Kummer, Michael; Schulte, Patrick

Working Paper

Money and privacy: Android market evidence

ZEW Discussion Papers, No. 14-131

Provided in Cooperation with:
ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Kummer, Michael; Schulte, Patrick (2014) : Money and privacy: Android market evidence, ZEW Discussion Papers, No. 14-131, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim,
https://nbn-resolving.de/urn:nbn:de:bsz:180-madoc-375394

This Version is available at:
https://hdl.handle.net/10419/106514

Discussion Paper No. 14-131

Money and Privacy – Android Market Evidence

Michael Kummer and Patrick Schulte

ZEW
Zentrum für Europäische Wirtschaftsforschung GmbH
Centre for European Economic Research

Discussion Paper No. 14-131

Money and Privacy – Android Market Evidence

Michael Kummer and Patrick Schulte

Money and Privacy - Android Market Evidence*

Michael Kummer†
Centre for European
Economic Research (ZEW)

Patrick Schulte‡
Centre for European
Economic Research (ZEW)

December, 2014

Abstract

We study the role of privacy in the market for mobile applications. For such programs used with smartphones and tablet PCs a very important market has emerged. Yet, neither the role of privacy on that market is well understood, nor do we have empirical evidence regarding its role therein. We exploit data on 300,000 mobile applications and almost 600 "applications-pairs" to analyze both sides of this market: First, we analyze the price that application suppliers charge for more privacy. Second, we study how users' installations are related to the "personal data greediness" of mobile applications.

We provide the first empirical evidence on the main assumptions of recent early models on suppliers' and consumers' strategies in this market. Our results show that (1) consumers take it into account when applications request rights to collect private information and (2) suppliers ask for more rights if they offer an app for free than if they offer it for a fee.

JEL Classification: D4, D83, L15

Keywords: Information Acquisition, Mobile Applications, Smartphones, Online Privacy, Permissions, Price, Privacy Regulation

*We are grateful to Irene Bertschek, Anindya Ghose, Avi Goldfarb, Sang-Pil Han, Andres Hervas-Drane, Arnold Picot, Imke Reimers, Rahul Telang, Bernd Theilen, Catherine Tucker, Hal Varian, Michael Ward, Manfred Wittenstein, Pinar Yildirim, Pai-Ling Yin, Michael Zhang and Christine Zulehner for valuable comments and helpful advice. We thank the participants of the 12th ZEW ICT conference 2014, EARIE 2014, WISE 2014. We thank Niklas Duerr, Florian Hofbauer and Steffen Viete for their extremely useful research assistance. For the authors' other projects please refer to http://www.zew.de.

†P.O. Box 103443, D–68034 Mannheim. Email: kummer@zew.de.

‡P.O. Box 103443, D–68034 Mannheim. Email: schulte@zew.de.

1 Introduction

In this paper, we exploit data on 300,000 mobile applications to analyze the role of privacy in this market. The recent rise of smartphones and tablet PCs has been the most significant change in the market of end user computational devices over the last decade. According to OECD's estimates, the number of users who access the internet via mobile devices is currently surpassing the number of users using a fixed line.[1] This is due to the emergence of a market for mobile applications (henceforth "apps") which allows users to tailor their devices to their needs. However, while apps have become one of the most important two-sided markets, little is known about how this market actually works.

Especially, little is known about the role of privacy in the market for mobile apps. How do suppliers trade direct revenues for more usage and the possibility of getting access to private data? Do users avoid "data greedy" apps consciously? We provide evidence on how consumers react to the (privacy-relevant) permissions that apps request upon installing them. Moreover, we use data on almost 600 "sister-pairs" of apps: two versions of the same app, where one is offered for free and the other one for pay. These pairs are used to analyze the supply side behavior of app developers and the price they charge for additional privacy. We can thus provide urgently needed empirical evidence to inform recently emerging economic theories that relate privacy to competition (Spiegel (2013), Casadesus-Masanell and Hervas-Drane (forthcoming)).

We use a cross-section of all apps available in 2012 to estimate the the effect of privacy-sensitive permissions on app success. The data was collected from one of the two major players for mobile applications, the Android market (in 2012). An app's success is measured by both, the number of installations and turnover (if applicable) Our results indicate a positive relationship between the number of installations and the number of requested permissions. However, at the same time, demand is lower for apps which request privacy-relevant rights. In our analysis of the supply side, we find strong evidence that free apps have a strong tendency to ask for more privacy-sensitive permissions than paid apps. We can confirm this finding using the matched dataset of app-pairs. This is the first evidence on the price at which suppliers are willing to forgo privacy-sensitive information.

We believe that our research provides important insights into consumers' privacy concerns when using smartphones as well as into suppliers valuation of user data. Also, our results shed light on stylized facts that are necessary for building theoretical models aimed

[1] See e.g. OECD Broadband Database (2012).

at understanding this topic and for further estimations.

The remainder of the paper is organized as follows. Section 2 reviews and summarizes the related literature. Section 3 describes our data set and presents descriptive evidence. Section 4 introduces the empirical framework we use and presents the results we obtain with the cross-section data and with the pairs data. Section 5 concludes and discusses limitations of the current approach.

2 Related Literature

We study how potentially data greedy mobile applications perform in the market relative to less demanding apps. First empirical evidence on this question was gathered by researchers at the OECD, who arrived at the conclusion that it is a decisive issue for the success or failure of an app whether and how it collects data of users.[2] According to their studies, it is of crucial importance for the consumers' trust in the entire market whether users know about the data that is being collected and how they will react to "data-greedy" apps. Clearly, the issue of consumer trust should not be underestimated, since a lack of trust could result in a massive hindrance for the development of the market.

To our knowledge, platforms where applications for mobile devices can be downloaded have hardly been investigated empirically. One exception is the paper by Ghose and Han (2014), where they estimate the demand for selected (and top-rated) apps using a structural empirical approach for the estimation (Berry et al. (1995), Nevo (2001)) of demand. Ghose and Han (2014) focus on the 300 top-rated apps on the platforms, and compare them for Android and Apple. The present study differs from theirs in three important ways. First, we have a different focus by analyzing the role of privacy in such markets, for both, demand and supply. Second, we observe the complete set of apps that was available in the Android Market in summer 2012 ($N = 300,000$). Third, we observe a categorical measure of downloads, rather than approximating demand via the sales ranks.

We evaluate if and how users react to "privacy greedy" permissions and whether they avoid installing apps that request more or very sensitive rights. Experimental and survey based research has investigated consumer's attitudes towards privacy in the smartphone market. Contrasting consumers' willingness to pay to protect privacy and their willingness to accept for giving away their personal information showed that the willingness to pay is much lower. (Grossklags and Acquisti (2007), Tsai et al. (2011)) Also a recent survey

[2]See OECD (2013).

based study (Savage and Waldman (2013)) found that consumers' self-reported willingness to accept, giving away the personal information that is typically shared with developers, is near 4 USD. The choice architecture of the platform affects smartphone users' willingness to pay premiums to limit their personal information exposure (Egelman et al. (2013)).

More technical analyses investigate the precise meaning of certain permissions and what they imply for the privacy of the device's owner (Chia et al. (2012)). Other studies investigated how dangerous apps can potentially be (z.B. Chia et al. (2012) or Fahl et al. (2012)). They carefully studied a smaller number of observations and they are mostly concerned with technical aspects. Yet, despite the fact that mobile apps are a relatively new phenomenon, it is possible to use existing and established methods to estimate demand for an app and to analyze how this demand is influenced by the rights an app asks for. Examples of such methodology can be found in existing demand estimations such as Danaher (2002), Iyengar et al. (2008) or Kim et al. (2010). Moreover, earlier studies that analyzed the software industry or the substitutability of fixed and mobile telephony (e.g. Ward and Woroch (2010) or Briglauer et al. (2011) and references therein) are a valuable source of theoretical predictions.

A final contribution lies in analyzing the suppliers' price for privacy, i.e. how much suppliers charge in exchange for requesting less privacy sensitive information of their customers. Almost nothing is known about how the supply side deals with privacy concerns. Preibusch and Bonneau (2013) analyze data collection policies of internet sites, and find that the intensity with which they collect data varies substantially. Several models have analyzed online privacy. In such models the knowledge about a personal preference of an agent can be used to price discriminate (Wathieu (2002), Taylor et al. (2010)). In such models firms can use customer information, such as the purchase history, to charge personalized prices in settings of electronic retailing (Taylor (2004), Acquisti and Varian (2005), Conitzer et al. (2012)). One of the more surprising findings showed, that it can be beneficial for consumers if the cost of hiding personalized information increases.

An alternative way to use the personal information is the context of direct marketing, which may result in costly efforts to avoid ads (Johnson (2013), Hann et al. (2004)). Increasing the cost of anonymity can benefit consumers, but only up to a point, after which the effect is reversed (Taylor et al. (2010)). Generally, these models see reduced privacy as a source of inefficiency. However, in the context of software production, providing free software in a bundle with (targeted) ads could also be welfare improving if the cost of

producing software is relatively low (Spiegel (2013)).

The ambiguity in the theoretical results provides a good description of the trade-off that is present on platforms for mobile apps. On the one hand, consumers might suffer disutility from potentially intrusive apps, on the other hand, many valuable services can be provided "for free" and create benefits for a much larger group of users. This ambiguity is also a dominant theme in a recent analysis of a situation where suppliers compete in privacy Casadesus-Masanell and Hervas-Drane (forthcoming). They analyze the effect of privacy sensitive information in a setting of competition in two-sided markets, as it was pioneered by Armstrong (2006) or Rochet and Tirole (2003). In the context of mobile apps, it is important to consider the potential repercussions of restricting the use of private information to both sides of the market.

We analyze the relationship between users' downloads and the privacy sensitive information they have to provide to the supplier. Baccara (2007) considers situations in which consumers are business clients who might suffer great losses when their private data are leaked. A series of recent studies analyzed how privacy policies affect users of social networks or the success of targeted advertisement (Goldfarb and Tucker (2011), Tucker (2012), Tucker (2014)). Restrictions on the usage of private data in advertisement substantially reduced targeting effectiveness. This resulted in lower revenues for the content site, but also highlighted that privacy policies have an important effect on consumer behavior.

The main contribution of our paper consists of analyzing the role of apps' "demand for access to personal information", about the users. We observe detailed information on the permissions (rights) that the app requests before installation. We provide evidence about how users' installations are related to the access permissions requested by an app. Moreover, we analyze the behavior of suppliers who offer the same app (i) for pay, but with limited access to personal information and (ii) for free, but with greater access to the user's personal data.

In addition, we provide insights into whether and how strongly the users of one of the most important platforms for mobile apps react to the current warning mechanisms about potentially malicious usages of permissions. If these mechanisms turn out to be insufficient this might hint to the necessity of considering other ways for warning users about the potential dangers of installing a specific app.

3 Data and Descriptive Evidence

In this section we briefly describe the data we use to analyze the role of privacy in app markets and provide some descriptive evidence on this issue.

3.1 Data

We collected data from one of the two largest platforms for mobile applications - namely Google's Play Store (formerly Android Market). In 2012 (and 2014) we extracted publicly available information on most apps available at that time.[3] Thus, our data set covers the full population of products available in 2012 (around 300,000). Figure 4 in the Appendix shows the design of Google's Play Store in 2012 and thus which information we were able to extract. The data set includes a rich set of app specific characteristics. In contrast to most cases of internet-based data sets, where demand variables have to be approximated via sales ranks (see e.g. Chevalier and Mayzlin, 2006, Garg and Telang, 2013 or Ghose and Han, forthcoming), our data set provides direct information on the total number of installations and sales for each app. This measure exists in a discrete form (17 levels, e.g. 1-5 installations, 6-10 installations, 11-50 installations, etc.). Next to the sales measure, we have information on many more app characteristics which are relevant for explaining app demand (see e.g. Ghose and Han, 2014), such as:

- app price,
- app version,
- required Android version,
- app size (in KB),
- number and level of ratings,
- length of app description (in number of characters),
- number of screenshots,
- dummy whether a video is available,
- the category an app belongs to (which include "games", "news & magazines", "communication", "books", in total 30 categories),
- top-developer dummy,
- number of apps by the same developer,
- average number of installations of apps by the same developer,
- content rating (everyone, low maturity, medium maturity, high maturity, not rated),

[3]The information was collected from April to October 2012,and once more in 2014, which allows us to analyze the development of these apps.

In addition, the play store provides for each application information on related apps via a section "users who viewed this also viewed", which can be used to easily identify groups of competing apps. We use this information to link competitors characteristics to each other, e.g. we construct the following three additional control variables:

- average price of competitor apps,
- average installations of competitor apps,
- average rating of competitor apps.

Most importantly, however, Google's Play Store provides precise insights into the permissions an app uses. This allows us, as well as app users, to understand in a detailed way which rights an app has, and thus which functions it can perform, including functions which allow an app to collect private information about the app user. In 2012, Google had defined 136 different of such permissions an app could use, such as e.g. "full internet access", "fine gps location", "read browser data", etc. (for some more examples, see Table 5). These permissions have to be declared in the app description and have to be accepted by the app user before installing the app.[4] Figure 5 in the Appendix illustrates the way the permissions in the app store are declared and described. That way, we have quite precise and detailed insights into the functionality of an app and into its ability to collect various types of information about a user.

To analyze the role of privacy-relevant permissions for demand and supply, we group the permissions according to their type of functionality and type of risk. We hereby follow 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, among which 13 are considered as being a risk to privacy.[5] Table 5 provides a short description for each of the permissions. In addition, it shows the remaining 13 critical permissions and how we group the set of critical permissions defined by Sarma et al. (2012) into various subsets. First, we construct a dummy which is equal to one if an app uses at least one permission of the privacy-relevant permissions. Next, we split this group into four subgroups: location-, profile-, communication- and ID-relevant permissions. Location-relevant permissions allow an app to identify a user's exact or coarse location. Profile-relevant permissions allow an app to get insights into the user's profile, i.e. into its

[4]Google also provides for each permission a standardized short explanation to inform users about the permission's meaning.

[5]These permissions, allowing an app to collect private information, include: "fine gps location", "intercept outgoing calls", "read calendar events", "read contact data", "read sms or mms", "receive sms", "receive mms", "receive wap", "coarse network based location", "read browser history and bookmark", "record audio", "read phone state and identity".

browsing behavior, its contacts, its calender data. Communication-relevant permissions are those which allow an app to read and to edit communication, that is e.g. to read sms or mms or to record audio. This privacy-relevant permission group is the least often used one. Last, we define an extra group for the permission "read phone state and ID", the ID-relevant permission group, which allows an app to identify the unique phone ID and thus to identify unambiguously the user's identity.

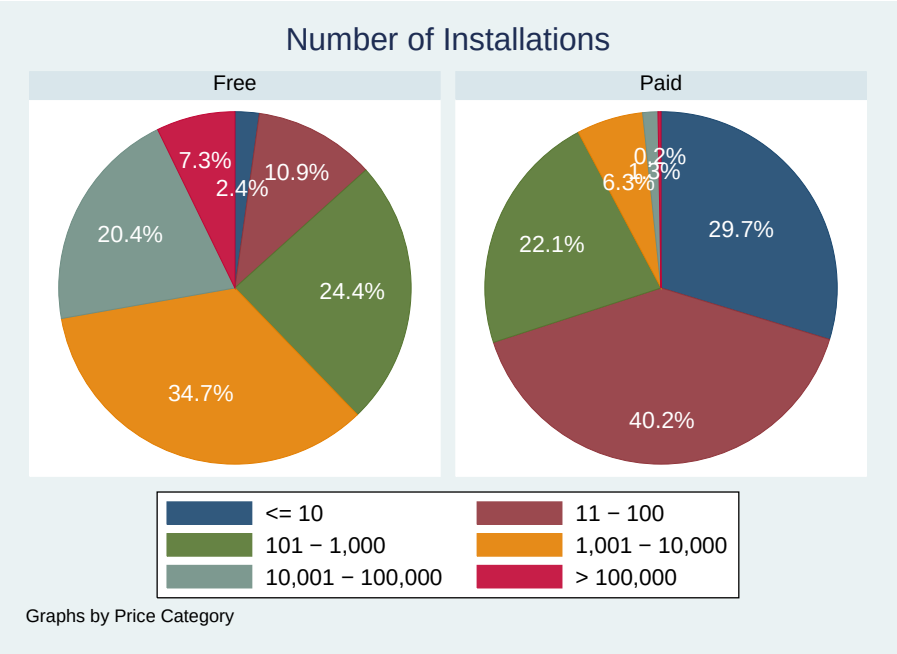
Next to the classification offered by Sarma et al. (2012), we use Google's own classification of 'potentially malicious apps' to classify permissions into subgroups. For 38 of the 136 available permissions, the official permission statement includes the note that the respective permissions might be 'potentially malicious', that is, it might be used to harm the user of this app. We combine this classification with that of Sarma et al. (2012) and form two groups: potentially malicious privacy-relevant permissions and not potentially malicious privacy-relevant permissions (see again Table 5).

3.2 Descriptive Evidence

In this section we provide first descriptive insights into the role of privacy in app markets. Figures 1 and 2 provide a summary of the distribution of installations and permissions of apps. Both figures distinguish between apps which are for free and apps which have a positive price. As can be seen, those two samples differ strongly with respect to the distribution of installations and the number of permissions. As Figure 1 shows, only around 13 percent of free apps have less or equal to 100 installations, whereas around 70 percent of apps for which a user has to pay a price have less or equal to 100 installations. Regarding the number of required permissions per app similar differences can be seen (see Figure 2). Apps which are for free typically require a higher number of permissions. For example, 65 percent of free apps use up to 5 permissions, whereas around 65 percent of paid apps require less than 3 permissions.

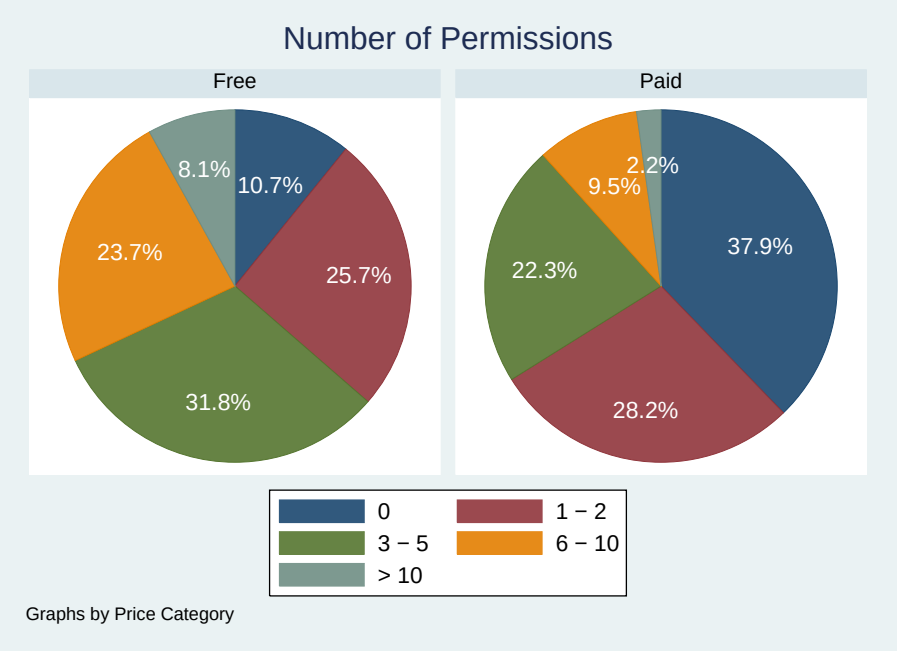
In addition, Table 6 (in the Appendix) summarizes our main variables for the cross-section and another data set, the pairs data, which we introduce later in more detail. There we also provide separate numbers for free and for paid apps. In line with the pie charts, free apps are installed more often than paid apps. At the same time they also use far more permissions (on average 4.65 vs. 2.22 permissions) and especially use more permissions allowing an app to collect private data about the user. On average, free apps use 1.16 of such permissions, whereas paid apps use on average only 0.39 of

Figure 1: Number of Installations by Free and Paid Apps



Notes: The Figure shows the distribution of installations by free and by paid apps. The underlying sample is the full cross-section of apps.

Figure 2: Number of Permissions by Free and Paid Apps



Notes: The Figure shows the distribution of the number of permissions requested by free and by paid apps. The underlying sample is the full cross-section of apps.

them. The same trend holds for all subgroups of privacy-relevant permissions. 38 (14) percent of free (paid) apps use the "read phone state and ID" permission (D_{ID}). 33 (10) percent of free (paid) apps use at least one permission of the location-relevant permission

group ($D_{Location}$). For communication-relevant permissions ($D_{Communication}$) the difference between free and paid apps is smaller (8 vs. 5 percent). In contrast, permissions of the profile-relevant permission group ($D_{Profile}$) are again used much more often by free apps (14 percent of the apps use at least one such permission) than by paid apps (5 percent). The average price of paid apps is equal to 2.15 US dollar.

Finally, Table 7 shows the distribution of app categories. In 2012, the most frequent category is the "personalization"-category, to which apps belong which facilitate the personalization of a smartphone (e.g. via individual wallpapers, ringtones, launchers etc.). 37708 apps or 12.4 percent of all apps belong to it. The next largest categories are "Entertainment", "Tools", "Brain & Puzzle" and "Books & Reference". All five categories which contain games together account for 12.3 percent of all apps.

4 Estimation and Results

This section discusses the estimation and our findings. The first subsection presents the results of the full cross-section. Subsequently, we discuss our findings on a dataset of matched "sister-pairs" of apps, where the same app is provided for pay and for free.

4.1 Cross-Section Data

We first present results of our baseline specification, where we analyze the relationship between app demand, approximated by the number of installations of an app, and the privacy-intrusiveness of an app. Subsequently, we present further results where we consider alternative proxies of demand as well as additional outcome variables of interest.

4.1.1 Baseline Specification

To analyze the relationship between app demand and permissions which allow an app to collect personal information we use econometric analysis. We apply a straightforward empirical demand specification, which models demand for an app as a function of its permissions, its price and other observable characteristics. Our main specification is based on the cross-section sample. The models were estimated using both, simple OLS and 2SLS that accounts for the endogeneity of the price. We estimate the following baseline model:

$$Demand_i = \alpha + \beta D_i + \gamma P_i + \theta X_i + \varepsilon_i \tag{1}$$

$Demand_i$ for app i is approximated by the number of installations of an app. Demand is a function of a vector of permission group dummies D_i, the app's price P_i and a set of observable characteristics X_i, including the log of the average rating, the app version, the app category, the length of the app description, the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the logarithmic number of apps of its developer, the average number of installations of the app's developer's apps, the minimum and the maximum compatible android version as well as information about the app's competitors' characteristics. ε_i is the error term. In what follows we shall focus on the simple OLS, because the estimated parameters of interest do not differ very much after instrumenting. The IV-estimations are available upon request. Despite controlling for a large variety of app characteristics having an influence on the demand for an app, we expect unobserved heterogeneity to potentially bias our estimates. Especially, we expect unmeasured quality of apps to be positively correlated with both, app demand and permission usage, and thus expect our permission-related estimates to be potentially upward biased. In reaction, subsequently, we check for robustness to alternative proxies of app success as well as to alternative econometric assumptions.

The results of our main specification are given in Table 1. Columns 1 to 3 show results for apps which are for free, whereas the remaining columns show results for apps one has to pay for before installing them. In all specifications, increasing the total number of permissions an app uses by one permission is related to an increased number of installations. For free apps, this increase is between 1.2 to 2.7 percent, whereas for payable apps the coefficient is a bit larger, between 2.8 and 4.4 percent. Specifications 1 and 4 include a dummy which takes the value of one, if an app uses at least one of the permissions allowing an app to collect private data about its user. In both specifications, this dummy shows no significant coefficient, whereas the remaining permission group dummies are significantly different from zero. Thus, permissions allowing an app to collect private data do not generally reduce demand for an app. However, as the remaining specifications will show, splitting the group of privacy-relevant apps into subsets shows, that certain types of permissions matter, while others do not. Specification 2 and 5 split them into four subgroups: ID, location, communication and profile. For free apps, we find a negative relationship for three out of the four groups. The permission allowing an app to identify the unique ID of the smartphone (D_{ID}) comes with a reduction of app demand by 3.5 percent, whereas the permissions allowing an app to collect information about its

user's profile ($D_{Profile}$) reduce app demand by 13.7 percent. The strongest coefficients are associated to the location group dummy ($D_{Location}$). Using at least one permission of this group comes with a demand reduction of 24 percent. This is a rather large effect, which, given our specification, should be considered with care. In contrast to those three groups, not related to a demand reduction are permissions controlling communication ($D_{Communication}$). Using such a permission is associated with increased demand (plus 5.7 percent). However, among the four subgroups of privacy-relevant permissions, the communication group is the most special which is used by only 8 percent of all free apps. That is, this result might be driven by rather specific apps.

Specification 3 and 6 contain an additional split of the privacy-relevant permissions into two subgroups using Google's definition of potentially malicious permissions. For the group of potentially malicious permissions we find a surprisingly large and significant negative coefficient. If a free app uses at least one permission, allowing an app to collect private data and being classified by Google as potentially malicious, this comes with a reduction in app demand by 32.9 percent. In contrast, permissions allowing to collect private information but not being classified as potentially malicious comes with a significant increase in app demand by 3.8 percent.

Specifications 4, 5 and 6 which analyze the relationship between privacy-relevant permissions and app demand for payable apps include the price of the app. The relationship is as one would expect negative, with a price elasticity of demand of around -0.06, that is, a price increase by one percent comes with reduced demand (0.06 percent). Regarding the role of the permissions, as for the free apps, we find mainly negative or insignificant coefficients. Specification 5 splits the privacy-relevant permissions into the same four subgroups as before, which results in negative coefficients for two of the four groups. The permission allowing an app to identify the unique phone ID comes with reduced demand, and also the communication permissions are negatively related to app demand. The permissions allowing an app to collect information about the user's location and profile are both insignificant. In specification 6, the privacy-relevant permissions are, like before, split into permissions which are flagged as potentially malicious by Google and those that are not. App demand for payable apps is only reduced by 8.4 percent in case a potentially malicious privacy-relevant permission is used. "Non malicious" privacy-relevant permissions are not significantly related to app demand.

In summary, we find either insignificant or significant negative coefficients. An app

Table 1: Cross Section - Relationship between Installations and Permissions

	Free Apps			Paid Apps		
Dep. Var.: Log. Installations	(1)	(2)	(3)	(4)	(5)	(6)
Log. Price				-0.062***	-0.060***	-0.062***
				(0.010)	(0.010)	(0.010)
Total Permissions	0.012***	0.025***	0.027***	0.028***	0.044***	0.035***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.007)	(0.006)
$D_{Privacy}$	-0.015			0.004		
	(0.014)			(0.023)		
$D_{Internet}$	-0.162***	-0.155***	-0.172***	0.073***	0.062**	0.070***
	(0.016)	(0.016)	(0.015)	(0.018)	(0.019)	(0.018)
D_{Ads}	0.136***	0.113***	0.107***	-0.154***	-0.163***	-0.162***
	(0.012)	(0.013)	(0.012)	(0.022)	(0.024)	(0.023)
D_{Other}	0.025*	0.023	0.013	0.051**	0.042*	0.050**
	(0.012)	(0.012)	(0.012)	(0.019)	(0.020)	(0.019)
D_{ID}		-0.035**			-0.116***	
		(0.013)			(0.026)	
$D_{Location}$		-0.240***			0.025	
		(0.020)			(0.039)	
$D_{Communication}$		0.057**			-0.192***	
		(0.021)			(0.035)	
$D_{Profile}$		-0.137***			-0.011	
		(0.017)			(0.040)	
$D_{MaliciousPrivacy}$			-0.329***			-0.084**
			(0.016)			(0.032)
$D_{NonmaliciousPrivacy}$			0.038**			-0.022
			(0.013)			(0.023)
Constant	2.802***	2.723***	2.668***	2.052***	2.081***	2.052***
	(0.221)	(0.221)	(0.221)	(0.331)	(0.331)	(0.331)
Category	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	191239	191239	191239	111687	111687	111687
Mean of dep. Var.	0.90	0.90	0.90	-3.51	-3.51	-3.51
SD of dep. Var.	2.84	2.84	2.84	2.84	2.84	2.84
Adjusted R^2	0.495	0.496	0.496	0.453	0.453	0.453

Notes: Dependent variable: log number of installations. The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group. Controls include: the log of the average rating, the app version, the app category, the length of the app description, the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the logarithmic number of apps of the developer, the average number of installations of the app's developer's apps, the minimum and the maximum compatible android version as well as information about the app's competitors' characteristics. All specifications are estimated using the Ordinary Least Squares estimator (OLS). Specifications (1) to (3) use only free apps, specifications (4) to (6) use only paid apps. Heteroscedasticity-robust standard erros in parentheses. ***, **, * significantly different from 0 at the 1%, 5%, and 10% levels, respectively.

developer using privacy-relevant permissions can expect to have lower installation numbers. Also the findings from columns 3 and 6 are noteworthy: If a permission is flagged as "potentially malicious" by Google, it is associated with massively lower demand, while without Google's flag the relationship is even *positive*. Obviously, the coefficient overestimates the effect of Google's flags, because flagged permissions should be expected to be more problematic and the concerned apps might also differ in reviews and in other ways. Nevertheless, this difference is striking: the coefficients suggest a large effect of warning

users about problematic permissions.

Comparing the results for free apps and their for-pay-counterparts, shows that we often find smaller coefficients for payable apps than for free apps. Moreover, the reaction to specific types of permissions typically is not the same. Tracking a user's ID is always associated with lower demand. However, tracking the location and building a profile appears to be more acceptable for a payable application than for a free app. In contrast, communication-related permissions on payable apps show a strong negative relationship with downloads despite showing a positive one for free apps.

The different role of permissions for free and for payable apps can also be seen when introducing cross-terms in a unified estimation. This is shown in columns 5 and 6 of Table 8 in the Appendix. Introducing cross-terms between privacy-relevant permissions and a price dummy, shows that using more than one privacy-relevant permission comes with a smaller demand reduction for payable apps than for free apps. In Columns 1-4 this is done for free and priced apps separately. The regressions show the effect of the number of privacy-relevant permissions. For both, free and paid apps, we find that if an app uses only one privacy-relevant permission, this is not significantly related to app demand. But if more than one permission is used, this is related to a lower number of installations.

The number of installations is, from our perspective, the most direct measure of app demand. However, alternative outcome measures which are of interest and which describe the success of an app are also available. These are analyzed in the next section.

4.1.2 Further Insights

In this section we show results describing the relationship between the privacy-relevant permission groups and the growth rate of the number of installations, the number of ratings an app receives and the probability of an app to exit the market. In addition, we analyze the relationship between the app developer's decision to offer an app for free or for pay (including its price level choice) and its decision to use privacy-relevant permissions. Results are given in Table 2.

Columns 1 and 2 show the results for our specifications analyzing growth of installations.[6] For free and for pay apps we find either a significant negative or an insignificant relationship. Especially, the location-relevant (for free apps) and the profile-relevant per-

[6]To analyze the relationship between permission use in 2012 and installation growth, we use a new cross-section of the installation data from 2014 and use it to compute the growth rate in the number of installations for the time between 2012 and 2014.

missions are strongly negatively related to installations growth.

In specifications 3 and 4, we analyze how the number of ratings is correlated with the use of the privacy-relevant permission groups. We consider this measure as a measure of usage intensity, and also as an alternative proxy of app demand, given that app installations and an app's number or ratings are highly correlated.[7] Again, location- and profile-relevant permissions show a strongly negative relationship for both free and paid apps, that is, they come with a reduced number of ratings. In general, for payable apps all four permission groups show a significant negative effect. In contrast, for free apps, communication-relevant permissions and the permission to access the unique phone ID are positively correlated with the number of ratings.

Also an app's exit can be interpreted as a sign of an app's (lacking) success. For free apps (column 5) we found a positive relationship between permission usage and app exits (three out of four permission groups). For payable apps we find more heterogeneous results. Location- and profile-relevant permissions are correlated negatively with app exits, that is, apps using such permissions are more likely to *survive*.

The final set of estimates analyzes a developer's decision to offer an app for free or for pay and their choice to ask for specific types of privacy sensitive permissions. We find very strong results showing that apps being offered for free ask for many more permissions than apps being offered for a positive price. This result holds for all types of permissions. Also, if an app developer decides to ask for a price upon installation, the price level is negatively correlated with two of the permission groups, namely profile- and location-relevant permissions. Communication-relevant permissions are not significantly related to the price level, whereas the permission allowing identifying the app user is positively correlated with the price of the app.

Overall, these results confirm our previous findings. We conclude from these results, that (1) privacy-relevant permissions and especially profile- and location-relevant permissions might be negatively related to demand for (free) apps, (2) that in general, privacy-relevant permissions are used much more often if an app is offered for free and (3) that in case an app is offered as a paid app, its price is negatively related to the prevalence of profile- and communication-relevant permissions. However, we are aware that we are using only a cross-section, due to data limitations.[8] Hence, those results have to be in-

[7] In our sample, regressing the log number of installations on the log number of ratings of a free (paid) app gives a highly significant coefficient of 0.41 (0.28) and shows an $adj.R^2$ of 0.56 (0.43).

[8] The information was originally collected at a weekly resolution from April to October 2012, with the intention of exploiting within-variation in the apps. However, in the variables of interest there is very

terpreted with great care. Several unobserved factors could drive our results. The next section provides one idea how to improve our identification.

4.2 Pairs Data

In this section we exploit a specific subsample of selected app pairs: App developers often offer two versions of the same app, one version can be downloaded for free and one version for which users have to pay a price before installation. The costly version typically offers some advantage over the free version: It may either offer additional functions, contain no or less in-app advertisement, and/or the costly version is also associated with fewer (privacy sensitive) permissions.[9] This feature of app pairs offers an opportunity for the researcher, because the paid version can serve as a technological counterfactual in a specific sense: Any permission that is not required by the paid version, is not necessary for the functioning of the app. Hence, any permission that is present in the free version but not in the paid sister, is definitely redundant for functionality, and, instead, related to monetization.[10] Moreover, free and paid sister applications are potentially very close substitutes, which only differ in their price and in displaying advertisements but otherwise having identical functionality.

We exploit these two features of app pairs to better understand to which extent developers are collecting user data as an alternative to monetary compensation. Moreover, the existence of app pairs, in principle, invites an estimation based on the differences between the two apps. The difference in the requested permissions could be used to predict the difference in installations when conditioning on price and the potential difference in the service quality. Preliminary results are provided in the end of this section.

To achieve a sample of more homogeneous pairs of applications, we manually identified pairs of apps that had no discernible difference in the amount of services it offered. For that, we manually identified app pairs which stated as only difference in their app descriptions that the free version uses advertisements, whereas the paid one does not use them. Note that this is necessary, because apps generally differ in the amount of services they provide, introduction the danger of confounding the estimation. We can account for these differences, by looking for pairs of apps, where the paid version does not provide

little weekly within variation, which is why we had to use the cross-section data.

[9]These pairs are identified using a word processing algorithm which identifies app pairs having the same name except for one of the following addings: 'free', 'paid', 'lite', 'full', 'demo', 'pro', 'premium', 'donate', 'trial', 'plus'.

[10]Note, that it is safe to assume that the free app does not offer more services than the paid.

Table 2: Cross Section - Further Insights

	Growth		User Assessment		App Survival		Supply	
Dep. Var.:	Δ Log. Installations		Log. Number of Ratings		$D_{Exit=1}$		$D_{Price>0}$	Log. Price
Sample	(Free)	(Paid)	(Free)	(Paid)	(Free)	(Paid)	(Both)	(Paid)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log. Price		0.734***		0.426***		-0.018***		
		(0.137)		(0.079)		(0.002)		
Total Permissions	0.007**	-0.017*	0.039***	0.071***	0.005***	0.009***	0.007***	0.027***
	(0.002)	(0.007)	(0.003)	(0.007)	(0.001)	(0.001)	(0.000)	(0.002)
D_{ID}	-0.015	-0.057*	0.100***	-0.188***	0.060***	0.071***	-0.059***	0.104***
	(0.011)	(0.028)	(0.011)	(0.019)	(0.003)	(0.005)	(0.002)	(0.009)
$D_{Location}$	-0.200***	-0.016	-0.182***	-0.249***	0.014***	-0.031***	-0.148***	-0.155***
	(0.017)	(0.037)	(0.016)	(0.025)	(0.004)	(0.007)	(0.002)	(0.009)
$D_{Communication}$	0.018	-0.068*	0.201***	-0.104***	-0.038***	0.012	-0.023***	0.001
	(0.018)	(0.031)	(0.018)	(0.024)	(0.004)	(0.007)	(0.003)	(0.013)
$D_{Profile}$	-0.072***	-0.108**	-0.175***	-0.084**	0.111***	-0.029***	-0.083***	-0.143***
	(0.018)	(0.042)	(0.014)	(0.033)	(0.003)	(0.008)	(0.003)	(0.014)
$D_{Internet}$	-0.030**	-0.130***	-0.266***	-0.046***	0.044***	0.024***	-0.189***	0.083***
	(0.011)	(0.018)	(0.013)	(0.012)	(0.003)	(0.004)	(0.002)	(0.006)
D_{Ads}	0.172***	0.240***	0.136***	-0.090***	0.029***	0.014**	-0.153***	-0.042***
	(0.010)	(0.022)	(0.011)	(0.016)	(0.003)	(0.004)	(0.002)	(0.007)
D_{Other}	-0.023*	-0.143***	-0.016	-0.114***	-0.002	-0.044***	0.044***	0.207***
	(0.011)	(0.034)	(0.010)	(0.019)	(0.002)	(0.004)	(0.002)	(0.006)
Log. Installations (in 1000)					-0.010***	-0.001		
					(0.000)	(0.001)		
Constant	-2.782***	-3.609***	1.761***	1.274***	-0.206***	-0.429***	-0.161***	0.010
	(0.167)	(0.244)	(0.172)	(0.192)	(0.045)	(0.070)	(0.034)	(0.101)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	129856	77960	191239	111687	191239	111687	302926	111687
Mean of dep. Var.	0.97	0.69	1.21	-4.28	0.32	0.30	0.37	0.27
SD of dep. Var.	1.45	1.61	5.15	6.57	0.47	0.46	0.48	0.76
Adjusted R^2	0.067	0.006	0.903	0.973	0.200	0.171	0.410	0.244

Notes: Δ Log. Installations represents the difference in the log number of installations between 2012 and 2014. The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group. Controls include: the log of the average rating, the app version, the app category, the length of the app description, the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the logarithmic number of apps of the developer, the average number of installations of the app's developer's apps, the minimum and the maximum compatible android version as well as information about the app's competitors' characteristics. All specifications are estimated using the Ordinary Least Squares estimator (OLS). Heteroscedasticity-robust standard errors in parentheses. ***, **, * significantly different from 0 at the 1%, 5%, and 10% levels, respectively.

additional services. This occurs naturally when the paid app is ad free, but otherwise does not offer any additional service. Sometimes, it is even the case that there is no discernible difference in the apps whatsoever, they only differ in the permissions and the price.

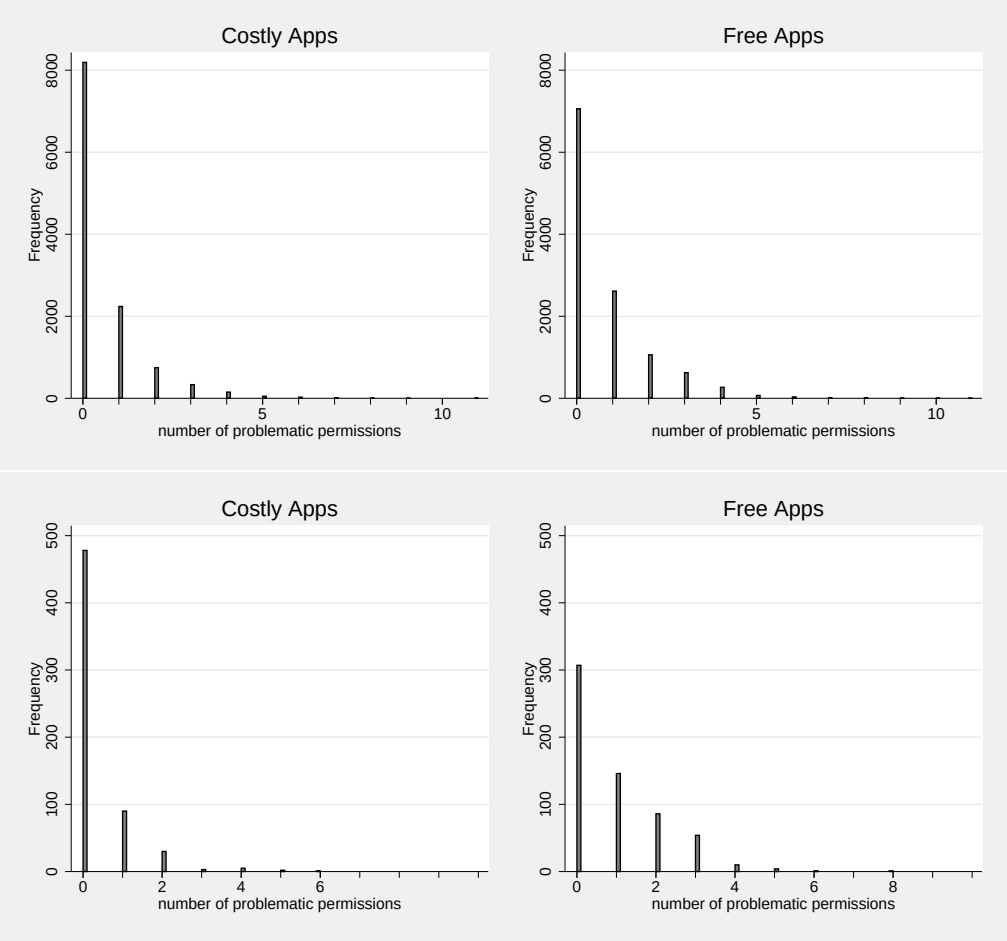
In line with the full cross-section, the number of installations and the number of permissions in the sample of free apps exceed the one of the paid apps. However, in the pairs sample we observe a higher average number of installations than in the full cross-section. This is not too surprising given that we here include only apps which are offered by developers capable and willing to provide two versions of an app, a free one and a priced one. This requires more resources and should be correlated with higher capabilities and higher quality of products and thus also with a higher number of installations.

A pairwise comparison of the privacy sensitive permissions reveals, first, that most of the pairs do not differ in the number of the permissions they request. Second, we find that the free version of an app is much more likely to request privacy sensitive permissions than the paid one. This can be seen in the histograms in Figure 3. The upper row contrasts all app pairs, and the lower row contrasts pairs which differ in nothing but the display of advertisements. In both samples the large majority of free apps does not request privacy-relevant permissions. However, if a privacy-relevant permission is introduced by one version and not by the other, it is by far more likely that it is the free version. In fact, once we condition on the same level of services (lower row), only in 2 out of more than 600 pairs the paid version asks for more permissions than the free one, while more than 200 free apps ask for one or more redundant permissions. This can be seen in Table 3, where we subtract the costly app's number of privacy sensitive permissions from the free app's number of permissions. A positive number indicates that the free app requests more privacy sensitive permissions. We do this for both definitions of privacy-relevant permissions, the one by Sarma et al. (2012) and Google's own criterion. Free apps ask for more permissions independently of which criterion is applied, but interestingly, the pattern is less striking for the "more visible" privacy-relevant permissions that are also flagged by Google.[11]

To conclude this section, we report the results from an estimation attempt that relates differences in the number of privacy-relevant permissions to differences in the installations. For these regressions we focus on the sample of pairs which only differ in advertisement or show no discernible at all. In principle, both, differentiating the app through the presence

[11] An interesting avenue for research would lie in analyzing potential shifts in developer behavior, after Google introduced these warnings. Unfortunately these data are not available to the authors.

Figure 3: The distributions of privacy sensitive permissions (free vs. paid services).



Notes: The figure contrasts the distribution of privacy sensitive permissions (according to Sarma et al.) in the free and paid versions of apps. The two upper histograms show this distribution for all pairs we found, the lower one for the selection of pairs, which are likely to differ only in that the free version uses ads. The left side of the graph shows paid apps, whereas the right one shows the distribution for free applications. For both samples the number of free apps are more likely to use privacy sensitive permissions.

Table 3: Privacy sensitive permissions used in the free, but not in the paid service.

No.	Privacy (Sarma et al.)	Privacy (Google)
-2	0.0	1.0
-1	2.0	0.0
0	383.0	515.0
1	130.0	68.0
2	55.0	19.0
3	35.0	4.0
4	4.0	0.0
5	0.0	2.0
Total	609.0	609.0

Notes: The table shows the difference in privacy sensitive permissions between the free and the paid version of the same app (free-paid). The unit of observation is a pair of two "sister-apps" (same app by the same developer), where the developer offers one for free and the other one for pay. A positive number means that the free version uses more privacy sensitive permissions, we show the difference according to both privacy criteria, the one based on Sarma et al. and permissions flagged as potentially malicious by Google. The table shows (i) that the free versions of the same app tend to use more privacy sensitive permissions, but(ii) that only a third of the apps differentiate.

or absence of ads, or trading a few dollars for more privacy, seem to be viable deals: the first would allow to identify the price that is being asked in exchange for the ability to show advertisements. The second group, which trades permissions for money, would shed light on the developers' price charged for fewer permissions. However, we end up with very few app pairs and we are slightly worried about the representativeness of the applications that we identify by our procedure, especially, about the apps that do not mention any additional service when charging money. Hence, we focus on the apps that differentiate via ads and we generally point out that the subsequent result should be taken with a grain of salt.

Table 4 shows the results of our estimations using selected app pairs. The first two regressions show the relationship of the "market share" in installations (share of paid installations in total installations of app pair) or reviews and the difference in privacy-relevant permissions. The third column analyzes the relationship of price and permissions. As is easily seen from these regressions, there is no statistically significant relationship of the privacy differential and the dependent variables. This may be driven by several factors: (i) we might simply have lost too much power to discern the effect, (ii) it might indicate that the developers do not use this dimension systematically to price discriminate and users do not systematically avoid free apps that "charge" redundant permissions.

Table 4: Relationship of demand for an app and the difference in permissions between paid and free - pairs, which differ only in the display of ads, but not in the services.

	ln(share installations)	ln(share ratings)	ln(price)
	(1)	(2)	(3)
free app demands D_{ID}	0.070	-0.202	-0.079*
	(0.170)	(0.182)	(0.047)
free app demands $D_{Location}$	-0.120	-0.030	0.000
	(0.183)	(0.193)	(0.051)
free app demands $D_{Communication}$	-1.815	-1.771	1.407***
	(1.541)	(1.344)	(0.427)
free app demands $D_{Profile}$	-0.476	-0.417	-0.166
	(0.688)	(0.599)	(0.193)
Log. Price	0.099	-0.343*	
	(0.164)	(0.179)	
Constant	-5.298***	-2.842***	0.023
	(0.084)	(0.090)	(0.024)
Observations	478	317	478
Adjusted R^2	-0.005	0.012	0.022

Notes: Dependent Variable: Column 1: log of the paid app's share of installations over total pair installations. column 2: log of the paid app's share of ratings over the total number of ratings. Column 3: log of the paid app's price. The independent variable is a dummy, which takes the value 1 if the free version requests a specific privacy sensitive permission, which is not present in the paid app (and price, where applicable). Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01.

Actually, preliminary results (not shown, but available upon request) point in the opposite direction.[12]

The descriptive evidence on app pairs clearly shows that developers have a tendency to request more permissions in the free version of their apps than in the paid version. However, from the regression results we cannot infer that there is systematic avoidance of privacy sensitive permissions. Moreover, we cannot see that users are willing to pay more (or developers do not charge more), for less privacy intrusion. Further research is aimed at digging deeper into these, a priori, conflicting results, as will be discussed in the next section.

5 Conclusion, Limitations and Further Research

In this paper, we analyze the role of privacy in the market for mobile apps. Specifically, we analyze the role of the permissions allowing apps to collect private information about its users and which mobile apps request upon installation of an app. Our results suggest that apps indeed request a lot of permissions for accessing the mobile device's primary functions when being installed. Moreover, free apps clearly request more access to users'

[12]If "differentiating apps" are systematically more popular (also on the free segment of the market), trying to work off the market share might lead to additional systematic problems, because of the sheer differences in market sizes. Yet, using the number of downloads directly, might lead to other problems in the specification.

personal data than paid apps. We analyze how users react to these permissions. Our analysis highlights two conflicting forces. On the one hand, we observe that the number of installations and the number of requested permissions seems to be positively correlated. This indicates that consumers do not refrain from installing apps that request many permissions. On the other hand, we see that the usage of permissions which allow collecting private information tends to be negatively correlated with installations. In continued research we analyze the behavior of the supply side in order to provide first empirical evidence to inform recently emerging economic theories relating privacy to competition in two-sided markets (Spiegel (2013), Casadesus-Masanell and Hervas-Drane (forthcoming)). We tackle the difficulties that stem from the large heterogeneity of apps (e.g. in quality and scope) by focusing on app-pairs (free and costly) of a single producer, which in addition allows us to analyze the price that suppliers offer for additional privacy. Doing so, we can confirm that free apps clearly request more access to users' personal data than paid apps.

Our approach suffers from several limitations, which force us to leave important questions to further research. Most importantly, its somewhat descriptive scope and the fact that we cannot account for unobserved product heterogeneity limit our approach. A panel analysis will be a first step to remedy these problems. Ongoing research is attempting to approach the issue via a better measure of downloads allowing us to use a balanced panel and a panel of newly published apps. Moreover, our analysis of the supply side does not exactly match the setting of the models we wish to test (Spiegel (2013), Casadesus-Masanell and Hervas-Drane (forthcoming)). We also lack a normative benchmark, against which to compare our results in order to evaluate whether the reaction of consumers is appropriate or not. However, the app-pairs are a first step of better understanding the role of redundant privacy sensitive apps. Even at the current stage, our research is able to provide first information on the role of permissions and their relation to the frequency of downloads. We believe that it is a crucial first step towards a better picture of these new markets, consumers' privacy concerns and their understanding of possibly problematic access permissions in mobile applications.

References

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

Armstrong, Mark, "Competition in Two-Sided Markets," *The Rand Journal of Economics*, 2006, *37* (3), 668–691.

Baccara, Mariagiovanna, "Outsourcing, Information Leakage, and Consulting Firms," *The Rand Journal of Economics*, 2007, *38* (1), 269–289.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Briglauer, Wolfgang, Anton Schwarz, and Christine Zulehner, "Is Fixed-Mobile Substitution Strong Enough to De-regulate Fixed Voice Telephony? Evidence from the Austrian Markets," *Journal of Regulatory Economics*, 2011, *39* (1), 50–67.

Casadesus-Masanell, Ramon and Andres Hervas-Drane, "Competing with Privacy," *Management Science*, forthcoming.

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 Hui, Yusuke Yamamoto, and N Asokan, "Is this App Safe?: A Large Scale Study on Application Permissions and Risk Signals," in "Proceedings of the 21st international conference on World Wide Web" ACM 2012, pp. 311–320.

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.

Danaher, Peter J, "Optimal Pricing of New Subscription Services: Analysis of a Market Experiment," *Marketing Science*, 2002, *21* (2), 119–138.

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

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

Hann, Il-Horn, Kai-Lung Hui, IPL Png, and Sang-Yong Tom Lee, "Direct Marketing: Privacy and Competition," 2004.

Iyengar, Raghuram, Kamel Jedidi, and Rajeev Kohli, "A Conjoint Approach to Multipart Pricing," *Journal of Marketing Research*, 2008, *45* (2), 195–210.

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

Kim, Youngsoo, Rahul Telang, William B Vogt, and Ramayya Krishnan, "An Empirical Analysis of Mobile Voice Service and SMS: A Structural Model," *Management Science*, 2010, *56* (2), 234–252.

Nevo, Aviv, "Measuring Market Power in the Ready-to-eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

OECD, "Broadband database," 2012.

_ , "The App Economy," *OECD Digital Economy Papers*, 2013.

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.

Rochet, Jean-Charles and Jean Tirole, "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 2003, *1* (4), 990–1029.

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," *Discussion Paper in Economics, University of Colorado at Boulder*, 2013.

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

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," *Economic Research*, 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.

Ward, Michael R and Glenn A Woroch, "The Effect of Prices on Fixed and Mobile Telephone Penetration: Using Price Subsidies as Natural Experiments," *Information Economics and Policy*, 2010, *22* (1), 18–32.

Wathieu, Luc, "Privacy, Exposure and Price Discrimination," *Harvard Business School Marketing Research Paper No. 02-03*, 2002.

6 Appendix

6.1 Tables

Table 5: Permission Group Definitions

Permissions (Group)	Description
$D_{Internet}$	
full internet access	Allows apps to open network sockets.
D_{Ads}	
view network state	Allows apps to access information about networks.
$D_{Privacy}$	
D_{ID}	
read phone state and ID	Allows read only access to phone state.
$D_{Location}$	
coarse location	Allows an app to access approximate location derived from network location sources such as cell towers and Wi-Fi.
fine gps location	Allows an app to access precise location from location sources such as GPS, cell towers, and Wi-Fi.
$D_{Communication}$	
intercept outgoing calls	Allows an app to see the number being dialed during an outgoing call with the option to redirect the call to a different number or abort the call altogether.
read sms or mms	Allows an app to read SMS and MMS messages.
receive sms	Allows an app to monitor incoming SMS messages, to record or perform processing on them.
receive mms	Allows an app to monitor incoming MMS messages, to record or perform processing on them.
record audio	Allows an app to record audio.
receive wap	Allows an app to monitor incoming WAP push messages.
$D_{Profile}$	
read calendar events	Allows an app to read the user's calendar data.
read contact data	Allows an app to read the user's contacts data.
read browser data	Allows an app to read (but not write) the user's browsing history and bookmarks.
read sensitive log data	Allows an app to read the low-level system log files.

Notes: The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group. $D_{MaliciousPrivacy}$ combines the definition of privacy-relevant permissions $D_{Privacy}$ and Google's definition of potentially malicious permissions. It consists of the following permissions: fine gps location, intercept outgoing calls, read calendar events, read contact data, read sms or mms, receive sms,receive mms, receive wap. $D_{NonmaliciousPrivacy}$ consists of: coarse network based location, read browser history and bookmark, recordaudio, read phone state and identity. D_{Other} consists of: mount and unmount file systems, add or modify calendar events and send, write contact data, write browser history and bookmark, edit sms or mms, modify delete usb storage contents, control near field communication, view configured accounts, create bluetooth connections, bluetooth administration, directly call any phone numbers, send sms messages.

Table 6: Summary Statistics Cross Section - Free & Paid Apps

	Cross Section		Pairs	
	Free	Paid	Free	Paid
Outcome Variables				
Installations (in 1000)	105.84	1.56	281.61	2.64
Average Rating	3.93	3.98	3.88	4.09
Permissions				
Total Permissions	4.65	2.22	4.28	1.89
Critical Permissions	3.14	1.40	3.20	0.96
Privacy Permissions	1.16	0.39	0.91	0.32
Malicious Privacy Permissions	0.43	0.16	0.26	0.12
Nonmalicious Privacy Permissions	0.71	0.23	0.64	0.20
$D_{Privacy}$	0.51	0.23	0.50	0.21
D_{ID}	0.38	0.14	0.36	0.13
$D_{Location}$	0.33	0.10	0.25	0.04
$D_{Communication}$	0.08	0.05	0.06	0.05
$D_{Profile}$	0.14	0.05	0.08	0.06
$D_{MaliciousPrivacy}$	0.33	0.12	0.21	0.09
$D_{NonmaliciousPrivacy}$	0.46	0.19	0.47	0.18
$D_{Internet}$	0.83	0.45	1.00	0.27
D_{Ads}	0.61	0.23	0.88	0.08
D_{Other}	0.42	0.29	0.36	0.27
App Characteristics				
Price	0.00	2.15	0.00	1.35
App Version	2.96	2.42	3.02	2.56
Size (in KB)	2677.99	5146.67	2272.48	1980.42
Length Description	759.14	993.19	978.40	858.13
Number Screenshots	3.46	3.62	4.03	3.99
Dummy: Video	0.10	0.10	0.14	0.13
Dummy: Top-Developer	0.01	0.01	0.01	0.01
Apps by Developer	105.72	193.15	13.36	13.36
Average Installations of Developer	88.24	40.17	69.52	130.43
Observations	191239	111687	610	610

Notes: The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group.

Table 7: Summary Statistics Cross Section - App Categories

	Total Number	Share
Personalization	37708	12.4
Entertainment	33350	11.0
Tools	27374	9.0
Brain & Puzzle	19642	6.5
Books & Reference	19426	6.4
Education	15274	5.0
Lifestyle	15175	5.0
Travel & Local	13109	4.3
Arcade & Action	11451	3.8
Casual	10802	3.6
Music & Audio	10581	3.5
Productivity	10332	3.4
Sports	10314	3.4
Business	8317	2.7
Health & Fitness	7241	2.4
Communication	7157	2.4
News & Magazines	6147	2.0
Social	5920	2.0
Finance	5417	1.8
Media & Video	4551	1.5
Photography	3941	1.3
Medical	3490	1.2
Shopping	2988	1.0
Cards & Casino	2892	1.0
Transportation	2690	0.9
Sports Games	2013	0.7
Comics	1756	0.6
Libraries & Demo	1519	0.5
Weather	1327	0.4
Racing	1022	0.3
Observations	302926	

Notes: This table shows the distribution of the 30 app categories available in the Android Market in 2012.

Table 8: Cross Section - Demand Side

	Free		Paid		Cross Terms	
	(1)	(2)	(3)	(4)	(5)	(6)
Log. Price			-0.062***	-0.066***		
			(0.010)	(0.010)		
Total Permissions	0.012***	0.034***	0.028***	0.051***	0.005**	0.038***
	(0.002)	(0.003)	(0.005)	(0.007)	(0.002)	(0.003)
Privacy Permissions		-0.100***		-0.124***		-0.165***
		(0.008)		(0.017)		(0.008)
$D_{Privacy}$	-0.015	0.088***	0.004	0.126***	-0.029*	0.166***
	(0.014)	(0.016)	(0.023)	(0.027)	(0.013)	(0.016)
$D_{Internet}$	-0.162***	-0.186***	0.073***	0.046*	-0.164***	-0.200***
	(0.016)	(0.016)	(0.018)	(0.019)	(0.015)	(0.015)
D_{Ads}	0.136***	0.110***	-0.154***	-0.182***	0.163***	0.126***
	(0.012)	(0.013)	(0.022)	(0.023)	(0.012)	(0.012)
D_{Other}	0.025*	-0.000	0.051**	0.030	0.052***	0.020
	(0.012)	(0.012)	(0.019)	(0.020)	(0.012)	(0.012)
D_{Price}					-3.323***	-3.333***
					(0.015)	(0.015)
$D_{Privacy} \times D_{Price}$					0.089***	-0.097**
					(0.021)	(0.032)
$D_{Internet} \times D_{Price}$					0.299***	0.303***
					(0.023)	(0.023)
$D_{Ads} \times D_{Price}$					-0.341***	-0.355***
					(0.023)	(0.023)
$D_{Other2} \times D_{Price}$					-0.041*	-0.055**
					(0.020)	(0.020)
Privacy Permissions $\times D_{Price}$						0.095***
						(0.014)
Constant	2.802***	2.733***	2.052***	2.048***	3.692***	3.589***
	(0.221)	(0.221)	(0.331)	(0.331)	(0.185)	(0.184)
Category	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	191239	191239	111687	111687	302926	302926
Mean of dep. Var.	0.90	0.90	-3.51	-3.51	-0.72	-0.72
SD of dep. Var.	2.84	2.84	2.84	2.84	3.55	3.55
Adjusted R^2	0.495	0.496	0.453	0.453	0.658	0.659

Notes: Dependent variable: log number of installations. The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group. Controls include: the log of the average rating, the app version, the app category, the length of the app description, the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the logarithmic number of apps of the developer, the average number of installations of the app's developer's apps, the minimum and the maximum compatible android version as well as information about the app's competitors' characteristics. All specifications are estimated using the Ordinary Least Squares estimator (OLS). Specifications (1) and (2) use only free apps, specifications (3) and (4) use only paid apps and specifications (5) and (6) use free and paid apps. Heteroscedasticity-robust standard erros in parentheses. ***, **, * significantly different from 0 at the 1%, 5%, and 10% levels, respectively.

Table 8 shows estimation results describing the relationship between app installations and the use of privacy-relevant permissions. In contrast to Table 1, here no split of the privacy-relevant permissions into subgroups is performed, instead we include in addition to the permission group dummies, the number of privacy-relevant permissions used by an app. Again, we split our sample into a sample of free apps and one of paid apps. However, here we also include two specifications (5) and (6) where we combine the two samples and make use of cross-terms.

Specification (1) and (3) are identical to specification (1) and (4) in Table 1. Specification (2) includes in addition the number of privacy-relevant permissions an app uses. Including it, leads to a positive coefficient of the dummy capturing the effect of the privacy-relevant permission group. However, the number of privacy-relevant permissions itself is negatively correlated with the number of app installations. That is, if an app uses one such permissions, the combined effect of the two variables is counterbalanced. But if more than one privacy-relevant permission is used, the latter effect dominates and thus an increasing number of privacy-relevant permissions is negatively related to app installations. The same holds for paid apps (specification 4). There also, with one permission, the two effects balance each other out, whereas with more than one permission, the negative effect dominates.

Specification (5) includes the same set of variables as specification (1) and (3) but adds a price dummy as well as cross-terms being one if at least one permission of a group is used and the price of the app is > 0. In this specification, the total number of permissions still shows a small positive relationship with app installations. In contrast, the variable of interest, the dummy for privacy-relevant permissions shows a negative significant relationship with app installations. It indicates that for free apps a negative relationship with app installations exists. However, for paid apps we see combined effects of this variable and the respective cross-term is positive. Thus, in contrast to free apps, for paid apps using privacy-relevant permissions is not related to a reduction in app demand. Specification (6) is extended by including as additional variable the number of privacy-relevant permissions and its cross-term with the price dummy. Including them changes the results of the privacy-related variables. The privacy-dummy becomes positive significant, whereas the cross-term of the privacy-dummy with the price dummy becomes negative significant. As in specification (2) and (3) the number of privacy-relevant permissions balances the effect of the privacy-dummy out. That is, for both, free and paid apps, one

privacy-relevant permission has no significant effect. However, if more than one privacy-relevant permission is used, for free apps this comes with a demand reduction, whereas for paid apps it comes with an increase in demand.

6.2 Difference in Permissions for all Pairs

When discussing pairs of apps in the main body of the paper we focused on a better matched sample of pairs, to ensure that the services offered are the same, except maybe for advertisement. In this appendix we provide the comparison of permissions requested by free and paid versions for all pairs of apps that we could identify. This highlights that the greater number of privacy sensitive permissions in the free apps are prevalent in any sample of pairs.

Table 9: Contrasting privacy sensitive permissions in free and paid app pairs (all pairs).

No.	Privacy (Sarma et al. (2012)	Privacy (Google)
-6	1.0	2.0
-5	0.0	3.0
-4	6.0	3.0
-3	12.0	18.0
-2	59.0	39.0
-1	252.0	153.0
0	9,710.0	10,744.0
1	991.0	485.0
2	362.0	147.0
3	224.0	118.0
4	115.0	9.0
5	14.0	21.0
6	1.0	4.0
7	3.0	4.0
8	4.0	1.0
≤ 9	1.0	4.0
Total	11,755.0	11,755.0

Notes: The table shows the difference in privacy sensitive permissions between the free and the paid version of the same app (free-paid). The statistics are shown for all apps that were found, including pairs where the free version might offer significantly fewer services (e.g. demo versions). The unit of observation is a pair of two "sister-apps" (same app by the same developer), where the developer offers one for free and the other one for pay. A positive number means that the free version uses more privacy sensitive permissions, we show the difference according to both privacy criteria, the one based on Sarma et al. and permissions flagged as potentially malicious by Google. The table shows (i) that the free versions of the same app tend to use more privacy sensitive permissions, but(ii) that only roughly 20% of the apps differentiate.

6.3 Figures

Figure 4: App Information in the Android Market 2012

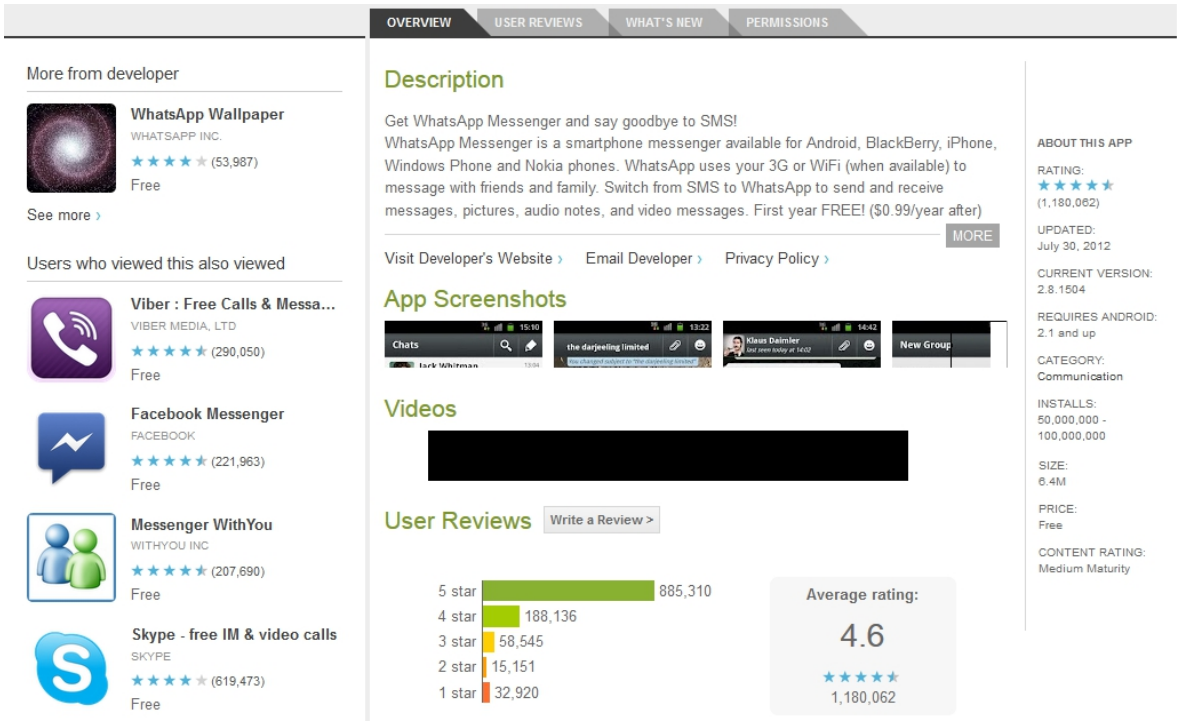


Figure 5: Permission Information in the Android Market 2012