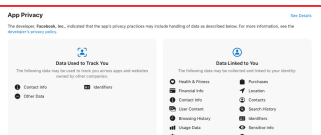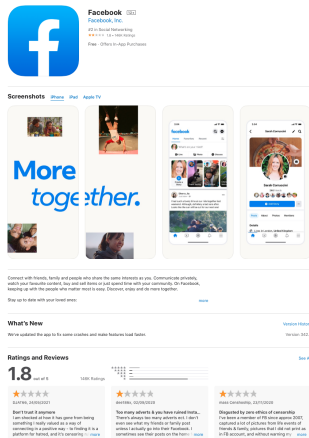 - Consumers in the US, UK, and France respond more negatively

Preview of Findings

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

- **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
  - Consumers in the US, UK, and France respond more negatively

- **Adverse stock market reactions**
  - -5% to -10% cumulative abnormal returns in 6 months
  - Driven by firms that harvest more data; Corroborated by decline in earnings

Introduction
ooo

supply of privacy
●ooooooooo

demand for privacy
ooooooooooooooooo

Laboratory, Sample, and Measurement

Introduction
○○○

supply of privacy
○●○○○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○○

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

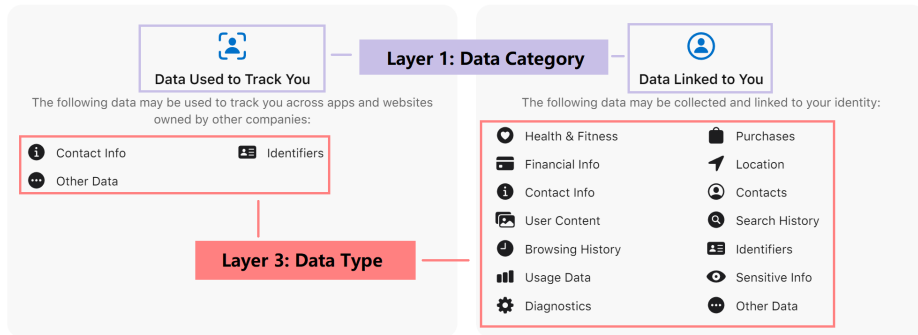▸ Data not Linked to You that the company generally aggregates into larger statistics

Introduction
○○○

supply of privacy
○○●○○○○○○

demand for privacy
○○○○○○○○○○○○○○○○

App Privacy: Facebook as an example
Front page

App Privacy

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.

Data Used to Track You

| | Layer 1: Data Category | | **Data Linked to You** |

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

| | | Contact Info | Identifiers |
| Other Data |

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

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

supply of privacy
○○○○●○○○○

demand for privacy
○○○○○○○○○○○○○○○○

Structure of Privacy Labels

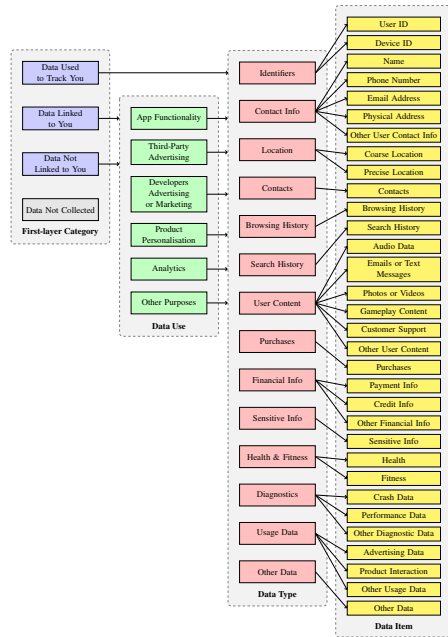1. Data Category: 3
   - Data used to track you
   - Data linked to you
   - Data not linked to you
   - Data not collected

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

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 >90% of store-wide downloads
- 60% of apps collect *data used to track you*
- Large variations across apps, especially within the same purpose

Introduction
○○○

supply of privacy
○○○○○○●○○

demand for privacy
○○○○○○○○○○○○○○○

Heterogeneity across app category
Layer 3: Data Types Collected
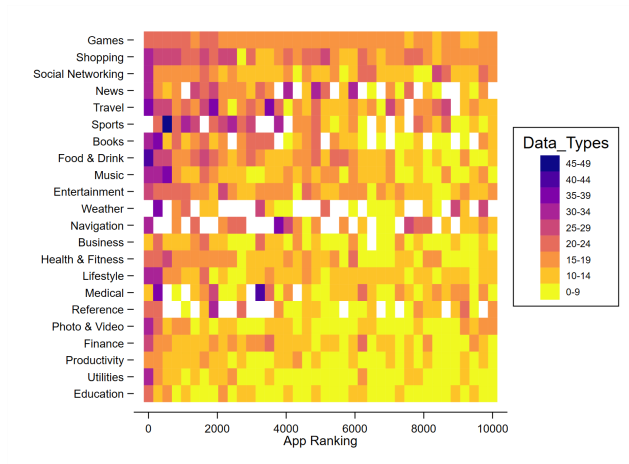
▸ High ranked apps collect more information
▸ Large within-category variation

Introduction
○○○

supply of privacy
○○○○○○●○●○

demand for privacy
○○○○○○○○○○○○○○○

Heterogeneity across app category
## Layer 3: Data Types Collected
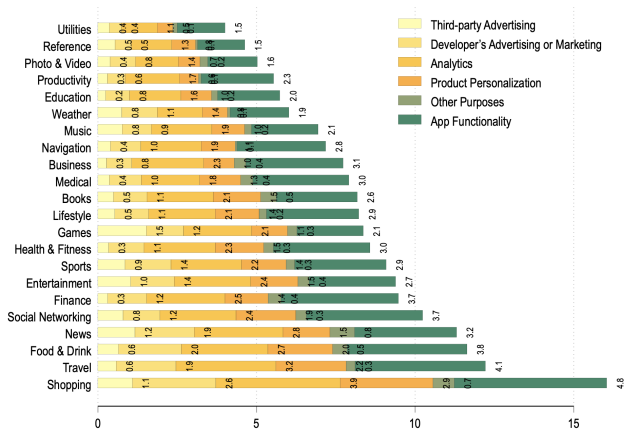
▸ 3rd party ad: Games, News, Entertainment, and Shopping

What determines data collection intensity?

	1(Data used to track you)		# Data Types Collected		# Data Items Collected	
	(1)	(2)	(3)	(4)	(5)	(6)
Within-category Market Share	-0.030**	-0.002	2.553***	3.148***	4.491***	5.746***
	(0.01)	(0.01)	(0.33)	(0.34)	(0.57)	(0.59)
In-app Purchase	0.239***	0.162***	1.187***	1.404***	0.699	0.916
	(0.01)	(0.01)	(0.28)	(0.33)	(0.50)	(0.57)
App Age	-0.065***	-0.034***	-0.430***	-0.127	-1.406***	-0.807***
	(0.01)	(0.01)	(0.16)	(0.15)	(0.27)	(0.27)
App Age^2	0.004***	0.002***	0.048***	0.031**	0.115***	0.083***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)
Ratings	0.114***	0.077***	3.025***	2.304***	4.690***	3.500***
	(0.01)	(0.01)	(0.24)	(0.24)	(0.42)	(0.42)
Public Developers	-0.002	-0.016	8.052***	8.004***	14.859***	14.664***
	(0.01)	(0.01)	(0.34)	(0.34)	(0.60)	(0.60)
Global App	-0.082***	0.045***	-3.299***	-1.403***	-5.412***	-1.852***
	(0.01)	(0.01)	(0.28)	(0.31)	(0.50)	(0.55)
Content Rating Age: 9+	0.131***	0.009	0.648	-0.701	0.802	-1.601
	(0.02)	(0.02)	(0.56)	(0.56)	(0.99)	(0.99)
Content Rating Age: 12+	0.160***	0.094***	3.925***	3.168***	6.769***	5.663***
	(0.01)	(0.01)	(0.32)	(0.33)	(0.56)	(0.57)
Content Rating Age: 17+	0.116***	0.101***	2.189***	1.943***	2.721***	2.492***
	(0.02)	(0.02)	(0.42)	(0.42)	(0.73)	(0.74)
Constant	0.133***	0.236***	0.099	1.889*	2.489	5.012**
	(0.05)	(0.05)	(1.15)	(1.13)	(2.01)	(2.00)
Category FE		Y		Y		Y
Observations	6,344	6,344	6,344	6,344	6,344	6,344
R-sq	0.195	0.295	0.193	0.238	0.189	0.230

Introduction
○○○

supply of privacy
○○○○○○○○○

demand for privacy
●○○○○○○○○○○○○○○

Consumers' Reaction to Privacy Label Release

Release Dates
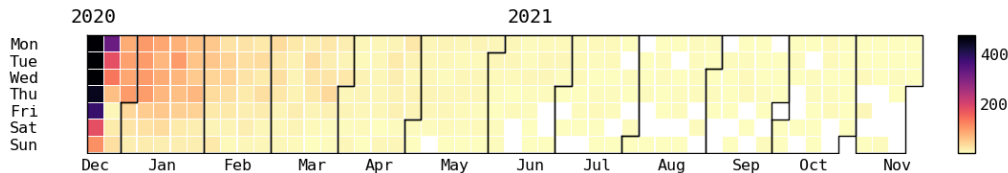
- ▸ Most apps release their privacy labels in a few weeks
- ▸ Privacy label policy only applies to iOS apps and affects iOS users
- ▸ Diff-in-Diff: pre vs. post; iOS (treatment) vs. Android (control)

Introduction
000

supply of privacy
000000000

demand for privacy
00●000000000000

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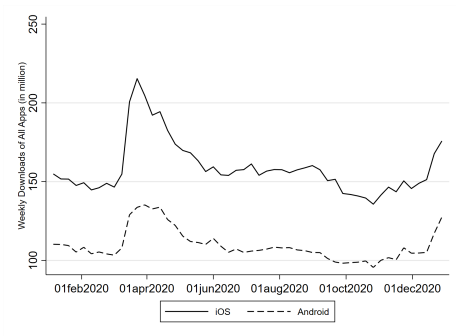
(a) weekly downloads
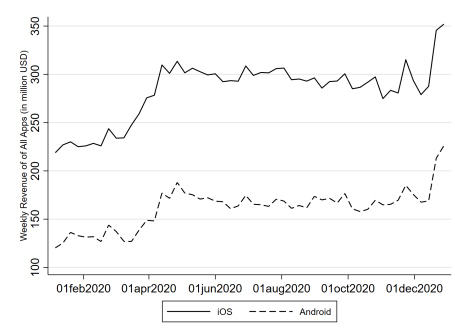
(b) weekly revenue

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

Introduction
○○○

supply of privacy
○○○○○○○○○

demand for privacy
○○○●○○○○○○○○○○○○○

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

Introduction
ooo

supply of privacy
ooooooooo

demand for privacy
ooooo●oooooooooo

Baseline DiD Estimation

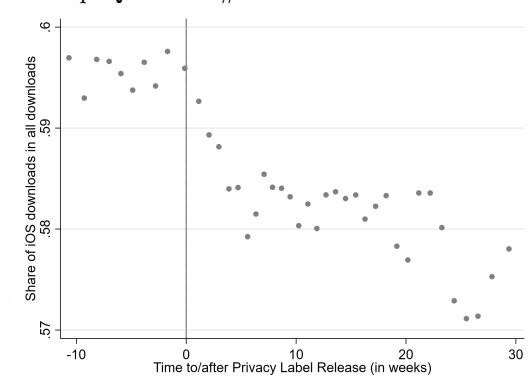|	Downloads			Revenue		
	(1) Main version	(2) All versions	(3) All versions	(4) Main version	(5) All versions	(6) All versions
Post	0.240***	0.238***	0.203***	0.255***	0.284***	0.207***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
iOS × Post	-0.117**	-0.117**	-0.138***	-0.135***	-0.195***	-0.151***
	(0.05)	(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: ∼ 11-14% drop
- Weekly revenue: ∼ 14-20% drop

Introduction
000

supply of privacy
000000000

demand for privacy
00000●000000000

Dynamics
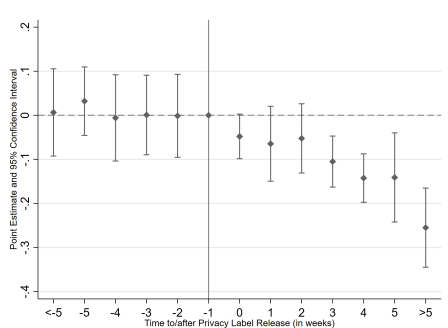
(a) # Data Types Collected - high
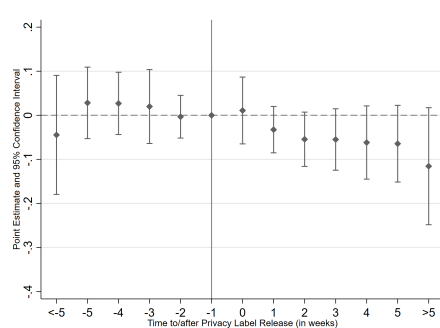
(b) # Data Types Collected - low

- Little pre-event effect: close to 0 and insignificant
- A larger decline for more privacy-intrusive apps

Heterogeneity
Data collection intensity

	Downloads (all versions)			Revenues (all versions)		
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.250***	0.189**	0.200**	0.227***	0.267***	0.283***
	(0.04)	(0.08)	(0.08)	(0.04)	(0.07)	(0.07)
iOS × Post	-0.063	0.064	0.070	-0.091**	-0.049	-0.029
	(0.05)	(0.09)	(0.09)	(0.04)	(0.07)	(0.06)
iOS × Post × 1(Data used to track you)	-0.103***			-0.109***		
	(0.03)			(0.04)		
iOS × Post × # Data Types Collected		-0.078***			-0.042*	
		(0.02)			(0.02)	
iOS × Post × # Data Items Collected			-0.071***			-0.043**
			(0.02)			(0.02)
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	0.570	0.569	0.569	0.920	0.920	0.920

‣ Apps that collect more data experience a larger decline in downloads and revenues

Introduction
000

supply of privacy
000000000

demand for privacy
0000000●0000000

Heterogeneity
Data Collection Purpose

	Downloads (all versions)					
	(1)	(2)	(3)	(4)	(5)	(6)
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

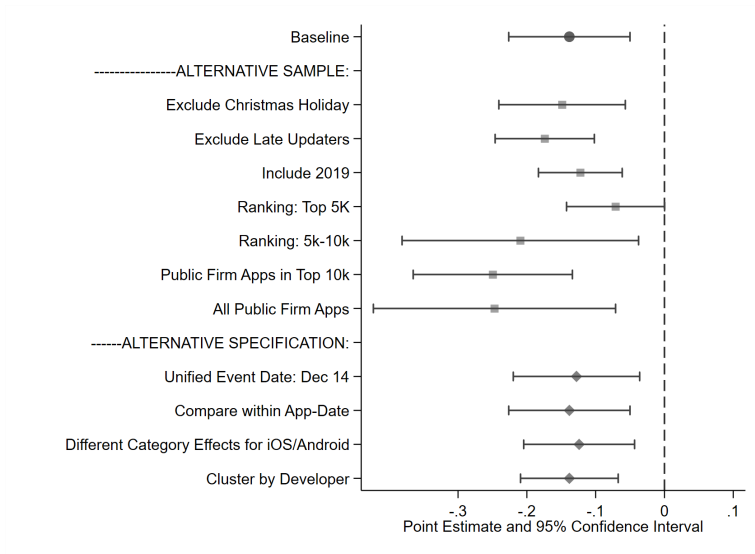‣ A larger drop in downloads for apps that collect data for more intrusive purposes

18 / 24

Introduction
000

supply of privacy
000000000

demand for privacy
000000000●000000

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 × 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 smaller drop when the app is less substitutable

Introduction
○○○

supply of privacy
○○○○○○○○○

demand for privacy
○○○○○○○○○●○○○○○

Robustness

Introduction
000

supply of privacy
000000000

demand for privacy
0000000**000**●**0000**

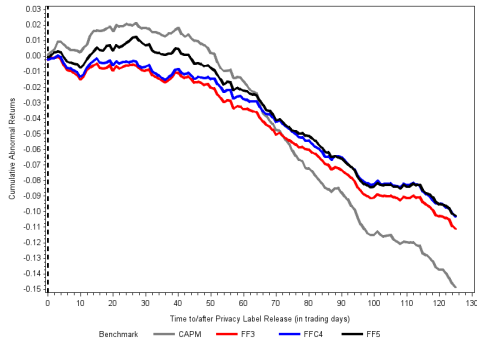# Cross-country comparison - Global Apps
Global Apps available in all 10 countries
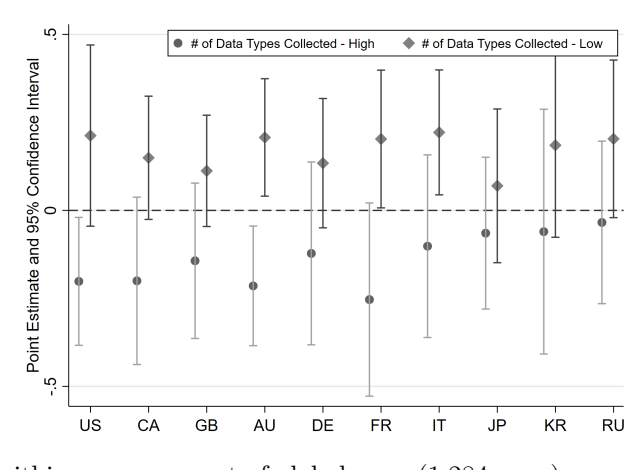
- Comparison within a common set of global apps (1,284 apps)
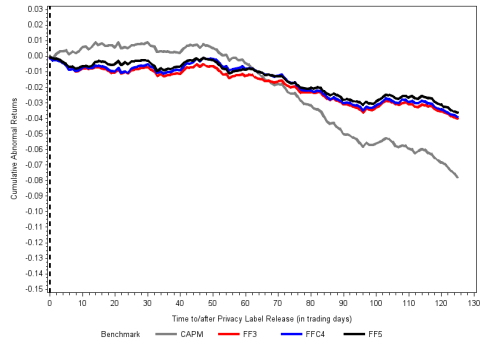- Consumers in the US, UK, and France are more averse to data collection

Introduction
ooo

supply of privacy
ooooooooo

demand for privacy
ooooooooooooo●ooo

Stock Market Reactions

Introduction

supply of privacy

demand for privacy

ooo

ooooooooo

ooooooooooooo●oo

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%

Stock Market Reaction

Sample split - # Data Types Collected

Firms	Group	Obs	CAPM	FF3	FFC4	FF5
All	H	226	-14.53	-10.95	-9.72	-9.88
			(-8.58)	(-7.67)	(-6.47)	(-7.31)
All	L	229	-7.70	-4.02	-3.41	-3.41
			(-5.50)	(-3.32)	(-2.82)	(-2.93)
All	H-L		-6.83	-6.93	-6.31	-6.47
			(-3.11)	(-3.70)	(-3.27)	(-3.62)
Retail & Service	H	82	-15.50	-11.53	-10.29	-10.31
			(-4.76)	(-4.13)	(-3.45)	(-4.06)
Retail & Service	L	82	-6.61	-3.07	-1.14	-2.39
			(-2.89)	(-1.47)	(-0.57)	(-1.20)
Retail & Service	H-L		-8.88	-8.46	-9.15	-7.92
			(-2.23)	(-2.43)	(-2.55)	(-2.45)

▸ More adverse stock market reactions when rely more on customer data

Introduction
000

supply of privacy
000000000

demand for privacy
00000000000000●

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

- **Policy implications**
  - Transparency and disclosure quality in firms' data collection policy is the key
  - Collaboration between regulators and platforms (iOS and Android)