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

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

October 24, 2018

Abstract

The large majority of digital goods such as smartphone apps are zero priced. To generate revenue, developers need to monetize their apps. However, little is known about their monetization strategies. We provide empirical evidence on the 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 purchases. In particular, personal data are essential for Internet companies but their involvement in the smartphone applications market remains relatively unexplored. We have unique data to measure how apps are monetized, based on information related to free applications available on the Google Play Store platform combined with data on applications' privacy-related behaviors provided by Privacy Grade. We provide information on the third party market related to mobile applications. We show that the economics of mobile applications and the third party economy should be considered together. Among the apps in our dataset, 9% collect personal data and use no other strategy. We show that a higher number of downloads is associated to the collection of more data, hence apps with more than 100 million downloads have more than 20% probability of exploiting users' personal data as a monetization strategy.

JEL CODE: D82, D83, M31, M37

*Institut Mines Telecom, Business School. Email: grazia.cecere@imt-bs.eu

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

‡Institut Mines Telecom, Business school, University of Paris Sud - Paris Saclay. Email: vincent.lefrere@u-psud.fr

We are grateful to Alessandro Acquisti, Alexandre de Corniere, Babur De los Santos, Anna D'Annunzio, Anja Lambrecht, Amedeo Piolatto, Patrick Schulte, Mirko Tonin, Catherine Tucker, 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), Royal economics society conference (April 2018), The Bordeaux JMA (June 2018), INFORM EURO 2018 Valencia Operational Research and Management Science (July 2018), the 45th EARIE Annual Conference (August 2018), the 16th IAOS Conference (September 2018), The French law and economics association conference (October 2018) for comments and suggestions. The work was supported by the DAPCODS/IOTics ANR 2016 project (ANR-16-CE25-0015).

Electronic copy available at: https://ssrn.com/abstract=3136661

1 Introduction

In 2018, free apps constitute 95% of total apps commercialized in the Google Play Store (App Brain, 2018[1]). Like other digital goods, apps are related to various revenue streams which include in-app purchases, advertising, e-shopping and users' personal data (Lambrecht *et al.*, 2014). While the theoretical literature shows that personal data are an essential input for digital platforms offering services at zero price (Chellappa and Shivendu, 2010; Casadesus-Masanell and Hervas-Drane, 2015), there is no clear understanding about how they interact with more traditional monetization strategies such as advertising and in-app purchases (also called freemium or integrated purchase), especially in the mobile applications market. We consider free apps those that are free to download for consumers without any paywall.

The present article fills a gap in the empirical literature by analyzing how developers combine different strategies to monetize free apps by focusing on the market for personal data. First, we investigate whether personal data are used to complement or substitute for advertising and in-app purchases. Second, we study the role of third parties in the definition of the monetization strategy. A third party is a software component developed by a third entity which provides specific services such as app analytics or advertising, or alternatively services that can be embedded into the app such as image management (Razaghpanah *et al.*, 2018). Although third parties are essential for the distribution of ads,[2] 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 monetization strategy choice. Third, we have a unique setting which allows us to measure whether or not personal data are collected for functional purposes. If they are not collected for technical reasons, we can assume different usage of user data. For instance, they can be

[1] `https://www.appbrain.com/stats`. Last retrieved September 2018.

[2] Yet, in 2017, for the first time, the revenue of mobile ads represented more than half of the digital advertising (Internet Advertising Revenue Report 2017).

Electronic copy available at: https://ssrn.com/abstract=3136661

used to improve ad targeting or they can be commercialized via the data broker market.

We have unique data based on matching the characteristics of apps evaluated by Privacy Grade[3] (Lin *et al.*, 2014) to app data collected via Google Play. Privacy Grade is an ongoing computer science project aimed at evaluating whether developers require more permissions than those required for functional purposes. For example, Uber requires the permission user locations since drivers need to know the user's location in order to pick them up but one application may also require more permissions than needed. Privacy Grade also identifies third parties associated to each app, offering additional services to developers, such as business analytics and advertising.

Developers of free apps have three possible monetization strategies which are not mutually exclusive: advertising, in-app purchases, and collection of personal data. We use a multivariate recursive probit model which allows us to estimate the joint probability that developers will choose more than one business model (Filippini *et al.*, 2018). This system of binary dependent variables with endogenous dummy variables requires an exclusion restriction to identify the coefficients (Jones *et al.*, 2013). Our exclusion restriction is the targeted age group which is likely to have a negative effect on the probability of collecting personal data (Cecere *et al.*, 2018) but is not likely to affect the probability of some other traditional monetization strategy. We use different methods to estimate the robustness of our results. First, we estimate the model with an alternative measure of advertising, namely Admob, the most important third parties in the mobile advertising in the Google Play Store. Second, the re-estimation of the model with an alternative dependent variable, measuring that users' data are not collected for functional purpose corroborates the main results that advertising is a complement of personal data collection and in-app purchases is likely to be a substitute. Third, we

[3] `http://www.privacygrade.org` Last retrieved September 2018.

Electronic copy available at: https://ssrn.com/abstract=3136661

also estimate the main equation for app in the Game category and apps in Lifestyle and Health & Fitness.

Our findings are directly relevant to the economics of mobile applications, economics of free digital goods, and economics of privacy literatures.

Works on the economics of mobile applications have mainly studied the demand for applications and the factors influencing the app's success. The literature shows that the market for applications has a long tail distribution (Garg and Telang, 2013; Gabaix, 2016) leading to huge competition over a number of downloads. Competition among developers aspiring to be top ranked, results in the implementation of different strategies. Li *et al.* (2016) show that the developers of new apps can buy downloads to increase their visibility in the Apple Store. Also, Comino *et al.* (2018) show that updates can be released strategically to increase downloads. Hence, developers on the iTunes platform compared to the Google Play platform, seem to rely mainly 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. In a study of the competition involved in achieving a top ranking, Yin *et al.* (2014) investigated the differences between game and non-game apps. They found 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 developer's experience. In an examination of the Google Play Store platform, Ershov (2017) shows that platform design influences consumer welfare and market competition. Using demand-ranked data for the Apple iOS market, Garg and Telang (2013) show that free apps are the most frequently downloaded apps.

Electronic copy available at: https://ssrn.com/abstract=3136661

This drives the competition among developers highlighting the challenges faced by developers that distribute free apps. Looking at free apps, Deng *et al.* (2018) analyze a sub-sample of free apps associated with paid apps and they show that launching a free version alongside a paid version increases the download of the paid version.

These results highlight the factors affecting the demand of apps but provide no insights into the monetization strategy related to free applications. Empirical work investigating firms' monetization strategies related to zero priced goods is scarce, especially in the app market.

Our article also contributes to the economics of privacy since we assess how personal data can complement revenue from free services. Work in this field provides evidence of the existence of different privacy markets (Bergemann *et al.*, 2018). First, there is one where individuals provide personal data in exchange for free services; second, there is a market that involves the commercialization of personal data by data brokers; and third, there is 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 investigates these markets for personal data in the context of mobile apps. The article by Kummer and Schulte (2018) shows that smartphone users take account of permissions when downloading applications but the literature tends to overlook the possible relationships between the three monetization strategies described. Indeed, personal data potentially could be related to advertising since it enables targeted advertising but could also be considered a monetization strategy in its own right. For example, personal data can be used by data brokers to infer socio-economic characteristics, e.g. to estimate the wealth of an individual (Blumenstock, 2018) or to assess consumer preferences, e.g. Athey *et al.* (2018) using mobile location data estimated both consumer

Electronic copy available at: https://ssrn.com/abstract=3136661

preferences for restaurants and the latent characteristics of each restaurant. Our research adds to this literature and identifies the role of third parties in the distribution of ads and the collection of users' data.

Overall, our results indicate that the collection of personal data is a strategy of monetization on its own. Furthermore, results suggest that applications with a high volume of downloads are more likely to rely on users' personal data as a source of revenue. We also find that the use of social networking third parties increases the probability of collecting personal data. Finally, our findings highlight the fact that personal data is used as a complement of advertising, whereas collecting user data decreases the probability of using integrated purchase.

The managerial and policy implications of these findings are threefold. First, they could help developers to identify the most profitable strategies for the distribution of free apps, and allow mobile analytics to implement more efficient marketing strategies. Indeed, personal data are extremely valuable to allow consumers to be targeted to improve the match between seller and buyer. Second, it should be informative for policy makers about the functioning of this competitive market. Third, our findings reveal the relationships between third parties and the app monetization strategies.

The article is organized as follows. Section 2 describes the data and key features of the app market for personal data, and section 3 presents the econometric models used to test our main assumption that developers can use personal data to monetize their applications. Section 4 discusses the econometric results, section 5 presents the discussion and section 6 concludes.

Electronic copy available at: https://ssrn.com/abstract=3136661

2 Data and main variables of interest

To build our dataset we match data from two websites - Privacy Grade data and Google Play Store - collected between May and July 2015. At this time, free apps represented 85% of all apps commercialized in the Google Play Store. Privacy Grade is an ongoing project of a group of computer science researchers at Carnegie Mellon University. Privacy Grade evaluates the relevance of the personal data required by permissions. Specifically, 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 a large sample of free apps and grade it based on this difference (Lin *et al.*, 2012, 2014) using information on the permission collected. In addition, the Privacy Grade data include information on the third parties related to each application through libraries.

In June 2015, we collected 475,787 free apps available in Privacy Grade data. This sample represented 36% of applications commercialized by the Google Play platform in 2015. Then, we matched Privacy Grade data to publicly available data from Google Playstore. These data include app characteristics such as the number of downloads, Google categories (Games, Health, Social, etc.) and user evaluations. We have only cross sectional data as Privacy grade evaluated only once apps commercialized in Google Play Store at the time we collected the data. Table 1 presents the main variables including summary statistics per type of monetization strategy. We can highlight that 17.7% of apps rely on personal data as a monetization strategy, 32.4% rely on advertising and 8.8% use in-app purchases. The games category is the largest category with about 19.1 % of apps, followed by the categories Education 8.5 %, Tools 8.1% and Lifestyle 6.7%.

Electronic copy available at: https://ssrn.com/abstract=3136661

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

Variables				Monetization strategies			
	Mean	Min.	Max.	Ads	In-app purchases	Personal data	None (No monetization strategy)
	(1)			(2)	(3)	(4)	(5)
Dependent variables							
Personal data	0.177	0	1	.	.	.	.
Advertising	0.324	0	1	.	.	.	.
In-app purchases	0.088	0	1	.	.	.	.
App characteristics							
Playstore rating	3.698	0	5	3.737	3.818	3.550	3.696
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
Everyone	0.583	0	1	0.559	0.563	0.188	0.665
Developers characteristics							
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
Category							
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	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	475,787			153,978	41,786	84,001	253,634

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) presents descriptive statistics for *In-app purchases*. Column (4) presents the descriptive statistics for *Personal data*. Column (5) presents statistics for developers with no monetization strategy. The monetization strategies are not mutual exclusive.

Electronic copy available at: https://ssrn.com/abstract=3136661

2.1 Third party libraries

We use a dataset collected by Privacy Grade. An important strength of this data is that they allow us to identify third parties (also called libraries) associated to each application.[4] Third parties can gain access to user data without the user being aware. Third parties can enable the inclusion of in-app advertising or offer tools to help developers create apps. They can gather personal data on app users in order to improve the app's functioning. Privacy Grade classifies these third party libraries into different groups. For the purpose of our analysis we use advertising, social networking, utility, and mobile analytics groups of third parties.[5] Our sample includes more than 170 third parties and 50.4% of apps have at least one third party embedded in their functioning. Table 2 indicates the percentage of apps that use each group of third parties.[6] While third parties are essential to enable certain app functionalities, little is known about the structure of this market or the actors involved.

The advertising third parties include different entities that allow the apps to deliver advertising, they are the largest group of third parties in our sample. They transfer a percentage of the revenues generated for the app developers. These third parties are used by 32.4 % of apps. To incorporate the fact that app developers use ad third parties, we create the variable *Advertising* (see section 2.3). It is important to underline that apps with their own ad platform do not use ad third parties such as Facebook.

[4]To identify third parties, computer science researchers use the code in the APK files of each app (Lin *et al.*, 2014).

[5]There are two other categories of thirds parties: Development aid and Payment. They represent small fraction of apps respectively 3.6% and 3.9%.

[6]Developers can use several libraries at the same time.

Electronic copy available at: https://ssrn.com/abstract=3136661

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

Category of third parties	Mean	Min	Max	Number of different third parties
Advertising third parties	0.324	0	1	79
Utility third parties	0.187	0	1	71
Social networking third parties	0.137	0	1	10
Mobile analytics third parties	0.078	0	1	12
Observations	475,787			172

Notes: This table provides summary statistics for different categories of third parties classified by Privacy Grade. The column 'Number of different third parties' indicates the number of different libraries in each category.

The utility third parties help developers to add functions not directly developed by themselves; these can be tools to manage images on the apps, such as Nostra 13, an open source program used by developers to manage images within the app.[7] Another example is Adobe third party, which enable apps to use their services. These third parties are used to construct the variable *Utility* and are employed by 18,7% of the applications in our sample.

The social network third party libraries link the app functioning to the services offered by the social network companies. This is used to build the dummy variable *Social Networking.* This group of third parties is used by 13.7% of applications (Table 2), Facebook and Twitter being examples of these libraries.

The mobile analytics libraries offer analysis of applications' usage (e.g. bug). This group of third parties is exploited by 7.8% of the applications and used to build the binary variable *Mobile Analytics.* An example is the third party Flurry owned by Yahoo which elaborates users' data to offer business analytics services to developers.

[7]For example, Nostra 13 helps developers with images, while Jsoup helps with HTLM language. Nostra 13 is the most widely used utility third party and consists of an open source program available on Github.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 3 shows the top 15 third parties related to each monetization strategy (Appendix Figure 3 presents the top 15 third parties). 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.[8] Facebook library is also widely used followed by Flurry and Twitter.

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	86.5%	Admob	31.1%	Facebook	33.4%
Facebook	20.9%	Facebook	25.8%	Admob	31.5%
Flurry	10.6%	Flurry	18.2%	Twitter4j	18.3%
Twitter4j	8.6%	Chartboost	9.5%	Flurry	16.9%
Millennial media	7.8%	Unity3d	8.3%	Paypal	10.4%
Inmobi	7.1%	Twitter4j	6.2%	Biznessapps	8.9%
Chartboost	5.9%	Tapjoy	5.9%	Nostra13	7.9%
Unity3d	5.3%	Inmobi	4.9%	Oauth	7.2%
Paypal	4.5%	Millennial media	4.6%	Millennial media	6.2%
Revmob	4.5%	Nostra13	4.4%	Inmobi	5.8%
Jsoup	4.3%	Oauth	4.2%	Acra	5.7%
Biznessapps	3.8%	Adobe	4.1%	Jsoup	4.8%
Nostra13	3.8%	Amazon	3.9%	Revmob	4.2%
Mopub	3.6%	Mopub	3.8%	Chartboost	4.1%
Oauth	3.3%	Loopj	3.3%	Ksoap2	4.1%

Notes: This Table presents summary statistics of the 15 biggest third parties by variables of interest - *Advertising, In-app purchases* and *Personal data*. The advertising third parties are: Admob, Chartboost, Inmobi, Millennial media, Mopub, Revmob Tapjoy. The social networking third parties are: Facebook, Twitter4J. The utility third parties are: Acra, Adobe, Amazon, Jsoup, Ksoap2, Loopj, Nostra13, Ouath, Unity3d. The mobile analytics third party is Flurry. The payment third party is Paypal. The mobile analtics third party is Flurry. The payment third party is Paypal. The development aid third party is Biznessapps.

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

Electronic copy available at: https://ssrn.com/abstract=3136661

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

In our empirical analysis, we want to model the monetization strategies of developers and so 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. For this purpose, our empirical approach allows us to use *Personal data* both as a dependent variable and as a regressor. Table 4 presents the statistics for different strategy combinations.

Table 4: Summary statistics: Combination of monetization strategies

Monetization strategy	Mean
None (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 presents all combinations of monetization strategies.

First, to measure whether the app collects personal data we use two sources of information: the Google Play permission system and data provided by Privacy Grade. The Android permission system allows developers to interact with the functionalities of the smartphone and potentially to collect personal 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 software permissions, only half of them are defined as dangerous by the Google Play Store, [9] thus,

[9]Defined by Google as: 'Permissions [...] considered as intrusive if they can affect the functioning

Electronic copy available at: https://ssrn.com/abstract=3136661

for the purposes of our study we consider only this sub-sample (detailed in appendix Table 11 and Table 12). Examples of the personal data collected are users' geo-location, contacts, and access to text messages. On average, apps have 3.2 dangerous permissions with a standard deviation of 2.9. About 10% of the apps in our sample use more than six dangerous permissions. We create the dummy variable *more than 6 permissions*

Alternatively, Privacy Grade measures the discrepancy between the required permission and the app's technical features.[10] More precisely, Privacy Grade data grades apps from D to A+ to rank their privacy intrusiveness by measuring the disparity between the app's functionality and the types of permissions required, with A+ referring to apps that collect personal data only needed for the functioning of the app. If an app receives a ranking between B and D, this means that at least one permission installed requests more information than is required for the app's functionality. Thus, we assume that apps which collect user data for other than technical reasons might commercialize them in the data broker market (Razaghpanah *et al.*, 2018). We created a dummy variable *Badgrade* which takes the value 1 if the app is graded between B and D, otherwise 0.

We use two measures of personal data because first not all permissions have been evaluated by Privacy Grade. Second, the collection of users' data is costly in terms of data management and legal requirements, thus apps requiring a large number of permission are likely to collect them not only for functional purpose. Thus, we use both the variable *Badgrade* and the number of permissions to construct the variable *Personal data*. This takes the value 1 if *Badgrade* is equal to 1, and/or if the app has more than six dangerous permissions and 0 otherwise. The descriptive statistics show that 9.2% of app use personal data as a monetization strategy, 6.2% of apps combine

of the device.' (November 2014 - October 2015). These permissions are those proposed within the Lollipop version of Google Play Store and they request users to provide explicit agreement. Our dataset includes 55 dangerous permissions.

[10]Appendix Figure 4 depicts an example of the grading system used by Privacy Grade.

Electronic copy available at: https://ssrn.com/abstract=3136661

personal data with advertising, and 1.2% combine all three monetization strategies (see Table 4). The robustness check includes the estimation of the main equation only with the variable *Badgrade*.

Second, *Advertising* is a dummy variable measuring whether developers provide ads to the apps through 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 through the app provided via third parties, as the option Contain Ad (informing consumers about the provision of ads in the app) was introduced only in January 2016 for all apps commercialized by the Google Play Store.[11]

Third, 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 games, upgrade, and bonuses. In this case, the platform remunerates the developers directly, and charges 15% of the amount spent. There are 4.1% of apps that use only in-app purchases.[12]

In our sample, 53.3% of apps have no monetization strategy. Based on the literature, we can propose several reasons for this. First, some apps are produced by non-profit organizations such as Wikipedia and Mozilla or government apps such as taxes. Second, developers can use their apps as 'visiting cards' to demonstrate their competences. 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, e.g. banking and TV channel apps. Fourth, apps can be created based on brands in order to adver-

[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 September 2018.

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

Electronic copy available at: https://ssrn.com/abstract=3136661

tise. Gupta (2013) explains that brands are aimed more at increasing interest in the product, e.g. Red Bull offers games associated to the brand.

2.3 Apps characteristics and developers

To measure the popularity of apps, we use the download category provided by Google Play which includes 19 discrete groups. The statistics for the number of downloads are presented in Table 5 and range from fewer than five downloads to over a thousand million downloads. To quantify whether the number of downloads affects the probability of choosing a particular business model, we include the vector of the variables measuring download intensity. It should be underlined that apps with more than 1k million downloads usually use an internal Ad platform and thus they are less likely to use advertising third parties. Facebook and Snapchat are an example.

Figure 1 depicts the monetization distribution by category of installations. For each download category (horizontal axis), we show the percentage of apps for each monetization strategy. While advertising is mostly used by apps with fewer 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 of the literature that the commercialization of personal data is valuable when huge amounts of data are collected (OECD, 2013; Lambrecht and Tucker, 2017).

Figure 2 shows the percentage of apps for each monetization strategy grouped by Google Play app category. As pointed out by Yin *et al.* (2014), who show that the strategy adopted by developers depends on their categorization, patterns of competition differ among categories. While the Games category is more likely to use advertising (see also Table 1 for descriptive statistics), four categories namely Communication, Business, Medical and Health & Fitness are more likely to use personal data as a business

Electronic copy available at: https://ssrn.com/abstract=3136661

model. The data collected in these categories are particularly valuable as they provide information on users' health and financial conditions.

The quality of the application and user satisfaction are measured using the variable *Playstore Rating*; app grading is provided 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 developers* 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 its own website. Third, dummy variable *Privacy policy* indicates if the developer has a privacy policy.

Electronic copy available at: https://ssrn.com/abstract=3136661

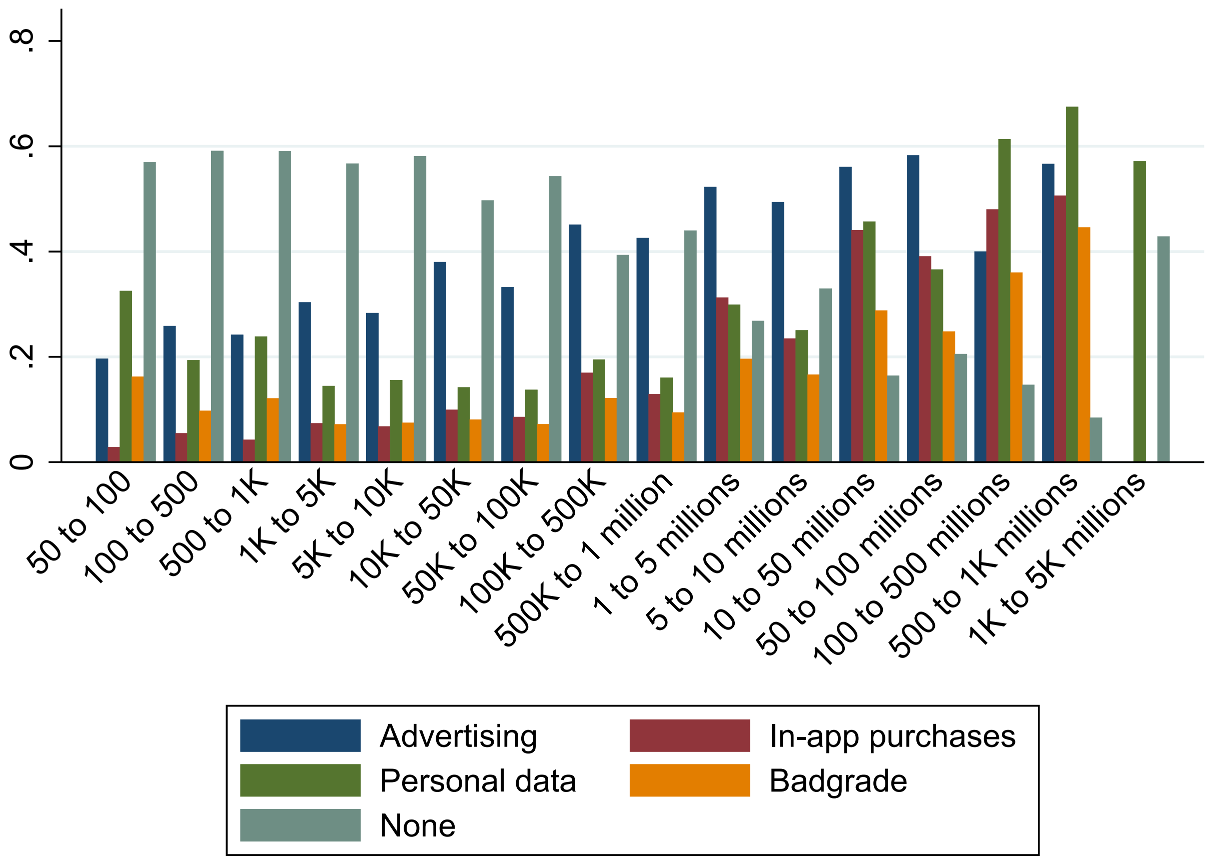
Table 5: Summary statistics: Number of downloads by monetization strategies

	Advertising	In-app purchases	Personal data	None	Overall
	(1)	(2)	(3)	(4)	(5)
Number installed 1-5	0.002	0.001	0.007	0.004	0.004
Number installed 5-10	0.000006	0.00002	0.00008	0.000004	0.000
Number installed 10-50	0.038	0.022	0.085	0.059	0.054
Number installed 50-100	0.003	0.001	0.009	0.005	0.005
Number installed 100-500	0.147	0.115	0.202	0.204	0.184
Number installed 500-1K	0.037	0.024	0.067	0.055	0.050
Number installed 1K-5K	0.197	0.176	0.172	0.223	0.210
Number installed 5K-10K	0.082	0.072	0.083	0.102	0.094
Number installed 10K-50K	0.193	0.186	0.132	0.153	0.164
Number installed 50K-100K	0.086	0.082	0.065	0.085	0.084
Number installed 100K-500K	0.092	0.128	0.073	0.049	0.066
Number installed 500K-1 million	0.066	0.073	0.045	0.041	0.050
Number installed 1-5 million	0.026	0.057	0.027	0.008	0.016
Number installed 5-10 million	0.021	0.038	0.02	0.009	0.014
Number installed 10-50 million	0.004	0.011	0.006	0.001	0.002
Number installed 50-100 million	0.005	0.012	0.006	0.001	0.003
Number installed 100-500 million	0.0002	0.0009	0.0005	0.00004	0.0002
Number installed 500-1K million	0.0003	0.001	0.0007	0.00003	0.0002
Number installed 1K-5K million	0	0	0.00009	0.00002	0.00003
Observations	153,978	41,786	84,001	253,634	475,787

Notes: The download ranges include 19 categories. Column (1) presents the percentage of apps that use *Advertising*. Column (2) presents the percentage of apps that use *In-app purchases*. Column (3) presents the percentage of apps that use *Personal data*. Column (4) presents the percentage of apps that do not use monetization strategies. Column (5) presents the percentages for the full sample. The monetization strategies are not mutual exclusive.

Electronic copy available at: https://ssrn.com/abstract=3136661

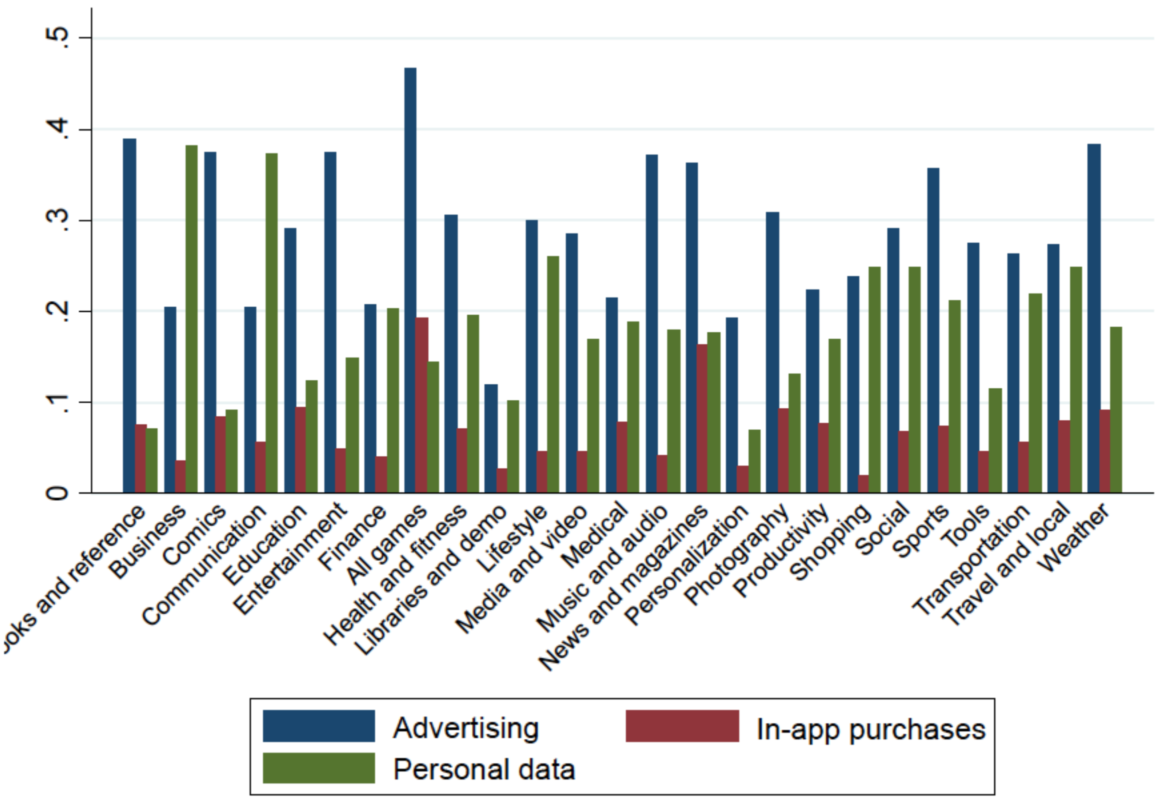
Figure 1: Strategies of monetization by volume of downloads



Notes: The vertical axis is the percentage of apps using a monetization strategy. The horizontal axis is the volume of downloads. The graph does not include the apps with less than 50 downloads for readability.

Electronic copy available at: https://ssrn.com/abstract=3136661

Figure 2: Strategies of monetization by Google category



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

3 Modeling the monetization choice

Advertising and in-app purchases are traditional business strategies in the digital 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 can be sold to data brokers. To model a developer's choice, we estimate a recursive multivariate probit which accounts for the endogeneity of personal data.

Electronic copy available at: https://ssrn.com/abstract=3136661

3.1 Recursive multivariate probit model

The literature suggests that there is a potential association between traditional business models, namely advertising, in-app purchases and personal data. In particular, users' personal data are likely to be used to run personalized ads. To address the potential endogeneity of the variable *Personal data*, we estimate a multivariate recursive probit model (Goy and Wang, 2015). Similar to the bivariate recursive probit, the multivariate recursive probit model is a system of three probit equations which allow the unobservables to be correlated (indeed, several unobservable factors such as the developer preference for easy monetization simultaneously could influence the choice of monetization strategy) with one binary dependent choice to be an endogenous explanatory variable in the other equations (Filippini *et al.*, 2018). This method is equivalent to an instrumental variable and is preferred once both the dependent variable and endogenous variable are dichotomous (Fairlie, 2006). In particular, personal data can be combined with other monetization strategies but it can also be used as a business model *per se*. The probabilities of *Advertising, In-app purchases*, and *Personal data* may not be independent and our empirical strategy allows us to measure the relationship among common unobservables which explain these three choices of monetization.

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 purchases* and *Personal data* of app i are estimated with a recursive multivariate probit as follows:

$$\begin{cases} y^*_{Ai} = \alpha_1 X_i + \beta_1 Personal\,Data_i + \epsilon_{1i} & (1) \\ y^*_{Ii} = \alpha_2 X_i + \beta_2 Personal\,Data_i + \epsilon_{2i} & (2) \\ Personal\,Data^*_i = \alpha_3 X_i + \gamma Z_i + \epsilon_{3i} & (3) \end{cases}$$

The errors terms $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})$ are distributed with a MVN(0,1). We can write the variance-covariance matrix Ω as follows:

Electronic copy available at: https://ssrn.com/abstract=3136661

$$\Omega = \begin{pmatrix} 1 & \rho_{AI} & \rho_{AP} \\ \rho_{AI} & 1 & \rho_{IP} \\ \rho_{AP} & \rho_{IP} & 1 \end{pmatrix}$$

where *Advertising* is denoted A, and *In-app purchases* is denoted I. y^*_{Ai}, y^*_{Ii} and $Personal\ Data^*_i$ are latent variables. X_i is a vector of app and developer characteristics as well as fixed effects for the Google Play Store categories. The ρ_{kl} with $k, l = A, I, P \ \ \forall \ \ k \neq l$ is the correlation between the error terms. Equation (1) and (2) represent respectively the choice of *Advertising* (y^*_{Ai}) and *In-app purchases* (y^*_{Ii}). Equation (3) represents the choice to use the *Personal data* ($Personal\ Data^*_i$). To improve identification, including at least one exclusion restriction Z_i is recommended (Goy and Wang, 2015) In this application, the variable *Everyone* is the exclusion restriction (see Section 3.2)

We normalized the residual and we use a 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; Cappellari *et al.*, 2003). This model allows consideration of the combination of the different choices with coefficients of correlation (ρ_{AI}, ρ_{AP}, ρ_{IP}). The rhos reflect (in part) the correlations the error terms the three equations but can be considered as a combination (a weighted average) of the error terms (Filippini *et al.*, 2018). If the monetization strategy decisions are dependent, the ρ are significantly different from zero.

In order to compute the magnitude of our regressors, we manually compute the average partial effect using the method proposed by Cameron and Trivedi (2010) and Jones *et al.* (2013). This method allows variations according to the scaling of each covariate; for the continuous variable we add a change equal to 1, and to compute the average partial effect of a dummy variable we use a change from 0 to 1. This method

Electronic copy available at: https://ssrn.com/abstract=3136661

allows us to compute only the standard deviation.

3.2 Exclusion restriction: Everyone

The identification of this model requires an exclusion restriction for robust identification of the parameters, for this purpose we use the age target group as the exclusion restriction. Indeed, the collection of users' personal data might be influenced by the Google parental control system which provides age guidelines based on app control. 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". They include apps targeting children and teens. We use the dummy variable *Everyone* as the exclusion restriction which takes the value 1 if the app is 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 (Cecere *et al.*, 2018). The underlying argument regarding the exclusion restriction is that the target group Everyone is likely to negatively influence the probability to collect data, but not the choice of ads and in-app purchases. Thus, the use of advertising or more traditional monetary transactions is less likely to be correlated to the (targeted) age group. We also compare the identification approaches with and without the exclusion restriction (see appendix Table 13). The model with the exclusion restriction is favored by the AIC and BIC criteria.

Electronic copy available at: https://ssrn.com/abstract=3136661

4 Estimation of the monetization strategies

Tables 6 and 7 present the results of the multivariate probit estimations. The rho values suggest strong unobserved correlations among the error terms of the *Advertising*, *In-app purchases* and *Personal data* variables which suggests that a multivariate is the appropriate estimation model.[13] 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 with no correction for endogeneity, estimates biased coefficients and justifies the choice of a multivariate probit. It means that collecting personal data (*Personal data*) can be understood as a separate monetization strategy and when studying app monetization should be taken account of systematically.

4.1 Main results

Table 6 presents the results of the main estimations. Column (1) presents the probabilities of *Advertising*; column (2) presents the probabilities of *In-app purchases*; column (3) reports the coefficients of the probability to collect *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 the regressions include the fixed effects for the Google Play Store categories and developer characteristics. To interpret the coefficients, we refer to the average partial effect. The results indicate that apps collecting *Personal data* have 15.7% of probability of using an *Advertising* strategy. This is in line with the traditional economics of privacy which consider collection of personal data as enabling personalized advertising. These strategies seem to be complementary. Conversely, the use of personal data is likely to reduce the probability of *In-app purchases* by 3.1% suggesting a substitution effect between *Personal data* and *In-app purchases*.

[13] Appendix Table 17 presents the three estimations independent probit estimations.

Electronic copy available at: https://ssrn.com/abstract=3136661

Monetization strategies are likely to be linked to download intensity. Most downloaded apps categories (more than 100 million downloads) are especially interesting as they increase the likelihood of collecting personal data by more than 20%. The signs of the coefficients confirm the intuition based on the graphical evidence that apps with more than 1 million downloads are likely to collect personal data, while those with less than 1 million downloads are likely to rely more on an advertising strategy. The literature highlights that the value of personal data is associated to the collection of huge amounts of data (OECD, 2013) which is in line with our findings. Also, the probability of *In-app purchases* increases in magnitude with the number of downloads (from 10,000 and 1K million categories).

Developers using *social networking* third parties are 17.7% more likely to use *Advertising*, 7.1% more likely to adopt *In-app purchases*, and 20.9% more likely to collect *Personal data*. This result suggests that there is a link between third parties and personal data. We also examined developer characteristics such as *Developer website* which is a measure of developer specialization. The indicator for the presence of a developer website decreases the probability of advertising by 3.4% but increases the probabilities of in-app purchases by 4.4% and personal data by 3.7%. The results indicate that less professional developers use the monetization strategy of advertising, while professional developers are more likely to choose personal data and in-app purchases.

Although the main focus of our estimation is the relationship between monetization strategies, the coefficients of the developer characteristics are interesting. Table 7 presents Google category coefficients and the average partial effects. These variables measures the competition and innovation among categories. In particular, belonging to *Communication* and *Business* categories increase the probability of collecting per-

Electronic copy available at: https://ssrn.com/abstract=3136661

sonal data respectively by 21.1% and 10.1% compared to the *Game* category (reference variable). Apps in *Lifestyle* category also are likely to rely on personal data and they are less likely of doing advertising and in-app purchases. The users' data collected in *Lifestyle* category might be valuable in the data broker market as they might collect users' health related behavior and hobbies.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 6: Multivariate probit estimations and Average partial effects with Advertising, Integrated Purchases and Personal data

	Estimations			Average Partial Effects		
Variable	(1) Advertising	(2) In-app purchases	(3) Personal data	(4) Advertising	(5) In-app purchases	(6) Personal data
Personal data	0.452***	-0.239***		0.157	-0.031	
	(0.053)	(0.046)				
Number installed 1K-5K	0.209***	0.213***	-0.128***	0.069	0.032	-0.024
	(0.009)	(0.017)	(0.012)			
Number installed 5K-10K	0.144***	0.153***	-0.110***	0.048	0.023	-0.021
	(0.010)	(0.017)	(0.012)			
Number installed 10K-50K	0.414***	0.368***	-0.102***	0.142	0.059	-0.019
	(0.011)	(0.018)	(0.016)			
Number installed 50K-100K	0.293***	0.299***	-0.133***	0.1	0.048	-0.025
	(0.012)	(0.020)	(0.016)			
Number installed 100K-500K	0.544***	0.627***	0.038**	0.191	0.119	0.007
	(0.015)	(0.021)	(0.019)			
Number installed 500K-1 million	0.515***	0.496***	-0.047**	0.18	0.089	-0.009
	(0.015)	(0.021)	(0.019)			
Number installed 1-5 million	0.630***	0.956***	0.224***	0.223	0.215	0.047
	(0.025)	(0.030)	(0.028)			
Number installed 5-10 million	0.596***	0.782***	0.154***	0.21	0.164	0.032
	(0.023)	(0.028)	(0.028)			
Number installed 10-50 million	0.629***	1.174***	0.484***	0.222	0.287	0.111
	(0.056)	(0.059)	(0.059)			
Number installed 50-100 million	0.740***	1.086***	0.334***	0.262	0.258	0.073
	(0.046)	(0.053)	(0.051)			
Number installed 100-500 million	0.253	1.298***	0.882***	0.086	0.33	0.222
	(0.169)	(0.169)	(0.208)			
Number installed 500-1000 million	0.569***	1.295***	1.015***	0.201	0.329	0.263
	(0.153)	(0.150)	(0.196)			
Number installed 1K-5K million	-4.573***	-3.936***	0.947***	-0.325	-0.09	0.242
	(0.139)	(0.148)	(0.166)			
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)			
Everyone			-1.074***			-0.233
			(0.014)			
Constant	-0.423***	-1.683***	-0.747***			
	(0.020)	(0.031)	(0.028)			
Developer characteristics	YES	YES	YES	YES	YES	YES
Apps category fixed effects	YES	YES	YES	YES	YES	YES
Log pseudolikelihood	-5.65e+05					
LR test chi2(3)	781.465					
ρ_{AI}	-0.045***	(0.011)				
ρ_{AP}	-0.144***	(0.031)				
ρ_{IP}	0.254***	(0.026)				
Number of draws	683					
Observations	475 787					

Notes: Recursive multivariate probit estimations. Columns (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app* and *Personal Data*. Columns (4) to (6) estimate the average partial effect respectively *Advertising*, *In-app purchases* and *Personal Data*. *Everyone* is the exclusion restriction variable. The coefficients of the app category variables are provided in Table 7. The omitted Google category is *Game*. The omitted number of installed download is *Number installed less than 1000 downloads*. Robust standard errors in parentheses are clustered at developer level. The full set of coefficients is available from the authors. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 7: Table 6 (continued) multivariate probit estimations with application category fixed effects

	Estimations			Average Partial Effects		
Apps' categories (Ref: Games)	(1) Advertising	(2) In-app purchases	(3) Personal data	(4) Advertising	(5) In-app purchases	(6) Personal 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 presents application category fixed effects of the recursive multivariate probit. Columns (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app purchases*, and *Personal Data*. Columns (4) to (6) estimate the average partial effects respectively for *Advertising*, *In-app*, and *Personal Data*. Robust standard errors in parentheses are clustered at developer level. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

4.2 Robustness checks

We check the robustness of our estimations to alternative dependent variables. First, we estimate the multivariate probit using the dependent variables: *Admob*, *In-app purchases* and *Personal data*. *Admob* is a dummy variable that measures whether the app uses Admob as an ad third party. It is used as an alternative measure of advertising as monetization strategy. Second, we estimate the multivariate probit using the dependent variables *Advertising*, *In-app purchases*, and *Badgrade*.[14] *Badgrade* measures whether the app uses personal data as monetization strategy. This binary variable is an alternative measures of users' data collection. Third, we estimate the main equation on two-subsamples of apps in the *Game* and the combination of *Health & Fitness* and *Lifestyle* categories.

4.2.1 Estimations with advertising third parties: Admob

Table 8 (and appendix Table 14) presents the model estimating the joint probability of using *Admob, In-app purchases* and *Personal data* where the binary variable *Admob* is an alternative measure of advertising as a monetization strategy. This addresses empirical concerns that Admob, the largest ad company in the group of ad third parties, might be driving our results. In particular, we measure whether the magnitude of the average partial effect of the variable personal data changes. We find that apps that collect personal data have an increasing probability of using Admob (7.4%) which results in a smaller coefficient compared to the main regression (15.7%). The other results are consistent with the main estimations.

[14]The binary variable *Badgrade* indicates an app received a grading between B and D.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 8: Multivariate probit estimations and Average partial effects with Admob, Integrated Purchases and Personal data

	Estimations			Average partial effects		
Variable	(1) Admob	(2) In-app purchases	(3) Personal data	(4) Admob	(5) In-app purchases	(6) Personal data
Personal data	0.223***	-0.239***		0.074	-0.030	
	(0.051)	(0.046)				
Number installed 1K-5K	0.174***	0.212***	-0.128***	0.056	0.032	-0.024
	(0.010)	(0.017)	(0.012)			
Number installed 5K-10K	0.118***	0.152***	-0.110***	0.038	0.022	-0.0207
	(0.010)	(0.017)	(0.012)			
Number installed 10K-50K	0.361***	0.367***	-0.102***	0.121	0.059	-0.019
	(0.012)	(0.018)	(0.016)			
Number installed 50K-100K	0.250***	0.299***	-0.134***	0.083	0.048	-0.025
	(0.012)	(0.020)	(0.016)			
Number installed 100K-500K	0.434***	0.626***	0.037*	0.150	0.118	0.007
	(0.015)	(0.021)	(0.019)			
Number installed 500K-1 million	0.437***	0.495***	-0.047**	0.151	0.089	-0.009
	(0.015)	(0.021)	(0.020)			
Number installed 1-5 million	0.422***	0.955***	0.223***	0.146	0.214	0.046
	(0.025)	(0.030)	(0.029)			
Number installed 5-10 million	0.437***	0.780***	0.153***	0.152	0.163	0.031
	(0.024)	(0.028)	(0.028)			
Number installed 10-50 million	0.300***	1.174***	0.484***	0.102	0.287151	0.110
	(0.056)	(0.059)	(0.059)			
Number installed 50-100 million	0.427***	1.086***	0.334***	0.148	0.257	0.072
	(0.046)	(0.053)	(0.051)			
Number installed 100-500 million	-0.110	1.296***	0.883***	-0.033	0.329	0.222
	(0.203)	(0.168)	(0.208)			
Number installed 500-1K million	0.048	1.295***	1.013***	0.015	0.328	0.262
	(0.164)	(0.150)	(0.197)			
Number installed 1K-5K million	-4.308***	-3.827***	0.945***	-0.280	-0.090	0.241
	(0.141)	(0.149)	(0.166)			
Playstore rating	-0.003	0.017***	-0.023***	-0.001	0.002	-0.232
	(0.003)	(0.005)	(0.003)			
Social networking	0.410***	0.423***	0.837***	0.141	0.070	0.209
	(0.023)	(0.027)	(0.018)			
Utility	0.137***	0.148***	0.241***	0.044	0.022	0.049
	(0.015)	(0.021)	(0.016)			
Developer website	-0.120***	0.351***	0.195***	-0.038	0.043	0.036
	(0.015)	(0.022)	(0.020)			
Everyone			-1.075***			0.00003
			(0.014)			
Constant	-0.535***	-1.682***	-0.751***			
	(0.020)	(0.031)	(0.028)			
Developer characteristics	YES	YES	YES	YES	YES	YES
Category fixed effects	YES	YES	YES	YES	YES	YES
Log pseudolikelihood	-5.57e+05					
LR test chi2(3)	880.895					
ρ_{AI}	-0.069***	(0.011)				
ρ_{AP}	-0.114***	(0.029)				
ρ_{IP}	0.253***	(0.026)				
Number of draws	683.000					
Observations	475,787					

Notes: Recursive multivariate probit estimations. Columns (1) to (3) estimate respectively the dependent variables *Admob*, *In-app*, and *Personal Data* . Columns (4) to (6) estimate the average partial effects of *Admob*, *In-app*, and *Personal Data*. *Everyone* is the exclusion restriction variable. The coefficients of the app category variables are provided in appendix Table 14. The omitted Google category is *Game*. The omitted number of installed download is *Number installed less than 1000 downloads.* Robust standard errors in parentheses are clustered at developer level. The full set of coefficients is available from the authors. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

4.2.2 Estimations using Badgrade as the dependent variable

Table 9 reports the estimation of the main equations where we use *Badgrade* to measure users' data collection (see also Table 15). This empirical strategy allows us to estimate whether a more 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. Also, we calculate the multivariate probit with *Advertising, In-app purchases, more than 6 permissions* as the dependent variable, the results are presented in Table 16.

The effects of *Badgrade* on the probabilities of advertising and in-app purchases are consistent with the previous estimations using the variable *Personal data*. The main results still hold. However, it should be underlined that apps in the top downloaded category 1K-5K million are not likely to have a *Badgrade* while in the main regression the sign was positive (see Table 6). This suggests that a given app might receive a favorable grade because Privacy Grade did not yet graded all permissions. Our variable *Personal data* allows to take in account that by looking to the number of dangerous permissions required.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 9: Multivariate probit estimations and Average partial effects with Advertising, Integrated Purchases and Badgrade

	Estimations			Average partial effects		
Variable	(1) Advertising	(2) In-app purchases	(3) Badgrade	(4) Advertising	(5) In-app purchases	(6) Badgrade
Badgrade	1.346***	-0.297***		0.474	-0.036	
	(0.077)	(0.089)				
Number installed 1K-5K	0.218***	0.214***	-0.108***	0.069	0.032	-0.013
	(0.009)	(0.017)	(0.015)			
Number installed 5K-10K	0.156***	0.152***	-0.118***	0.049	0.023	-0.014
	(0.009)	(0.017)	(0.014)			
Number installed 10K-50K	0.412***	0.371***	-0.033*	0.134	0.060	-0.004
	(0.011)	(0.019)	(0.019)			
Number installed 50K-100K	0.299***	0.301***	-0.088***	0.097	0.048	-0.011
	(0.011)	(0.020)	(0.019)			
Number installed 100K-500K	0.523***	0.629***	0.100***	0.175	0.119	0.013
	(0.016)	(0.021)	(0.023)			
Number installed 500K-1 million	0.506***	0.499***	0.015	0.170	0.089	0.002
	(0.015)	(0.021)	(0.023)			
Number installed 1-5 million	0.580***	0.954***	0.264***	0.197	0.214	0.039
	(0.026)	(0.030)	(0.032)			
Number installed 5-10 million	0.552***	0.783***	0.219***	0.187	0.164	0.032
	(0.024)	(0.028)	(0.033)			
Number installed 10-50 million	0.550***	1.163***	0.417***	0.186	0.283	0.068
	(0.057)	(0.060)	(0.061)			
Number installed 50-100 million	0.667***	1.084***	0.363***	0.228	0.257	0.057
	(0.048)	(0.053)	(0.056)			
Number installed 100-500 million	0.101	1.284***	0.737***	0.032	0.325	0.139
	(0.169)	(0.174)	(0.189)			
Number installed 500-1K million	0.396**	1.285***	0.869***	0.132	0.325	0.173
	(0.155)	(0.152)	(0.202)			
Number installed 1K-5K million	-4.271***	-4.095***	-4.229***	-0.324	-0.090	-0.093
	(0.150)	(0.165)	(0.169)			
Playstore rating	0.001	0.016***	-0.033***	.0002	0.002	-0.004
	(0.003)	(0.005)	(0.003)			
Social networking	0.302***	0.441***	0.936***	0.099	0.074	0.178
	(0.026)	(0.034)	(0.020)			
Utility	0.197***	0.141***	0.193***	0.063	0.020	0.027
	(0.014)	(0.021)	(0.020)			
Developer website	-0.099***	0.343***	0.084***	-0.031	0.042	0.010
	(0.015)	(0.021)	(0.024)			
Everyone			-0.814***			-0.111
			(0.019)			
Constant	-0.520***	-1.666***	-0.867***			
	(0.021)	(0.035)	(0.031)			
Developer characteristics	YES	YES	YES	YES	YES	YES
Category fixed effects	YES	YES	YES	YES	YES	YES
Log pseudolikelihood	-5.10e+05					
LR test chi2(3)	1233.465					
ρ_{AI}	-0.081***	(0.013)				
ρ_{AP}	-0.393***	(0.042)				
ρ_{IP}	0.319***	(0.048)				
Number of draws	683					
Observations	475,787					

Notes: Recursive multivariate probit estimations. Columns (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app*, and *Badgrade*. Columns (4) to (6) estimate the average partial effects respectively *Advertising*, *In-app* and, *Badgrade*. *Everyone* is the exclusion restriction variable. The coefficients of the app category variables are presented in appendix Table 15. The omitted Google category is *Game*. The omitted number of installed download is *Number installed less than 1000 downloads*. Robust standard errors in parentheses are clustered at developer level. The full set of coefficients is available from the authors. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

4.2.3 Estimations on subsamples

We estimate the main equation on the subsample of apps in *Games* category because it is the largest category in the Google Play Store. Columns 1 to 3 of Table 10 present the estimation results. In the case of the *Games* category, *Personal data* is associated positively to integrated purchase compared to a negative relationship in the main results. Also, *Playstore rating* is correlated positively to *Advertising* and *In-app purchases.* The use of social network third parties is correlated positively to all three monetization strategies while utility third parties is correlated only to advertising and personal data. In addition, we estimate the main equation by merging the *Health & Fitness* with *Lifestyle* categories. The descriptive statistics (Table 1) suggest that apps in these categories will be more likely to collect personal data compared to the other categories (apart from Games). Apps in these categories potentially can collect data on users' health and lifestyles (such as pregnancy or hours spent on a bicycle). Columns 4 to 6 of Table 10 present the estimations. Number of downloads is negatively correlated to personal data which suggests that small amounts of data collected by the apps in these categories may be used commercially and have some intrinsic value. Both the Utility and Social network third parties are likely to be correlated to all the monetization strategies which confirms our main findings.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 10: Multivariate probit estimations for the categories *Games*, *Health & Fitness* and *Lifestyle*.

	Games all			Health & Fitness and Lifestyle		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Advertising	In-app purchases	Personal data	Advertising	In-app purchases	Personal data
Personal data	0.311***	0.798***		0.347***	-0.533***	
	(0.051)	(0.043)		(0.053)	(0.044)	
Number installed 1K-5K	0.214***	0.216***	0.108***	0.218***	0.249***	-0.408***
	(0.013)	(0.018)	(0.018)	(0.020)	(0.031)	(0.022)
Number installed 5K-10K	0.193***	0.144***	0.000	0.156***	0.120***	-0.252***
	(0.017)	(0.023)	(0.024)	(0.025)	(0.041)	(0.027)
Number installed 10K-50K	0.324***	0.391***	0.232***	0.470***	0.352***	-0.476***
	(0.014)	(0.017)	(0.018)	(0.021)	(0.033)	(0.026)
Number installed 50K-10K	0.250***	0.320***	0.122***	0.306***	0.291***	-0.476***
	(0.018)	(0.022)	(0.024)	(0.027)	(0.041)	(0.032)
Number installed 100K-500K	0.453***	0.656***	0.370***	0.613***	0.500***	-0.309***
	(0.017)	(0.020)	(0.020)	(0.030)	(0.042)	(0.038)
Number installed 500K-1 million	0.439***	0.560***	0.287***	0.621***	0.421***	-0.410***
	(0.020)	(0.023)	(0.024)	(0.032)	(0.047)	(0.043)
Number installed 1-5 million	0.638***	0.976***	0.443***	0.662***	0.920***	-0.207***
	(0.028)	(0.028)	(0.029)	(0.067)	(0.075)	(0.078)
Number installed 5-10 million	0.581***	0.794***	0.475***	0.684***	0.642***	-0.294***
	(0.030)	(0.030)	(0.032)	(0.066)	(0.085)	(0.081)
Number installed 10-50 million	0.825***	1.155***	0.519***	0.473**	1.661***	0.261
	(0.070)	(0.067)	(0.068)	(0.236)	(0.225)	(0.259)
Number installed 50-100 million	0.868***	1.086***	0.511***	0.766***	0.894***	-0.522***
	(0.062)	(0.057)	(0.059)	(0.171)	(0.194)	(0.189)
Number installed 100-500 million	0.642**	1.547***	0.469*	-3.508***	-2.934***	-3.096***
	(0.289)	(0.356)	(0.276)	(0.223)	(0.211)	(0.215)
Number installed 500-1K million	1.573***	0.953***	0.708***			
	(0.389)	(0.258)	(0.254)			
Playstore rating	0.016***	0.115***	0.003	-0.007	0.007	-0.038***
	(0.005)	(0.008)	(0.006)	(0.005)	(0.008)	(0.005)
Social networking	0.743***	0.485***	0.692***	0.561***	0.238***	0.894***
	(0.016)	(0.016)	(0.014)	(0.026)	(0.031)	(0.018)
Utility	0.401***	-0.018	0.155***	0.212***	0.277***	0.147***
	(0.013)	(0.014)	(0.014)	(0.017)	(0.025)	(0.019)
Developer website	0.142***	0.420***	0.112***	-0.132***	0.348***	0.301***
	(0.010)	(0.013)	(0.013)	(0.016)	(0.030)	(0.022)
Everyone			-0.738***			-1.367***
			(0.011)			(0.019)
Constant	-0.766***	-2.263***	-1.119***	-0.846***	-2.237***	-0.405***
	(0.020)	(0.035)	(0.029)	(0.026)	(0.047)	(0.028)
Dvp characteristics	YES	YES	YES	YES	YES	YES
Log pseudolikelihood	-1.27e+05			-5.22e+04		
LR test chi2(3)	270.645			80.993		
ρ_{AI}	-0.101***	(0.007)		-0.005	(0.014)	
ρ_{AP}	-0.034	(0.029)		-0.047	(0.034)	
ρ_{IP}	-0.161***	(0.024)		0.271***	(0.026)	
Number of draws	300			213		
Observations	90960			45346		

Notes: Columns (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app*, and *Personal data* for the subsample of apps in the games category. Columns (4) to (6) estimate respectively the dependent variables *Advertising*, *In-app*, and *Personal Data* for the subsample of apps in the Health and Lifestyle category. *Everyone* is the exclusion restriction variable. The omitted number of installed download is *Number installed less than 1000 downloads*. Robust standard errors in parentheses are clustered at developer level. The full set of coefficients is available from the authors. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

5 Discussions

This paper has economic and managerial implications for developers, platforms, business analytics and regulators. Developers need a better understanding of the market and what matters 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 adapt existing ones. Personal data may be required to both enable app functionality and as a monetization strategy. While the collection of personal data might have a negative impact on demand (Kummer and Schulte, 2018), we show that it favors the development of free services such as free apps.

How developers obtain revenue is a critical issue for digital platforms such as Google Play Store which need to encourage the entry of new innovative developers, and increase their visibility. In particular, business analytics should consider the role of third parties to identify effective strategies. We highlight that the presence of third parties is important for the provision of enhanced services and features which promote the creation of new innovative companies.

We show that the app economy and the third party economy are evolving markets which should be considered together. Our paper underlines this link by highlighting the different strategies of monetization. More precisely, we find that the link between third parties and personal data is likely to be positive. As shown by Razaghpanah *et al.* (2018) third parties can gain access to 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 the personal data self-regulatory approach, we need a more thorough investiga-

Electronic copy available at: https://ssrn.com/abstract=3136661

tion of how developers and platforms could improve their transparency through their permission systems by encouraging developers to declare which third parties they use. Our paper provides a preliminary examination of the third party market related to apps. The descriptive evidences show that while developers can choose among multiple third parties, market share seems to be highly concentrated in a few third parties which are chosen by the majority of apps. Since the third party market includes some dominant players, this raises concern over competition policy. User data could be concentrated among a few actors at different levels. This raises questions about the market power of killer apps and the competition dynamics related to platforms.

6 Conclusion

From a managerial perspective, the mobile app economy is characterized by the pace of innovation and reduced barriers to the market entry of new developers (Davis *et al.*, 2016), with the majority of apps "sold" at zero price, especially in the Google Play Store. We show that the apps and the third parties are co-evolving markets and should be considered together.

Our approach differs from that taken in previous empirical investigations; we focused on the monetization strategies of developers, and analyzed how personal data are combined with more traditional monetization strategies, such as advertising and in-app purchases. First, our results suggest that overall, personal data are used to monetize applications thus favoring the development of free services. Second, we show that the monetization strategy differs 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 with more specialized developers, and can be applied once the app achieves a certain level of market power. Third, we find that

Electronic copy available at: https://ssrn.com/abstract=3136661

the monetization strategy also depends non-linearly on the download category. Apps with a large number of downloads collect personal data without the use of ads or in-app purchases strategies. We held informal discussions with professional developers. They corroborate that the collection of personal data is associated with data management investment and they confirm that users' data are usually commercialized when there are large amount of data. Health data are an exception, they can be commercialized also in relatively small quantities. Fourth, we also observed that the use of social networking third parties increases the probability of using personal data.

Our study has some limitations. First, our results should be interpreted with caution since we use only cross-sectional data, and thus can estimate only correlations not causality. 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 to calculate the thresholds where strategies might change dramatically.

Electronic copy available at: https://ssrn.com/abstract=3136661

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.

Bergemann, D., Bonatti, A. *et al.* (2018). *Markets for Information: An Introduction*. Technical report. Cowles Foundation for Research in Economics, Yale University.

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

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

Cappellari, L., Jenkins, S. P. *et al.* (2003). Multivariate probit regression using simulated maximum likelihood. *The Stata Journal*. 3(3), 278–294.

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

Cecere, G., Le Guel, F., Lefrere, V., Tucker, C. and Yin, P. (2018). Child apps, personal data regulation and home-country compliance. Working paper.

Chellappa, R. K. and Shivendu, S. (2010). Mechanism design for "free" but "no free disposal" services: The economics of personalization under privacy concerns. *Management Science*. 56(10), 1766–1780.

Comino, S., Manenti, F. M. and Mariuzzo, F. (2018). *To upgrade or not to upgrade? the release of new versions to survive in the hypercompetitive app market*. 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.

Deng, Y., Lambrecht, A. and Liu, Y. (2018). *Spillover Effects and Freemium Strategy in Mobile App Market*. Working paper.

Ershov, D. (2017). *The Effect of Consumer Search Costs on Entry and Quality in the Mobile App Market*. Working Paper.

Fairlie, R. W. (2006). The personal computer and entrepreneurship. *Management Science*. 52(2), 187–203.

Electronic copy available at: https://ssrn.com/abstract=3136661

Filippini, M., Greene, W. H., Kumar, N. and Martinez-Cruz, A. L. (2018). A note on the different interpretation of the correlation parameters in the Bivariate Probit and the Recursive Bivariate Probit. *Economics Letters*. 167, 104–107.

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.

Gupta, S. (2013). For Mobile Devices, Think Apps, Not Ads. *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. (2018). When private information settles the bill: Money and privacy in Google's market for smartphone applications. *Management Science, forthcoming*.

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.

Electronic copy available at: https://ssrn.com/abstract=3136661

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

Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C. and Gill, P. (2018). Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In *Network and Distributed Systems Security (NDSS)*.

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.

Electronic copy available at: https://ssrn.com/abstract=3136661

7 Appendix

Table 11: Permissions and Google group of permissions

Name	Purpose	Google group of permissions
Access fine location	Precise location (gps and network-based)	Location
Access coarse location	Approximate location (network-based)	Location
Use credentials	Use accounts on the device	Accounts
Voicemail	Add voicemail	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
Bind device admin	Interact with a device admin	Network
Internet	Full network access	Network
Vpn service	Bind to a vpn service	Network
Nfc service	Bind to a nfc service	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
Phone states	Read precise phone states	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
Tablet	Permanently disable tablet	System tools
Read user dictionary	Read terms you added to the dictionary	User dictionary

Notes: This Table presents the 55 dangerous permissions classified by Google, by name, purpose and group of the permissions.

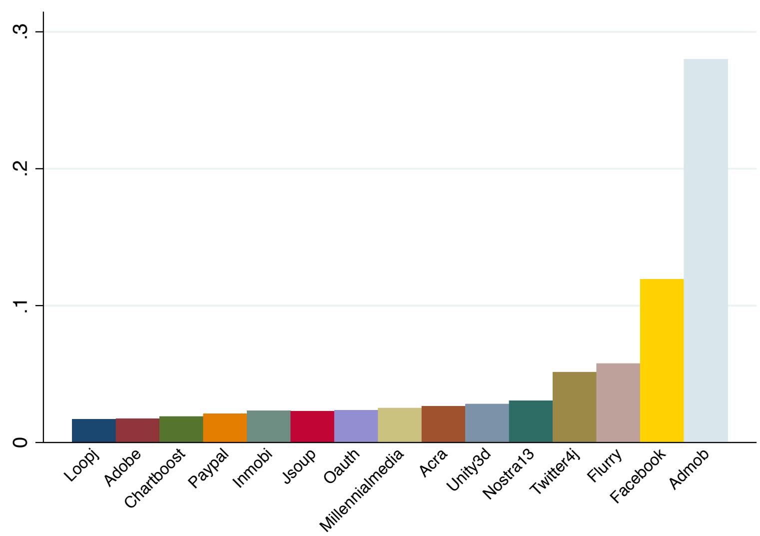
Electronic copy available at: https://ssrn.com/abstract=3136661

Table 12: Use of permissions Google group by strategy of monetization

Advertising		In app Purchases		Personal Data		None	
Permissions	%	Permissions	%	Permissions	%	Permissions	%
Network	99,8%	Network	96,6%	Network	99,4%	Network	79,0%
Storage	55,4%	Storage	74,1%	Storage	90,2%	Storage	45,2%
Phone calls	40,6%	Phone calls	45,6%	Phone calls	88,7%	Phone calls	24,4%
Location	28,7%	Accounts	33,0%	Location	73,4%	Location	17,0%
Accounts	16,0%	Location	23,8%	Accounts	50,6%	Accounts	9,9%
Camera	12,7%	App info	12,4%	Camera	43,8%	Camera	7,8%
Microphone	8,5%	Camera	9,0%	Microphone	29,1%	App info	3,6%
Messages	6,4%	Social info	6,6%	Messages	26,5%	Microphone	2,6%
App info	5,7%	Microphone	6,1%	Social info	25,8%	Display	2,4%
Social info	4,1%	Messages	4,2%	App info	20,6%	Social info	2,2%
Display	3,3%	Display	3,6%	Personal info	9,5%	Messages	1,6%
Personal info	1,8%	Screenlock	2,8%	Bluetooth network	8,0%	Bluetooth network	1,5%
Bluetooth network	1,6%	Bluetooth network	2,6%	Display	7,2%	System tools	1,1%
System tools	1,6%	Personal info	2,4%	System tools	6,5%	Screenlock	1,1%
Screenlock	1,3%	System tools	1,4%	Screenlock	5,9%	Personal info	0,9%
Affects battery	0,2%	Affects battery	0,7%	Affects battery	1,2%	Affects battery	0,4%
User dictionary	0,1%	User dictionary	0,2%	User dictionary	0,2%	User dictionary	0,2%
Signature	0,0%	Bookmarks	0,1%	Bookmarks	0,2%	Signature	0,0%
Bookmarks	0,0%	Signature	0,0%	Signature	0,1%	Bookmarks	0,0%

Notes: This Table presents the use of permissions by monetization strategy.

Figure 3: Distribution of top 15 third parties



Source: Drawn up by the author

Electronic copy available at: https://ssrn.com/abstract=3136661

Figure 4: Screen shot of Privacy Grade permissions

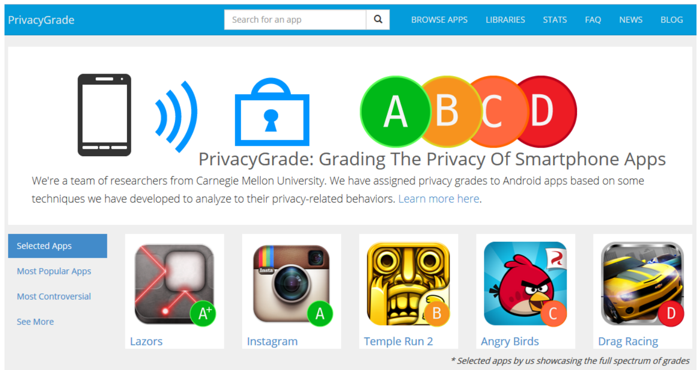


Table 13: Fitting test with and without exclusions restriction

Model	Obs	ll(model)	df	AIC	BIC
Without exclusion restriction	475,787	-587131.2	135	1174532	1176027
With Everyone	475,787	-564523.3	138	1129323	1130851

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 14: Table 8 (continued) multivariate probit estimations for application category fixed effects

	Estimations			Average partial effects		
Variable	(1) Admob	(2) In-app purchases	(3) Personal data	(4) Admob	(5) In-app purchases	(6) Personal data
Following Table 8	[...]	[...]	[...]	[...]	[...]	[...]
Books and reference	-0.037	-0.545***	-0.356***	-0.011	-0.057	-0.061
	(0.041)	(0.059)	(0.053)			
Business	-0.559***	-0.798***	0.456***	-0.150	-0.073	0.102
	(0.023)	(0.041)	(0.027)			
Comics	-0.114	-0.495***	-0.302**	-0.035	-0.051	-0.052
	(0.072)	(0.096)	(0.136)			
Communication	-0.614***	-0.642***	0.855***	-0.159	-0.062	0.212
	(0.031)	(0.037)	(0.039)			
Education	-0.292***	-0.386***	-0.042	-0.086	-0.045	-0.008
	(0.023)	(0.036)	(0.033)			
Entertainment	-0.159***	-0.744***	-0.102***	-0.048	-0.071	-0.019
	(0.018)	(0.037)	(0.026)			
Finance	-0.507***	-0.847***	0.152***	-0.137	-0.073	0.031
	(0.027)	(0.037)	(0.035)			
Health and fitness	-0.306***	-0.563***	0.016	-0.088	-0.057	0.003
	(0.025)	(0.055)	(0.032)			
Libraries and demo	-0.897***	-0.893***	-0.060	-0.204	-0.072	-0.011
	(0.072)	(0.103)	(0.086)			
Lifestyle	-0.367***	-0.771***	0.144***	-0.105	-0.073	0.029
	(0.020)	(0.030)	(0.026)			
Media and video	-0.357***	-0.817***	0.165*	-0.101	-0.070	0.034
	(0.037)	(0.042)	(0.093)			
Medical	-0.532***	-0.466***	0.007	-0.141	-0.049	0.001
	(0.040)	(0.067)	(0.043)			
Music and audio	-0.186***	-0.899***	0.043	-0.056	-0.076	0.008
	(0.038)	(0.051)	(0.055)			
News and magazines	-0.205***	-0.144**	-0.118***	-0.061	-0.018	-0.022
	(0.032)	(0.057)	(0.042)			
Personalization	-0.684***	-0.980***	-0.130**	-0.175	-0.080	-0.024
	(0.055)	(0.069)	(0.059)			
Photography	-0.383***	-0.502***	-0.135***	-0.108	-0.052	-0.025
	(0.034)	(0.042)	(0.049)			
Productivity	-0.452***	-0.483***	0.216***	-0.125	-0.051	0.045
	(0.022)	(0.039)	(0.031)			
Shopping	-0.462***	-1.282***	0.063	-.126	-.084	0.012
	(0.038)	(0.052)	(0.057)			
Social	-0.362***	-0.659***	0.050	-0.103	-0.063	0.009
	(0.026)	(0.047)	(0.038)			
Sports	-0.196***	-0.583***	-0.046	-0.059	-0.058	-0.008
	(0.032)	(0.048)	(0.040)			
Tools	-0.298***	-0.702***	0.081***	-0.087	-0.069	0.016
	(0.016)	(0.022)	(0.021)			
Transportation	-0.384***	-0.643***	-0.039	-0.108	-0.061	-0.007
	(0.033)	(0.067)	(0.038)			
Travel and local	-0.376***	-0.526***	-0.149***	-0.107	-0.055	-0.027
	(0.035)	(0.078)	(0.036)			
Weather	-0.098	-0.553***	-0.093	-0.030	-0.055	-0.017
	(0.066)	(0.082)	(0.125)			

Notes: This Table presents application category fixed effects for the recursive multivariate probit. Columns (1) to (3) estimate respectively the dependent variables *Admob*, *In-app*, and *Personal Data*. Columns (4) to (6) estimate the average partial effects respectively for *Admob*, *In-app*, and *Personal Data*. Standard errors in parentheses. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 15: Table 9 (continued) multivariate probit estimation for application category fixed effects with *Advertising*, *In-app purchases* and *Badgrade* dependent variables.

	Estimations			Average partial effects		
Variable	(1) Advertising	(2) In-app purchases	(3) Badgrade	(4) Advertising	(5) In-app purchases	(6) Badgrade
Following Table 9	[...]	[...]	[...]	[...]	[...]	[...]
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***	-0.182	-0.076	-0.026
	(0.021)	(0.039)	(0.033)			
Comics	-0.155**	-0.509***	-0.473***	-0.046	-0.053	-0.048
	(0.068)	(0.096)	(0.124)			
Communication	-0.616***	-0.732***	-0.523***	-0.162	-0.068	-0.052
	(0.029)	(0.034)	(0.039)			
Education	-0.331***	-0.405***	-0.384***	-0.096	-0.0472	-0.0428
	(0.023)	(0.036)	(0.037)			
Entertainment	-0.191***	-0.760***	-0.343***	-0.057	-0.073	-0.0389
	(0.018)	(0.037)	(0.028)			
Finance	-0.556***	-0.894***	-0.690***	-0.149	-0.075	-0.064
	(0.026)	(0.036)	(0.043)			
Health and fitness	-0.372***	-0.584***	-0.338***	-0.105	-0.059	-0.038
	(0.026)	(0.055)	(0.038)			
Libraries and demo	-0.888***	-0.922***	-0.620***	-0.210	-0.074	-0.0584
	(0.082)	(0.103)	(0.080)			
Lifestyle	-0.438***	-0.801***	-0.292***	-0.123	-0.075	-0.034
	(0.020)	(0.030)	(0.030)			
Media and video	-0.417***	-0.849***	-0.387***	-0.116	-0.072	-0.041
	(0.034)	(0.043)	(0.045)			
Medical	-0.573***	-0.499***	-0.519***	-0.152	-0.052	-0.052
	(0.040)	(0.068)	(0.059)			
Music and audio	-0.246***	-0.917***	-0.253***	-0.072	-0.077	-0.030
	(0.037)	(0.050)	(0.069)			
News and magazines	-0.219***	-0.156***	-0.272***	-0.065	-0.020	-0.031
	(0.030)	(0.057)	(0.049)			
Personalization	-0.721***	-0.991***	-0.390***	-0.185	-0.082	-0.043
	(0.052)	(0.069)	(0.070)			
Photography	-0.421***	-0.528***	-0.540***	-0.117	-0.054	0-.054
	(0.035)	(0.042)	(0.059)			
Productivity	-0.503***	-0.517***	-0.485***	-0.137	-0.055	-0.050
	(0.025)	(0.039)	(0.050)			
Shopping	-0.531***	-1.325***	-0.581***	-0.143	-0.086	-0.056
	(0.036)	(0.052)	(0.053)			
Social	-0.393***	-0.712***	-0.659***	-0.110	-0.066	-0.061
	(0.026)	(0.047)	(0.037)			
Sports	-0.256***	-0.603***	-0.344***	-0.074	-0.060	-0.038
	(0.031)	(0.048)	(0.045)			
Tools	-0.339***	-0.728***	-0.519***	-0.098	-0.072	-0.053
	(0.016)	(0.022)	(0.026)			
Transportation	-0.435***	-0.690***	-0.658***	-0.121	-0.065	-0.061
	(0.032)	(0.067)	(0.058)			
Travel and local	-0.468***	-0.557***	-0.564***	-0.129	-0.058	-0.057
	(0.037)	(0.077)	(0.048)			
Weather	-0.270***	-0.553***	-0.071	-0.078	-0.056	-0.009
	(0.058)	(0.082)	(0.144)			

Notes: This Table presents application category fixed effects of the recursive multivariate probit. Column (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app*, and *Badgrade* . Columns (4) to (6) estimate the respective average partial effects for *Advertising*, *In-app*, and *Badgrade*. Robust standard errors in parentheses are clustered at developer level. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 16: Multivariate probit estimations with *Advertising*, *Integrated Purchases* and *More than 6 permissions*

Variable	(1) Advertising	(2) In-app purchases	(3) More than 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)
Number installed 1K-5K	0.212***	0.193***	-0.221***
	(0.010)	(0.017)	(0.014)
Number installed 5K-10K	0.146***	0.137***	-0.165***
	(0.010)	(0.017)	(0.014)
Number installed 10K-50K	0.422***	0.341***	-0.273***
	(0.012)	(0.018)	(0.017)
Number installed 50K-100K	0.298***	0.275***	-0.254***
	(0.012)	(0.020)	(0.018)
Number installed 100K-500K	0.559***	0.600***	-0.169***
	(0.015)	(0.021)	(0.021)
Number installed 500K-1 million	0.525***	0.469***	-0.227***
	(0.015)	(0.021)	(0.022)
Number installed 1-5 million	0.655***	0.936***	0.003
	(0.024)	(0.030)	(0.030)
Number installed 5-10 million	0.620***	0.756***	-0.080***
	(0.023)	(0.028)	(0.031)
Number installed 10-50 million	0.658***	1.178***	0.290***
	(0.057)	(0.058)	(0.066)
Number installed 50-100 million	0.762***	1.079***	0.182***
	(0.045)	(0.053)	(0.055)
Number installed 100-500 million	0.284	1.347***	0.734***
	(0.175)	(0.166)	(0.218)
Number installed 500-1K million	0.632***	1.308***	0.588***
	(0.164)	(0.149)	(0.160)
Number installed 1K-5K 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)
Developer characteristics	YES	YES	YES
Category fixed effects	YES	YES	YES
Log pseudolikelihood	-5.22e+05		
LR test chi2(3)	1631.910		
ρ_{AI}	-0.052***	(0.011)	
ρ_{AP}	-0.321***	(0.030)	
ρ_{IP}	0.335***	(0.025)	
Number of draws	683.000		
Observations	475,787		

Notes: Recursive multivariate probit estimations. Columns (1) to (3) estimate respectively the dependent variable *Advertising*, *In-app*, and *More than 6 permissions* . *Everyone* is the exclusion restriction variable. The full set of coefficients is available from the authors. Robust standard errors in parentheses are clustered at developer level. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661

Table 17: Estimation of three different probits

	(1) Advertising	(2) In-app purchases	(3) Personal data
Personal data	0.223***	0.158***	
	(0.015)	(0.022)	
Number installed 1K-5K	0.201***	0.232***	-0.125***
	(0.009)	(0.017)	(0.012)
Number installed 5K-10K	0.138***	0.169***	-0.107***
	(0.010)	(0.017)	(0.012)
Number installed 10K-50K	0.408***	0.386***	-0.096***
	(0.011)	(0.018)	(0.016)
Number installed 50K-100K	0.286***	0.320***	-0.129***
	(0.012)	(0.020)	(0.016)
Number installed 100K-500K	0.548***	0.629***	0.043**
	(0.015)	(0.021)	(0.019)
Number installed 500K-1 million	0.512***	0.508***	-0.042**
	(0.015)	(0.021)	(0.019)
Number installed 1-5 million	0.649***	0.935***	0.227***
	(0.024)	(0.030)	(0.028)
Number installed 5-10 million	0.609***	0.768***	0.157***
	(0.023)	(0.028)	(0.028)
Number installed 10-50 million	0.671***	1.116***	0.489***
	(0.054)	(0.059)	(0.058)
Number installed 50-100 million	0.769***	1.051***	0.333***
	(0.045)	(0.052)	(0.050)
Number installed 100-500 million	0.323*	1.192***	0.871***
	(0.165)	(0.180)	(0.208)
Number installed 500-1K million	0.648***	1.177***	1.010***
	(0.151)	(0.146)	(0.195)
Number installed 1K-5K million	0.000	0.000	0.950***
	(.)	(.)	(0.164)
Playstore rating	-0.007**	0.021***	-0.023***
	(0.003)	(0.005)	(0.003)
Social networking	0.573***	0.312***	0.835***
	(0.017)	(0.024)	(0.018)
Utility	0.226***	0.117***	0.241***
	(0.015)	(0.022)	(0.016)
Developer website	-0.096***	0.340***	0.198***
	(0.015)	(0.022)	(0.020)
Everyone			-1.079***
			(0.014)
Constant	-0.401***	-1.737***	-0.737***
	(0.020)	(0.031)	(0.028)
Developer characteristics	YES	YES	YES
Category fixed effects	YES	YES	YES
Observations	475,787	475,787	475,787

Notes: Columns (1) to (3) estimate respectively the dependent variables *Advertising*, *In-app*, and *Personal data*. Robust standard errors in parentheses are clustered at developer level. The full set of coefficients is available from the authors. Significance level: $*: p < .10$, $**: p < .05$, $***: p < .01$.

Electronic copy available at: https://ssrn.com/abstract=3136661