Economics of free mobile applications: Personal data as a monetization strategy

Cecere, G. Institut Mines Telecom, Business School
Le Guel F. *Université Paris Sud*, Lefrere, V. *Institut Mines Telecom, Business School*

25 septembre 2019

Motivation and Research Question

Motivation : Collection of data is persasive but there are few empirical studies to monetize free digital goods.

How do developers combine personal data with more traditional strategies to monetize free apps ?

The underlining questions are :

- How do third parties influence the choice of monetization of smartphone developers ?
- How does the number of downloads influence the choice of monetization ?

Android Market and free apps

Worldwide market share of Operating system (Gartner, 2018)

- **Android 85.9%**
- Apple 14.%
- Microsoft and others 0.1%

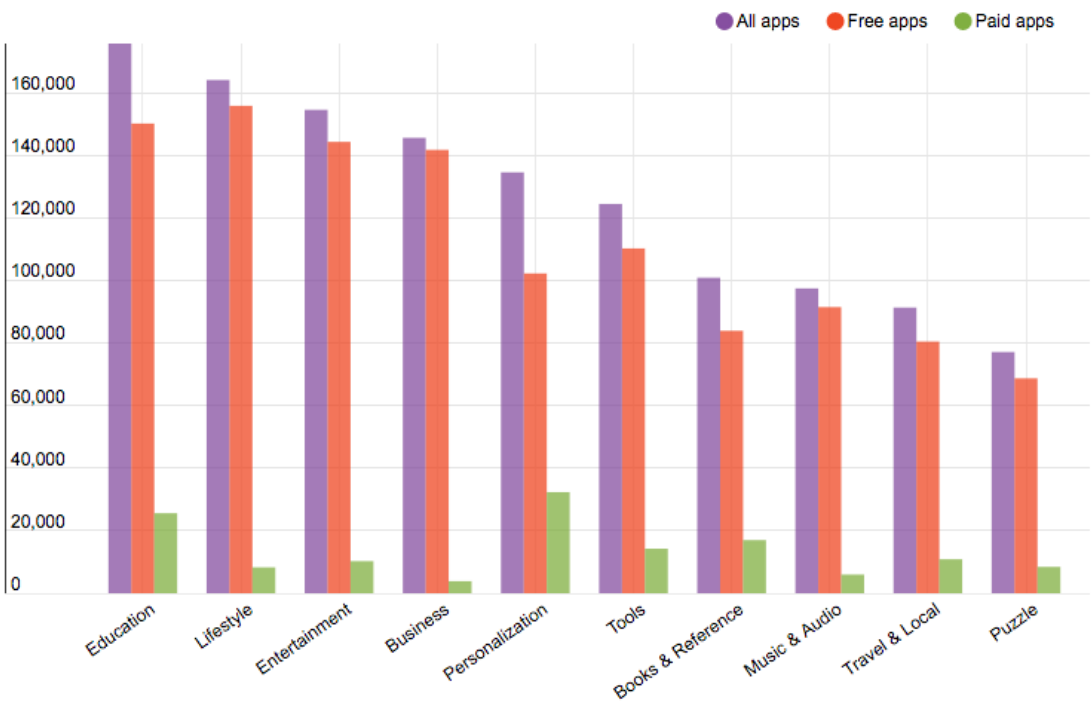
In 2015, when we collect our data... (Appbrain, 2015)

- **Free apps : 1,292,029 (85,6%)**
- Paid apps : 216,358 (14,4%)

January 2019

- Apple market counts about 2,000,000 apps (Statista, 2018)
- Android market counts 2 574 383 apps with **2,449,720 (95.2%)** (Appbrain, 2018)

Google Play : Free vs Paid apps worldwide



Methods and Main Findings

Method : Two-Stage Least Squares (2SLS) estimations

Main findings

- **Collection of users' data** can substitute advertising and freemium business models suggesting that it is a monetary strategy *per se*,
  - *Empirical evidences of personal data market previous highlighted by theoretical paper*
- Personal data are largely collected by by applications with high volume of downloads.
  - *Regulation of personal data market concentrated on big companies*
- The use of **third parties** combined with the collection of personal data can be used as alternative monetization strategy.
  - *Uncover the market of third parties*

Contribution to the literature

Economics of mobile applications

- Developers aim to increase downloads (Yin et al. 2014, 2016 ; Comino et al., 2019) and search costs influence developers' strategies (Ershov, 2019)
- *Contribution : Free apps and third parties*

Economics of free digital goods

- Economics of platform : one side of the market receive free services (Parker and Van Alysten, 2005) ; Monetization of free digital goods (Evans, 2011 ; Lambrecht et al. 2014 ; Bresnahan et al. 2015)
- *Contribution : Identify the combination of business models*

Economics of privacy

- Acquisti et al. (2016) identify three different markets of personal data. Theoretical literature models the market of personal data (Casadesus-Masanell and Hervas-Drane, 2015 ; Lefouili and Lei Toh, 2017)
- *Contribution : How personal data used as business model per se*

Monetization of applications

- Ad is the most used strategy but in-apps purchase can increase income of developers (Bresnahan et al. 2015)
- Kummer and Schulte (2019) highlight a trade off between price and personal data

The previous empirical literature has investigated the differences among free and paid applications...

Contribution

- Understand how personal data are combined with more traditional monetization strategies of free digital goods and add to empirical literature on mobile apps

Data : Google Play applications

- Public available data of free Google Play apps collected in May and June 2015
- 475 857 free applications Google play apps classified by Privacy Grade

Our sample includes :

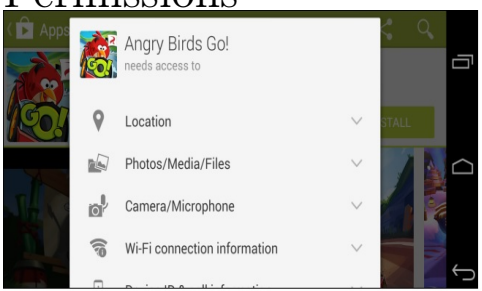
- Apps characteristics, Developers characteristics, Permissions, Google Play Category (Education, Games...)

Not used in the previous empirical literature

- Thirds parties
- Data elaborated by Privacy Grade

Angry Bird permissions are visible to users, third parties not

Permissions



Thirdparties

Libraries

Ad network libraries

- AdMob
- Facebook Audience Network
- Unity Ads
- AdColony
- Vungle
- Integral Ad Science
- Ironsource

Social libraries

- Facebook
- Google Play Games Services

Development tools

- Android Support library
- Firebase
- Google Cloud Messaging (GCM)
- Android Architecture Components
- Google Analytics
- Google gson
- Okio
- Picasso
- BoltsFramework
- Crashlytics
- Retrofit
- FMOD Ex Programmers API
- Unity 3D
- ReactiveX
- Flurry Analytics
- greenrobot EventBus
- Dagger
- Moat Analytics
- HockeyApp
- Disk LRU Cache

Thirds parties or libraries

Third parties include libraries (program, framework by other developers to access some functionalities) or companies (doing mobile analytics or connection between social networks and apps) :

	mean	sd	min	max	Number of third parties
Advertising	0.324	0.468	0	1	79
Payment	0.036	0.186	0	1	8
Social networking	0.137	0.344	0	1	10
Utility	0.187	0.390	0	1	71
Development aid	0.039	0.194	0	1	9
Mobile analytics	0.078	0.268	0	1	12
Observations	475867				

Two main variables of interest

We need to model the two choices of monetization :

- Advertising
  - App uses an advertising thirds parties
- Freemium

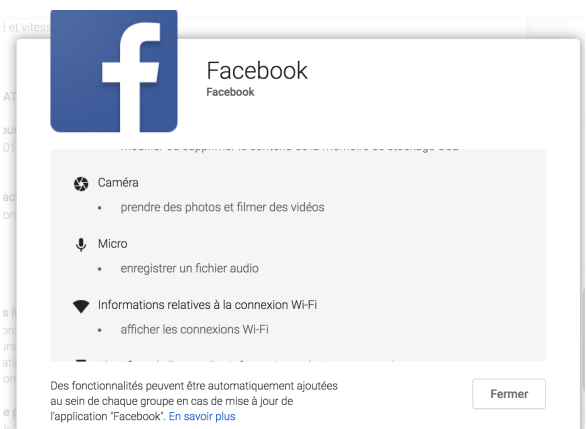
We investigate whether the **collection of sensitive data** can be also considered as business model

- We use two alternative measures of personal data collection : Sensitive data and Badgrade.

Permissions

Permissions system is the mechanism for a developer to interact with functionality or data from the users'smartphone.

- *"Permissions considered intrusive or which can affect the device"* Examples : Access to contact, Camera, Geo-localization...



List of permissions used to construct the variable *Sensitive Data*

Sensitive data takes value 1 if the app collects at last one sensitive data.

	Sarma *et al.*, 2012	Cecere *et al.*, 2019
Approximate location (network-based)	✓	✓
Precise location (gps and network-based)	✓	✓
Read phone status and identity		✓
Reroute outgoing calls	✓	
Read calendar events plus confidential information	✓	
Read your contacts	✓	
Read your web bookmarks and history	✓	
Read precise phone states	✓	
Read your text messages (sms or mms)	✓	
Receive text messages (mms)	✓	
Receive text messages (sms)	✓	
Record audio	✓	✓
Receive text messages (wap)	✓	
Read call log	✓	
Take pictures and videos		✓

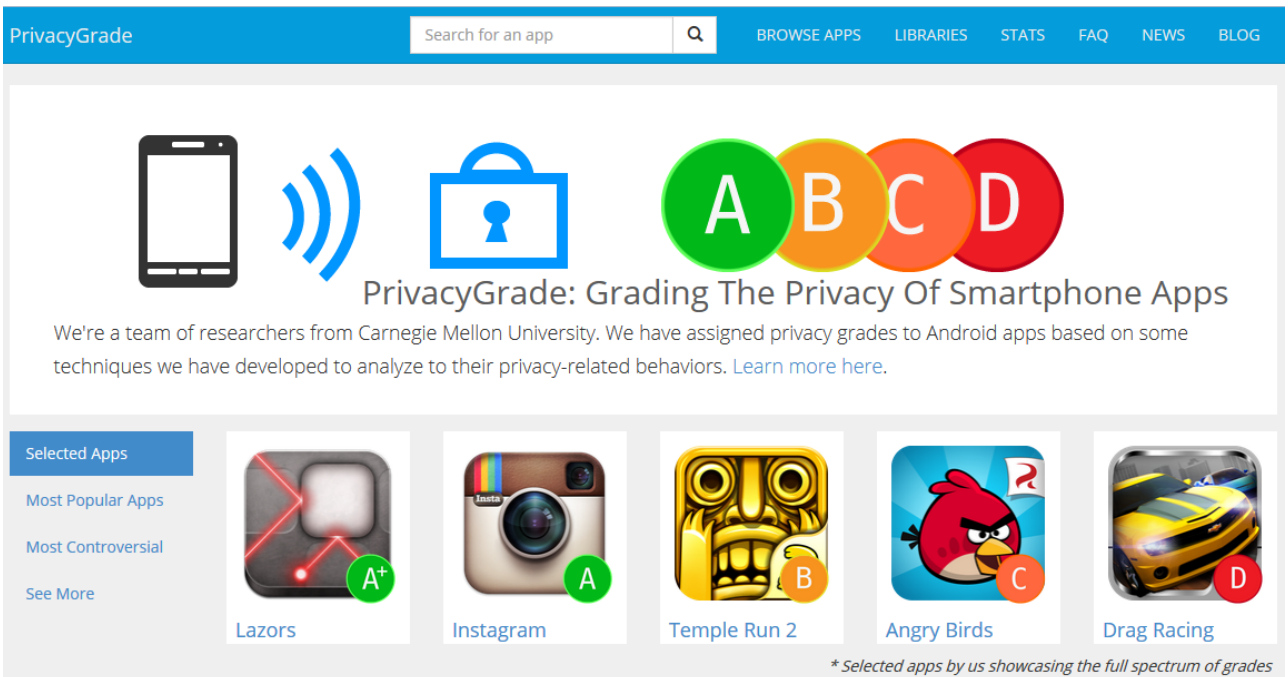
How we construct the variable Badgrade

- Privacy grade measures the gap between users' expectations about an app's behaviour and the app's actual behaviour in terms of privacy reeling on permissions
- It's on going project
- The grade ranges from A+ to D

Badgrade= Dummy variable which takes value 1 if the app has a note between B, C and D.

Example of Privacy grade rating system

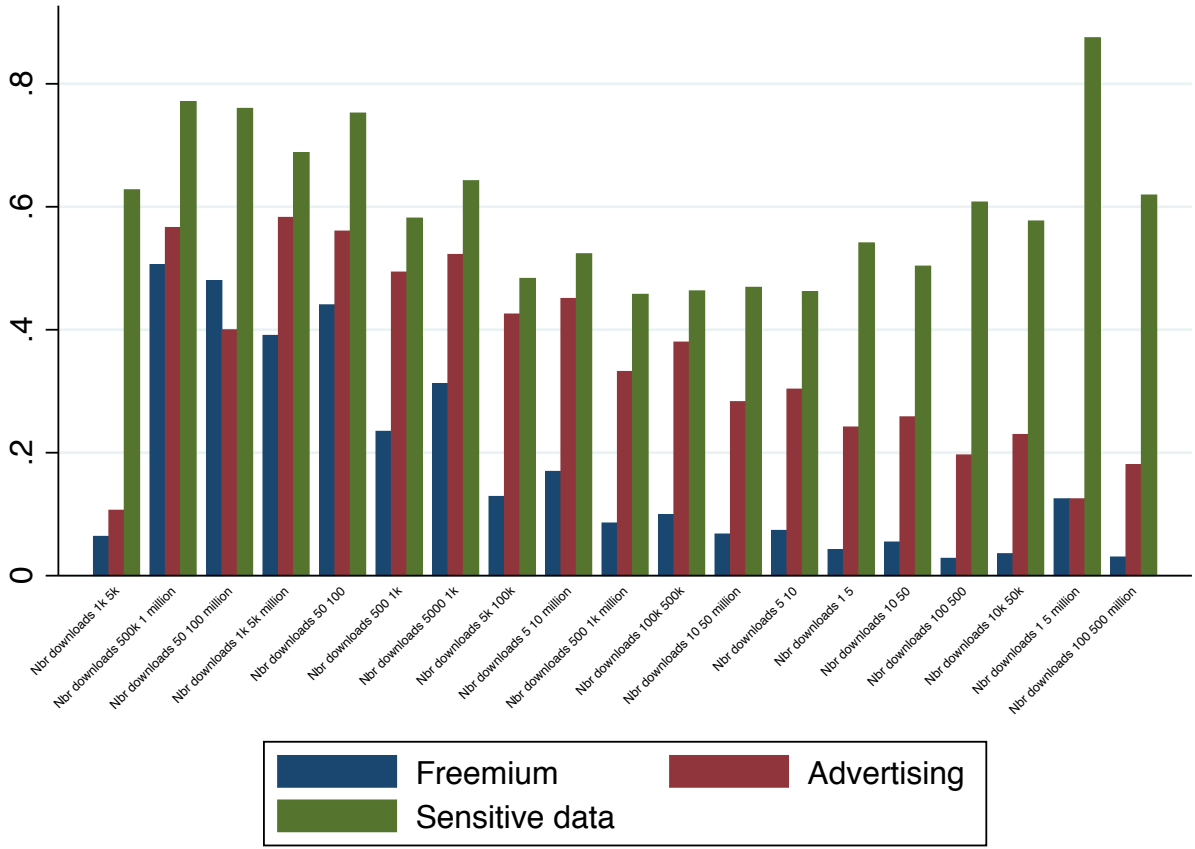
PrivacyGrade



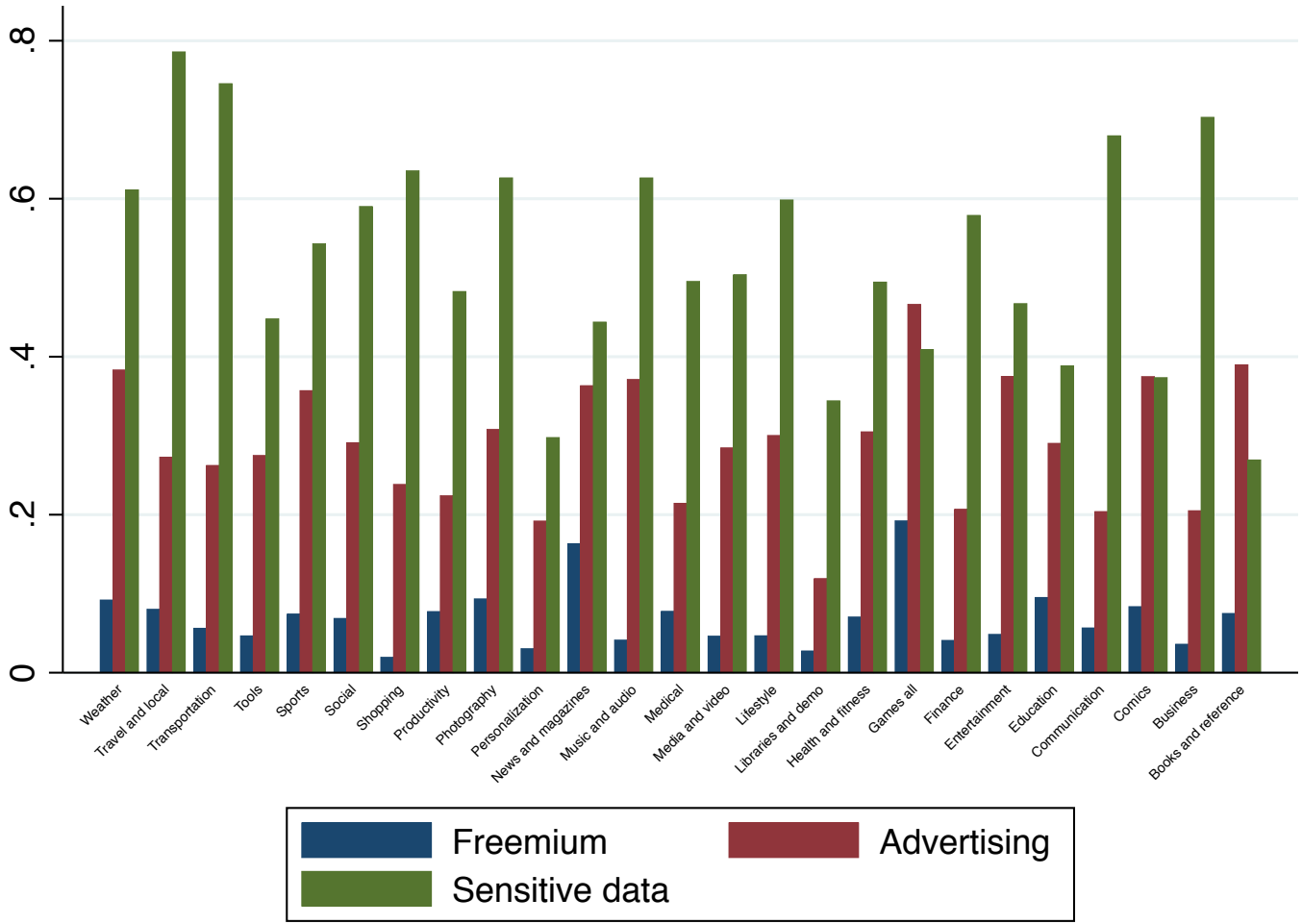
Top 15 thirds by strategy of monetization

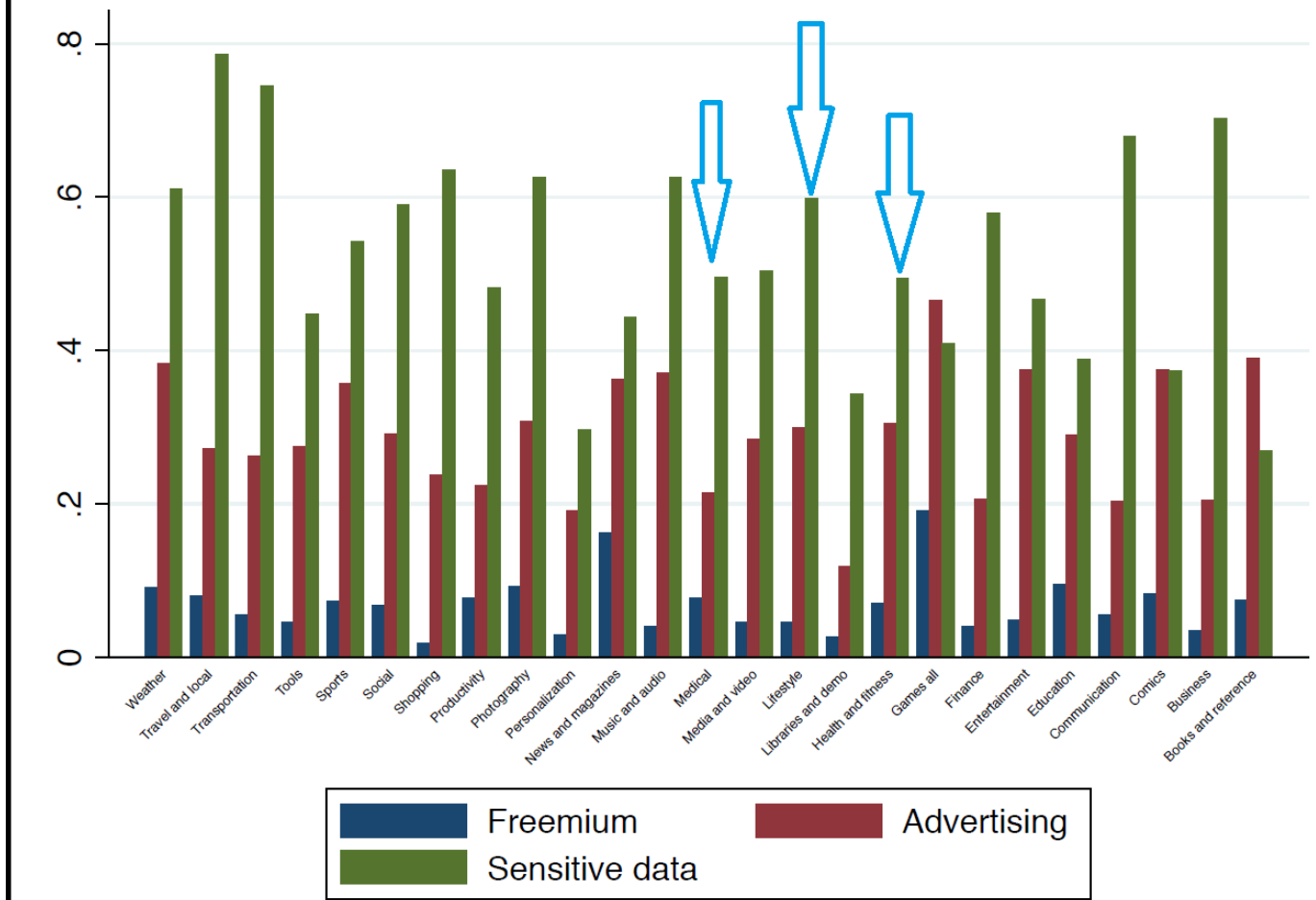
Advertising (1)		Freemium (In-app purchases) (2)	
Thirds	Percentage	Thirds	Percentage
Admob	86.5%	Admob	31.1%
Facebook	20.9%	Facebook	25.8%
Flurry	10.6%	Flurry	18.2%
Twitter4j	8.6%	Chartboost	9.5%
Millennial media	7.8%	Unity3d	8.3%
Inmobi	7.1%	Twitter4j	6.2%
Chartboost	5.9%	Tapjoy	5.9%
Unity3d	5.3%	Inmobi	4.9%
Paypal	4.5%	Millennial media	4.6%
Revmob	4.5%	Nostra13	4.4%
Jsoup	4.3%	Oauth	4.2%
Biznessapps	3.8%	Adobe	4.1%
Nostra13	3.8%	Amazon	3.9%
Mopub	3.6%	Mopub	3.8%
Oauth	3.3%	Loopj	3.3%

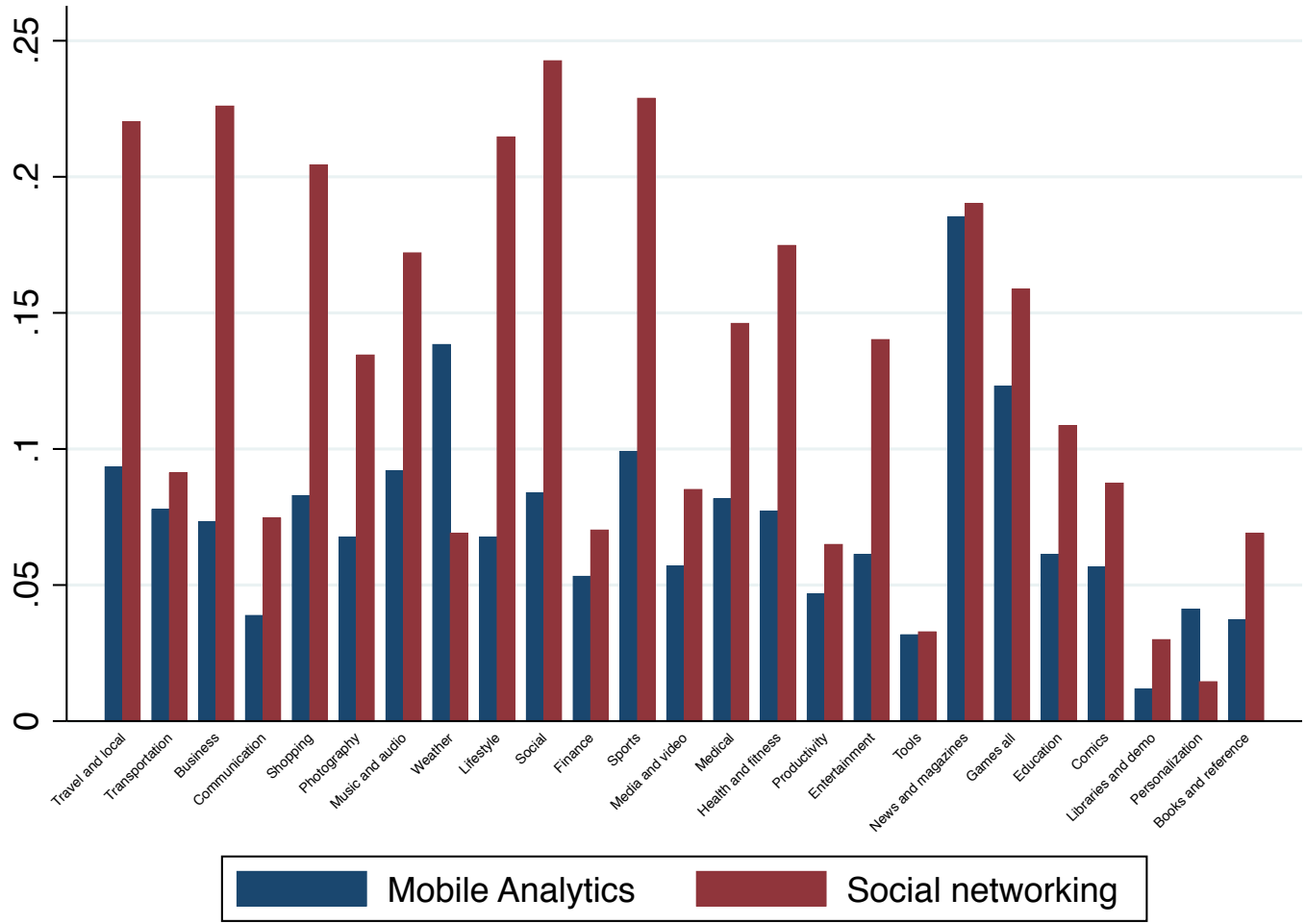
Monetization strategies and volume of downloads



Business models and sensitive data by category




0
.2
.4
.6
.8
Weather
Travel and local
Transportation
Tools
Sports
Social
Shopping
Productivity
Photography
Personalization
News and magazines
Music and audio
Medical
Media and video
Lifestyle
Libraries and demo
Health and fitness
Games all
Finance
Entertainment
Education
Communication
Comics
Business
Books and reference
Freemium
Advertising
Sensitive data


0
.05
.1
.15
.2
.25
Travel and local
Transportation
Business
Communication
Shopping
Photography
Music and audio
Weather
Lifestyle
Social
Finance
Sports
Media and video
Medical
Health and fitness
Productivity
Entertainment
Tools
News and magazines
Games all
Education
Comics
Libraries and demo
Personalization
Books and reference
Mobile Analytics
Social networking

Empirical model : 2SLS estimation

The IV equation takes the form of the OLS equation, the endegenous variable is replaced by the predicted value :

$$Advertising_i = \beta_0 + \beta_1 \widehat{SensitiveData}_i + \beta_2 X_i + \epsilon_i$$
$$Freemium_i = \beta_0 + \beta_1 \widehat{SensitiveData}_i + \beta_2 X_i + \epsilon_i$$

where $\widehat{SensitiveData}$ is generated by the first stage regression in the IV framework :

$$SensitiveData_i = \beta_0 + \beta_1 \overline{SensitiveData_i} + \beta_2 \overline{Everyone_i} + \beta_3 X_i + \epsilon_i$$

Main regressions : IV estimations

	Freemium		Advertising		Sensitive data
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) First stage
Sensitive data	0.0199***	-0.2861***	-0.0142***	-0.3567***	
	(0.001)	(0.043)	(0.002)	(0.047)	
Average Sensitive Data (IV)					-45.6637***
					(2.296)
Average Everyone (IV)					-20.1311***
					(2.488)
Constant	-0.0334***	0.0572***	0.1595***	0.2610***	34.5153***
	(0.004)	(0.017)	(0.009)	(0.020)	(2.215)
Third parties	Yes	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes	Yes
Developer characteristic	Yes	Yes	Yes	Yes	Yes
Category fixed effects	Yes	Yes	Yes	Yes	Yes
Installation fixed effects	Yes	Yes	Yes	Yes	Yes
Underidentification (LM)		420.3815		420.3815	
P-value (LM-Stat)		0.0000		0.0000	
Weak identification (F-stat)		198.4614		198.4614	
Hansen's J statistic		2.8869		1.2690	
P-value (J-Stat)		0.0893		0.2600	
N	475867	475867	475867	475867	475867

Badgrade estimation

	Freemium		Advertising		Badgrade
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) First stage
Badgrade	0.0244***	-1.0613***	0.2232***	-1.1928***	
	(0.003)	(0.273)	(0.003)	(0.384)	
Average Everyone					-7.6418***
					(1.294)
Constant	-0.0283***	0.0106	0.1473***	0.1981***	4.5069***
	(0.005)	(0.018)	(0.009)	(0.025)	(0.756)
Third parties	Yes	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes	Yes
Developer characteristics	Yes	Yes	Yes	Yes	Yes
Category fixed effects	Yes	Yes	Yes	Yes	Yes
Underidentification (LM)		35.2233		35.2233	
P-value (LM-Stat)		0.0000		0.0000	
Weak identification		34.8572		34.8572	
N	475867	475867	475867	475867	475867

Estimations with interaction terms

	Freemium		Advertising		Sensitive data
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) First stage
Social networking	0.0959***	0.5566***	0.1396***	0.5205***	
	(0.009)	(0.054)	(0.013)	(0.049)	
Sensitive data	0.0275***	-0.2782***	-0.0137***	-0.3069***	
	(0.003)	(0.044)	(0.005)	(0.048)	
Social networking × Sensitive data	-0.0361***	-0.6167***	0.0028	-0.4712***	
	(0.011)	(0.074)	(0.014)	(0.067)	
Average sensitive data					-0.1268***
					(0.004)
Social networking × Average sensitive data					0.9492***
					(0.001)
Average Everyone					-0.1123***
					(0.003)
Social networking × Average Everyone					0.9112***
					(0.001)
Constant	-0.0283***	0.0625***	0.1435***	0.2304***	0.1280***
	(0.007)	(0.017)	(0.012)	(0.020)	(0.004)
Third parties	Yes	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes	Yes
Developer characteristics	Yes	Yes	Yes	Yes	Yes
Category fixed effects	Yes	Yes	Yes	Yes	Yes
Installation fixed effects	Yes	Yes	Yes	Yes	Yes
Underidentification (LM)		456.4639		456.4639	
P-value (LM-Stat)		0.0000		0.0000	
Weak identification (F-stat)		105.6745		105.6745	
Hansen's J statistic		0.3689		3.3698	
P-value (J-Stat)		0.5436		0.0664	
N	475867	475867	475867	475867	475867

Discussion

- Developers have a better understanding about this market
- Platform can also improve transparency in forcing app to declare the thirds parties they used
- Third parties market is skewed
  - One of the main actors of Advertising thirds parties, it's also one of the major actor of smart phone application and also the owner of the platform
- There is link between thirds parties and personal data, highlighting the fact that "Killer apps" use more personal data and can also be presented on the library market raise some concerns

Limitation

- Only cross sectional data but the IV allows to estimate casuality
  - estimate correlations rather than causalities
- Further research is needed to investigate applications with no clear monetization strategy
- More details investigation between thirds, apps and personal data is needed

Thank you for your attention !