Economics of free mobile applications: Personal data *

Grazia Cecere†, Fabrice Le Guel‡, Vincent Lefrere §

July 10, 2018

Abstract

The large majority of smartphone apps are zero priced. To generate revenue, developers have to monetize theirs apps, however little is known about their strategies. The theoretical literature underlines the importance of personal data for Internet companies' strategies but their implication in the smartphone applications market remains relatively unexplored. We provide empirical evidence of the monetization strategies related to free apps by studying how the collection of personal data is combined with more traditional revenue sources such as advertising and in-app purchase. We have unique data to measure how apps are monetized based on information related to 475,867 free applications available on the Google Play platform combined with data on applications' privacy-related behaviors provided by PrivacyGrade. Among the apps in our dataset, 9% collect personal data and use no other monetization strategy. Social networking and utility third parties are largely used by apps that rely on personal data as a monetization strategy. Apps with more than 1 million downloads rely more on personal data.

JEL CODE: D82, D83, M31, M37

*We are grateful to Alessandro Acquisti, Alexandre de Corniere, Babur De los Santos, Anna D'Annunzio, Antoine Dubus, Anja Lambrecht, Amedeo Piolatto, Mirko Tonin, Catherine Tucker and Pai-Ling Yin and participants at the TSE Digital Seminar (June 2017), the AFSE Conference (June 2017), the ZEW ICT Conference (June 2017), the Peex Meeting Seminar (November 2017), Universita di Padova Seminar (March 2018), 2018 Royal economics society conference,The Bordeaux JMA (June 2018), INFORM EURO 2018 Valencia Operational Research and Management Science (July 2018), and the 45th EARIE Annual Conference (August 2018), for comments and suggestions. Work supported by the DAPCODS/IOTics ANR 2016 project (ANR-16-CE25-0015).

†Institut Mines Telecom, Business School and DigitalSociety Institute. Email: grazia.cecere@telecom-em.eu

‡RITM-University of Paris Sud - Paris Saclay. Email: fabrice.le-guel@u-psud.fr

§Telecom Ecole de Management, Institut Mines Telecom-RITM-University of Paris Sud - Paris Saclay. Email: vincent.lefrere@u-psud.fr

1 Introduction

The smartphone apps market has experienced an exceptional growth in the recent years with an increase of free applications. Like other digital goods, free smartphone applications are related to different revenue streams. We can mentioned in-app purchase, advertising, and e-shopping even though the personal data collected by these apps is also a source of revenue (Lambrecht *et al.*, 2014). The literature on the economics of mobile applications investigates the factors influencing the success of mobile applications by looking at differences between paid and free applications, or comparing the Google Play and Apple app markets (Ghose and Han, 2014; Yin *et al.*, 2014; Kummer and Schulte, 2016). But, the increased number of free compared to paid applications is drawing attention to the business models of free apps. The present article fills in a gap in the literature by analyzing how developers combine different strategies to monetize free apps focusing in particular on the market of personal data. In particular, it investigates whether personal data are used to complement or substitute for an advertising and/or an in-app purchase. An examination of the different types of monetization strategies allows us to identify the third parties or libraries associated to each application. However, our understanding of the economics and managerial implications of these third parties is limited, although we know that they are essential for the distribution of ads, business analytics and the connecting of apps to social networking services. Although third parties are essential for the distribution of ads, business analytics and the connection of apps to social networking services, to our knowledge there is no previous work that assesses their role in the choice of monetization strategies.

Our sample includes 475,867 free Google Play apps evaluated by Privacy Grade in 2015 which correspond to 36% of the total market [1]. Thus, we combine two sets of data: publicly available data from Google Play, and data on the apps tested and ranked by

[1] This sample has been downloaded by a crawler develop by the CMU researcher

PrivacyGrade[2] (Lin *et al.*, 2012, 2014). At the time of our study in 2015, the Google Play platform included 1,292,029 free apps [3] **which represent about 85% of the total available applications, je ne sais pas ce qu'on voulait dire ici mais le chiffre est faux ou mal exprimés (85% c'est la part de marche du playsotre mais on le dit plus bas**. **To our knowledge, there is a lack of empirical investigation on the business models of these free apps (repetitif).** This paper investigates the differences related to monetizing apps, and the link between advertising, in-app purchase, and personal data. The availability of large amounts of data has enabled the development of different revenue combinations, which is challenging our understanding of both the business models related to free services and the competition.

Our article contributes to three streams of literature: the economics of free digital goods, the economics of mobile applications, and the economics of privacy. The growing share of free services in digital economics (Brynjolfsson and Saunders, 2010) is challenging economics approaches to measuring market power and antitrust analysis. While it is recognized that competition at zero prices is a special case of price competition (Smith and Telang, 2009; Evans, 2011), empirical work investigating firms' monetization strategies related to zero priced goods is scarce, especially in the application market. Also, similar to other digital goods, mobile applications can combine various revenue streams. Bresnahan *et al.* (2015) show that advertising is the most frequent revenue stream for developers on the Android platform, and usually, is combined with in-app purchase. However, the number of free applications has increased over time which is highlighting the need for studying innovative business models to create new revenue streams.

[2] http://www.privacygrade.org Last retrieved February 2018

[3] http://www.appbrain.com/stats/free-and-paid-android-applications Last retrieved February 2015

The literature on the economics of mobile applications shows that the market for applications has a long tail distribution (Garg and Telang, 2013; Li *et al.*, 2016; Gabaix, 2016) i.e. very few apps attract the majority of downloads, leading to competition over numbers of downloads. Competition among developers aspiring to be top ranked, results in the implementation of different strategies. Bresnahan *et al.* (2015) describe top players as 'killer apps', and show that demand for apps is concentrated on a very small number due, perhaps, to the increasing returns from adoption. Li *et al.* (2016) estimate the characteristics and the rationales of app developers that buy downloads to increase their visibility in the market. Comino *et al.* (2016) show that updates can be released strategically to increase downloads. Hence, developers on the iTunes platform compared to the Google Play platform, seem to mainly rely on updates to increase their rankings. The study by Ghose and Han (2014) uses a structural model to estimate the factors influencing consumer demand for apps. This demand increases with the precision of the app description and the number of its previous versions, and decreases with in-app purchasing options and advertising. This drives the competition among developers, and highlights the challenges faced by developers that distribute free apps. Yin *et al.* (2014) investigate the differences between game and non-game apps aiming to achieve killer app status. They find that developers of non-game apps have a higher chance of developing a killer app if they focus on a single app and improve it via updates. In the case of game apps, the probability of a particular app being successful increases with the developers' experience. Yin *et al.* (2014) show that the strategy adopted by developers depends on their categorization, since the patterns of competition differ among categories. Using demand-ranked data for the Apple iOS market, Garg and Telang (2013) show that while free apps are the most frequently downloaded apps, in the subsample of paid apps, hte top ranked ones are more downloaded, regardless of the price. While price has an impact on the demand for apps, there are other factors that influence this last one. All of these results are of interest in the context of our study as they highlight the factors affecting the demand of apps but they do not give insight on the strategy of monetization of free applications.

Our article contributes to the economics of privacy and industrial organization literature as we assess how personal data can be used to complement the revenue from free services. The economics and marketing literature identifies various strategies used to monetize digital goods sold at zero price, namely advertising, personal data, and subscription (Lambrecht *et al.*, 2014). There are several examples of two-sided markets, and particularly, in the case of digital markets where one side of the market receives free services or products aimed at increasing the demand for complementary products (Parker and Van Alstyne, 2005). A cornerstone of Internet companies business model is the personal information provided by users which can be used to improve the quality of the services provided, and to allow personalized advertising (Casadesus-Masanell and Hervas-Drane, 2015).

Spiegel (2013) provides an interesting theoretical contribution which models the distribution of free software bundled with ads, defined as 'adware'. The software is able to collect data and display ads. Although it involves some loss of privacy, the software provides information to consumers. The increased quality of the information provided to consumers is associated to increased benefits for them. Prior to the installation of this software, the user is informed of the permissions required by the developer (which are displayed on the smartphone screen before the user downloads and installs the particular application). The seminal contribution of Kummer and Schulte (2016) shows that Android OS smartphone users take account of these permissions when downloading applications. The authors show that applications developers require more intrusive permissions in the case of free applications compared to paid applications, and that consumers seem to prefer less intrusive apps. Work in the economics of privacy provides evidence of the existence of different markets of privacy. First, a market where individuals provide personal data in exchange for free services; second, a market that involves commercialization of personal data by data brokers; and third, a market where individuals pay to protect their data (Acquisti *et al.*, 2016). In particular, the literature on the economics of privacy suggests that personal data can be exchanged among data brokers (Lambrecht and Tucker, 2017). The present article aims to provide insights into these markets for personal data related to mobile apps.

However, the literature overlooks the relations that might exist between these three monetization strategies. Indeed, personal data can potentially be related to advertising as it allows targeted advertising but it could also be considered a monetization strategy in its own right. Personal data can then be used to complement these traditional strategies or can be considered a monetization strategy *per se* in the data broker market (Acquisti *et al.*, 2016). Personal data gathered from mobile phone can be used to infer socioeconomic characteristics, e.g. to estimate the wealth of individuals (Blumenstock, 2018) or to assess consumer preferences, e.g. using mobile location data Athey *et al.* (2018) estimated both consumer preferences for restaurants and latent characteristics for each restaurant.

The managerial and policy implications of these findings are threefold. First, a study of the apps market could help developers identify the most profitable strategies for the distribution of free apps, and allow mobile analytics to implement more efficient marketing strategies. Second, it can be informative for policy makers on the functioning of this competitive market. Personal data are required to customize services and provide personalized advertising; alternatively, they can be commercialized by data brokers. Indeed, personal data are extremely valuable in allowing the targeting of con-

sumers to improve the match between seller and buyer. Third, it reveals the relations between third parties and the app monetization strategy. On peut mieux faire, je repasserai dessus

The article is organized as follows. The next section describes the data and key features of the app market for personal data, followed by a section devoted to the econometric models used to test our main assumption that developers can use personal data to monetize their applications. The fourth section discusses the econometrics results and the paper concludes with a final section.

2 Data and main variables of interest

2.1 Data sources

Our study focuses exclusively on free applications commercialized in the Google play market which is nowadays the largest OS worldwide with a market share of around 87,7% worldwide [4] in 2015 it represented 82.8% worldwide [5]. We examine the monetization of free applications focusing on developers' strategies. We match data from two websites - Privacy Grade and Google Play Store (during May and July, 2015). First, we collected publicly available Privacy Grade data. Privacy Grade is an ongoing project of a group of computer science researchers at Carnegie Mellon University. We then restrict our sample to apps that have been evaluated by Privacy Grade. The project is aimed at measuring the gap between users' expectations about an app's behavior in terms of privacy, and the app's actual behavior. The researchers evaluate every app and grade them based on this difference. In June 2015, Privacy Grade had evaluated a random

[4] https://www.statista.com/statistics/266136/global-market-share-held-by-smartphone-operating-systems/ Last retrieved February 2018

[5] https://www.idc.com/promo/smartphone-market-share/os Last retrieved August 2015

sample of 475 867 apps. These data are originally compared to publicly available data, since Privacy Grade evaluates the relevance of the personal data required by permissions. Figure 4 (in the appendix) shows an example of the grading system used by Privacy Grade. In addition, the data provided by Privacy Grade includes information on the third parties related to each application through libraries.

Second, we match Privacy Grade data with publicly available data from Google Playstore. Moreover we collected available data including detailed characteristics of the apps, such as number of downloads, Google category (Games, Health, Social, etc.), type of permissions required, and user evaluations. Table 1 describes the main variables, including the summary statistics per type of monetization strategy.

Table 1: Descriptive statistics of all sample and summary statistics by monetization strategy

Variable	Mean (1)	Min.	Max.	Ads (2)	In-app (3)	Pers. data (4)	None (5)
Personal data	0.177	0	1	.	.	.	.
Advertising	0.324	0	1	.	.	.	.
In-app purchases	0.088	0	1	.	.	.	.
Playstore rating	3.698	0	5	3.737	3.818	3.550	3.696
Everyone	0.583	0	1	0.559	0.563	0.188	0.665
Social networking	0.137	0	1	0.230	0.275	0.377	0.059
Utility	0.187	0	1	0.251	0.275	0.339	0.132
Apps by dvp	15.769	1	455	16.823	16.278	17.655	15.259
Developer website	0.768	0	1	0.756	0.885	0.872	0.745
Privacy Policy	0.146	0	1	0.137	0.309	0.206	0.126
Books and reference	0.049	0	1	0.060	0.042	0.020	0.049
Business	0.055	0	1	0.035	0.023	0.119	0.055
Comics	0.003	0	1	0.004	0.003	0.002	0.003
Communication	0.023	0	1	0.015	0.015	0.049	0.021
Education	0.085	0	1	0.076	0.092	0.059	0.092
Entertainment	0.074	0	1	0.086	0.041	0.062	0.074
Finance	0.025	0	1	0.016	0.012	0.028	0.028
Games all	0.191	0	1	0.275	0.418	0.156	0.148
Health and fitness	0.029	0	1	0.027	0.023	0.032	0.029
Lifestyle	0.067	0	1	0.062	0.035	0.098	0.065
Media and video	0.014	0	1	0.012	0.007	0.013	0.015
Medical	0.014	0	1	0.009	0.012	0.014	0.015
Music and audio	0.036	0	1	0.042	0.017	0.037	0.037
News and magazines	0.035	0	1	0.039	0.064	0.035	0.029
Personalization	0.050	0	1	0.029	0.017	0.019	0.069
Photography	0.014	0	1	0.014	0.015	0.011	0.015
Productivity	0.032	0	1	0.022	0.028	0.030	0.036
Shopping	0.015	0	1	0.011	0.003	0.022	0.017
Social	0.020	0	1	0.018	0.016	0.028	0.019
Sports	0.024	0	1	0.027	0.020	0.029	0.023
Tools	0.081	0	1	0.069	0.043	0.053	0.095
Transportation	0.014	0	1	0.012	0.009	0.018	0.015
Travel and local	0.042	0	1	0.035	0.038	0.059	0.043
Weather	0.004	0	1	0.005	0.005	0.005	0.004
Observations	475,867			153,988	41,792	84,035	25,3672

Notes: This table presents the descriptive statistics for the overall sample. Column (1) shows the statistics of the whole sample. Column (2) presents descriptive statistics for *Advertising*. Column (3) depicts descriptive statistics for *In-app purchases*. Column (4) shows the descriptive statistics for *Personal data*. Column (5) presents statistics for developers without monetization strategy.

2.2 Third parties libraries

An important strength of our data set is that we are able to identify third parties associated to each application. These data are collected by Privacy Grade.[6] Third party libraries are developed by companies and professional developers to offer different functionalities to app developers. While third parties are essential to enable certain app functionalities, little is known about the structure of this market, or the actors involved. In particular, they can enable the inclusion of advertising in an app or offer tools to help developers create apps. They are also able to gather personal data on app users in order to improve app. We use this information to construct the dependent variable *Advertising* described in Section 2.3. functioning.[7] Privacy Grade classifies these third parties' libraries into six groups - *advertising, payment, social networking, utility, development aid, and mobile analytics*, presented in Table 2. Developers can use several libraries at the same time. Our sample includes more than 182 different third parties, and 50.4% of apps with at least one third party.

Table 2 displays the number of different third parties indicating the percentage of apps that use each group of third parties. The advertising third parties include 79 different entities that enable apps to deliver advertising; they transfer a percentage of the revenues generated to developers. These third parties are used by 32.4 % of apps.

The social network third party libraries link the app's functioning to the services offered by the social network companies. This is used to build up the dummy variable *Social Networking*. This group of third parties is used by 13.7% of applications, with Facebook and Twitter examples of these libraries (Table 2). The mobile analytics group includes 12 different libraries that offer an analysis of applications usage (e.g.

[6]They use the content contained in the APK files of each app.

[7]https://www.theguardian.com/technology/2017/nov/28/android-apps-third-party-tracker-google-privacy-security-yale-university and https://privacylab.yale.edu/press/android-trackers Last retrieved 6 Mars 2018

Table 2: Breakdown statistics of the third parties presented in our sample

Category of thirds parties	Mean	Min	Max	Number of different third parties
Advertising third parties	0.324	0	1	79
Payment third parties	0.036	0	1	8
Social networking third parties	0.137	0	1	10
Utility third parties	0.187	0	1	71
Mobile analytics third parties	0.078	0	1	12
Development aid third parties	0.039	0	1	9
Observations	475,867			

Notes: This table depicts the summary statistics of different categories of thirds parties classified by Privacy Grade. The last column 'Number of different thirds parties' indicates the number of different libraries in each category. Standard deviations are in parentheses.

bug) service. This group of third parties is exploited by 7.8% of the applications and used to build up the binary variable *Mobile Analytics*.

The utility third parties help the developer to add functions or a framework to their code. For example, they can be used to manage images on the apps. These third parties are used to construct the variable *Utility* and are employed by 18.7% of the applications in our sample. The utilities third parties include 71 heterogeneous companies.[8] *Development aid third parties* are used by 3.9% of apps, and this group includes 12 different third parties.[9]

Table 3 shows the top 15 third parties related to each monetization strategy. Admob, which belongs to the Advertising third parties, is used by 86.52% of apps using advertising as a business strategy, and is used by only 31.08% of apps using in app purchase.[10]

[8]E.g., Nostra 13 helps developers with images, while Jsoup helps with HTLM languages. Nostra 13 is the most widely used utility third party and consists of an open source program available on Github.

[9]For the purposes of our analysis, we do not include this measure in our estimations since only non-professional developers employ this type of tool, thus use of development aid third parties is likely to be negatively correlated to the professional developers measures.

[10]Admob is the Google's advertising service. The company was created in 2006 and was bought by Google in 2009 for 750$ millions. More than 1 million applications use Admob, resulting in payments of US 1 billion to developers since 2012.

Table 3: Top 15 third parties by strategy of monetization

Advertising (1)		In-app purchases (2)		Personal Data (3)	
Thirds	Percentage	Thirds	Percentage	Thirds	Percentage
Admob	86.52%	Admob	31.08%	Facebook	33.36%
Facebook	20.88%	Facebook	25.81%	Admob	31.55%
Flurry	10.57%	Flurry	18.19%	Twitter4j	18.29%
Twitter4j	8.57%	Chartboost	9.50%	Flurry	16.86%
Millennial	7.77%	Twitter4j	6.19%	Paypal	10.41%
Inmobi	7.11%	Tapjoy	5.86%	Biznessapps	8.94%
Chartboost	5.85%	Inmobi	4.91%	Nostra13	7.86%
Unity3d	5.29%	Millennial	4.58%	Oauth	7.20%
Revmob	4.51%	Nostra13	4.41%	Millennial	6.21%
Paypal	4.46%	Adobe	4.08%	Inmobi	5.80%
Jsoup	4.34%	Amazon	3.89%	Acra	5.66%
Nostra13	3.81%	Nostra13	5.18%	Jsoup	4.76%
Biznessapps	3.75%	Mopub	3.81%	Revmob	4.19%

Notes: This table depicts the summary statistics of the 15 biggest thirds parties by variables of interests, *Advertising, In-app purchases and Personal data*. Column (1) shows the distribution of third parties used by apps doing *Advertising*. Column (2) shows the distribution of third parties used by apps doing *In-app purchases*. Column (3) reports the distribution of third parties used by apps collecting *Personal data*.

2.3 The dependent variables: *Advertising, In-app purchases, Personal data*

As we aim to model the monetization strategies of developers, we estimate three variables of interest *Advertising*, *In-app purchases* and *Personal data*. These three monetization strategies are not mutually exclusive; developers can combine more than one strategy and we do not have any assumptions on the order of the choices. Our empirical strategy permits to use *Personal data* as regressors. Table 4 presents the statistics for different strategy combinations.

First, *Advertising* is a dummy variable measuring whether developers provide ad to the apps through the third parties that act as ad networks, 22.5% of apps use advertising only. At the time of our data collection, we measured only advertising provided via third parties as developers are required to declare ad status in Google Play Store from January 2016.[11]

Second, the dummy variable *In-app purchases* measures whether the apps allow integrated purchases which enable the purchase of services and digital goods within the applications such as boosts, life in game, upgrade, and bonus. There are 4.1% of apps that use only in-app purchases. In this case, the platform remunerates the developers directly and takes 15% of the amount spent.[12]

Third, in order to measure whether the app collects personal data we use two sources of data: the Google Play permissions system and the data provided by Privacy Grade. The Android permissions system allows developers to interact with the functionalities of the smartphone and potentially to collect data. Therefore, before downloading an app, users are informed about the permissions attached to its use. Permissions allow developers to gather different sets of information related to the functioning of the smartphone and users' behaviors. While the Android system includes 138 permissions only 56 are defined as dangerous[13], thus, for the purposes of our study we consider this subsample (detailed in appendix Table 10). Examples of the personal data collected are users' geo-location, contacts and access to text messages.

PrivacyGrade measures the discrepancy between permissions and app's technical features. Privacy Grade data grades apps from D to A+ to rank their privacy intrusiveness by measuring the disparity between the functionality of the apps and the types of permissions required, *with A+ referring to apps that collect personal data needed for the functioning of the app.* The evaluation of permissions by Privacy Grade is ongoing.

[11]The Contain Ad includes ads delivered through third party ad networks, display ads, native ads and banner ads http://support.visiolink.com/hc/en-us/articles/206050941-Action-required-declare-ad-status-for-your-Google-Play-apps Last retrieved 6 March 2018

[12]https://www.theverge.com/2017/10/19/16502152/google-play-store-android-apple-app-store-subscription-revenue-cut Last retrieved 6 March 2018

[13]Define by Google as:"Permissions are considered as intrusive if they can affect the functioning of the device." These permissions request users to provide an explicit agreement.

If an app receives a ranking between B and D this means that at least one permission installed requests more information than required for app functionality. **Indeed, it is costly to collect data so we assume that apps that collected information for an other purpose than functionality have a goal**. We create a dummy variable *Badgrade* which takes the value 1 if the app is graded between B and D.

To measure the collection of personal data, we create the variable *Personal Data* which takes the value 1 if the app is ranked by Privacy Grade between B and D, and/or if the app has more than six dangerous permissions and 0 otherwise. On average, apps have 3.2 permissions with a standard deviation of 2.9. Note that only 10% of the samples use more than six dangerous permissions. Since not all permissions have been evaluated by Privacy Grade, to alleviate measurement problems we consider two measures to build the dependent variable *Personal data*. The statistics show that 9.2% of apps use personal data as a monetization strategy, 6.2% of apps combine personal data with advertising, and 1.2% combine all three monetization strategies (see Table 4).

In our sample, 53.3% of apps have no monetization strategy. Based on the literature, we identify several reasons for this. First, some apps are produced by non-profit organizations such as Wikipedia and Mozilla. Second, developers can use their apps as 'visiting cards' to demonstrate their competencies. For example, Xu *et al.* (2014) show that developers use the forum platform to improve their job opportunities. Third, some apps are produced by corporate groups, like banking and TV channel apps. Fourth, apps can be created based on brands in order to advertise. Gupta (2013) explains that brands are aimed more at increasing interest in the product, e.g. Red Bull offers games associated to the brand.

Table 4: **Summary statistics: Combination of monetization strategies**

Exclusive Monetization strategies	Mean
No monetization strategy	0.533
Only Personal data	0.092
Only In-app purchases	0.041
Only Advertising	0.225
In-app purchases & Personal data	0.010
Advertising & Personal data	0.062
Advertising & In-app purchases	0.024
Advertising & Personal data & In-app purchases	0.012

Notes: This table indicates all combination of monetization strategies.

2.4 Apps characteristics and developers

To measure the popularity of apps, we use the download category provided by Google Play which includes 19 discrete distinctions. The statistics for number of downloads are presented in Table 5, and range from less than five downloads, to over a thousand million downloads. To measure whether the number of downloads affects the probability of choosing a particular business model, we include the vector of the variables measuring download intensity.

The quality of the application and the user's satisfaction are measured using the variable *Playstore Rating*, app grading is given by users and goes from 0 to 5. In order to measure whether the developer has professional experience, we include in the regression three sets of dummy variables. First, *App by developer* indicates the number of apps produced by each developer in all categories in our sample. Second, the binary variable *Developer website* indicates whether the developer has declared a website. Third, *Privacy policy* indicates if the developer has a privacy policy and 0 otherwise.

We first provide a graphical depiction of the distribution of monetization strategies per number of downloads, and second we describe their distribution by Google categories. Figure 1 depicts the monetization distribution by category of installations.

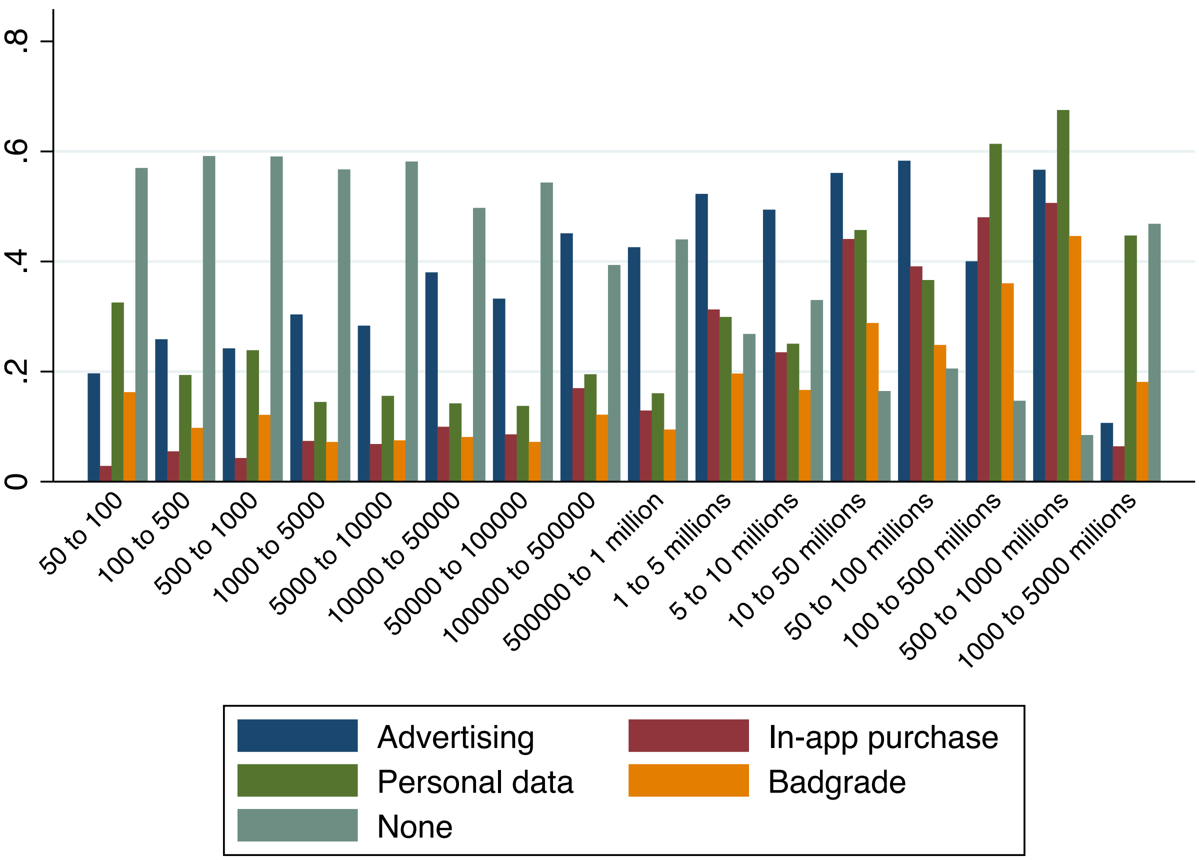
For each download category (vertical axis), we show the percentage of apps for each monetization strategy. While advertising is mostly used by apps with less than 100 million downloads, the percentage of apps using personal data increases for the top downloaded apps. These raw data patterns are consistent with the intuition that the commercialization of personal data is valuable for huge amount of data are collected (OECD, 2013; Lambrecht and Tucker, 2017). Figure 2 shows the percentage of apps for each monetization strategy grouped by app Google Play category. While Game category is more likely to use advertising (see also Table 1 for descriptive statistics), both Communication and Business categories are more likely to use personal data as a business model, and Medical and Health & Fitness categories also tend to collect personal data. The data collected in this categories are particular valuable as they have information on users' health and financial condition.

Table 5: Summary statistics: Number of downloads by monetization strategies

Install categories	Advertising (1)	In-app (2)	Pers data (3)	None (4)
Number install 1-5	0.002	0.001	0.007	0.004
Number install 5-10	0.000	0.000	0.000	0.000
Number install 10-50	0.038	0.022	0.085	0.059
Number install 50-100	0.003	0.001	0.009	0.005
Number install 100-500	0.147	0.115	0.202	0.204
Number install 500-1000	0.037	0.024	0.067	0.055
Number install 1000-5000	0.197	0.176	0.172	0.223
Number install 5000-10000	0.082	0.072	0.083	0.102
Number install 10000-50000	0.193	0.186	0.132	0.153
Number install 50000-100000	0.086	0.082	0.065	0.085
Number install 100000-500000	0.092	0.128	0.073	0.049
Number install 500000-1 million	0.066	0.073	0.045	0.041
Number install 1-5-million	0.026	0.057	0.027	0.008
Number install 5-10 million	0.021	0.038	0.020	0.009
Number install 10-50-million	0.004	0.011	0.006	0.001
Number install 50-100 million	0.005	0.012	0.006	0.001
Number install 100-500-million	0.000	0.001	0.001	0.000
Number install 500-1000 million	0.000	0.001	0.001	0.000
Number install 1000-5000 million	0.000	0.000	0.000	0.000
Observations	153,978	41,786	84,001	253,634

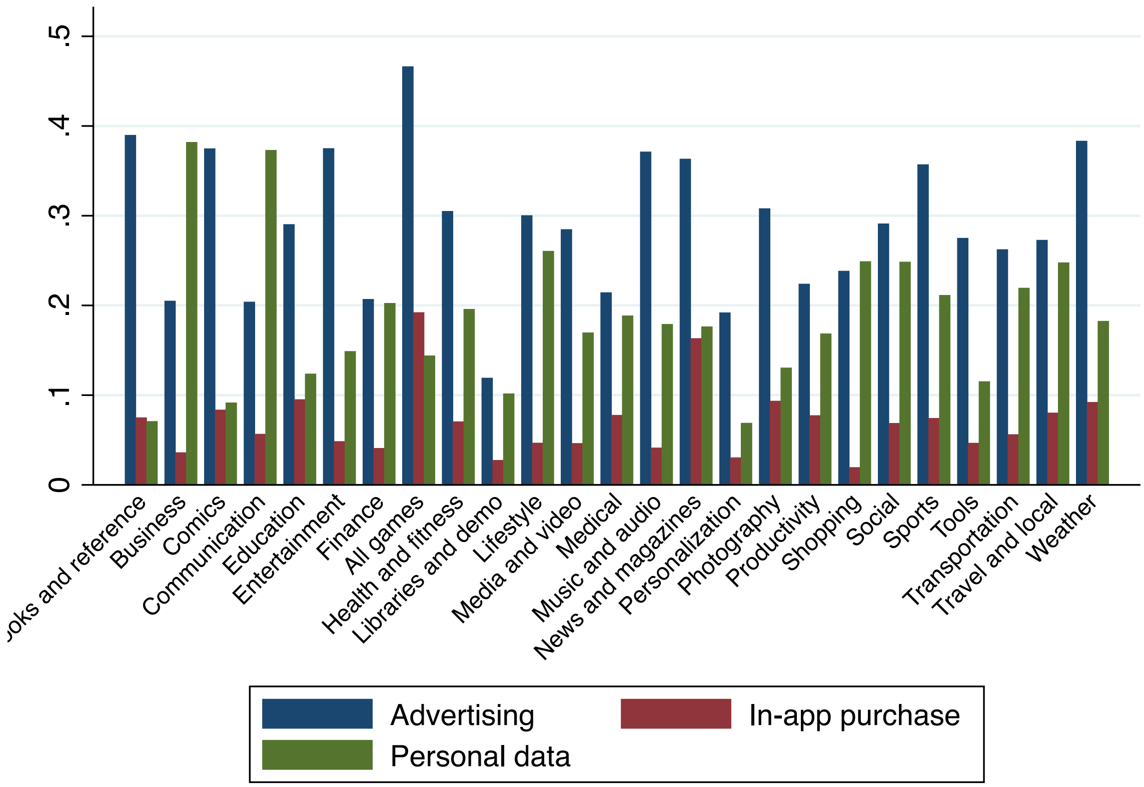
Notes: The downloads on the Google categories are divided in 19 category. Column (1) represents the percentage of apps that use Advertising. Column (2) represents the percentage of apps that use In-app purchases. Column (3) represents the percentage of apps that use Personal data. Column (4) indicates the percentage of apps that do not use monetization strategies.

Figure 1: Strategies of monetization by volume of downloads



Notes: The vertical axis is the percentage of apps using a monetization strategy. The horizontal axis is the volume of downloads.

Figure 2: Strategies of monetization by category of applications.



Notes: The vertical axis is the percentage of apps using a monetization strategy. The horizontal axis is the apps category.

2.5 Exclusion restriction: Everyone

The instrumental variable is a dummy variable *Everyone* which measures the Age restriction on the app defined by its developers. It follows Google's parental control system which provides guidelines about age based on app content. Google Play uses four levels of maturity: "Everyone", "Low Maturity", "Medium Maturity", and "High Maturity". Apps that contain suggestive or sexual references are defined as "Medium maturity" or "High maturity". Apps with content suitable for all individuals are categorized "Everyone". We create the variable *Everyone* which takes the value 1 if the app is designed for all the user. This variable allows us to identify the content proposed by the apps and the population targeted. It is used to instrument the choice of monetization *Personal data* since 'Everyone' apps are aimed at children, teens, and adults. According to Google Play guidelines and COPPA legislation to protect children and teens, these apps are supposed to collect relatively less personal data because they are likely to be downloaded by children and teens. In contrast, the use of advertising or more traditional monetary transactions is less likely to be correlated to the (targeted) age group.

3 Modeling the monetization choice

The literature on the economics of privacy shows that app developers can choose among three main monetization strategies: advertising, in-app purchase and personal data. To our knowledge, this study provides the first empirical evidence of app characteristics related to the use of personal data as a monetization strategy for free goods. Advertising and in-app purchase are traditional business strategies in the Internet economy, and personal data may complement or substitute for these business models. For example, personal data can be used to display targeted ads (i.e. to complement advertising), or it can be sold to data brokers. To model the developer's choice, we estimate a recursive trivariate probit that accounts for the endogeneity of personal data.

3.1 Recursive multivariate probit model

To obtain consistent estimates of the advertising, In-app purchase and the personal data equations, the explanatory variables should be exogenous (Maddala, 1986). However, the literature suggests that there is a potential association between the traditional business models and personal data collection. In particular, personal data are likely to be used to run personalized ads. To address the potential endogeneity of the variables of *Personal data*, we rely on the methodology in Goy and Wang (2015) which uses a recursive multivariate probit to estimate the probability of not mutually exclusive choices. The probability of doing *Advertising, In-app purchase*, and *Personal data* may not be independent and our empirical strategy allows us to measure the relation among common unobservables that explain three choices.

Building on our conceptual framework, we estimate the joint probability to implement one of the three monetization strategies. Therefore, the latent probabilities to use *Advertising, In-app purchase* and *Personal data* of app i are estimated with recursive trivariate probit as follow:

$$y^*_{ji} = X'_{ji}\beta_j + \gamma_j Personal\ Data_i + \epsilon_{ji}, \quad j = A, I \qquad (1)$$

$$Personal\ Data^*_i = X'_i\alpha + Z_i\phi + u_i \qquad (2)$$

Where *Advertising* is denoted A, and *In-app purchase* is denoted *I*. $Personal\ Data^*_i$ is a latent variable and $Personal\ Data_i = 1$ if $Personal\ Data^*_i > 0$, X_i is a vector of the regressors affecting the choice to monetize an app. u_i is the error term which is trivariate normal with $\epsilon_{ji}, (j = A, I)$ such that $var(\epsilon_{Ai} = 1)$, $var(\epsilon_{Ii} = 1)$, $var(u_i = 1)$, $cov(\epsilon_{Ai}, \epsilon_{Ii}) = \rho_{AI}$, $cov(\epsilon_{Ai}, u_i) = \rho_{PA}$ and $cov(\epsilon_{Ii}, u_i) = \rho_{PI}$ (where P is the $PersonalData$). Equation (1) represents the choice of *Advertising* and *In-app purchase* according to the vector X_{ji} of the exogenous variables, and whether or not the developer decides to use $Personal\ Data_i$.

Equation (2) represents the choice to use the *Personal data* strategy where the vector X_i of exogenous variables. To identify the model, we use the instrumental variable Z_i as the exogenous variable to include in the equation $Personal\ Data$ but not in the y^*_{ji} (Wilde, 2000). The exclusion restriction is the variable *Everyone* as discussed in section 2.5.

We normalized the residual and we use maximum likelihood estimator. We employ a GHK (Geweke-Hajivassiliou-Keane) algorithm and set the square root of the number of observations as the number of draws (Hajivassiliou and Ruud, 1994). Then, we use the maximum solution of the log-likelihood on our simulate probabilities to obtain the estimators (Greene, 2003; Train, 2009). This model allows to take into consideration the combination of the different choices with a coefficient of correlation (ρ_{AI}, ρ_{AP}, ρ_{PI}). The rho reflects the correlations between the errors (ϵ_{ji}) of the two equations. If the decisions of monetization strategies are dependent the ρ are significantly different from zero. We manually compute the average partial effect using the method proposed by Cameron and Trivedi (2010) and Jones *et al.* (2013). This method allows us to vary with the scaling of each covariates, for the continuous variable we add a change equal to 1 and to compute the average partial effect of a dummy variable use a change from 0 to 1. This method computes only the standard deviation.

4 Estimation of the monetization strategies

The results of the trivariate probit estimations are reported in Table 6. The Rho values suggest strong unobserved correlations among the *Advertising*, *In-app purchases* and *Personal data* variable error terms, supporting use of a trivariate model as the appropriate estimation model. Also, the LR test is statistically significant which rejects the null hypothesis that the three equations should be estimated separately. In other words, the probit model without correction for endogeneity, estimates biased coefficients, and justifies the choice of a trivariate probit model. This implies that *Personal data* can be a monetization strategy in its own right, and should be systematically taken into account when studying monetization. All of this evidence suggests that unobservable factors that influence the probability of *Advertising* and *In-app purchases* are also likely to affect the probability of choosing *Personal Data* strategy.

Table 6 presents the results of the main estimations. Column (1) shows the estimation results for *Advertising*; Column (2) presents the results of the equation for *In-app purchases*; Column (3) presents the results for *Personal data* which includes the exclusion restriction *Everyone*. Columns (4), (5) and (6) report the average partial effects computed at the mean for each equation. All regressions include Google Category fixed effects. To interpret the coefficients, we refer to the average partial effect. The results indicate that apps collecting *Personal data* have a 15.8% probability of using an *Advertising* strategy. This is in line with the traditional economics of privacy approach which models personal data as enabling personalized advertising. The use of personal data is likely to reduce the probability of *In-app purchases* by 3% suggesting a substitution effect between *Personal data* and the more traditional freemium business model.

Monetization strategies are likely to be linked to the download intensity. Average partial effects tend to be strong depending on the number of installations. The influence of the category of 1,000-5,000 million downloads (the top one) is especially interesting as it is likely to increase the likelihood of collecting personal data by 24.1%, and to decrease the probability of advertising by 32.5% and of in-app purchase by 9% (Table 6). The signs of the coefficients confirm the intuition based on the graphical evidences that apps enjoying more than 1 million downloads, are likely to collect personal data while those experiencing less than 1 million downloads are likely to use an advertising strategy. Previous literature has already highlighted that personal data value is associated to the collection of huge amount of data (OECD, 2013). The probability of *In-app purchase* increases also with the number of downloads.[14]

[14] Apps with more than 1 thousand million of download include Facebook, Snapchat and Whatsapp. These top apps companies use internal Ad platform thus they are less likely to use advertising third parties

In the case of third parties, developers that allow the presence of *social networking* third parties are 17.7% more likely to use *Advertising*, 7.1% more likely to adopt an *In-app purchase* strategy, and 20.9% more likely to collect *Personal data*. The use of *utility* third parties increases the probability of advertising by about 7%, the probability of an *In-app purchases* strategy by around 2.2% and the probability of collecting personal data by about 5%. It is interesting to see that apps that adopt a personal data strategy and also advertising are likely to receive negative user feedback (variable *Playstore rating*).

The indicator for a privacy policy decreases the probability of doing advertising by 6.8% but increases the probability of an in-app purchase strategy by 6.6%. It has only a small positive effect on the probability of collecting personal data. We also examined developer characteristics such as *Developer website* and the number of developer application (*Apps by dvp*). We found that those variables can be considered a measure of the developer's specialization. The results indicate that less professional developers use the monetization strategy of advertising, whereas professional developers are more likely to choose personal data and in-app purchase.

Table 11 in Appendix, shows that Google Category affect also the choice of monetization strategies. In particular, Business, Lifestyle and Productivity Categories are more likely to rely on Personal data strategy compared to apps in the Game Category.

Table 6: Trivariate estimations and Average partial effect with Advertising, Integrated Puchase and Personal data

	Estimations			APE		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Advertising	I-P	Pers data	Advertising	I-P	Pers data
Personal data	0.452***	-0.239***		0.157	-0.031	
	(0.053)	(0.046)				
Playstore rating	-0.005*	0.017***	-0.023***	-0.002	0.002	-0.005
	(0.003)	(0.005)	(0.003)			
Social networking	0.500***	0.425***	0.838***	0.177	0.071	0.209
	(0.024)	(0.027)	(0.018)			
Utility	0.208***	0.149***	0.243***	0.07	0.022	0.05
	(0.015)	(0.021)	(0.016)			
Developer website	-0.103***	0.351***	0.195***	-0.034	0.044	0.037
	(0.015)	(0.022)	(0.020)			
Privacy policy	-0.215***	0.401***	0.044**	-0.068	0.066	0.009
	(0.019)	(0.024)	(0.020)			
Apps by dvp	0.001*	0.001*	0.000	0	0	0
	(0.001)	(0.001)	(0.001)			
Number install 1000-5000	0.209***	0.213***	-0.128***	0.069	0.032	-0.024
	(0.009)	(0.017)	(0.012)			
Number install 5000-10000	0.144***	0.153***	-0.110***	0.048	0.023	-0.021
	(0.010)	(0.017)	(0.012)			
Number install 10000-50000	0.414***	0.368***	-0.102***	0.142	0.059	-0.019
	(0.011)	(0.018)	(0.016)			
Number install 50000-100000	0.293***	0.299***	-0.133***	0.1	0.048	-0.025
	(0.012)	(0.020)	(0.016)			
Number install 100000-500000	0.544***	0.627***	0.038**	0.191	0.119	0.007
	(0.015)	(0.021)	(0.019)			
Number install 500000-1 million	0.515***	0.496***	-0.047**	0.18	0.089	-0.009
	(0.015)	(0.021)	(0.019)			
Number install 1-5-million	0.630***	0.956***	0.224***	0.223	0.215	0.047
	(0.025)	(0.030)	(0.028)			
Number install 5-10 million	0.596***	0.782***	0.154***	0.21	0.164	0.032
	(0.023)	(0.028)	(0.028)			
Number install 10-50-million	0.629***	1.174***	0.484***	0.222	0.287	0.111
	(0.056)	(0.059)	(0.059)			
Number install 50-100 million	0.740***	1.086***	0.334***	0.262	0.258	0.073
	(0.046)	(0.053)	(0.051)			
Number install 100-500-million	0.253	1.298***	0.882***	0.086	0.33	0.222
	(0.169)	(0.169)	(0.208)			
Number install 500-1000 million	0.569***	1.295***	1.015***	0.201	0.329	0.263
	(0.153)	(0.150)	(0.196)			
Number install 1000-5000 million	-4.573***	-3.936***	0.947***	-0.325	-0.09	0.242
	(0.139)	(0.148)	(0.166)			
Everyone			-1.074***	-0.233		
			(0.014)			
Constant	-0.423***	-1.683***	-0.747***			
	(0.020)	(0.031)	(0.028)			
Log pseudolikelihood	-5.65e+05					
LR test chi2(3)	781.465					
rho21	-0.045***	0.011				
rho31	-0.144***	0.031				
rho32	0.254***	0.026				
Number of draw	683					
Observations	475 867					

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Advertising*, *In-app* and *Personal Data* . Column (4) to column (6) estimate the average partial effects respectively *Advertising*, *In-app* and *Personal Data*. *Everyone* is the exclusion restriction variable. The coefficient of the app category variables are displayed in annexe, table 11. The omitted category is *Game all.* Robust standard errors at category levers are in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

4.1 Robustness check

We estimate two robustness checks. We show the robustness of our estimations to alternative dependent variables. First, we estimate the trivariate probit that uses the dependent variables: *Admob, In-app purchase and Personal data*, and estimate the dummy variable *Admob* instead of *Advertising*. Second, we estimate the trivariate probit using the dependent variables *Admob, In-app purchase, Badgrade*, and replace *Personal data* by *Badgrade*.

4.1.1 Estimations with advertising third parties: Admob

Table 7 reports the model that includes *Admob* as a monetization strategy instead of *Advertising*. This addresses empirically concerns as the estimations are driven by Admob which is the largest ad company in the group of ad third parties. Thus, we measure whether the magnitude of the average partial effects of the variable personal data changes. In fact, apps that collect personal data are likely to show a 7.4% higher probability of using Admob which results in a smaller coefficient compared to the main regression (15.8%). The results of the estimations are consistent with the main estimations.

Table 7: Trivariate estimations with Admob, Integrated Puchase and Personal data

	Estimations			**Average partial effect**		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Admob	I-P	Pers data	Admob	I-P	Pers data
Personal data	0.223***	-0.239***		.0740034	-.0305478	
	(0.051)	(0.046)				
Playstore rating	-0.003	0.017***	-0.023***	-.0010256	.002452	-.2329865
	(0.003)	(0.005)	(0.003)			
Social networking	0.410***	0.423***	0.837***	.1411472	.0709955	.2090116
	(0.023)	(0.027)	(0.018)			
Utility	0.137***	0.148***	0.241***	.0446349	.0219767	.049998
	(0.015)	(0.021)	(0.016)			
Developer website	-0.120***	0.351***	0.195***	-.0387156	.0436082	.0367252
	(0.015)	(0.022)	(0.020)			
Privacy policy	-0.220***	0.401***	0.044**	-.0668485	.0662746	.0087256
	(0.020)	(0.024)	(0.020)			
Apps by dvp	0.001**	0.001*	0.000	.0004357	.0001913	-.0045144
	(0.001)	(0.001)	(0.001)			
Number install 1000-5000	0.174***	0.212***	-0.128***	.0566341	.0321011	-.0243349
	(0.010)	(0.017)	(0.012)			
Number install 5000-10000	0.118***	0.152***	-0.110***	.0383899	.0229798	-.0207428
	(0.010)	(0.017)	(0.012)			
Number install 10000-50000	0.361***	0.367***	-0.102***	.1215729	.0593296	-.019454
	(0.012)	(0.018)	(0.016)			
Number install 50000-100000	0.250***	0.299***	-0.134***	.0834578	.0483675	-.025033
	(0.012)	(0.020)	(0.016)			
Number install 100000-500000	0.434***	0.626***	0.037*	.1503972	.1183064	.0073287
	(0.015)	(0.021)	(0.019)			
Number install 500000-1 million	0.437***	0.495***	-0.047**	.1515659	.0890124	-.0090707
	(0.015)	(0.021)	(0.020)			
Number install 1-5-million	0.422***	0.955***	0.223***	.1469994	.2148683	.0468392
	(0.025)	(0.030)	(0.029)			
Number install 5-10 million	0.437***	0.780***	0.153***	.1523574	.1631962	.0314241
	(0.024)	(0.028)	(0.028)			
Number install 10-50-million	0.300***	1.174***	0.484***	.1023814	.287151	.1105882
	(0.056)	(0.059)	(0.059)			
Number install 50-100 million	0.427***	1.086***	0.334***	.148886	.2577796	.0729526
	(0.046)	(0.053)	(0.051)			
Number install 100-500-million	-0.110	1.296***	0.883***	-.0339706	.3291236	.2226507
	(0.203)	(0.168)	(0.208)			
Number install 500-1000 million	0.048	1.295***	1.013***	.0155338	.3286448	.2623054
	(0.164)	(0.150)	(0.197)			
Number install 1000-5000 million	-4.308***	-3.827***	0.945***	-.2806006	-.0900433	.2413867
	(0.141)	(0.149)	(0.166)			
Everyone			-1.075***			.0000332
			(0.014)			
Constant	-0.535***	-1.682***	-0.751***			
	(0.020)	(0.031)	(0.028)			
Log pseudolikelihood	-5.57e+05					
LR test chi2(3)	880.895					
rho21	-0.069***	0.011				
rho31	-0.114***	0.029				
rho32	0.253***	0.026				
Number of draw	683.000					
Observations	475 867					

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Admob*, *In-app* and *Personal Data* . Column (4) to column (6) estimate the average partial effects respectively *Admob*, *In-app* and *Personal Data*. *Everyone* is the exclusion restriction variable. The coefficients of the app category variables are provided in annex table 12. The omitted category is *Game all.* Robust standard errors are in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

4.1.2 Estimations with Badgrade

Table 8 reports the model with *Badgrade* instead of *Personal data* as a monetization strategy. The binary variable *Badgrade* indicates if the app received a grade between B and D. This empirical strategy allows us to estimate whether the use of a conservative definition of personal data might affect our results. If unobserved heterogeneity associated to the choice of permissions is affecting our results, we can measure any changes in our estimations.

Badgrade is likely to be correlated with the probability to do advertising and to be complement of freemium strategy. The effects of *Badgrade* on the probability of Advertising and In-app purchase are consistent with the previous estimations using the variable *Personal data*. As in the main regression, apps with small numbers of installations (under 500,000) are less likely to use personal data and more likely to use advertising as their monetization strategy.

Table 8: Trivariate estimations and APE with Advertising, Integrated Puchase and Badgrade

	Estimations			APE		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Advertising	I-P	Badgrade	Advertising	I-P	Badgrade
Badgrade	1.346***	-0.297***		.4744184	-.0361762	
	(0.077)	(0.089)				
Playstore rating	0.001	0.016***	-0.033***	.0002274	.0023186	-.0042778
	(0.003)	(0.005)	(0.003)			
Social networking	0.302***	0.441***	0.936***	.0995689	.0746515	.1786521
	(0.026)	(0.034)	(0.020)			
Utility	0.197***	0.141***	0.193***	.0632001	.0208286	.0272703
	(0.014)	(0.021)	(0.020)			
Developer website	-0.099***	0.343***	0.084***	-.0310292	.042868	.0108097
	(0.015)	(0.021)	(0.024)			
Privacy policy	-0.198***	0.389***	-0.057**	-.0595072	.064026	-.0073318
	(0.019)	(0.024)	(0.025)			
Apps by dvp	0.001	0.001*	0.001*	.0003305	.0001982	.0001535
	(0.001)	(0.001)	(0.001)			
Number install 1000-5000	0.218***	0.214***	-0.108***	.0693422	.032362	-.0137948
	(0.009)	(0.017)	(0.015)			
Number install 5000-10000	0.156***	0.152***	-0.118***	.0498898	.0230393	-.0148539
	(0.009)	(0.017)	(0.014)			
Number install 10000-50000	0.412***	0.371***	-0.033*	.1348444	.0602218	-.0043503
	(0.011)	(0.019)	(0.019)			
Number install 50000-100000	0.299***	0.301***	-0.088***	.0974234	.0489114	-.0111803
	(0.011)	(0.020)	(0.019)			
Number install 100000-500000	0.523***	0.629***	0.100***	.1758769	.1191645	.0139114
	(0.016)	(0.021)	(0.023)			
Number install 500000-1 million	0.506***	0.499***	0.015	.1702537	.0899516	.0020473
	(0.015)	(0.021)	(0.023)			
Number install 1-5-million	0.580***	0.954***	0.264***	.1972233	.2146634	.0398837
	(0.026)	(0.030)	(0.032)			
Number install 5-10 million	0.552***	0.783***	0.219***	.1873387	.164145	.0324559
	(0.024)	(0.028)	(0.033)			
Number install 10-50-million	0.550***	1.163***	0.417***	.1865742	.2834873	.0682868
	(0.057)	(0.060)	(0.061)			
Number install 50-100 million	0.667***	1.084***	0.363***	.2282217	.2572187	.057893
	(0.048)	(0.053)	(0.056)			
Number install 100-500-million	0.101	1.284***	0.737***	.0320002	.3252065	.1392825
	(0.169)	(0.174)	(0.189)			
Number install 500-1000 million	0.396**	1.285***	0.869***	.1321132	.325558	.1733927
	(0.155)	(0.152)	(0.202)			
Number install 1000-5000 million	-4.271***	-4.095***	-4.229***	-.3242858	-.0903304	-.093314
	(0.150)	(0.165)	(0.169)			
Everyone			-0.814***			-.1114871
			(0.019)			
Constant	-0.520***	-1.666***	-0.867***			
	(0.021)	(0.035)	(0.031)			
Log pseudolikelihood	-5.10e+05					
LR test chi2(3)	1233.465					
rho21	-0.081***	0.013				
rho31	-0.393***	0.042				
rho32	0.319***	0.048				
Number of draw	683					
Observations	475,787			475,787		

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Advertising*, *In-app* and *Badgrade* . Column (4) to column (6) estimate the average partial effects respectively *Advertising*, *In-app* and *Badgrade*. *Everyone* is the exclusion restriction variable. The coefficient of the app category variables are displayed in annexe, table 11. The omitted category is *Game all*. Robust standard errors at category levers are in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

5 Discussions

This paper has some economics and managerial implications for the regulator, developers, platforms, and it also investigates new research questions. In particular, developers need a better understanding of the market, and what is crucial for the development of a competitive app in a winner-takes-all market structure. Our results should help developers to identify the right monetization strategy, or to adapt existing ones. We show that some applications have a significant number of third parties which are able to access user data without the user's awareness. This raises questions about the access' management by third parties. We show that personal data can be requested for later use and not just to enable app functionality. Further more we highlight that data are collected not just to improve a service but the literature shows that requesting more personal data has a negative impact on the demand (Kummer and Schulte, 2016), **so some developers choose to support this cost for collecting personal data**.

How developers obtain revenue is a critical issue for platforms which need to encourage the entry of new innovative developers, and help them increase their visibility. In particular, business analytics should consider the role of third parties to identify effective strategies. While the presence of third parties is important for the provision of enhanced services and features, it also allows the collection of individuals' data for business analytics purposes, or detection of technical problems. Some applications have significant numbers of libraries which third parties can access to obtain user data without the user being aware. Platforms should design more transparent systems that allow users to be better informed about the presence of third parties, and at the same time allow developers to improve the technical and economic performance of their apps.

In a self-regulatory approach to personal data, we need to investigate further how developers and platforms might improve transparency through their permissions systems by encouraging developers to declare which third parties they use. It would be interesting to also investigate the market for third party libraries more thoroughly. Our paper provides a preliminary examination of the third party market related to apps. Whereas developers have the choice among multiple third parties, market shares will become very concentrated on a few third parties chosen by the majority of apps. In addition, if we consider that the third party market includes some dominant players, concerns related to competition policy are yet to be done. User data could be concentrated among a few actors at different levels, and at the applications level for killer apps in particular. This raises questions about the market power of killer apps and the competition's dynamics of platforms. However, this needs careful examination in this very competitive market where network effects are quite strong, which could counteract this increasing returns effect.

6 Conclusion

From an industrial organization perspective, the sector is characterized by the pace of innovation and reduced barriers to market entry by new developers (Davis *et al.*, 2016). The majority of apps are "sold" at zero price.

Our approach differs from previous empirical investigations by focusing on the monetization strategies of developers, and analyzing how personal data are combined with more traditional monetization strategies such as advertising and in-app purchase. First, our results suggest that overall, personal data are used to monetize applications. Second, we show that monetization strategies differ depending on the app category which has important managerial implications for patterns of innovation and development in this sector. Also, a personal data strategy seems to be associated to more specialized developers, and can be used once the apps achieve a certain level of market power.

Third, we find that monetization strategies depend non-linearly also on the download category. Killer apps collect personal data without using ads or in-app purchase strategies. It is interesting to highlight that apps that apply a personal data strategy seem to receive the worst feedback suggesting that the collection and use of personal data are perceived negatively by consumers, and only the big developers (apps) can harmlessly implement this strategy. We also observed that the use of social networking third parties increases the probability of using in-app purchase.

Our study has some limitations. First, our results should be interpreted with caution since we use only cross sectional data, and thus can only estimate correlations rather than causalities. Second, it seems that there are threshold effects related to the number of downloads and the choice of monetization strategy. It would be desirable to obtain precise numbers of downloads per app instead of a range of downloads to calculate the thresholds where strategies might change dramatically.

References

Acquisti, A., Taylor, C. and Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*. 54(2), 442–92.

Athey, S., Blei, D., Donnelly, R., Ruiz, F. and Schmidt, T. (2018). Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data. *AEA Papers and Proceedings*. 108, 64–67.

Blumenstock, J. E. (2018). Estimating Economic Characteristics with Phone Data. *AEA Papers and Proceedings*. 108, 72–76.

Bresnahan, T., Davis, J., Jaconette, T. and Yin, P.-L. (2015). Mobile Applications, the Economics of. In *The New Palgrave Dictionary of Economics*. (pp. 1–8). Springer.

Brynjolfsson, E. and Saunders, A. (2010). Wired for innovation. *How Information technology in reshaping the economy. Massachusetts Institute of Technology. USA.*

Cameron, A. C. and Trivedi, P. K. (2010). *Microeconometrics using stata*. vol. 2. Stata press College Station, TX.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015). Competing with privacy. *Management Science*. 61(1), 229–246.

Comino, S., Manenti, F. M. and Mariuzzo, F. (2016). *Updates management in mobile applications. iTunes vs Google Play*. Working paper.

Davis, J. P., Chhabra, Y. and Yin, P.-L. (2016). *Experimentation Strategies and Entrepreneurial Innovation: Parallel and Sequential Innovation in the iPhone App Ecosystem*. Working paper.

Evans, D. (2011). The Antitrust Economics of Free. *CPI Journal*. 7, 1.

Gabaix, X. (2016). Power laws in economics: An introduction. *The Journal of Economic Perspectives*. 30(1), 185–205.

Garg, R. and Telang, R. (2013). Inferring App Demand from Publicly Available Data. *Management Information Systems Quarterly*. 37(4), 1253–1264.

Ghose, A. and Han, S. P. (2014). Estimating demand for mobile applications in the new economy. *Management Science*. 60(6), 1470–1488.

Goy, F. and Wang, C. (2015). Does knowledge tradeability make secrecy more attractive than patents? An analysis of IPR strategies and licensing. *Oxford Economic Papers*. 68(1), 64–88.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Gupta, S. (2013). For Mobile Devices, Think Apps, Not Ads.(cover story). *Harvard Business Review*. 91(3), 70–75.

Hajivassiliou, V. A. and Ruud, P. A. (1994). Classical estimation methods for LDV models using simulation. *Handbook of econometrics*. 4, 2383–2441.

Jones, A. M., Rice, N., d'Uva, T. B. and Balia, S. (2013). *Applied health economics*. Routledge.

Kummer, M. and Schulte, P. (2016). *When private information settles the bill: Money and privacy in Google's market for smartphone applications*. Working Paper.

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N. and Yao, S. (2014). How do firms make money selling digital goods online? *Marketing Letters*. 25(3), 331–341.

Lambrecht, A. and Tucker, C. E. (2017). Can Big Data Protect a Firm from Competition? *Competition Policy International*.

Li, X., Bresnahan, T. and Yin, P.-L. (2016). *Paying incumbents and customers to enter an industry: Buying downloads*. Working paper.

Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J. and Zhang, J. (2012). Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy Through Crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 501–510.

Lin, J., Liu, B., Sadeh, N. and Hong, J. I. (2014). Modeling Users Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. USENIX Association, 199–212.

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. 3. Cambridge university press.

OECD (2013). *Exploring the Economics of Personal Data*. OECD Digital Economy Papers 220. OECD Publishing.

Parker, G. G. and Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management science*. 51(10), 1494–1504.

Smith, M. D. and Telang, R. (2009). Competing with free: the impact of movie broadcasts on DVD sales and internet piracy 1. *mis Quarterly*. 33(2), 321–338.

Spiegel, Y. (2013). Commercial software, adware, and consumer privacy. *International Journal of Industrial Organization*. 31(6), 702–713.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics letters*. 69(3), 309–312.

Xu, L., Nian, T. and Cabral, L. (2014). *What Makes Geeks Tick? A Study of Stack Overflow Careers.* Technical report. Working paper.

Yin, P.-L., Davis, J. P. and Muzyrya, Y. (2014). Entrepreneurial innovation: Killer apps in the iPhone ecosystem. *The American Economic Review.* 104(5), 255–259.

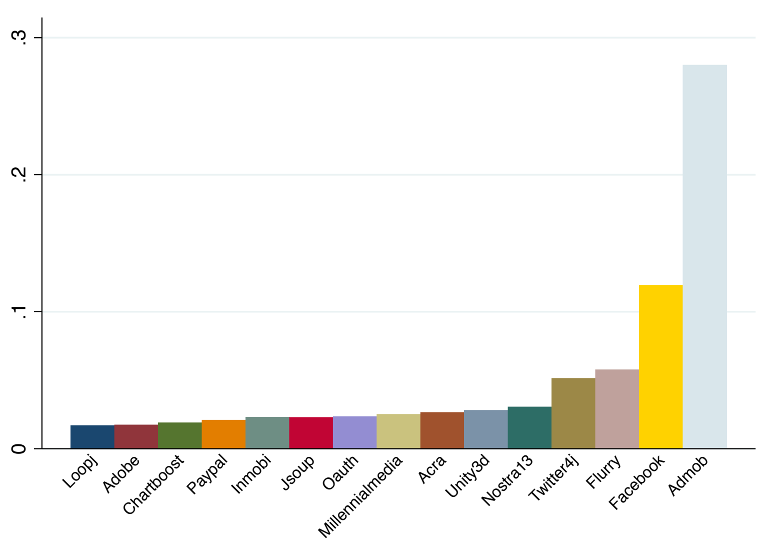
7 Appendix

Table 9: Use of permissions Google group by strategy of monetization

Advertising		In app Purchase		Personal Data		None	
Variable	%	Variable	%	Variable	%	Variable	%
Network	99.76%	Network	97.46%	Network	99.41%	Network	78.75%
Storage	55.42%	Storage	78.75%	Storage	90.23%	Storage	44.81%
Phone calls	40.60%	Phone calls	55.62%	Phone calls	88.66%	Signature	31.20%
Signature	28.79%	Signature	41.58%	Signature	73.44%	Phone calls	24.20%
Accounts	16.02%	Accounts	36.99%	Accounts	50.63%	Accounts	9.38%
Camera	12.69%	Camera	24.79%	Camera	43.79%	Camera	7.59%
Microphone	8.45%	Microphone	18.62%	Microphone	29.13%	App info	3.59%
Messages	6.41%	Messages	17.11%	Messages	26.45%	Microphone	2.65%
App info	5.73%	App info	10.67%	Social info	25.83%	Display	2.40%
Social info	4.12%	Social_info	7.10%	App info	20.56%	Social info	2.19%
Display	3.31%	Display	3.02%	Personal info	9.51%	Messages	1.56%
Personal info	1.83%	Personal info	2.92%	Bluetooth	8.04%	Bluetooth	1.51%
Bluetooth	1.64%	Screenlock	2.33%	Display	7.16%	Screenlock	1.12%
System tools	1.60%	Bluetooth	2.32%	System_tools	6.49%	Systemtools	1.05%
Screenlock	1.33%	System tools	1.74%	Screenlock	5.95%	Personal info	0.82%
Affects battery	0.20%	Affects battery	0.57%	Affects battery	1.16%	Affects battery	0.37%
User dictionary	0.05%	User dictionary	0.12%	Bookmarks	0.24%	User dictionary	0.17%
Bookmarks	0.05%	Bookmarks	0.10%	User dictionary	0.17%	Bookmarks	0.00%
Voicemail	0.00%	Voicemail	0.00%	Voicemail	0.00%	Voicemail	0.00%

Notes:

Figure 3: Distribution of top 15 thirds parties



Source: Elaborated by the author

Table 10: Permissions and Google group of permissions

Name	Permission	Google group of permissions
Access fine location	Precise location (gps and network-based)	Location
Access coarse location	Approximate location (network-based)	Location
Bind device admin	Interact with a device admin	
Use credentials	Use accounts on the device	Accounts
Manage accounts	Add or remove accounts	Accounts
Get accounts	Find accounts on the device	Accounts
Authenticate accounts	Create accounts and set passwords	Accounts
Change wifi multicast state	Allow wi-fi multicast reception	Affects battery
Get tasks	Retrieve running apps	App info
Kill background processes/restart packages	Close other apps	App info
Bluetooth admin	Access bluetooth settings	Bluetooth network
Bluetooth	Pair with bluetooth devices	Bluetooth network
Read history bookmarks	Read your web bookmarks and history	Bookmarks
Write history bookmarks	Write web bookmarks and history	Bookmarks
Camera	Take pictures and videos	Camera
System alert window	Draw over other apps	Display
Send sms	Send sms messages	Messages
Write sms	Edit your text messages (sms or mms)	Messages
Receive sms	Receive text messages (sms)	Messages
Receive wap push	Receive text messages (wap)	Messages
Read sms	Read your text messages (sms or mms)	Messages
Receive mms	Receive text messages (mms)	Messages
Record audio	Record audio	Microphone
Internet	Full network access	Network
Change network state	Change network connectivity	Network
Nfc	Control near field communication	Network
Change wifi state	Connect and disconnect from wi-fi	Network
Change wimax state	Change wimax state	Network
Write profile	Modify your own contact card	Personal info
Read profile	Read your own contact card	Personal info
Write calendar	Add or modify calendar events and send	Personal info
Read calendar	Read calendar events plus confidential	Personal info
Call phone	Directly call phone numbers	Phone calls
Read phone state	Read phone status and identity	Phone calls
Process outgoing calls	Reroute outgoing calls	Phone calls
Use sip	Make/receive internet calls	Phone calls
Disable keyguard	Disable your screen lock	Screenlock
Write contacts	Modify your contacts	Social info
Write social stream	Write to your social stream	Social info
Read call log	Read call log	Social info
Write call log	Write call log	Social info
Read social stream	Read your social stream	Social info
Read contacts	Read your contacts	Social info
Write external storage	Modify or delete the contents of your u	Storage
Install shortcut	Install shortcuts	System tools
Uninstall shortcut	Uninstall shortcuts	System tools
Access mock location	Mock location sources for testing	System tools
Subscribed feeds write	Write subscribed feeds	System tools
Clear app cache	Delete all app cache data	System tools
Read user dictionary	Read terms you added to the dictionary	User dictionary

Notes:

Figure 4: Screen shot of PrivacyGrade permissions

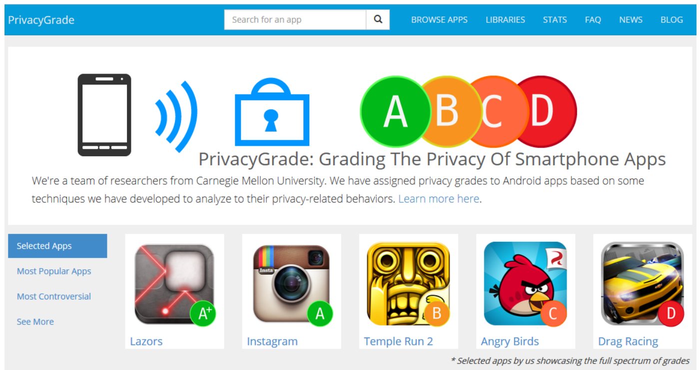

PrivacyGrade
Search for an app
BROWSE APPS
LIBRARIES
STATS
FAQ
NEWS
BLOG
A B C D
PrivacyGrade: Grading The Privacy Of Smartphone Apps
We're a team of researchers from Carnegie Mellon University. We have assigned privacy grades to Android apps based on some techniques we have developed to analyze to their privacy-related behaviors. Learn more here.
Selected Apps
Most Popular Apps
Most Controversial
See More
Lazors
Instagram
Temple Run 2
Angry Birds
Drag Racing
* Selected apps by us showcasing the full spectrum of grades

Table 11: Table 6 (continued) Trivariate probit with application category fixed effects

	Estimations			Average Partial Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
Apps' categories (Ref: Games)	Advertising	In-app	Pers data	Advertising	In-app	Pers data
Following table 6	[...]	[...]	[...]	[...]	[...]	[...]
Books and reference	-0.148***	-0.545***	-0.359***	-0.047	-0.057	-0.062
	(0.040)	(0.059)	(0.054)			
Business	-0.768***	-0.799***	0.451***	-0.205	-0.073	0.101
	(0.023)	(0.041)	(0.027)			
Comics	-0.212***	-0.496***	-0.308**	-0.066	-0.051	-0.053
	(0.070)	(0.096)	(0.136)			
Communication	-0.804***	-0.643***	0.849***	-0.209	-0.062	0.211
	(0.031)	(0.037)	(0.039)			
Education	-0.401***	-0.386***	-0.047	-0.12	-0.045	-0.009
	(0.023)	(0.036)	(0.033)			
Entertainment	-0.242***	-0.746***	-0.107***	-0.075	-0.072	-0.02
	(0.018)	(0.037)	(0.026)			
Finance	-0.679***	-0.847***	0.142***	-0.184	-0.073	0.029
	(0.027)	(0.037)	(0.035)			
Health and fitness	-0.428***	-0.563***	0.010	-0.126	-0.057	0.002
	(0.026)	(0.055)	(0.032)			
Libraries and demo	-0.979***	-0.894***	-0.063	-0.235	-0.073	-0.012
	(0.083)	(0.103)	(0.086)			
Lifestyle	-0.493***	-0.773***	0.138***	-0.143	-0.073	0.028
	(0.020)	(0.030)	(0.025)			
Media and video	-0.501***	-0.818***	0.158*	-0.143	-0.071	0.032
	(0.039)	(0.041)	(0.093)			
Medical	-0.665***	-0.467***	0.000	-0.18	-0.05	0
	(0.041)	(0.068)	(0.043)			
Music and audio	-0.291***	-0.900***	0.039	-0.089	-0.077	0.008
	(0.038)	(0.051)	(0.055)			
News and magazines	-0.252***	-0.144**	-0.126***	-0.078	-0.019	-0.024
	(0.031)	(0.057)	(0.042)			
Personalization	-0.794***	-0.979***	-0.132**	-0.21	-0.081	-0.025
	(0.053)	(0.069)	(0.059)			
Photography	-0.509***	-0.503***	-0.142***	-0.145	-0.053	-0.026
	(0.034)	(0.042)	(0.049)			
Productivity	-0.602***	-0.484***	0.210***	-0.168	-0.052	0.044
	(0.023)	(0.039)	(0.031)			
Shopping	-0.643***	-1.283***	0.056	-0.176	-0.085	0.011
	(0.037)	(0.052)	(0.057)			
Social	-0.522***	-0.660***	0.042	-0.149	-0.063	0.008
	(0.026)	(0.047)	(0.038)			
Sports	-0.312***	-0.584***	-0.051	-0.095	-0.059	-0.01
	(0.031)	(0.048)	(0.040)			
Tools	-0.428***	-0.703***	0.076***	-0.128	-0.07	0.015
	(0.015)	(0.022)	(0.021)			
Transportation	-0.542***	-0.643***	-0.044	-0.153	-0.062	-0.008
	(0.033)	(0.068)	(0.038)			
Travel and local	-0.548***	-0.527***	-0.155***	-0.156	-0.056	-0.029
	(0.034)	(0.078)	(0.036)			
Weather	-0.248***	-0.555***	-0.097	-0.076	-0.056	-0.018
	(0.063)	(0.082)	(0.125)			

Notes: This table gives details on application category fixed effects of the recursive trivariate probit. Column (1) to Column (3) estimate respectively the dependent variable *Advertising*, *In-app* and *Personal Data* . Column (4) to column (6) estimate the average partial effects respectively *Advertising*, *In-app* and *Personal Data*. Robust standard errors clustered at category level are in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Table 12: Table 7 (continued) Trivariate Probit for application category fixed effects

	Estimations			Average partial effect		
Variable	(1) Admob	(2) In-app	(3) Pers data	(4) Admob	(5) In-app	(6) Pers data
Following table 7	[...]	[...]	[...]	[...]	[...]	[...]
Books and reference	-0.037	-0.545***	-0.356***	-.0117994	-.0573167	-.0614853
	(0.041)	(0.059)	(0.053)			
Business	-0.559***	-0.798***	0.456***	-.1504176	-.0731803	.1026196
	(0.023)	(0.041)	(0.027)			
Comics	-0.114	-0.495***	-0.302**	-.035233	-.0514687	-.0526193
	(0.072)	(0.096)	(0.136)			
Communication	-0.614***	-0.642***	0.855***	-.1595246	-.0622135	.2128943
	(0.031)	(0.037)	(0.039)			
Education	-0.292***	-0.386***	-0.042	-.0861077	-.0452056	-.0080624
	(0.023)	(0.036)	(0.033)			
Entertainment	-0.159***	-0.744***	-0.102***	-.0486852	-.0719145	-.0193316
	(0.018)	(0.037)	(0.026)			
Finance	-0.507***	-0.847***	0.152***	-.1373803	-.0730289	.031113
	(0.027)	(0.037)	(0.035)			
Health and fitness	-0.306***	-0.563***	0.016	-.0888329	-.0574529	.0030472
	(0.025)	(0.055)	(0.032)			
Libraries and demo	-0.897***	-0.893***	-0.060	-.2043546	-.0726581	-.0114286
	(0.072)	(0.103)	(0.086)			
Lifestyle	-0.367***	-0.771***	0.144***	-.1055312	-.0730232	.0293694
	(0.020)	(0.030)	(0.026)			
Media and video	-0.357***	-0.817***	0.165*	-.1015631	-.0706243	.0340383
	(0.037)	(0.042)	(0.093)			
Medical	-0.532***	-0.466***	0.007	-.141976	-.0497583	.0014217
	(0.040)	(0.067)	(0.043)			
Music and audio	-0.186***	-0.899***	0.043	-.0563143	-.0764695	.0086093
	(0.038)	(0.051)	(0.055)			
News and magazines	-0.205***	-0.144**	-0.118***	-.0615781	-.0186566	-.0221565
	(0.032)	(0.057)	(0.042)			
Personalization	-0.684***	-0.980***	-0.130**	-.1755962	-.0809146	-.0243761
	(0.055)	(0.069)	(0.059)			
Photography	-0.383***	-0.502***	-0.135***	-.1080368	-.0524062	-.0250797
	(0.034)	(0.042)	(0.049)			
Productivity	-0.452***	-0.483***	0.216***	-.1252372	-.0518059	.045223
	(0.022)	(0.039)	(0.031)			
Shopping	-0.462***	-1.282***	0.063	-.1266895	-.0848848	.0126651
	(0.038)	(0.052)	(0.057)			
Social	-0.362***	-0.659***	0.050	-.1030429	-.0630757	.0098492
	(0.026)	(0.047)	(0.038)			
Sports	-0.196***	-0.583***	-0.046	-.0590367	-.0586469	-.008856
	(0.032)	(0.048)	(0.040)			
Tools	-0.298***	-0.702***	0.081***	-.0878836	-.0697538	.0161456
	(0.016)	(0.022)	(0.021)			
Transportation	-0.384***	-0.643***	-0.039	-.1084456	-.0617155	-.0074294
	(0.033)	(0.067)	(0.038)			
Travel and local	-0.376***	-0.526***	-0.149***	-.1070324	-.055579	-.027588
	(0.035)	(0.078)	(0.036)			
Weather	-0.098	-0.553***	-0.093	-.0304461	-.0555725	-.0175638
	(0.066)	(0.082)	(0.125)			

Notes: This table gives details on application category fixed effects of the recursive trivariate probit. Column (1) to Column (3) estimate respectively the dependent variable *Admob*, *In-app* and *Personal Data* . Column (4) to column (6) estimate the average partial effects respectively *Admob*, *In-app* and *Personal Data*. Standard errors in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$..

Table 13: Table 8 (continued) Trivariate Probit for application category fixed effects with *Advertising*, *In-app purchases* and *Badgrade* dependent variables.

	Estimations			Average partial effect		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Advertising	In-app	Badgrade	Advertising	In-app	Badgrade
Following table 8	[...]	[...]	[...]	[...]	[...]	[...]
Books and reference	-0.079**	-0.554***	-0.546***	-.024	-.0581	-.0551
	(0.040)	(0.059)	(0.064)			
Business	-0.706***	-0.842***	-0.218***	-.182	-.076	-.026
	(0.021)	(0.039)	(0.033)			
Comics	-0.155**	-0.509***	-0.473***	-.046	-.053	-.048
	(0.068)	(0.096)	(0.124)			
Communication	-0.616***	-0.732***	-0.523***	-.162	-.068	-.0526804
	(0.029)	(0.034)	(0.039)			
Education	-0.331***	-0.405***	-0.384***	-.096	-.0472	-.0428
	(0.023)	(0.036)	(0.037)			
Entertainment	-0.191***	-0.760***	-0.343***	-.057	-.073	-.0389
	(0.018)	(0.037)	(0.028)			
Finance	-0.556***	-0.894***	-0.690***	-.149	-.075	-.064
	(0.026)	(0.036)	(0.043)			
Health and fitness	-0.372***	-0.584***	-0.338***	-.105	-.059	-.038
	(0.026)	(0.055)	(0.038)			
Libraries and demo	-0.888***	-0.922***	-0.620***	-.210	-.074	-.0584
	(0.082)	(0.103)	(0.080)			
Lifestyle	-0.438***	-0.801***	-0.292***	-.123	-.075	-.034
	(0.020)	(0.030)	(0.030)			
Media and video	-0.417***	-0.849***	-0.387***	-.116	-.072	-.041
	(0.034)	(0.043)	(0.045)			
Medical	-0.573***	-0.499***	-0.519***	-.152	-.052	-.052
	(0.040)	(0.068)	(0.059)			
Music and audio	-0.246***	-0.917***	-0.253***	-.072	-.077	-.030
	(0.037)	(0.050)	(0.069)			
News and magazines	-0.219***	-0.156***	-0.272***	-.065	-.020	-.031
	(0.030)	(0.057)	(0.049)			
Personalization	-0.721***	-0.991***	-0.390***	-.185	-.082	-.043
	(0.052)	(0.069)	(0.070)			
Photography	-0.421***	-0.528***	-0.540***	-.117	-.054	-.054
	(0.035)	(0.042)	(0.059)			
Productivity	-0.503***	-0.517***	-0.485***	-.137	-.055	-.050
	(0.025)	(0.039)	(0.050)			
Shopping	-0.531***	-1.325***	-0.581***	-.143	-.086	-.056
	(0.036)	(0.052)	(0.053)			
Social	-0.393***	-0.712***	-0.659***	-.110	-.066	-.061
	(0.026)	(0.047)	(0.037)			
Sports	-0.256***	-0.603***	-0.344***	-.074	-.060	-.038
	(0.031)	(0.048)	(0.045)			
Tools	-0.339***	-0.728***	-0.519***	-.098	-.072	-.053
	(0.016)	(0.022)	(0.026)			
Transportation	-0.435***	-0.690***	-0.658***	-.121	-.065	-.061
	(0.032)	(0.067)	(0.058)			
Travel and local	-0.468***	-0.557***	-0.564***	-.129	-.058	-.057
	(0.037)	(0.077)	(0.048)			
Weather	-0.270***	-0.553***	-0.071	-.078	-.056	-.009
	(0.058)	(0.082)	(0.144)			

Notes: This table gives details on application category fixed effects of the recursive trivariate probit. Column (1) to Column (3) estimate respectively the dependent variable *Badgrade*, *In-app* and *Personal Data* . Column (4) to column (6) estimate the average partial effects respectively *Badgrade*, *In-app* and *Personal Data*. Standard errors in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Table 14: Trivariate estimations with Advertising, Integrated Puchase and Personal data for the category Education

Variable	(1) Advertising	(2) I-P	(3) Pers data
Personal data	0.790***	-0.878***	
	(0.061)	(0.040)	
Playstore rating	-0.007	0.014*	-0.025***
	(0.006)	(0.009)	(0.007)
Social networking	0.357***	0.651***	0.899***
	(0.030)	(0.028)	(0.024)
Utility	-0.008	0.192***	0.240***
	(0.018)	(0.023)	(0.023)
Developer website	-0.168***	0.535***	0.408***
	(0.017)	(0.030)	(0.028)
Privacy policy	-0.238***	0.482***	0.101***
	(0.020)	(0.022)	(0.026)
Apps by dvp	-0.001***	-0.000	-0.002***
	(0.000)	(0.000)	(0.000)
Number install 1000-5000	0.352***	0.295***	-0.255***
	(0.020)	(0.028)	(0.025)
Number install 5000-10000	0.275***	0.146***	-0.233***
	(0.025)	(0.036)	(0.033)
Number install 10000-50000	0.697***	0.570***	-0.285***
	(0.021)	(0.029)	(0.031)
Number install 50000-100000	0.518***	0.446***	-0.260***
	(0.027)	(0.036)	(0.038)
Number install 100000-500000	0.948***	0.742***	-0.252***
	(0.033)	(0.042)	(0.052)
Number install 500000-1 million	0.752***	0.654***	-0.370***
	(0.033)	(0.042)	(0.055)
Number install 1-5-million	0.862***	0.969***	0.037
	(0.091)	(0.093)	(0.133)
Number install 5-10 million	0.798***	0.800***	-0.336***
	(0.079)	(0.090)	(0.130)
Number install 10-50-million	-4.431***	0.707	0.576
	(0.146)	(1.076)	(0.996)
Number install 50-100 million	1.816***	1.291***	-5.327***
	(0.493)	(0.425)	(0.219)
Everyone			-1.245***
			(0.020)
Constant	-0.862***	-2.220***	-0.766***
	(0.027)	(0.046)	(0.037)
Log pseudolikelihood	-4.51e+04		
LR test chi2(3)	402.455		
rho21	-0.042***	0.014	
rho31	-0.317***	0.036	
rho32	0.534***	0.022	
Number of draw	200		
Observations	40332		

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Admob*, *In-app* and *Personal Data* . We only focus on the subsample of apps belonging to the Education category. *Everyone* is the exclusion restriction variable.

Table 15: Trivariate estimations with Advertising, Integrated Puchase and Personal data for All games

Variable	(1) Advertising	(2) I-P	(3) Pers data
Personal data	0.311***	0.798***	
	(0.051)	(0.043)	
Playstore rating	0.016***	0.115***	0.003
	(0.005)	(0.008)	(0.006)
Social networking	0.743***	0.485***	0.692***
	(0.016)	(0.016)	(0.014)
Utility	0.401***	-0.018	0.155***
	(0.013)	(0.014)	(0.014)
Developer website	0.142***	0.420***	0.112***
	(0.010)	(0.013)	(0.013)
Privacy policy	-0.224***	0.520***	0.097***
	(0.014)	(0.014)	(0.016)
Apps by dvp	0.009***	-0.003***	0.001***
	(0.000)	(0.000)	(0.000)
Number install 1000-5000	0.214***	0.216***	0.108***
	(0.013)	(0.018)	(0.018)
Number install 5000-10000	0.193***	0.144***	0.000
	(0.017)	(0.023)	(0.024)
Number install 10000-50000	0.324***	0.391***	0.232***
	(0.014)	(0.017)	(0.018)
Number install 50000-100000	0.250***	0.320***	0.122***
	(0.018)	(0.022)	(0.024)
Number install 100000-500000	0.453***	0.656***	0.370***
	(0.017)	(0.020)	(0.020)
Number install 500000-1 million	0.439***	0.560***	0.287***
	(0.020)	(0.023)	(0.024)
Number install 1-5-million	0.638***	0.976***	0.443***
	(0.028)	(0.028)	(0.029)
Number install 5-10 million	0.581***	0.794***	0.475***
	(0.030)	(0.030)	(0.032)
Number install 10-50-million	0.825***	1.155***	0.519***
	(0.070)	(0.067)	(0.068)
Number install 50-100 million	0.868***	1.086***	0.511***
	(0.062)	(0.057)	(0.059)
Number install 100-500-million	0.642**	1.547***	0.469*
	(0.289)	(0.356)	(0.276)
Number install 500-1000 million	1.573***	0.953***	0.708***
	(0.389)	(0.258)	(0.254)
Everyone			-0.738***
			(0.011)
Constant	-0.766***	-2.263***	-1.119***
	(0.020)	(0.035)	(0.029)
Log pseudolikelihood	-1.27e+05		
LR test chi2(3)	270.645		
rho21	-0.101***	0.007	
rho31	-0.034	0.029	
rho32	-0.161***	0.024	
Number of draw	300		
Observations	90960		

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Admob*, *In-app* and *Personal Data* . We only focus on the subsample of apps belonging to the Game category. *Everyone* is the exclusion restriction variable.

Table 16: Trivariate estimations with Advertising, Integrated Puchase and Personal data for the catefgories Healh and Lifestyle

Variable	(1) Advertising	(2) I-P	(3) Pers data
Personal data	0.347***	-0.533***	
	(0.053)	(0.044)	
Playstore rating	-0.007	0.007	-0.038***
	(0.005)	(0.008)	(0.005)
Social networking	0.561***	0.238***	0.894***
	(0.026)	(0.031)	(0.018)
Utility	0.212***	0.277***	0.147***
	(0.017)	(0.025)	(0.019)
Developer website	-0.132***	0.348***	0.301***
	(0.016)	(0.030)	(0.022)
Privacy policy	-0.305***	0.316***	0.044**
	(0.019)	(0.024)	(0.021)
Apps by dvp	-0.000	0.002***	-0.001***
	(0.000)	(0.000)	(0.000)
Number install 1000-5000	0.218***	0.249***	-0.408***
	(0.020)	(0.031)	(0.022)
Number install 5000-10000	0.156***	0.120***	-0.252***
	(0.025)	(0.041)	(0.027)
Number install 10000-50000	0.470***	0.352***	-0.476***
	(0.021)	(0.033)	(0.026)
Number install 50000-100000	0.306***	0.291***	-0.476***
	(0.027)	(0.041)	(0.032)
Number install 100000-500000	0.613***	0.500***	-0.309***
	(0.030)	(0.042)	(0.038)
Number install 500000-1 million	0.621***	0.421***	-0.410***
	(0.032)	(0.047)	(0.043)
Number install 1-5-million	0.662***	0.920***	-0.207***
	(0.067)	(0.075)	(0.078)
Number install 5-10 million	0.684***	0.642***	-0.294***
	(0.066)	(0.085)	(0.081)
Number install 10-50-million	0.473**	1.661***	0.261
	(0.236)	(0.225)	(0.259)
Number install 50-100 million	0.766***	0.894***	-0.522***
	(0.171)	(0.194)	(0.189)
Number install 100-500-million	-3.508***	-2.934***	-3.096***
	(0.223)	(0.211)	(0.215)
Everyone			-1.367***
			(0.019)
Constant	-0.846***	-2.237***	-0.405***
	(0.026)	(0.047)	(0.028)
Log pseudolikelihood	-5.22e+04		
LR test chi2(3)	80.993		
rho21	-0.005	0.014	
rho31	-0.047	0.034	
rho32	0.271***	0.026	
Number of draw	213		
Observations	45346		

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Admob*, *In-app* and *Personal Data* . We only focus on the subsample of apps belonging to the Health and Lifestyle category. *Everyone* is the exclusion restriction variable.

Table 17: Trivariate estimations with Advertising, Integrated Puchase and More tahn 6 permissions

	Estimations		
	(1)	(2)	(3)
Variable	Advertising	I-P	More 6 perms
More Than 6 permission	0.506***	-0.584***	
	(0.055)	(0.046)	
Playstore rating	-0.006**	0.015***	-0.017***
	(0.003)	(0.005)	(0.003)
Social networking	0.509***	0.469***	0.790***
	(0.022)	(0.026)	(0.021)
Utility	0.216***	0.158***	0.155***
	(0.014)	(0.021)	(0.019)
Developer website	-0.102***	0.358***	0.256***
	(0.015)	(0.021)	(0.022)
Privacy policy	-0.205***	0.399***	0.008
	(0.019)	(0.024)	(0.024)
Apps by dvp	0.001*	0.001*	-0.001
	(0.001)	(0.001)	(0.001)
Number install 1000-5000	0.212***	0.193***	-0.221***
	(0.010)	(0.017)	(0.014)
Number install 5000-10000	0.146***	0.137***	-0.165***
	(0.010)	(0.017)	(0.014)
Number install 10000-50000	0.422***	0.341***	-0.273***
	(0.012)	(0.018)	(0.017)
Number install 50000-100000	0.298***	0.275***	-0.254***
	(0.012)	(0.020)	(0.018)
Number install 100000-500000	0.559***	0.600***	-0.169***
	(0.015)	(0.021)	(0.021)
Number install 500000-1 million	0.525***	0.469***	-0.227***
	(0.015)	(0.021)	(0.022)
Number install 1-5-million	0.655***	0.936***	0.003
	(0.024)	(0.030)	(0.030)
Number install 5-10 million	0.620***	0.756***	-0.080***
	(0.023)	(0.028)	(0.031)
Number install 10-50-million	0.658***	1.178***	0.290***
	(0.057)	(0.058)	(0.066)
Number install 50-100 million	0.762***	1.079***	0.182***
	(0.045)	(0.053)	(0.055)
Number install 100-500-million	0.284	1.347***	0.734***
	(0.175)	(0.166)	(0.218)
Number install 500-1000 million	0.632***	1.308***	0.588***
	(0.164)	(0.149)	(0.160)
Number install 1000-5000 million	-5.908***	-5.078***	1.111***
	(0.567)	(0.543)	(0.196)
Everyone			-1.225***
			(0.016)
Constant	-0.382***	-1.694***	-1.699***
	(0.020)	(0.030)	(0.035)
Log pseudolikelihood	-5.22e+05		
LR test chi2(3)	1631.910		
rho21	-0.052***	0.011	
rho31	-0.321***	0.030	
rho32	0.335***	0.025	
Number of draw	683.000		
Observations			

Notes: Recursive Trivariate probit estimations. Column (1) to Column (3) estimate respectively the dependent variable *Advertising*, *In-app* and *More than 6 permissions* . *Everyone* is the exclusion restriction variable