Mobile Applications and Access to Private Data

The Supply Side of the Android Ecosystem

Reinhold Kesler*
Centre for European
Economic Research (ZEW)

Michael Kummer†
Georgia Institute of Technology &
Centre for European
Economic Research (ZEW)

Patrick Schulte‡
Centre for European
Economic Research (ZEW)

November 17, 2017

Abstract

We use data on mobile applications to analyze the costs and benefits from using privacy sensitive permissions for data collection in the market for mobile applications. We tracked 200,000 apps over four years to analyze the data collection policies of mobile app developers. We surprisingly find that single app developers are less likely to run intrusive apps, and generally obtain better privacy grades. Moreover we show that strong competition results in a small but positive pressure to produce more intrusive apps, and that a privacy policy is highly predictive of problematic levels of access to user data. Finally, intrusive data collection is most likely in apps that target the 13+ category, which might raise concerns for the protection of young consumers. We shed important new light on the possibility of collecting and monetizing user data for the supply side in the market for mobile applications. Moreover, our findings can serve as a basis for the development of a structural approach that allows a counterfactual analysis of the market without access to sensitive user data.

Keywords: Digital Economy, Mobile Applications, Supply, Privacy, Innovation.
JEL Classification Numbers: O3, L1, D62, D85, D29)

*P.O. Box Mannheim, Germany. Email:
†Corresponding Author. , U.S.A; Phone: ; Email:
‡P.O. Box Mannheim, Germany. Email:

1 Research Questions and Motivation

The value of private data has become a central theme of the (economic) public debate. Experts had issued warnings about the possibilities of tracking individuals online already for many years. While these were widely ignored before 2013, the news of the U.S. National Security Administration's (NSA) data gathering and analysis has increased public awareness and raised considerable concerns. These recently have been further increased when it became known that the Central Information Agency (CIA) was using similar methods. It might be one of the decisive challenges of our time to find the right balance between "too little" and "too much" privacy protection for new digital technologies and online markets. While the market success of new digital technologies may depend on the services' ability to collect and analyze *enough* personal information (Aziz and Telang, 2015; Goldfarb and Tucker, 2011; Johnson, 2013b), consumers might loose their trust in the market if *too much* personal data is stored (Acquisti et al., forthcoming, 2015; Miller and Tucker, 2009).[1] On the one hand, the new technologies bear enormous potential to increased welfare by facilitating better information flows, choices, and more efficiency. Yet, if too much data ends up in the wrong hands this may also imply significant societal risks, such as totalitarian structures, or persecution.[2]

We use market-transaction based evidence to study developers' strategies with respect to privacy sensitive information in this market. Specifically, we focus on predicting which types of apps, and which types of developers are more likely to request intrusive access to user information. To this end we analyze data on 300,000 smartphone applications (henceforth "apps") from almost 30,000 developers, on Google Playstore. The app-market is relevant for our question, because apps have *transformed* information exchanges in less than eight years, and they offer unseen potentials for collecting private user information at low cost. Moreover,

[1] According to the Pew Research Center, 68% of adults believe current laws to insufficiently protect individuals' online privacy (Rainie et al., 2013).

[2] Recent events (in the United States and elsewhere) show that governments continue to have strong incentives to access the data collected by new online technologies.

app market data are informative, because Android's Operating System highlights an app's ability to access private information of users. This introduces a real cost to developers, who request excessive access to user information. Users might avoid granting these excessive rights and it might generate negative publicity.

We combine this information on each app's ability to collect personal information with a rich data set on 300,000 apps from Google's Android Market. The data covers all publicly available app-specific information including each app's number of installation, its price and even each app's closest competitors. We collected the data repeatedly in 2012 (over 6 months) and once in 2014. Moreover we collected additional datawaves in 2015 and 2016. We augmented these data with information from *alexa.com, privacygrade.org* to add further background information about app providers and permissions.

We provide the first large scale and market-transaction based evidence to help understand developers' privacy strategies and how they influence app success. Our research informs the debate about optimal levels of privacy protection, and is crucial for regulating the supply side in this market. Additionally the insights from our project will serve as the foundational input to developing a structural approach to estimating the value of privacy sensitive data for app developers.

2 Contribution to the Literature

A large stream of literature studied the demand for privacy. Generally users were found to demand privacy (Marthews and Tucker, 2014; Turow et al., 2009), and demand was found to change over time (Goldfarb and Tucker, 2012), and depended on context or framing (Acquisti et al., 2013; Gross and Acquisti, 2005). Existing research on private information in app markets is based on experimental and survey data, and the estimated valuations ranged from zero to very large numbers (Beresford et al., 2012; Carrascal et al., 2013; Grossklags and Acquisti, 2007; Racherla et al., 2011; Tsai et al., 2011). Savage and Waldman (2014)

find an average valuation of $4 self-reported willingness to accept for giving away personal information that is typically shared with app developers of mobile apps, and Ghose and Han (2014) structurally estimate the demand for selected apps. More technical studies investigated the implications of certain permissions for the privacy of the device's owner (see e.g. Chia et al., 2012; Egelman et al., 2013; Sarma et al., 2012), their potential intrusiveness (e.g. Chia et al., 2012; Fahl et al., 2012), and positive effects of locally storing sensitive user data Sutanto et al. (2013).[3]

A small but growing stream of research focused on private information and supply in online markets. Preibusch and Bonneau (2013) analyze data collection policies of internet sites, and a series of related studies analyzes how privacy policies affect users of social networks or the success of targeted advertisement (Aziz and Telang, 2015; Goldfarb and Tucker, 2011; Johnson, 2013a; Tucker, 2012, 2014). [4]

Hardly any of these studies provides transaction based empirical evidence on the role of personal data in the market for mobile applications, and none of them focuses on developers' strategies to obtain private user data. We fill this gap by providing the first large scale quantitative evidence on the role of privacy sensitive information for app developers' strategies and their success in the market for mobile apps. We provide the first evidence on this question based on large and detailed panel data that covers all apps in the Android Market from 2012-2016. Our data set is unique, because we observe app ownership, and we see which permissions developers requested for each of their individual app, before they

[3]For a survey of theoretical and empirical studies on the economics of privacy, see Acquisti et al. (forthcoming).

[4]Theory models suppliers who use their knowledge about an agent's preferences to price discriminate (Acquisti and Varian, 2005; Conitzer et al., 2012; Taylor, 2004; Taylor et al., 2010; Wathieu, 2002). Alternatively, they may attempt direct marketing, which consumers my seek to avoid avoid (Hann et al., 2008; Johnson, 2013b; Taylor et al., 2010). Taken together, these models see reduced privacy as disadvantageous for consumers. However, reduced privacy may allow to provide valuable services "for free" and can create benefits for users.(Casadesus-Masanell and Hervas-Drane, 2015; Spiegel, 2013) Several papers studied the general functioning of app markets, and focused on innovation Yin et al. (2014), Davis et al. (2014), the role of bestseller ranks (Carare, 2012), or large scale promotions on sales (Askalidis, 2015; Chaudhari, 2015). Most closely related to this paper is (Kummer and Schulte, 2016, R&R at Management Science) which analyzed the role of privacy as a second currency in app markets. Their limited analysis of supply side behavior lays the foundation for this study. Finally, Casadesus-Masanell and Hervas-Drane (2015) analyzed privacy as a "second currency" in a model where suppliers compete in both price and privacy.

could be installed. This information allows us to investigate developers' strategies with regard personal information and privacy, and to which extent app success depends on personal user information. We are the first to analyze multiple apps of the same supplier and how they negotiate the trade-offs surrounding privacy sensitive information.

3 Background

Mobile Applications from Google's Play Store are both, relevant an suitable for the purposes of our study. Google's Android Operating System (Android OS) was released in 2008 and now dominates the market (in most countries). According to the International Data Corporation (IDC, 2015) the Android OS reached a market share of around 75% in 2012. Google's platform for the distribution of apps, the "*Google Play Store*." The platform, which serves as a distribution channel for apps, books, movies, music and newspapers featured approximately 400,000 apps in 2012, and 1.5 million apps in 2015. In 2015, the revenue of the mobile app store was near $40 billion and is expected to reach $100 billion in 2020 (App Annie, 2016).

Moreover, app market data are informative, because we can exploit Google's unique policy of highlighting an app's ability to access private information of users. Android's Operating System confronts all users with the complete list of rights (henceforth "permissions") that an app requests. Users must grant these permissions before installation.

Hence, the permission system is not only a central feature of the Android app ecosystem, it is also the enabling feature of this study. This system, which is specific to the Android OS, and provides the setting in which the money-for-privacy trade-off can be meaningfully studied. When setting up their app, developers can choose among standardized permissions. Examples are permissions which enable access to a user's location, communication, browsing behavior etc. Apps must declare *before* installation which permissions they use and must request the permission from the users. More precisely, the system provides a list of per-

mission names alongside a short explanation for each permission. Users must accept this list and explicitly acknowledge that they are granting these permissions to proceed with the installation. Users can also cancel the installation if they feel uncomfortable about the set of permissions requested. It is worth noting that such explicit consumer consent to the set of permissions does not exist in Apple's iOS. There the information remains implicit before installation. In its essence this procedure remained stable since 2012, and is still in place today despite the fast growth of the Android Market.

Very recently, Google allows users to withdraw individual permissions from an app after the installation. However, this affects only the most recent version of the Android OS (Version 6.0, named "Marshmallow"). The resulting effects cannot be evaluated in this paper since yet only very customers use this version of the OS. However, in 2012 developers could choose among 136 predefined permissions.[5] This large number illustrates the quantity and diversity of information app developers can potentially collect about app users. Figure A2 illustrates the way the permissions were displayed in the Android Market in 2012. Since then, Google introduced several small modifications to how permissions are displayed to the user. Before 2014, the list of permissions provided permission names next to short explanations of the permissions. Since 2014 the system shows only the names of aggreagated permission *groups* (but users can open a more detailed dialogue for each permission group). Still, users must approve of the permission list before proceeding with the installation process.

Finally, developers can monetize their apps via four important channels. According to AppBrain (2016), around 20 percent of the apps are paid apps, whereas the remaining apps are for free.[6] Alternative revenue channels are in-app advertisment, in-app purchases and data trade. The importance of these alternative revenue channels was relatively stable since 2012 except for in-app purchases, which were introduced shortly before our period of

[5]Today the count stands at 137 permissions (although the precise contents of some permissions changed; see `http://developer.android.com/reference/android/Manifest.permission.html`).

[6]Developers receive 70 percent of the app price, and 30 percent go to distribution partners and operating fees (see `https://support.google.com/googleplay/android-developer/answer/112622?hl=en`).

observation.[7] In 2012, when we collected our data, the "freemium" model based on in-app purchases hardly existed. Since then the market has seen a marked increase of this model, where users may install the apps for free, but must pay a fee to unlock important functions. The two other channels, in-app advertisement and data trade were already common. Data trading is deemed the more privacy-sensitive way of creating revenue from an app, whereas in-app advertisement is deemed more acceptable by many users.

Yet, data trading for monetization purposes is very widely common in the mobile app industry. Christl and Spiekermann (2016) survey several studies who which have shown that apps very commonly shared data with third parties.[8] They also provide a brief discussion of the data sharing business model, which was also confirmed in our conversations with industry experts. Apps Developers have several ways of exploiting their app usage data to generate revenues. Most importantly, they can simply use the information and to sell anonymized targeted advertisements. However, doing so is very costly and only pays off for very large apps with sufficient traffic. Smaller apps can achieve better targeting by sharing their data with a third party broker, who can provide advertisers with access to users from a bigger pool of multiple apps (and developers to more advertisers). While this type of information brokerage potentially offers greater matching efficiency of the ads, sharing the data implies that the app users' information is passed on, and their safety depends only on the integrity of the developer's advertisement partner.

In addition, there are several other very common ways that app developers can trade in their data with third parties. First, they can exchange their data for direct monetary

[7] Only in 2011, Google added in-app billing to Android Market, allowing apps to sell in-app products (see `http://android-developers.blogspot.de/2011/04/new-carrier-billing-options-on-android.html`).

[8] The most common transmitted pieces of information about the user are identifying information (name, user's mail address, phone ID, gender, age or birthdate), location data, contacts, or usage data. Sometimes 'data input' (such as search terms etc.) are transmitted as well. In some contexts, like health, this information is more sensitive than in others. (Reference 1, 2, 3 , and 4). Most of these studies are based on very small samples. E.g. (Seneviratne et al, 2015) studied the apps installed on the phones of 338 users and identified 124 different trackers in 509 unique apps in Australia, Brazil, Germany and the United States. Trackers were categorized as: 'advertising' (e.g. Google, Ads, Millennial Media, Inmobi, Mopub), 'analytics' (e.g. Flurry, Google Analytics, Comscore, Health and Amazon Insights, Localytics, Kontagent, Apsalar) and 'utilities' (e.g. Crashlytics, Bugsense). Moreover they found that 50

benefit. A second more prevalent channel of data flows arises when developers trade their data for valuable third party services. The most important example of such services are app analytics. App analytics help the developer to gain insight on who uses their app when and where and together with which other processes that take place on the phone. Combined with the developer's own knowledge of the user's in-app behavior, this can be a useful input for improving the app or, again, for advertisement purposes. Like with third party advertisement, the secondary usage of the user data depends on the analytics site's own policy and integrity. There is no mechanism to enforce that developers choose a careful and trustworthy provider of analytics or other third party services.

4 Data

We extracted all publicly available information on as many apps as we could find on the English Android Market website in 2012, 2014 (later "Google Play Store"), 2015, and 2016. We collected the data monthly from April to October 2012, once in 2014, and again monthly since October 2015. The repeated data collection in 2012 and since 2015 allows us to use panel data methodology. The additional wave from 2014 was gathered to use a low-density panel for long-term outcomes, such as installation growth over 2 years. Our data set covers nearly the full population of products available in 2012 (around 300,000 apps), and after merging our data with the data from *privacygrade.org* we still observe more than 150,000 apps. In 2012, we could discern 136 distinct permissions in apps and record the permission requirements of each app. Moreover we expand our data with additional datasources: We evaluate privacy intrusiveness be combining our data with information from *privacygrade.com*, and data from *alexa.com* provides additional information on an app's popularity.

Figure A1 in the Appendix shows the design of Google's Android Market in 2012 which corresponds to the information we were able to collect. To study our research questions we need three types of information: a demand measure, a price measure and a measure of apps'

ability to collect private information. In the following section we introduce each of these measures as well as the core control variables.

Main Outcome Variables: Our main demand measure is the number of installations of an app. Our data set contains direct information on the total number of installations (i.e. sales) for each app. It is available in discrete form (17 levels, e.g. 1-5 installations, 6-10 installations, 11-50 installations, etc.). This is an improvement over most previous internet-based data sets, where demand variables had to be approximated, which was achieved using an app's sales rank, but not its real downloads (see e.g. Chevalier and Mayzlin, 2006; Garg and Telang, 2013; Ghose and Han, 2014). In addition we can use the strategies developed in these papers to improve our baseline demand measure, by exploiting information on the number of ratings of an app, which is available as a continuous measure. We exploit the continuous number of ratings to predict a continuous number of installations per app, which we then use in our panel analysis. The second main outcome variable is the price of apps, for which we have precise information (in Euros) for each app. Our other main outcome are the choices of developers to build privacy sensitive features into their apps.

Identifying the Privacy-Sensitiveness of Apps: To measure apps' ability to collect private information, we take advantage of the fact that, as described before, Google provides precise insights into the permissions an app uses. This feature allows us to understand in a detailed way which functions an app can perform, including functions which allow an app to collect private information about the app user. In 2012, developers could choose from 136 different permissions, which included e.g. 'read SMS or MMS', 'fine (GPS) location', 'read browser data', etc. All of these permissions have to be declared in the app description and have to be accepted by the app user before installing the app.[9] Among the 136 permissions, some can be considered innocuous with respect to the privacy of the user, while others grant an app access to sensitive information. To identify such privacy-sensitive permissions, we

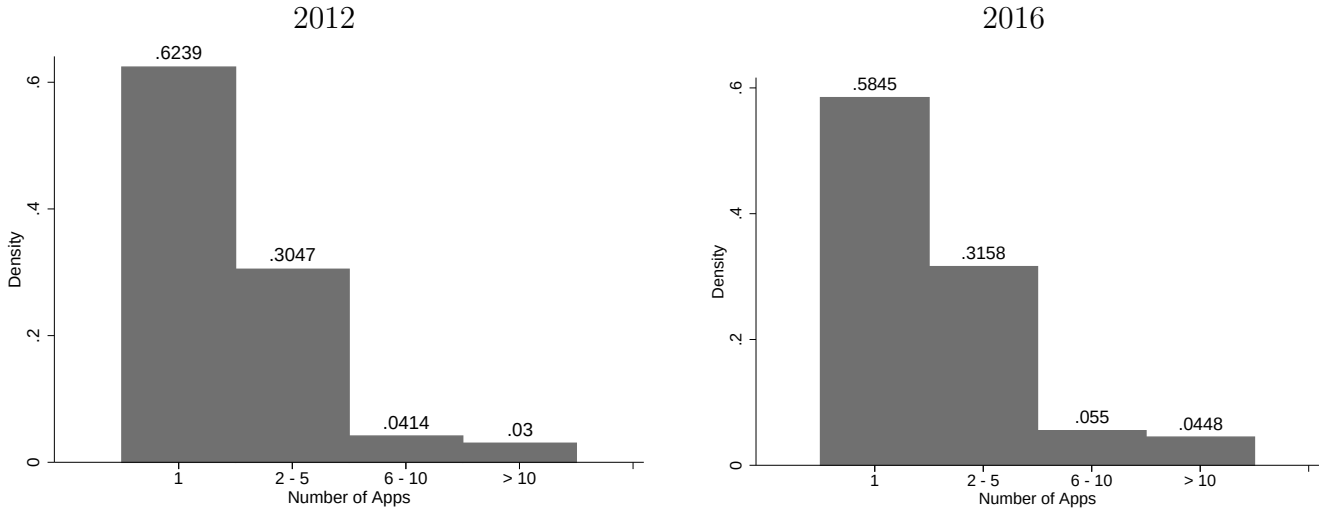
[9] We use the standardized short explanation to inform users about the permission's meaning by Google.

Table 1: Summary statistics for the main variables and privacy measures in 2012.

	mean	sd	min	p50	max
Outcome Variables					
Installations (in 1000)	67.00	1252.77	0	0	300000
Average Rating	3.91	0.83	1	4	5
Permissions					
$\#_{TotalPerm.}$	3.62	3.90	0	3	114
$\#_{CriticalPerm.}$	2.43	2.30	0	2	23
$\#_{PrivacyPerm.}$	0.86	1.33	0	0	12
$\#_{MaliciousPrivacyPerm.}$	0.34	0.69	0	0	7
$\#_{NonmaliciousPrivacyPerm.}$	0.51	0.79	0	0	4
$D_{Privacy}$	0.40	0.49	0	0	1
$D_{PrivCatSpec}$	0.18	0.39	0	0	1
$D_{MaliciousPrivacy}$	0.25	0.43	0	0	1
$D_{NonmaliciousPrivacy}$	0.35	0.48	0	0	1
$D_{Internet}$	0.68	0.47	0	1	1
D_{Ads}	0.45	0.50	0	0	1
D_{Other}	0.36	0.48	0	0	1
App Characteristics					
Price	0.80	3.29	0	0	157
App Version	16.88	148.32	0	2	9561
Size (in KB)	3048.15	7326.62	4	960	809000
Length Description	805.08	809.78	1	504	12285
Number Screenshots	3.40	1.83	0	3	8
Dummy: Video	0.10	0.30	0	0	1
Dummy: Top-Developer	0.01	0.08	0	0	1
Apps by Developer	157.36	412.24	1	10	2963
Average Installations of Developer	79.08	1050.53	0	5	300000
Observations	233813				

NOTES: List of core variables: For **developers** we observe the name of the developer, the top developer status (yes/no), number of its apps, the set of its available apps. The main variables at the **app-level** are: total number of installations, (monthly) downloads of an app, information on updates (date, textual information on what is new, version number), names and IDs of similar apps, the permissions requested upon installation, number and values of quantitative ratings (from 1 to 5 stars), is the app an editor's choice (yes/no), text of reviews (date, rating from 1 to 5 stars, content, availability of a developer-response), price (in Euro),in-app purchases (yes/no) and price-range of items, in-app advertisements (yes/no), app category (e.g. Personalization, Traveling, Weather, etc.), code size (in KB), apps' description (length, content) and its illustration in the Play Store (video and screen-shot availability), content rating (USKs), availability of interactive elements (e.g. 'users interact', 'digital purchases' etc.), Android version required for installation.

Figure 1: Number of Apps per Developer.



NOTES: The figure shows the share of developers by their respective number of apps in the market. Approximately 60% had one app, 30% had 2-5 apps and the remaining 10% of developers had 6 or more apps. While these shares were similar in 2012 and 2016, it can be seen that the share of 'experienced' developers with more than one app has increased over the four years. Note that this distribution of developers implies that the majority of apps had a 'sibling' from the same developer (82% in 2016; 77% in 2012).

use four alternative permission classifications. Our main classification (1) is derived from previous research by Sarma et al. (2012). The three alternative classifications are (2) a category-specific modification thereof, (3) a classification based on Google's assessment, (4) one derived by hiring 400 classifiers at Amazon Mechanical Turk, and (5) a classification from *privacygrade.com*, which was carried out in 2014 and gives us an outlook on the apps' behavior in the future.

Our baseline definition of privacy-sensitive permissions follows Sarma et al. (2012) who analyze the benefits and risks of Android app permissions and classify them according to different risk types. 26 permissions are classified as critical, and among these 13 are considered as being a risk to privacy.[10] Based on this classification, we construct our main variable of interest ($D_{Privacy}$), which is a dummy equal to one if an app uses at least one of the 13 privacy-sensitive permissions and zero otherwise. To capture the intensity of an

[10]For the permission *read calender* we were not able to collect information, such that we only have information on twelve privacy-sensitive permissions.

apps's ability to collect private information, in addition, we make use of the number of privacy-sensitive permissions per app. While this strategy allows us a deep insight into the permissions an app requests, it does not provide us with an evaluation of the app's actual privacy intrusiveness. To acquire this information we combine our data with information from *privacygrade.com*, which is the thus far most comprehensive effort of computer scientists to evaluate potential privacy intrusiveness of apps. They analyzed detailed information about more than one million apps' privacy-related behaviors (Lin et al., 2014., 2012). They summarize these behaviors in the form of a grade, ranging from A+ (most privacy sensitive) to D (least privacy sensitive).[11] Grades were assigned using a privacy model that measures the gap between people's expectations of an app's behavior and the app's actual behavior. For example, most people don't expect games like Cut the Rope to use location data, but many of them actually do. This kind of surprise is represented as a penalty to an app's overall privacy grade. In contrast, most people do expect apps like Google Maps to use location data. This lack of surprise was represented as a small or no penalty. [12] Privacygrades were computed in 2014 and 2016, which allows us to get a hard measure of the app's future behavior. Moreover, we can run a cluster analysis to see which permissions are typically associated with bad privacygrades, and can apply this prediction ex-post to our 2012 wave.

5 Methodology and Empirical Approach

We aim at answering two overarching questions: (i) What strategies do developers use w.r.t. private data? (ii) Which app developers are more prone to privacy intrusive choices? Hence, we pursue our analysis in three steps. First we use reduced form analysis of developer behavior to understand the role of data in developers' strategies. Second we analyze apps' data collection strategies over time. In the third and last step we focus on developers' behavior over time.

[11] Their results are provided publicly: see: http://privacygrade.org/

[12] The description was taken from http://privacygrade.org/faq [retrieved: 11/05/2016]

Model Free Analysis: While the analysis is not guided by a model, we believe that the disciplines knowledge about app developers (let alone privacy abusive developers) is so limited at the current point that a model guided analysis is hard to achieve. Developers might attempt hit and run strategies, or try to build reputation and increase their permission requests later. They might target valuable consumers or aim at tricking "easy victims."

Hence, rather than deriving the analysis from theory, we hope to generate valuable stylized facts that can be used in theoretical or structural analysis.

5.1 Cross Sectional Analysis

The main goal of our analysis is to gain insights into the drivers of privacy intrusive app design. Moreover, we want to shed light on both the first order and the second order effect of data access for the success of an app. First, access to data enables a healthy eco system, in which developers have a better understanding of their user's needs, and can monetize their services. Both are critical ingredients for further improving the app. Second however, excessive requests for access to data cannot easily translate to additional benefit. Hence, based on our study of the literature, we expect positive but diminishing marginal returns to more data access, and we wish to test the following two hypotheses:

While we suspect access to more user information to influences app success (measured in downloads and survival) positively, we would expect this relationship to take an inverted 'U-shape.' That is, the marginal improvement in success decreases when requesting additional access to user data, especially if the requested data is not associated with improved functionality. Finally, we would suspect that better access to more user information (directly and indirectly) facilitates innovative activity, which could be a major channel to explain any positive relationship.

To test these hypothesis we will first analyze the correlational dynamics in the data to understand the global patterns and relationships. This will help us to understand whether access to data plays a primary role in the market for mobile apps or whether other aspects

in this market may be more important. For this broad and descriptive analysis we will focus on the cross sectional variation.

$$Success_i = \beta D^{Privacy_i} + \gamma Pricing_i + \theta X_{1,i} + \varepsilon_i. \tag{1}$$

Most importantly, we can also analyze what drives excessive data use. Hence, this first part will shed light on what types of developers use 'excessive' data strategies, and which patterns in an app's development render such strategies more likely? Among other things, we will analyze whether developers of single apps are different, in the sense that they ask for more or less privacy intrusive permissions.

$$Privacy_i = \beta D_i^{SingleAppDeveloper} + \gamma Pricing_i + \theta X_i + \varepsilon_i. \tag{2}$$

In this regression $Privacy_i$ represents the app's potential threat to a user's privacy. We measure that by the app's use of permissions, and its *privacygrade*. The indicator $D_i^{SingleAppDeveloper}$ is the variable of interest and we control for $Pricing_i$ and other control variables (X_i).

5.2 Fixed Effects Analysis at the Developer Level

Beyond the plain cross-section analysis, we can exploit a special feature in our data. We can observe developers with multiple apps, which facilitates a fixed effects analysis. This analysis will allow us to understand how the data strategies of the same developer vary with the environment.

We observe a series of important factors that influence a developer's data strategy for a given factor. Among these factors are the app's competitive environment, the app category, the app's past success, or whether the app is offered for free or for pay. In further research we investigate ways to analyze how app success varies for a developer's different data strategies. Thus we focus on developers (j) and use permissions or the *privacgrade* to measure their

apps' (i) data behavior, which results in the following regression:

$$Privacy_{ij} = \alpha_j + \gamma P_{ij} + \theta X_{2,ij} + \varepsilon_{ij}. \tag{3}$$

In this regression $Privacy_{ij}$ measures the privacy sensitivity of developer j's app i. The developer's fixed effect is measured by α_j, P_{ij} measures the app's price (which is potentially 0), and $X_{2,ij}$ are other control variables, such as the app's competitive environment or its category.

Thus analyzing a developers' data strategy across different apps, we can highlight which factors drive the same developer to use more problematic data strategies?

5.3 App *Developers* over Time

In the core part of the analysis, however, we want to exploit the fact that many developers have multiple apps in this market. Frequently we see that apps from the same developer vary in how much access to private information they request. Using a fixed effects design, we can leverage this variation to study how varying access to personal data influences the success of apps by the *same* developer. The ability of comparing apps by the same developer serves two purposes: First, it massively strengthens our research design to effectively analyze learning of developers. For example we can understand whether the first app is different from subsequent apps. Secondly, we can answer the questions with regards to the dynamics within an app.

Being able to track developers over time allows us to analyze how they experiment with access to data and how their strategies change with more experience. For this last step of the analysis we exploit our ability to run our analyses within and between apps of the same developer (as he introduces new apps). Specifically, we can analyze which factors can predict that a developer will choose a strategy that extracts high intensity personal data, and how data strategies of the same developer vary with the competitive environment.

Specifically, ongoing research could take the static analysis in the highlight questions a level further and ask when the same developers use more problematic dynamic data strategies? Again focusing on developers with multiple apps such an analysis would start by asking whether the first apps are different, and then move on to characterize dynamic strategies of developers. Similarly we can characterize the relative hostility, or dynamism of the app's competitive environment, based on app category, initial app demand, or the 'initial' data requirements.

Using permissions and the *privacgrade* as outcomes of interest to quantify data behavior, we can analyze dynamic data strategy over time. This analysis can shed light on how the dynamic data strategies of the same developers, and vary with the environment, or the initial app success. Continued success and continued innovation on the app as outcome are the other crucial economic outcomes in this big picture.

Understanding the relationship between data collection, developers' privacy strategies and how they influence app success and other outcomes will provide invaluable insights into the role of data availablilty/privacy for the supply side. Additionally the insights from our project will serve as the foundational input to developing a structural approach to estimate the value of accessing privacy sensitive data for the developers. This will allow us to evaluate the harm (or its absence) of more rigorous privacy regulation in the market of mobile apps. Specifically, it will also shed light on which levels of data access are necessary for the market to thrive, and which levels are not.

6 Data and Description of the Variables

6.1 Key Variables

As discussed in the key dataset, the main variables were obtained from our full 2012 cross-section, which contains up to 300,000 apps. These data were merged with the data from *privacygrade.org*, to add the the privacy grade, and a derived dummy *failed* which takes

the value 1, if an app received grades C or D. , After merging our data with the 2014 data from *privacygrade.org* we still observe more than 150,000 apps.

6.2 Stylized Facts

We first compare the apps of single-app developer to the apps of multi-app developers, and then use a fixed effects analysis to analyze the behavior of multi-app developers. Single app developers might be more prone to reverting to data-selling if their app is unsuccessful, whereas multi-app developer might have a reputation to loose. Tables 2 and 3 show the descriptives for single app developers' apps and multi app developers seperately. Moreover, in Figure 2 we visualize this comparison. It is easy to see from this raw comparison, that the apps of multi-app developers appear to be more professional and more demanded (more installs, ratings, and higher price). Yet the apps of single app developers use slightly more permissions, and still have similar ratings.

Figures A3 and A4 complete the descriptive picture. They highlight that single-app developers are more active in Entertainment, Health, and Business, but relatively less in Education and Games. IN terms of required maturity levels single, and multi-app developers behave very similarly. The vast majority of apps are for everyone or for low maturity (i.e. Kids).

7 Results

7.1 Regression Results

Single- vs. Multi App Developers: We use the new variable that measures the privacy intrusiveness in 2014 via the *privacygrade*, to shed light on which types of apps are more likely to be problematic. We first ask whether apps of single app developers or multi-app developers are more likely to become abusive. In Table 4 we show the results of this descriptive cross section. Surprisingly, multi-app developers are more than 2 % more likely

Table 2: Single App Developers

	(1)						
	mean	sd	min	p10	p50	p90	max
Number of Installs	55.39548	812.772	.003	.075	3	30	75000
Number of Ratings	347.1522	6174.736	1	1	8	134	439719
Rating	3.995473	.8979601	1	2.8	4.1	5	5
Price	.2794387	1.778883	0	0	0	.75	105.43
Number of Permissions	3.884763	3.915905	0	0	3	9	114
Number of Malicious Perm.	2.613125	2.243873	0	0	2	6	21
Observations	35275						

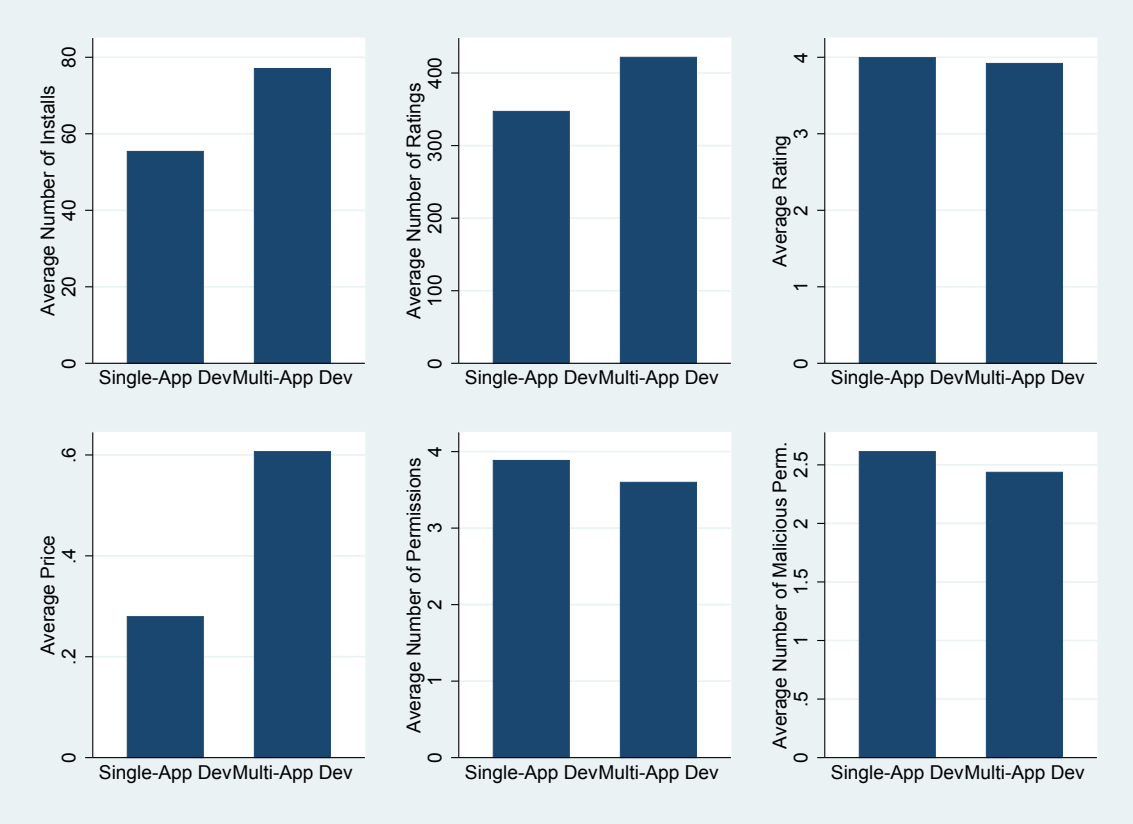
NOTES: The table shows the averages and distribution for the most important variables of single app developers. It shows the mean, standard deviation, minimum value, median, maximum value, and the 10th and 90th percentile of the distribution.

Table 3: Multi App Developers

	(1)						
	mean	sd	min	p10	p50	p90	max
Number of Installs	77.07398	1432.493	.003	.03	3	30	300000
Number of Ratings	421.6965	7780.803	1	1	7	174	1101035
Rating	3.919851	.9778243	1	2.6	4.1	5	5
Price	.6066419	2.521879	0	0	0	1.5	136.49
Number of Permissions	3.598694	3.778822	0	0	3	9	114
Number of Malicious Perm.	2.436763	2.22157	0	0	2	5	23
Observations	144765						

NOTES: The table shows the averages and distribution for the most important variables of multi app developers. It shows the mean, standard deviation, minimum value, median, maximum value, and the 10th and 90th percentile of the distribution.

Figure 2: Comparison of Single and Multi App Developers



NOTES: The figure compares the apps of single app developers and multi app developers. Each plot shows a variable of interest in the 2012 data (from topp left to bottom right): (i) average number of installations, (ii) average number of ratings, (iii) average rating, (iv) average price, (v) avg. number of permissions and (vi) avg. number of malicious permissions. On each panel the left bar shows the value for single-app developers and the right bar shows the average value of multi app developers' apps.

Table 4: Privacygrade: Comparing Single App and Multi App Developers - privacygrade

	(1) failing	(2) failing	(3) failing	(4) failing	(5) app grade
Single App Developer	-0.027***	-0.024***	-0.027***	-0.028***	-0.062***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)
Dummy: Top-Developer		0.048***	0.048***	0.036**	0.114**
		(0.015)	(0.017)	(0.017)	(0.044)
Apps by Developer		0.000***	0.000***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)	(0.000)
Price		-0.003***	-0.003***	-0.003**	-0.012**
		(0.001)	(0.001)	(0.001)	(0.006)
Number Screenshots		0.005***	0.005***	0.004***	0.028***
		(0.000)	(0.000)	(0.000)	(0.001)
Dummy: Video		0.012***	0.011***	0.008**	0.012
		(0.003)	(0.003)	(0.003)	(0.009)
Number Ratings		0.000***	0.000***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)	(0.000)
Average Rating		-0.002***	-0.002**	-0.002**	-0.042***
		(0.001)	(0.001)	(0.001)	(0.003)
Installations (in M)			0.000	-0.000	0.002
			(0.001)	(0.001)	(0.004)
Competitors' Inst.			0.000**	0.000*	0.002***
			(0.000)	(0.000)	(0.000)
Avg. Install Developer			-0.000	-0.000**	-0.000**
			(0.000)	(0.000)	(0.000)
#ExtremPrivPerm				0.004	0.037
				(0.015)	(0.040)
#UnusualPrivPerm				0.029***	0.162***
				(0.010)	(0.023)
Privacy Policy				0.042***	0.178***
				(0.005)	(0.014)
Constant	0.059***	0.047***	0.046***	0.047***	2.092***
	(0.001)	(0.004)	(0.004)	(0.004)	(0.012)
Observations	88893	88893	67685	64731	64731
Adjusted R^2	0.002	0.010	0.010	0.012	0.024

NOTES: The table compares the permission usage of single and multi app developers and analyzes whether it lead to bad privacygrades. The regressions are cross sectional OLS regressions. The dependent variable in columns 1-4 is a dummy that indicates that the app will receive a failgrade (C or F). In columns 5 we use the grade that was obtained in 2014 (5=A+, 1=F). Col. 1-4 start from the raw correlation and gradually add control variables. Col. 5 shows the result with all controls, but for the privacygrade directly. Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

to obtain a fail grade in 2014. This relationship remains stable as we introduce control variables that account for (a) the app's characteristics (b) developer characeristics and the competitive environment, and (c) the app's privacy policy. None of these controls affects the finding that single-app developers are less likely to obtain a failgrade, and the relationship is equally true when using the 5 scale (where 5 corresponds to A+) app grade instead of the dummy.

When looking at the control variables we find interesting results. First, successful apps are more at risk of having high privacy requirements and receiving a fail grade. Second, apps with a privacy policy were by a substantial 5% more likely to fail the 2014 test, and obtained a lower grade (0.17 gradepoints on average).

App Developers' over time: The next step in our analysis focuses on developers development over time. To answer this question we identified an app's launch date and analyzed whether developers use more or less intrusive permissions on their first apps.

The results of this analysis are shown in Table 5. The table controls for developer heterogeneity by using developer fixed effects (cf. below). It shows two major results: First, developers ask for less intrusive permissions on their first two apps, independent of whether we look at sensitive permissions, google's "potentially malicious" permissions, or the *privacygrade*.[13] Even for the likelihood of finding a failgrade we observe a small and negative effect for the first, but not the second app, which indicates that "rookie-apps," like single-developer apps, ask for fewer data access. We also see that the second app is slightly less moderate for all measures.

The second finding comes from the analysis of the maturity ratings. The analysis shows that rated apps are generally more likely to be intrusive. Most importantly, for all measures, including a failgrade from *privacygrade.com* we see that apps for teenagers are most intrusive and potentially abusive. This suggests that teenagers face the greatest risk of being

[13]In this dataset A+ takes the value of 1 and F takes the value 5, and a negative coefficient means better protection.

Table 5: Multi App Developers Learning - - Permission Usage and Experience

	sensitive	potential malicious	# pot. malic.	privacygrade	failgrade
	(1)	(2)	(3)	(4)	(5)
Developer's 1st App	-0.015***	-0.016***	-0.078***	-0.015*	-0.005*
	(0.006)	(0.005)	(0.021)	(0.009)	(0.003)
Developer's 2nd App	-0.013**	-0.014***	-0.060***	-0.015*	-0.003
	(0.005)	(0.005)	(0.020)	(0.009)	(0.003)
Average Rating	-0.006**	-0.006**	-0.003	-0.005	0.002
	(0.003)	(0.002)	(0.009)	(0.006)	(0.002)
# Ratings in 1000	0.001***	0.000	0.005***	0.003**	0.001*
	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)
Dummy: Video	0.031***	0.043***	0.220***	0.079***	0.023***
	(0.011)	(0.010)	(0.041)	(0.019)	(0.007)
Dummy: Website	0.011	0.045***	0.133***	0.026	0.004
	(0.014)	(0.016)	(0.051)	(0.028)	(0.008)
Privacy Policy	0.076***	0.064**	0.451***	0.091**	0.007
	(0.022)	(0.027)	(0.101)	(0.044)	(0.016)
Log. Price	0.024***	0.069***	0.241***	-0.128*	-0.017
	(0.009)	(0.008)	(0.030)	(0.067)	(0.015)
$DPrice$	-0.321***	-0.952***	-3.324***	1.428*	0.205
	(0.098)	(0.098)	(0.350)	(0.778)	(0.174)
Maturity Rating n.a.	0.001	-0.076***	-0.232***	-0.019	-0.001
	(0.016)	(0.017)	(0.054)	(0.025)	(0.007)
Medium Maturity	0.205***	0.094***	0.801***	0.118***	0.007
	(0.021)	(0.015)	(0.076)	(0.026)	(0.010)
Low Maturity (13+)	0.455***	0.173***	1.669***	0.229***	0.041***
	(0.014)	(0.010)	(0.052)	(0.016)	(0.006)
High Maturity (17+)	0.180***	0.077***	0.641***	0.098***	0.027**
	(0.025)	(0.020)	(0.086)	(0.030)	(0.012)
Local Market Share	0.020	-0.006	0.149**	0.052**	0.000
	(0.013)	(0.015)	(0.061)	(0.023)	(0.009)
Competitors' Inst.	0.001***	0.001***	0.007***	0.001**	-0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Constant	0.586***	1.516***	4.576***	0.479	-0.170
	(0.099)	(0.096)	(0.351)	(0.770)	(0.169)
Observations	47300	47300	47300	29465	29465
Developers	15748	15748	15748	12882	12882
Adjusted R^2	0.162	0.081	0.198	0.036	0.012

NOTES: The table analyzes whether app developers' permission usage changed over time, and whether this affects the resulting *privacygrade* in 2014. The regressions are panel OLS regressions with a developer fixed effect. The dependent variables in the five columns are: (1) usage of sensitive permissions (2) usage of potentially malicious permissions (according to Google) (3) number of potentially malicious permissions (according to Google) (4) the *privacygrade* that was obtained in 2014 (1=A+, 5=F), and (5) a dummy that indicates that the app will receive a failgrade (C or F). The reference content rating are apps rated as "Everyone." Robust Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

confronted with intrusive apps. Since this specification includes a developer fixed effect, this means that the same developer will use more intrusive permissions and libraries in an app for teenagers.

Finally, in Table A2 we analyze the correlation of developer characteristics with requesting intrusive permissions. Specifically, we include developers' initial success on their first two apps. While most developer characteristics have little value for predicting intrusive behaviors, we find a very small positive association of previous app demand (installations on the first, ratings on the second app) with future apps' intrusiveness, but this correlation does not control for the first app's intrusiveness, and might thus be confounded by developer heterogeneity.

Developer Fixed Effects Analysis Our main strategy to control for developers' varying abilities, strategies and intentions relies on using a developer fixed effects analysis. We used this strategy in Table 5, and will use it in all subsequent result tables. Doing so, we can contrast the intrusiveness of apps by the *same* developer in different market environments. Moreover, we can also investigate the other patterns that emerged from the analysis of control variables, we run a fixed-effects regression for all multi-app developers (including those where we could not identify first and second app). These regressions compare apps of the same developer and thus allow to study how the use of permissions depends on an app's environment. In Table 6 we show the results for the developers' permission usage when restricting our attention to multi-app developers. The dependent variable in columns 1-3 is a dummy indicating the usage of sensitive permissions and, in columns 4-6, a dummy that indicates potentially malicious permissions (as flaggef by google). Col. 1&4 analyze paid and free apps seperately. The negative coefficient implies that sensitive permissions occur far more likely in free apps. Col. 2&5 analyze categories seperately, and highlight a tendency for fewer sensitive permissions in educational apps and games. However, the other patterns for categories are not consistent. Col. 3&6 add variables that measure (i)

the strength of the competitors, and (ii) the app's market share, measured by its strength relative to it's closest substitutes. The results show a significant but very small positive effect of competitive pressure on permission usage. They also show that market share is strongly correlated with using intrusive permissions. Stronger apps seem to use their market power for acquiring more data.

In Table A1 we go one step further and analyze if the analyzed factors are also predicitve of a bad *privacygrade* in 2014. In general terms we find consistent patterns for the grade, both for the control variables and the effect of competitive pressure. We would like to highlight the fairly strong effect of having a privacy policy, which continues o be associated with a bad privacy grade. However, hardly any of the variables matters for predicting failgrade once we add the developer FE. Only competitive pressure has a significant, but negligible effect on privacy abusive behavior. The scarcity of significant coefficients in the FE regression with failgrade as dependent variable could be due to smaller sample sizes. However, it is more plausible that the developer fixed effect is the most important factor when accounting for privacy-endangering behaviors. Similarly, finding no effect for market share suggests that strong apps use their market power for acquiring more data, but without taking it too far.

Robustness: To investigate the robustness and external validity of the patterns that emerged from the main results, we run all fixed-effects regressions from above, but exclude developers' first and second apps. This ensures that the apps are made by developers who all have more than 3 apps, and who have gathered initial experience. The results are shown in Tables A3 and A4. These tables confirm all previous finding, on a smaller but more rigorously compared set of apps.

Ongoing work and Limitations: At this point many limitations remain, since the results are largely based on the augmented 2012 app data. Future work will validate the findings we have shown by using the more rigorous identification strategies that are outlined in the methodological section. Especially focussing on the apps that were newly created during

Table 6: Multi App Developers and Environment - Permission Usage

	sensitive			potential malicious		
	(1)	(2)	(3)	(4)	(5)	(6)
Average Rating	-0.005***	-0.009***	-0.009***	-0.004***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
# Ratings in 1000	0.001***	0.001***	0.001***	0.000**	0.001***	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dummy: Video	0.044***	0.044***	0.035***	0.045***	0.038***	0.039***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Dummy: Website	0.024***	0.025***	0.024**	0.037***	0.037***	0.037***
	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)	(0.012)
Privacy Policy	0.110***	0.114***	0.098***	0.066***	0.073***	0.058***
	(0.019)	(0.020)	(0.023)	(0.015)	(0.018)	(0.020)
$DPrice$	-0.673***			-1.093***		
	(0.086)			(0.091)		
Log. Price	0.049***			0.078***		
	(0.007)			(0.007)		
Categ: Education		-0.031**			-0.058***	
		(0.012)			(0.011)	
Categ: Entertain.		0.033***			0.006	
		(0.011)			(0.008)	
Categ: Games		-0.022*			-0.035***	
		(0.012)			(0.010)	
Categ: Tools/Perso.		0.046***			-0.053***	
		(0.011)			(0.008)	
Categ: Lifestyle		0.066***			-0.006	
		(0.012)			(0.009)	
Categ: Health		0.003			-0.032***	
		(0.013)			(0.010)	
Local Market Share			0.086***			0.110***
			(0.013)			(0.013)
Competitors' Inst.			0.002***			0.001***
			(0.000)			(0.000)
Constant	1.018***	0.415***	0.430***	1.687***	0.783***	0.739***
	(0.083)	(0.013)	(0.010)	(0.087)	(0.012)	(0.012)
Observations	136078	136078	105371	136078	136078	105371
Developers	29530	29530	26857	29530	29530	26857
Adjusted R^2	0.018	0.009	0.008	0.058	0.007	0.009

NOTES: The table analyzes the driving factors of developers' permission usage. The regressions are panel fixed OLS regressions with a developer fixed effect. The dependent variable in columns 1-3 is a dummy indicating the usage of sensitive permissions and, in columns 4-6, a dummy that indicates potentially malicious permissions. Col. 1&4 analyze paid and free apps seperately. Col. 2&5 analyze categories seperately and col. 3&6 add variables that measure the strength of the competitors. Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

2012 will allow to understand how early-life factors influenced later privacy choices. Also analysing the behavior of multi-app developers between 2012 and 2016 will shed additional light on the relationships of interest.

Moreover, we do not observe in-app purchases. While this is arguably no problem for 2012, and hence the present cross section analysis in app purchases are an important consideration for dynamic analysis of the data. In future research we will attempt to use the information from the *privacygrade*-data to solve this issue. Next, it is challenging to cleanly disentangle monetization from functionality. Future research could parse updated descriptions and reviews, or alternatively use free/paid pairs to spot typically redundant permissions. Finally, there is additional potential in seeking additional sources of exogenous variation, such as the roll out of Android 6

8 Conclusions

In this paper we analyze mobile applications to analyze the data collection policies of mobile app developers. We tracked more than 300,000 apps over 4 years, and combine these data with additional information from *privacygrade.org* on the apps' behavior. We find that single app developers are less likely to run intrusive apps, and generally obtain better privacy grades. For multi-app developers, we find that they use less intrusive permissions in their earlier apps. USing a developer-fixed effects analysis, we show that strong competition results in a small but positive pressure to produce more intrusive apps, and that developers ask for more intrusive permissions on their apps with larger market share (without becoming abusive, however). Moreover we find that developers are more likely to use intrusive (and potentially abusive) permissions when their apps are targeted at the 13+ age category, which raises considerable concerns for consumer protection. Finally, a privacy policy and a video in the description are highly predictive of problematic levels of access to user data. These findings are confirmed when focusing exclusively on the third and later apps of multi-app

developers in conjunction with a developer fixed effect.

The last decade alone has seen the creation of more than a million applications, which generate high total user value. Given this recent rise of smart phones and mobile apps and their transformative power on human interaction and the economy as a whole, the issue of privacy concerns in this market is arguably a very important one. We argue, that it is important to anticipate the risk of excessive collection of private user data early, to take well informed measures to protect consumers and control these risks without unnecessarily curtailing the market's impressive potentials for generating user welfare.

Our paper is a first step towards understanding the collection and sharing of personal user data. We analyze which user groups are most exposed, how developers gain access to sensitive information, and how they react to the factors that motivate them to engage in excessive data-collection. Understanding thse questions could help decision makers to successfully design the market and protect users. Moreover our paper generates additional value, by creating a unique database on privacy in the market for mobile applications. Future work should build on this foundation to develop a structural model that allows for a counterfactual analysis of potential regulatory or other measures.

References

Acquisti, Alessandro and Hal R Varian, "Conditioning Prices on Purchase History," *Marketing Science*, 2005, *24* (3), 367–381.

_ , Curtis R Taylor, and Liad Wagman, "The Economics of Privacy," *Journal of Economic Literature*, forthcoming.

_ , Laura Brandimarte, and George Loewenstein, "Privacy and Human Behavior in the Age of Information," *Science*, 2015, *347* (6221), 509–514.

_ , Leslie K John, and George Loewenstein, "What is Privacy Worth?," *The Journal of Legal Studies*, 2013, *42* (2), 249–274.

App Annie, "App Annie Mobile App Forecast: The Path to $100 Billion," 2016. Available at https://www.appannie.com/en/landing/forecast.

AppBrain, "Google Play Stats," 2016. Available at http://www.appbrain.com/stats/free-and-paid-android-applications.

Askalidis, Georgios, "The Impact of Large Scale Promotions on the Sales and Ratings of Mobile Apps: Evidence from Apple's App Store," Technical Report 2015.

Aziz, Arslan and Rahul Telang, "What is a Cookie Worth?," Technical Report 2015.

Beresford, Alastair R, Dorothea Kübler, and Sören Preibusch, "Unwillingness to Pay for Privacy: A Field Experiment," *Economics Letters*, 2012, *117* (1), 25–27.

Carare, Octavian, "The Impact of Bestseller Rank on Demand: Evidence from the App Market," *International Economic Review*, 2012, *53* (3), 717–742.

Carrascal, Juan Pablo, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," in "Proceedings of the 22nd international conference on World Wide Web" International World Wide Web Conferences Steering Committee 2013, pp. 189–200.

Casadesus-Masanell, Ramon and Andres Hervas-Drane, "Competing With Privacy," *Management Science*, 2015, *61* (1), 229–246.

Chaudhari, Harshal, "The Impact of Zero-Price Promotions on Sales: A Case Study of Amazon Appstore," Technical Report 2015.

Chevalier, Judith A and Dina Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354.

Chia, Pern H, Yusuke Yamamoto, and N Asokan, "Is this App Safe? A Large Scale Study on Application Permissions and Risk Signals," in "in" ACM 2012, pp. 311–320.

Christl, Wolfie and Sarah Spiekermann, *Networks of Control - A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*, Facultas, Wien, 2016.

Conitzer, Vincent, Curtis R Taylor, and Liad Wagman, "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases," *Marketing Science*, 2012, *31* (2), 277–292.

Davis, Jason P, Yulia Muzyrya, and Pai-Ling Yin, "Experimentation Strategies and Entrepreneurial Innovation: Inherited Market Differences in the iPhone Ecosystem," Technical Report 2014.

Egelman, Serge, Adrienne Porter Felt, and David Wagner, "Choice Architecture and Smartphone Privacy: There is a Price for That," in "The Economics of Information Security and Privacy," Springer, 2013, pp. 211–236.

Fahl, Sascha, Marian Harbach, Thomas Muders, Lars Baumgärtner, Bernd Freisleben, and Matthew Smith, "Why Eve and Mallory Love Android: An Analysis of Android SSL (in) Security," in "Proceedings of the 2012 ACM Conference on Computer and Communications Security" ACM 2012, pp. 50–61.

Garg, Rajiv and Rahul Telang, "Inferring App Demand from Publicly Available Data," *MIS Quarterly*, 2013, *37* (4), 1253–1264.

Ghose, Anindya and Sang Pil Han, "Estimating Demand for Mobile Applications in the New Economy," *Management Science*, 2014, *60* (6), 1470–1488.

Goldfarb, Avi and Catherine Tucker, "Privacy Regulation and Online Advertising," *Management Science*, 2011, *57* (1), 57–71.

_ **and** _ , "Shifts in Privacy Concerns," *American Economic Review: Papers and Proceedings*, 2012, *102* (3), 349–353.

Gross, Ralph and Alessandro Acquisti, "Information Revelation and Privacy in Online Social Networks," in "Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society" ACM 2005, pp. 71–80.

Grossklags, Jens and Alessandro Acquisti, "When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information," in "Sixth Workshop on the Economics of Information Securtiy (WEIS)" 2007.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong T Lee, and Ivan PL Png, "Consumer Privacy and Marketing Avoidance: A Static Model," *Management Science*, 2008, *54* (6), 1094–1103.

IDC, "Worldwide Quarterly Mobile Phone Tracker," 2015. Available at http://www.idc.com/prodserv/smartphone-os-market-share.jsp.

Johnson, Garrett A, "The Impact of Privacy Policy on the Auction Market for Online Display Advertising," Technical Report 2013.

Johnson, Justin P, "Targeted Advertising and Advertising Avoidance," *The Rand Journal of Economics*, 2013, *44* (1), 128–144.

Kummer, Michael E and Patrick Schulte, "When private information settles the bill: Money and privacy in Google's market for smartphone applications," *ZEW-Centre for European Economic Research Discussion Paper*, 2016, (16-031).

Lin, J., J.I. Hong, and N. Sadeh., "Modeling UsersŠ Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings.," 2014.

_ **, S. Amini, J.I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang.**, "Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing.," 2012, pp. 501–510.

Marthews, Alex and Catherine Tucker, "Government Surveillance and Internet Search Behavior," Technical Report 2014.

Miller, Amalia R and Catherine Tucker, "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records," *Management Science*, 2009, *55* (7), 1077–1093.

Preibusch, Sören and Joseph Bonneau, "The Privacy Landscape: Product Differentiation on Data Collection," in "Economics of Information Security and Privacy III," Springer, 2013, pp. 263–283.

Racherla, Pradeep, Jeffry S Babb, and Mark J Keith, "Pay-What-You-Want Pricing for Mobile Applications: The Effect of Privacy Assurances and Social Information," in "Conference for Information Systems Applied Research Proceedings," Vol. 4 2011, pp. 1–13.

Rainie, Lee, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish, "Anonymity, Privacy, and Security Online," Technical Report 2013.

Sarma, Bhaskar Pratim, Ninghui Li, Chris Gates, Rahul Potharaju, Cristina Nita-Rotaru, and Ian Molloy, "Android Permissions: A Perspective Combining Risks and Benefits," in "Proceedings of the 17th ACM symposium on Access Control Models and Technologies" ACM 2012, pp. 13–22.

Savage, Scott J. and Donald M. Waldman, "The Value of Online Privacy: Evidence from Smartphone Applications," Technical Report 2014.

Spiegel, Yossi, "Commercial Software, Adware, and Consumer Privacy," *International Journal of Industrial Organization*, 2013, *31* (6), 702–713.

Sutanto, Juliana, Elia Palme, Chuan-Hoo Tan, and Chee Wei Phang, "Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users.," *Mis Quarterly*, 2013, *37* (4), 1141–1164.

Taylor, Curtis R, "Consumer Privacy and the Market for Customer Information," *The Rand Journal of Economics*, 2004, *35* (4), 631–650.

_ **, Vincent Conitzer, and Liad Wagman**, "Online Privacy and Price Discrimination," Technical Report 2010.

Tsai, Janice Y, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, "The Effect

of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research*, 2011, *22* (2), 254–268.

Tucker, Catherine, "The Economics of Advertising and Privacy," *International Journal of Industrial Organization*, 2012, *30* (3), 326–329.

_ , "Social Networks, Personalized Advertising and Privacy Controls," *Journal of Marketing Research*, 2014, *51* (5), 546–562.

Turow, Joseph, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy, "Americans Reject Tailored Advertising and Three Activities that Enable It," Technical Report 2009.

Wathieu, Luc, "Privacy, Exposure and Price Discrimination," Technical Report 2002.

Yin, Pai-Ling, Jason P Davis, and Yulia Muzyrya, "Entrepreneurial Innovation: Killer Apps in the iPhone Ecosystem," *American Economic Review: Papers and Proceedings*, 2014, *104* (5), 255–259.

A Additional Tables and Figures

Figure A1: App Information in the Android Market 2012

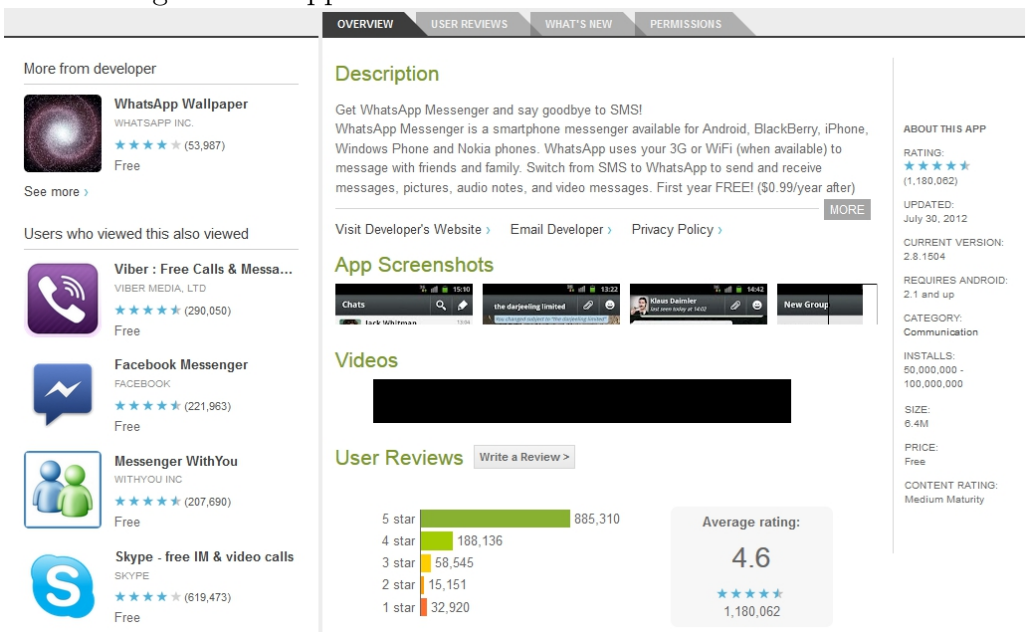


Figure A2: Permission Information in the Android Market 2012

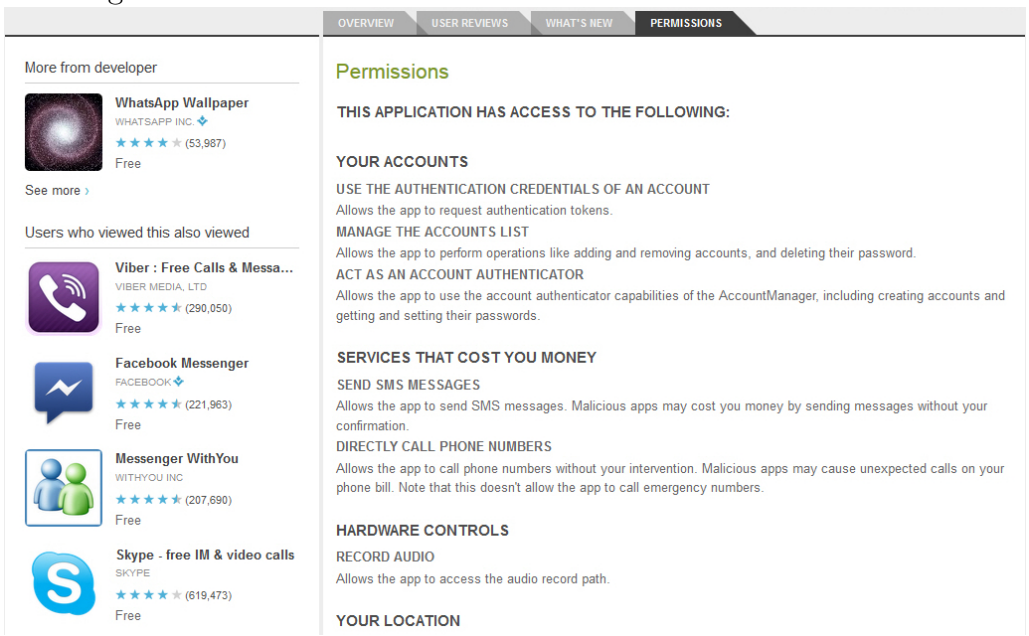


Figure A3: Comparison of Categories of Single and Multi App Developers

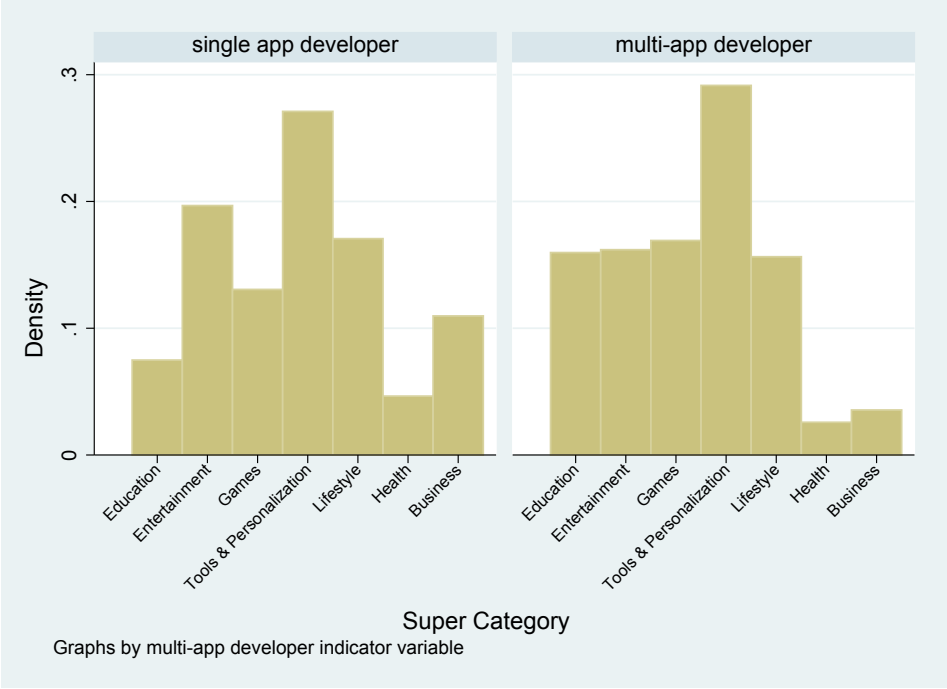


Figure A4: Comparison of Maturity Requirements of Single and Multi App Developers

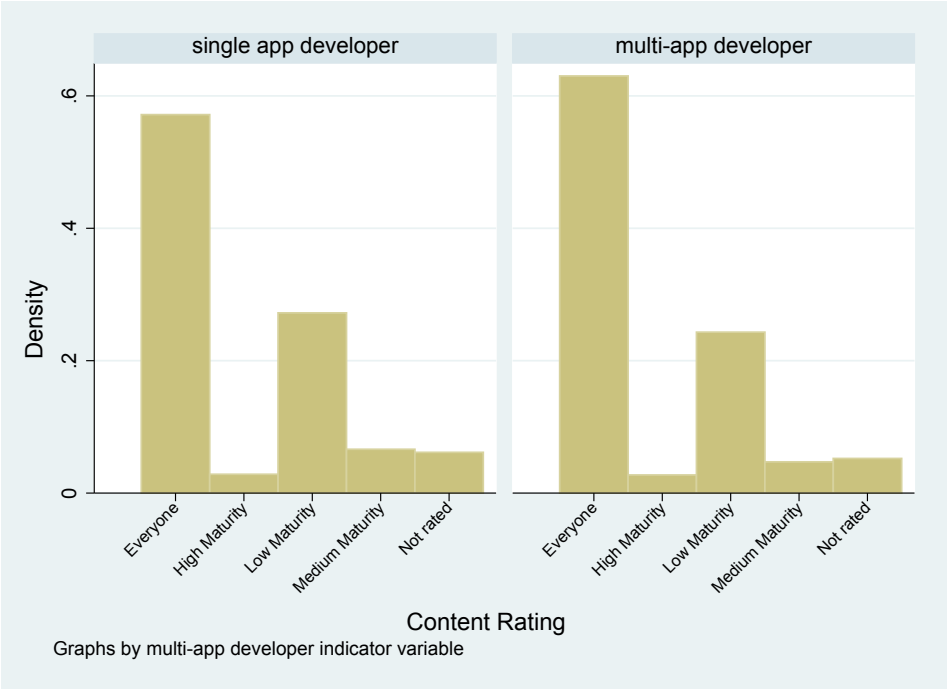


Table A1: Multi App Developers and Environment - Future Outlook *privacygrade*

	privacygrade			failgrade		
	(1)	(2)	(3)	(4)	(5)	(6)
Average Rating	-0.006**	-0.005	-0.005	-0.000	0.000	0.001
	(0.003)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)
# Ratings in 1000	0.002***	0.002***	0.003***	0.001***	0.001***	0.001**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Dummy: Video	0.072***	0.070***	0.077***	0.020***	0.019***	0.025***
	(0.014)	(0.014)	(0.016)	(0.005)	(0.005)	(0.006)
Dummy: Website	0.037**	0.038**	0.036*	0.008	0.008	0.008
	(0.017)	(0.017)	(0.021)	(0.005)	(0.005)	(0.006)
Privacy Policy	0.075***	0.075***	0.065**	0.013	0.013	0.010
	(0.026)	(0.026)	(0.032)	(0.009)	(0.009)	(0.012)
$DPrice$	-0.640			0.054		
	(0.453)			(0.096)		
Log. Price	0.047			-0.005		
	(0.039)			(0.008)		
Categ: Education		-0.015			0.003	
		(0.018)			(0.006)	
Categ: Entertain.		0.029*			0.005	
		(0.016)			(0.006)	
Categ: Games		0.011			0.010	
		(0.018)			(0.006)	
Categ: Tools/Perso.		-0.031*			0.001	
		(0.016)			(0.006)	
Categ: Lifestyle		0.036**			0.004	
		(0.017)			(0.006)	
Categ: Health		0.007			0.002	
		(0.026)			(0.010)	
Local Market Share			0.044**			0.001
			(0.019)			(0.007)
Competitors' Inst.			0.002***			0.000**
			(0.000)			(0.000)
Constant	2.579***	2.025***	2.021***	-0.004	0.044***	0.043***
	(0.451)	(0.023)	(0.022)	(0.098)	(0.008)	(0.008)
Observations	67404	67404	51504	67404	67404	51504
Developers	22410	22410	19755	22410	22410	19755
Adjusted R^2	0.005	0.007	0.008	0.003	0.003	0.003

NOTES: The table analyzes the driving factors of developers' permission usage and the resulting *privacygrade* that was obtained in 2014. The regressions are panel fixed OLS regressions with a developer fixed effect. The dependent variable in columns 1-3 is the grade that was obtained in 2014 (1=A+, 5=F) and, in columns 4-6, a dummy that indicates that the app will receive a failgrade (C or F). Col. 1&4 analyze paid and free apps seperately. Col. 2&5 analyze categories seperately and col. 3&6 add variables that measure the strength of the competitors. Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table A2: Multi App Developers Learning - First App Success

	sensitive	potential malicious	# pot. malic.	privacygrade	failgrade
	(1)	(2)	(3)	(4)	(5)
Dummy: Top-Developer	0.118	0.150***	0.276	0.100	0.017
	(0.092)	(0.057)	(0.326)	(0.126)	(0.036)
Apps by Developer	0.000**	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
app1 installs	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
app1 ratings	0.000**	0.000**	0.000**	0.000***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
app1 rel strength	-0.095	-0.263**	-0.588	-0.448***	-0.062*
	(0.109)	(0.127)	(0.539)	(0.128)	(0.037)
app2 installs	0.000	0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
app2 ratings	-0.000	-0.000*	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
app2 rel strength	0.037	-0.021	-0.042	0.057	0.034
	(0.082)	(0.082)	(0.402)	(0.145)	(0.061)
Average Rating	-0.019**	-0.039***	-0.063*	-0.028**	0.001
	(0.008)	(0.007)	(0.032)	(0.013)	(0.005)
# Ratings in 1000	0.001***	0.001**	0.008**	0.001	0.000
	(0.001)	(0.000)	(0.003)	(0.001)	(0.000)
Dummy: Video	0.024	0.062	-0.098	-0.056	-0.008
	(0.044)	(0.046)	(0.181)	(0.041)	(0.014)
Dummy: Website	-0.021	-0.007	-0.216	0.080	0.027
	(0.041)	(0.045)	(0.200)	(0.053)	(0.020)
Privacy Policy	0.099**	0.083**	0.677***	0.091	0.020
	(0.045)	(0.040)	(0.256)	(0.098)	(0.026)
Constant	0.506***	0.894***	2.866***	2.163***	0.054**
	(0.044)	(0.045)	(0.209)	(0.069)	(0.025)
Observations	71491	71491	71491	29743	29743
Developers					
Adjusted R^2	0.024	0.022	0.013	0.019	0.007

NOTES: The table analyzes whether app developers' success on the first and second apps is associated with their permission usage and the resulting (2014) *privacygrade* on subsequent apps. The regressions are panel OLS regressions clustered standard errors at the developer level. The dependent variables in the five columns are: (1) usage of sensitive permissions (2) usage of potentially malicious permissions (according to Google) (3) number of potentially malicious permissions (according to Google) (4) the *privacygrade* that was obtained in 2014 (1=A+, 5=F), and (5) a dummy that indicates that the app will receive a failgrade (C or F). Robust Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table A3: Robustness: Multi App Developers and Environment excl. Developers' first two apps.

	sensitive				potential malicious			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Rating	-0.003	-0.005***	-0.003**	-0.005***	-0.004***	-0.008***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
# Ratings in 1000	0.001***	0.001***	0.001***	0.001***	0.000	0.001**	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dummy: Video	0.034***	0.031***	0.024**	0.031***	0.029***	0.021**	0.019*	0.021**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Dummy: Website	0.025**	0.025**	0.020**	0.022*	0.018	0.018	0.016	0.015
	(0.011)	(0.011)	(0.009)	(0.012)	(0.015)	(0.016)	(0.016)	(0.017)
Privacy Policy	0.104***	0.111***	0.075***	0.102***	0.059**	0.071**	0.052*	0.059*
	(0.030)	(0.031)	(0.027)	(0.035)	(0.024)	(0.029)	(0.028)	(0.031)
DPrice	-0.819***				-1.147***			
	(0.148)				(0.160)			
Log. Price	0.058***				0.079***			
	(0.012)				(0.013)			
Categ: Education		-0.038**				-0.050***		
		(0.017)				(0.016)		
Categ: Entertain.		0.025*				0.010		
		(0.014)				(0.011)		
Categ: Games		-0.013				-0.007		
		(0.015)				(0.014)		
Categ: Tools/Perso.		0.023				-0.050***		
		(0.014)				(0.011)		
Categ: Lifestyle		0.029*				-0.016		
		(0.016)				(0.013)		
Categ: Health		-0.007				-0.035***		
		(0.016)				(0.013)		
Maturity Rating n.a.			-0.004				-0.039	
			(0.020)				(0.025)	
Medium Maturity			0.203***				0.099***	
			(0.017)				(0.012)	
Low Maturity (13+)			0.401***				0.188***	
			(0.018)				(0.019)	
High Maturity (17+)			0.175***				0.070***	
			(0.022)				(0.017)	
Local Market Share				0.086***				0.104***
				(0.018)				(0.018)
Competitors' Inst.				0.001***				0.001***
				(0.000)				(0.000)
Constant	1.141***	0.430***	0.324***	0.436***	1.714***	0.766***	0.692***	0.728***
	(0.141)	(0.017)	(0.011)	(0.012)	(0.152)	(0.018)	(0.016)	(0.016)
Observations	71502	71502	71502	56082	71502	71502	71502	56082
Developers	8720	8720	8720	7887	8720	8720	8720	7887
Adjusted R^2	0.033	0.006	0.150	0.007	0.084	0.006	0.045	0.008

NOTES: The table analyzes the driving factors of developers' permission usage. The regressions are panel fixed OLS regressions with a developer fixed effect. The dependent variable in columns 1-4 is a dummy indicating the usage of sensitive permissions and, in columns 5-8, a dummy that indicates potentially malicious permissions. Col. 1&5 analyze paid and free apps seperately. Col. 2&6 analyze categories seperately (baseline: "Business"). Col. 3&7 shows how intrusiveness varies for different maturity ratings (baseline: "Everybody"), and col. 4&8 add variables that measure the strength of the competitors. Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table A4: Robustness: Multi App Developers and Environment excl. Developers' first two apps. *privacygrade*

	privacygrade				failgrade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Rating	-0.002	-0.001	-0.001	-0.002	0.002	0.002	0.002	0.003
	(0.004)	(0.004)	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
# Ratings in 1000	0.002*	0.002*	0.002*	0.003	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Dummy: Video	0.067***	0.065***	0.060***	0.074***	0.021**	0.021**	0.020**	0.030***
	(0.022)	(0.021)	(0.021)	(0.025)	(0.009)	(0.008)	(0.008)	(0.010)
Dummy: Website	0.037	0.038	0.034	0.052	0.012	0.011	0.010	0.018
	(0.030)	(0.030)	(0.030)	(0.036)	(0.009)	(0.009)	(0.009)	(0.011)
Privacy Policy	0.043	0.045	0.026	0.034	0.004	0.004	0.000	0.000
	(0.036)	(0.036)	(0.035)	(0.044)	(0.013)	(0.013)	(0.013)	(0.017)
DPrice	-1.725*				-0.128			
	(0.968)				(0.214)			
Log. Price	0.142*				0.012			
	(0.083)				(0.019)			
Categ: Education		-0.006				0.006		
		(0.028)				(0.010)		
Categ: Entertain.		0.002				-0.000		
		(0.025)				(0.011)		
Categ: Games		-0.000				0.007		
		(0.029)				(0.010)		
Categ: Tools/Perso.		-0.044				0.007		
		(0.027)				(0.011)		
Categ: Lifestyle		-0.008				0.000		
		(0.026)				(0.009)		
Categ: Health		-0.052				-0.009		
		(0.042)				(0.017)		
Maturity Rating n.a.			0.023				0.001	
			(0.030)				(0.010)	
Medium Maturity			0.156***				0.035***	
			(0.023)				(0.009)	
Low Maturity (13+)			0.243***				0.061***	
			(0.020)				(0.009)	
High Maturity (17+)			0.102***				0.025***	
			(0.025)				(0.009)	
Local Market Share				0.039				-0.000
				(0.024)				(0.009)
Competitors' Inst.				0.001***				0.000
				(0.000)				(0.000)
Constant	3.723***	2.097***	2.006***	2.063***	0.208	0.062***	0.047***	0.054***
	(0.965)	(0.039)	(0.030)	(0.036)	(0.218)	(0.013)	(0.010)	(0.013)
Observations	29753	29753	29753	22888	29753	29753	29753	22888
Developers	6404	6404	6404	5698	6404	6404	6404	5698
Adjusted R^2	0.003	0.004	0.029	0.005	0.001	0.001	0.012	0.003

NOTES: The table analyzes the driving factors of developers' permission usage and the resulting *privacygrade* that was obtained in 2014. The regressions are panel fixed OLS regressions with a developer fixed effect. The dependent variable in columns 1-4 is the grade that was obtained in 2014 (1=A+, 5=F) and, in columns 5-8, a dummy that indicates that the app will receive a failgrade (C or F). Col. 1&5 analyze paid and free apps seperately. Col. 2&6 analyze categories seperately (baseline: "Business"). Col. 3&7 shows how intrusiveness varies for different maturity ratings (baseline: "Everybody"), and col. 4&8 add variables that measure the strength of the competitors. Robust standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

B Additional Methodological Approach: Apps over Time

B.1 Analyzing Apps over Time

To deepen this analysis we exploit the fact that we can study apps over time. To shed light on developers' main strategies of gaining/managing access to personal user data over time, we can analyze two strategies. We can show to which extent developers added permissions as function of success, by investigating the slope with which permissions are added. Second we can analyze whether developers initially role out their apps at low levels of intrusiveness, and add in more intrusive permissions on later updates.

Specifically we can study the apps on a weekly interval 2012, and then observe it in 2014/2016, in combination with the data from *privacygrade.com*. This allows us to move beyond the cross section and highlight how the development of an app over time can predict its behavior with respect to privacy. For example, we study whether it is possible to predict the 2014 privacy grade based on 2012 behavior, and specifically we can study whether bad 2014 grade is a function of 2012-6-month success and the competitive situation in the market. Moreover, we can study what happens when apps switch from the paid to the free model. Finally, we want to highlight at what point in an app's life cycle developers tend to introduce high intensity permissions (if at all).

$$Privacy_{it+1} = \alpha_i + \gamma P_{it} + \theta X_{2,it} + \varepsilon_{it}. \tag{4}$$

In this regression $Privacy_{it+1}$ measures the app's future privacy sensitivity, α_i is a fixed effect, P_{it} measures the app's price (which is potentially 0), and $X_{2,it}$ are other control variables, such as the app's competitive environment or the app's past success (=user base).

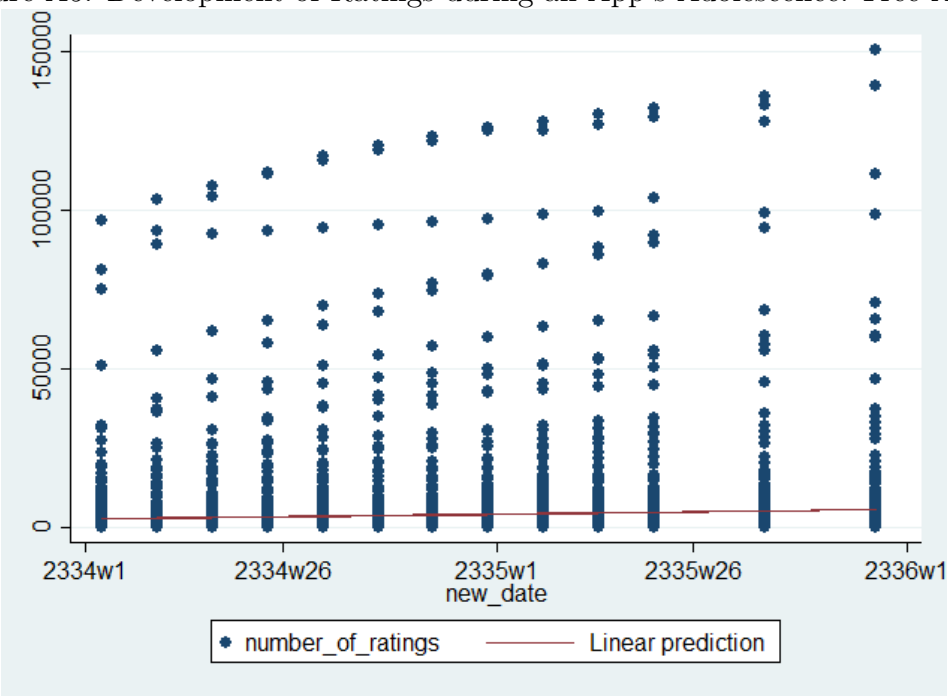
Exploiting the time dimension within apps also allows us to gain deeper insight on how developers' data strategies influence their apps' success and continued innovation? This question is actually hard to analyze in cross sectional data and can only be understood if the temporal dynamic in the data can be exploited. Our strategy is to look for major shifts in the permissions, and/or for sensitive permissions that were introduced without improved functionality. We want to use such shifts in a regression discontinuity design, where we estimate the counterfactual growth path of the app absent the new permissions based on its previous growth path, or based on the growth path of comparable app. Any kinks in the growth path would be attributed to the change in permissions. While permissions may be negatively correlated with downloads, the data might be a valuable input for continued innovation. Hence, we will also study the frequency of updates as a measure of an app's innovativeness.

$$AppSuccess_{it+1} = \alpha_i + \beta D_{it}^{Privacy} + \gamma P_{it} + \theta X_{1,it} + \varepsilon_{it}. \tag{5}$$

We measure *AppSuccess* via continued downloads (demand), innovativeness and long-term survival as indicator of viability. Privacy, price and the control variables have been described above. The key challenge in this analysis will be capturing redundant permissions. We will use three strategies to tackle this challenge: First, we exploit the Play Store categorical information, to identify permissions which are hardly ever used by apps in the same category. Second, we can exploit pairs of free and paid apps in the data, where the paid app can serve as "technological benchmark," that allows the researcher to see what is needed for the app's functionality. Third and finally, we can use the information on *privacygrade.com* to filter out which apps are malicious and which permissions they typically use.

Special Identification Strategy - Early App Development and "Excessive" Data Use? Our main strategy for identification expoits apps that were newly launched during our first data acquisition in 2012. Our data structure allows us to identify apps that were launched between April and September 2012 and to focus on this inflow sample of apps. For these apps, we can analyze and categorize their early success on the market, and their continued development. Important dimensions for this issue are variables like initial growth, the competitive environment, the app's initial permission usage, and the developer's experience. Once we characterized the app's early properties and success patterns, we can predict the app's grade on *privacgrade.org* as primary outcome that measures data behavior, and we can identify early-stage patterns that are more likely to result in problematic privacy grade.

Figure A5: Development of Ratings during an App's Adolescence: Free Apps



B.2 Initial Descriptive Results

In this section we show results for a selected sample of 750 apps which were all launched at the same point in time (Spring 2012) and were all featured by Google as a "top new app." Conditioning on a such narrowly defined set of apps allows us to study their development over time. Moreover, being able to observe the app's adloescence allows us to leverage greater variation in our measure of installations, because Google's step function has much smaller intervals initially. Ongoing work evaluates the timing of when permissions were introduced and how initial dynamics translate into better or worse privacy grades in 2014.

Figure A6: Development of Ratings during an App's Adolescence: Paid Apps

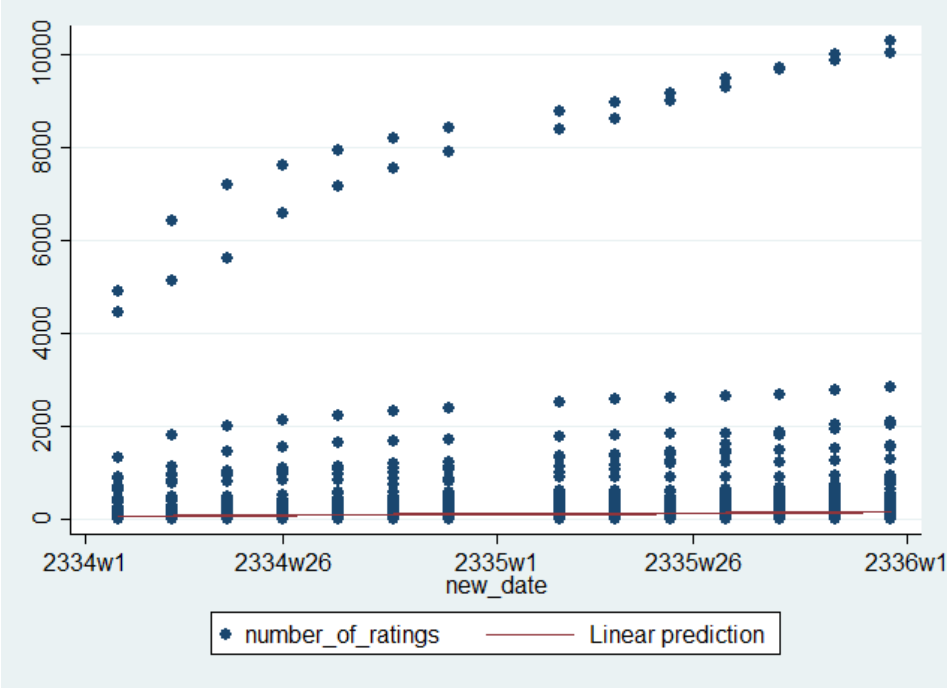


Figure A7: Development of Installations during an App's Adolescence: Paid Apps

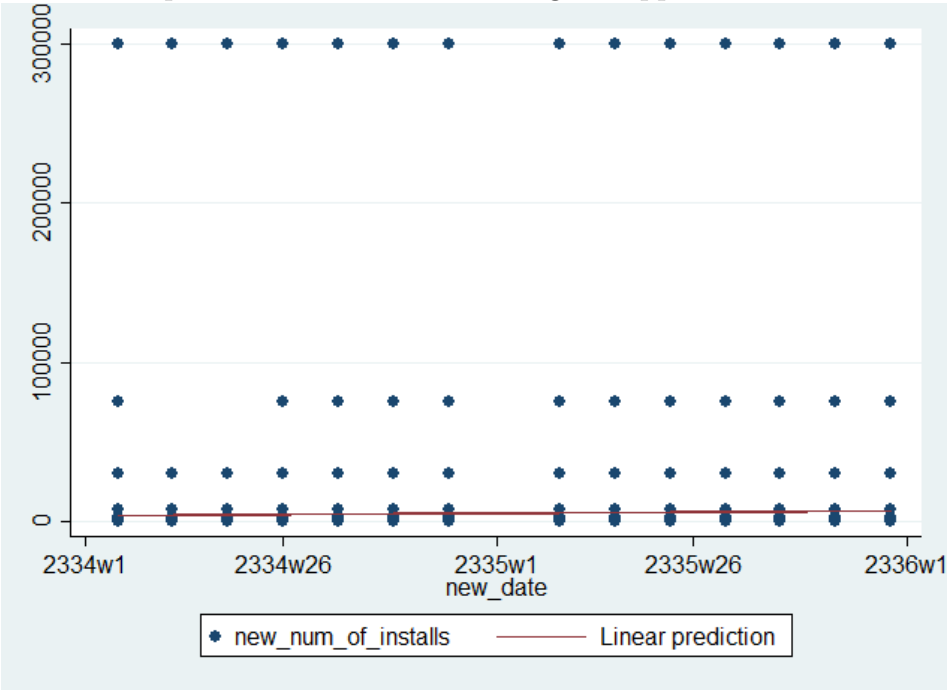


Table A5: App Adolescence and Future *privacygradde*/2016 Permissions

	failgrade		privacygrade		# permissions 2016	
	(1)	(2)	(3)	(4)	(5)	(6)
Average Rating	0.034	0.028	0.054	0.035	-0.447	-0.689
	(0.029)	(0.034)	(0.075)	(0.087)	(0.397)	(0.430)
# Ratings in 1000	-0.001	0.002	-0.001	0.002	-0.030	0.007
	(0.005)	(0.003)	(0.011)	(0.007)	(0.046)	(0.019)
Dummy: Video	0.090	0.101	0.297*	0.320**	0.415	1.135
	(0.064)	(0.065)	(0.152)	(0.156)	(0.668)	(0.744)
Dummy: Website	0.062	0.076	0.156	0.188	0.668	0.349
	(0.063)	(0.065)	(0.148)	(0.152)	(0.725)	(0.722)
Dummy: Privacy Policy	-0.050	-0.062	-0.080	-0.114	2.972***	2.486**
	(0.063)	(0.063)	(0.145)	(0.146)	(0.952)	(0.975)
Initial Growth: Ratings	0.000		0.000		0.000	
	(0.000)		(0.000)		(0.000)	
Initial Growth: Inst. Levels	-0.072**		-0.146		0.822**	
	(0.035)		(0.090)		(0.376)	
Initial Growth: # Instal.	-0.000		-0.000		0.000	
	(0.000)		(0.000)		(0.000)	
Early Growth Acceleration		0.149		0.342		4.642***
		(0.109)		(0.286)		(1.170)
Constant	0.096	-0.103	2.306***	1.886***	6.414***	4.253*
	(0.118)	(0.189)	(0.314)	(0.487)	(1.853)	(2.398)
Observations	274	265	274	265	372	310
Adjusted R^2	0.012	0.007	0.009	0.011	0.065	0.099

NOTES: The table analyzes the driving factors of developers' permission usage and the resulting *privacygrade* that was obtained in 2014. The regressions are panel fixed OLS regressions with a developer fixed effect. The dependent variable in columns 1- a dummy that indicates that the app will receive a failgrade (C or F) and, in columns 3 & 4, is the grade that was obtained in 2014 (1=A+, 5=F). Col. 5 & 6 show the number of permissions used in 2016. Col. 1, 3 & 5 analyze app growth in the app's first months, and col. 2, 4 & 6 analyze growth acceleration (based on ratings). Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01