When Private Information Settles the Bill:

Money and Privacy in Google's Market for Smartphone Applications*

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

Patrick Schulte‡
Centre for European
Economic Research (ZEW) &
Deutsche Bundesbank

May 16, 2018

Abstract

We shed light on a money-for-privacy trade-off in the market for smartphone applications ("apps"). Developers offer their apps at lower prices in return for greater access to personal information, and consumers choose between low prices and more privacy. We provide evidence for this pattern using data from 300,000 apps obtained from the Google Play Store (formerly Android Market) in 2012 and 2014. Our findings show that the market's supply and demand sides both consider an app's ability to collect private information, measured by the apps's use of privacy–sensitive permissions: (1) cheaper apps use more privacy–sensitive permissions; (2) given price and functionality, demand is lower for apps with sensitive permissions; (3) the strength of this relationship depends on contextual factors, such as the targeted user group, the apps's previous success and its category. Our results are robust and consistent across several robustness checks, including the use of panel data, a difference-in-differences analysis, the use of "twin" pairs of apps, or the use of various measures of privacy-sensitivity and app demand.

JEL Classification: D12, D22, L15, L86

Keywords: Android; Mobile Applications; Privacy; Permissions; Supply and Demand for Private Information.

---

*We thank three anonymous referees and an anonymous editor for extremely careful and constructive feedback. We are grateful to Alessandro Acquisti, Irene Bertschek, Tibor Besedes, Jörg Claussen, Daniel Erdsiek, Chris Forman, Anindya Ghose, Avi Goldfarb, Shane Greenstein, Sang-Pil Han, Andres Hervas-Drane, Erik Johnson, Reinhold Kesler, Tobias Kretschmer, David Laband, Fernando Luco, Bertin Martens, Markus Mobius, Thomas Niebel, Matthew Oliver, Martin Peitz, Arnold Picot, Imke Reimers, Olga Slivko, Rahul Telang, Bernd Theilen, Catherine Tucker, Hal Varian, Frank Verboven, Joel Waldfogel, Michael Ward, Simon Wilkie, Manfred Wittenstein, Pinar Yildirim, Pai-Ling Yin, Michael X. Zhang, Nick Zubanov and Christine Zulehner for valuable comments and helpful advice. We thank the participants in the 12th ZEW ICT conference 2014, EARIE 2014, WISE 2014, IIOC 2015, 6th SEARLE Internet Search and Innovation 2015, SEEK Digital Economy Workshop Torino 2015, NBER Summer Institute 2015, and Sevilla Apps Economy Workshop 2015. We thank Niklas Duerr, Julián Hidalgo, Florian Hofbauer, Alaina Totten, and Steffen Viete for their extremely useful research assistance.This paper represents the authors' personal opinions and does not necessarily reflect the position or official policy of the Deutsche Bundesbank.

†221, Bobby Dodd Way, #208 Atlanta, GA, 30332, U.S.A Email: `michael.kummer@econ.gatech.edu`.

‡Wilhelm-Epstein-Str. 14; 60431 Frankfurt am Main, Germany. Email: `patrick.schulte@bundesbank.de`.

1 Introduction

Striking a balance between "too little" and "too much" privacy protection in digital technologies and markets poses a non-trivial challenge for regulators and society as a whole. On the one hand, the market success of new digital technologies may greatly depend on the services' ability to collect and analyze *enough* personal information (Aziz and Telang, 2015; Goldfarb and Tucker, 2011; Johnson, 2013b). The resulting products have enormous potential for offering better information flows, better choices, efficiency and increased welfare. On the other hand, widespread unease and loss of trust in the market could result when providers store *too much* personal data (Acquisti, Brandimarte, and Loewenstein, 2015; Acquisti, Taylor, and Wagman, 2016; Miller and Tucker, 2009).[1] Ultimately, too much data collection may also carry significant societal risks.[2]

In this paper, we analyze data from around 300,000 smartphone applications ("apps") to inform the debate about optimal levels of privacy protection. Using data from Google's Android Market, we study the extent to which private information resembles a second "means of payment" on both sides of the market for apps.[3] Specifically, we use our data to analyze three research questions:

1. Do app developers request access to more sensitive data in exchange for lower prices?
2. Do users avoid installing apps with permissions that access privacy-sensitive data?
3. Are app users' privacy concerns context dependent?

Our results provide the first large-scale empirical evidence on a money-for-privacy trade-off on both sides in the app market: App developers trade greater access to personal information for lower prices, and consumers choose between lower prices and greater privacy. More precisely, our results suggest that: (1) App developers ask for more privacy-sensitive permissions if they offer a free app than if they offer a paid app; (2) consumers take this trade-off into account and show up to 6 percent less demand for apps that ask for privacy-sensitive permissions; (3) the negative relationship between permissions and downloads depends on the app's context. Factors such as trust, the targeted user group (e.g., mature users) or the sensitivity of the app's context (e.g., in health-related apps) moderate the relationship's strength.

[1] According to the Pew Research Center, 68% of adults believe current laws to protect individuals' online privacy are insufficient (Rainie, Kiesler, Kang, Madden, Duggan, Brown, and Dabbish, 2013).

[2] We refer to real historical experience, e.g., in Europe or Asia, where too much data in the wrong hands contributed to totalitarianism, persecution, mass murder, and war.

[3] We use "second means of payment" metaphorically, because both parties agree to exchange the developer's service for the user's data, and then developers exchange user information for money or services of third parties (comparable to a complementary medium of exchange, a tradable asset, or a valuable resource). We do not suggest that private data satisfies the definition of a full-fledged and widely accepted currency, with established exchange rates.

Our data are relevant and informative for analyzing the money-for-privacy trade-off. The Android Market for apps has a large number and variety of products and is of high economic value. In most app categories, users can choose among many alternatives. Privacy is salient, as users are shown the list of permissions requested by an app prior to installation. This policy creates a setting in which the money-for-privacy trade-off can be meaningfully studied, because we can discern 136 distinct permissions and record each app's permission requirements in 2012. We combine this information on each app's ability to collect personal information with a rich dataset on 300,000 apps from Google's Android Market. The data cover publicly available app-specific information, including each app's number of installations, its price, and even each app's closest competitors. We collected data repeatedly in 2012 (over five months) and once in 2014. We augmented these data with information from AppAnnie.com and the Apple's App Store, to add further information about apps and contrast performance among operating systems.[4]

Our inference is based on a full cross section, on panel datasets, on "app siblings" consisting of a free and paid version of the same app, and on a difference-in-differences-style analysis of demand for apps. Our results emerge consistently for both the supply and demand side across these different datasets and in various specifications. The findings are robust to using alternative measures of demand, other outcomes such as survival, different ways of measuring privacy-sensitive permissions, and considering subsegments of the market. They are also robust to an IV approach that instruments for app prices or to leveraging our data on "app siblings" – a free version and a paid version of the same app – to identify and account for category-specific permissions (e.g., running apps that need to track location). A second source of identification comes from a cross-platform comparison between Android and Apple's iOS. We leverage the difference between Apple's and Google's notification policies regarding permissions and show that an app's download ranks in Android are worse than in Apple's iOS if they require sensitive permissions. This finding corroborates our baseline conclusions.

Our paper follows the usual structure. Two online appendices contain additional supporting materials and cover the data in more detail.

2 Related Literature

By highlighting the money-for-privacy trade-off on both sides (demand and supply) of the market for mobile apps, we contribute to two streams of the literature: A large stream on the value of

[4]For several robustness checks, we further augmented our database with information from Alexa.com and data collected using Amazon's Mechanical Turk.

privacy and a growing stream on the mobile app market. Additionally, we hope to inform both policy makers and legal scholars concerned with privacy in mobile apps about the trade-offs in this sensitive market.

A central stream of the literature has studied the value of privacy both for the demand and the supply side of markets. The key idea in the literature about the value of privacy is that consumers or users can be compensated for sharing information about their data or their activities with service providers and other parties. Whereas the large majority of the papers in the previous literature have investigated this idea in theoretical models, experiments, or surveys, we use a large observational dataset. Moreover, we are the first to use such data to analyze the role of privacy in the market for mobile apps, a market with unseen potentials for collecting information about users. A major advantage of our setting is that the privacy permissions are unusually clearly communicated. Whereas past work found little demand for privacy, especially with permissive default settings (e.g. Gross and Acquisti (2005)), the observed choices in the present setting reflect many deliberate choices by consumers to download apps with privacy being a component of that bundle. Thus, our setting adds an additional angle to the existing debate, because our inference is based on the users' actual choices in a new and highly privacy-sensitive market.

To motivate our research questions we can draw from a wealth of existing theoretical studies of privacy. Specifically, in most theoretical papers that model the demand side, the cost of privacy to consumers arises either as nuisance from too aggressively targeted ads that may trigger costly avoidance efforts (Hann, Hui, Lee, and Png, 2008; Johnson, 2013b), or from a firm's ability to use knowledge about a user's preferences to price discriminate (Taylor, Conitzer, and Wagman, 2010; Wathieu, 2002).[5] Existing literature identifies three key moderators of aggregate privacy concerns: (1) the composition of consumers, (2) trust-inspiring measures by firms, and (3) the nature and amount of the requested private information (Acquisti, Taylor, and Wagman, 2016; Chellappa and Sin, 2005; Chellappa, Sambamurthy, and Saraf, 2010). Similar factors arise in our empirical context and will be analyzed in section 6.3, where we used the model by Chellappa, Sambamurthy, and Saraf (2010) as a framework to guide our analysis of moderating factors.[6] Note that the empirical

[5]Acquisti, Taylor, and Wagman (2016) and Tucker (2012) provide surveys. Taken together, these models find that reduced privacy is disadvantageous for consumers. Firms can use customer information, such as purchase history, to charge personalized prices in electronic retailing settings (Acquisti and Varian, 2005; Conitzer, Taylor, and Wagman, 2012; Taylor, 2004). However, increasing the cost of anonymity can benefit consumers up to a point, after which the effect is reversed (Taylor, Conitzer, and Wagman, 2010). Spiegel (2013) shows in the context of software production that free software in a bundle with (targeted) ads could improve welfare if the cost of producing software is low. Reduced privacy enables the provision of valuable services "for free" and can create benefits for users.

[6]Chellappa, Sambamurthy, and Saraf (2010) set up a mechanism design problem in which a monopolist faces consumers who have heterogeneous privacy concerns and decides on an optimal menu of personalization with possible

literature also studied other sources of privacy concern than targeting and price discrimination. Such additional concerns arise from direct discrimination (e.g., in hiring, housing, health-care or credit markets), or from fears of errors, improper access, or secondary uses (e.g., identity theft) of personal information (Acquisti, Taylor, and Wagman, 2016). Compared to price discrimination and targeted advertising, consumers' concerns about direct discrimination or identity theft could be higher.[7] Such concerns might be among the key factors that moderate the demand for privacy in our analysis. For example, we find evidence that consumers' concerns vary with an app's context, by showing that users are more cautious in especially sensitive settings, such as health-related apps.

More generally, users demand privacy and attribute value to not being targeted or "seen" in many settings (Marthews and Tucker, 2017; Turow, King, Hoofnagle, Bleakley, and Hennessy, 2009). The demand for privacy was found to change over time and to depend on context or framing (Acquisti, John, and Loewenstein, 2013; Goldfarb and Tucker, 2012; Gross and Acquisti, 2005). Such findings are related to the so-called "privacy paradox" (cf. Acquisti, Taylor, and Wagman, 2016; Norberg, Horne, and Horne, 2007), which is a major topic of debate in the existing literature, and refers to the large difference between the estimates from observational vs. experimental/survey-based evidence. We contribute an additional piece of evidence to this debate, because we use observational data on the app market, in which the use of personal data is a clearly communicated part of the bundle.

In other words, we contribute to the empirical literature on the value of privacy by measuring the managerially relevant magnitude of consumers' demand for privacy, using large transaction-based data that reflect revealed preferences in the market for mobile apps. Existing empirical estimates of the value of private information in online markets (rather than mobile apps) are based on experimental and survey data. The estimated valuations ranged from zero to very large numbers (Beresford, Kübler, and Preibusch, 2012; Carrascal, Riederer, Erramilli, Cherubini, and de Oliveira, 2013; Grossklags and Acquisti, 2007; Racherla, Babb, and Keith, 2011; Tsai, Egelman, Cranor, and Acquisti, 2011). Our results suggest that the consumers' revealed valuations for privacy in the 2012 mobile app market were so low that a business model which offers a privacy preserving app for small payment is not viable for most contexts or applications.

subsidization. Casadesus-Masanell and Hervas-Drane (2015) studied the money vs. privacy trade-off in a model in which suppliers compete on both dimensions (price and privacy) of a two-sided market, as pioneered by Armstrong (2006) and Rochet and Tirole (2003).

[7]For example, by 2008, the self-reported valuations for protection against improper access and secondary use were found to range from USD 30 to USD 50 (Hann, Hui, Lee, and Png, 2008, 2007). High exposure of personal information through a smartphone can lead to identity theft, and users' behavior can even affect the privacy of uninvolved friends, or the information security of their company by increasing the risk of successful "Choice and Chance" attacks. (Pu and Grossklags, 2016; Ransbotham and Mitra, 2009).

For the supply side, that is for firms, a similarly rich literature has documented that user data are valuable. We extend this notion to the market for mobile apps. Data collection on the internet is inexpensive, but it enhances the effectiveness of targeted advertisement, which affects revenues for content sites (Aziz and Telang, 2015; Goldfarb and Tucker, 2011; Tucker, 2012, 2014).[8] Moreover, digital services, such as apps or websites, can directly track their customers to learn who uses their service and how they use it (Hu, Zeng, Li, Niu, and Chen, 2007; Mobasher, Cooley, and Srivastava, 2000). Thus, user data can help service providers to better understand their customer's needs and how to improve their service (Bertschek and Kesler, 2017; Tambe, Hitt, and Brynjolfsson, 2012). The overwhelming insight from this literature is that user data can be a valuable resource for the firm. We add to it, by highlighting, for the first time, that developers offer their products at lower prices in exchange for additional data from their users.

Our paper also closes a gap in a second more recent stream of research, which used observational data to shed light on the market for mobile apps but did not focus on the role of privacy (Askalidis, 2015; Bresnahan, Orsini, and Yin, 2014; Carare, 2012; Chaudhari and Byers, 2017; Davis, Muzyrya, and Yin, 2014; Ershov, 2016; Ghose and Han, 2014; Li, Bresnahan, and Yin, 2016; Liu, Nekipelov, and Park, 2014; Liu, 2017; Yin, Davis, and Muzyrya, 2014).[9] Their findings have increased our understanding of the app market but left room for studies that focus on the role of private user data. The paper most similar to ours (Savage and Waldman, 2014) uses survey-based data to estimate users' average valuation of the data that users typically share with developers. They find a valuation of USD 4, which implies that offering more privacy could be a profitable market positioning for app developers.[10] Our results based on observational data suggest that privacy-preserving app strategies may not be economically viable.[11] This difference both reaffirms the privacy paradox and shows the importance of considering observational data based on market transactions to understand the role of private user data in the market for mobile apps.[12]

[8] Specifically, restricting the use of private data in advertising can reduce the success of targeted advertising and the revenue generated. More generally, websites differ considerably in how intensely they collect user data, and privacy policies can affect user behavior (Johnson, 2013a; Preibusch and Bonneau, 2013).

[9] Pioneering papers focused on identifying key drivers of app demand and estimating their importance. Other recurring themes are the relationship between innovation and quality in app markets, developers' strategic behavior, or their strategies for establishing a new app in the market (e.g., paying users to download the app onto their devices in order to obtain a position on the list of the most downloaded apps). Structural models analyzed demand and app developers' resulting platform choice (Android or iOS).

[10] They use a valuation scheme based on providing information about privacy relevant aspects of the data and ask users their willingness to accept. They also cannot say much about the developer dimension.

[11] Specifically, we show that consumer demand responds very little to potentially privacy-endangering permissions, which is in line with considerably lower consumer valuation for privacy and implies our conclusion.

[12] More technical studies have been done on the precise meaning of certain permissions and what they imply for the privacy of the device's owner (see, e.g., Chia, Yamamoto, and Asokan, 2012; Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy, 2012), or investigate the potential intrusiveness of apps (e.g., Chia, Yamamoto, and Asokan,

To summarize, little is known about the role of private data in the market for mobile apps. We close this gap by highlighting the money-for-privacy trade-off on both sides (demand and supply) of this highly data-driven market. We show that users prefer and value apps with greater respect for their privacy, which confirms earlier research.[13] In addition, we highlight app developers' willingness to trade forgone revenue for personal user information. Thus, user data are a valuable resource for developers, by which an emerging market can subsidize the creation of innovative products. We complete the picture by showing the context dependence of privacy concerns for mobile apps. Among these context-dependent factors are a developer's outside reputation, the existence of privacy policies, or an app's category (cf. Acquisti, Brandimarte, and Loewenstein, 2015).

3 Background: Data Sharing in the Mobile App Industry

We begin this section by discussing the app market's background and setup with regard to private data and then turn to the developers' ability to resell an app's data.

3.1 The App Market

In 2007, Apple introduced the iPhone. The device triggered a radical transformation of mobile communication in which screen-only smartphones replaced their predecessors. One of the main competitive advantages of the iPhone and its successors was a large app ecosystem. Apps allow users to tailor their devices to their needs by enabling multiple uses in addition to traditional phone calls and text-based apps. In 2008, the first smartphone using Google's Android Operating System (Android OS) was released. Although users adopted initial versions slowly, Android gained popularity in 2010 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 percent in 2012. In 2015, revenue from the mobile app store was nearly $40 billion and is expected to reach $100 billion in 2020 (App Annie, 2016).

In addition to the Android OS, Google also introduced its own platform for the distribution of apps, originally called *Android Market* but renamed *Google Play Store* in mid-2012. It serves as a distribution channel for apps, books, movies, music, and newspapers. By the end of 2012, the Google

2012; Fahl, Harbach, Muders, Baumgärtner, Freisleben, and Smith, 2012). Other studies investigate how platform architecture affects smartphone users' willingness to pay premiums to limit exposure of their personal information (Egelman, Felt, and Wagner, 2013). Sutanto, Palme, Tan, and Phang (2013) suggest storing user information on the device to overcome the personalization vs. privacy paradox and document a positive effect on demand. Two studies by Lin, Amini, Hong, Sadeh, Lindqvist, and Zhang (2012) and Lin, Hong, and Sadeh. (2014) analyze more than one million apps' privacy-related behaviors. They summarize these behaviors using a grade scale, ranging from A+ (most privacy sensitive) to D (least privacy sensitive; cf. http://privacygrade.org/faq [retrieved: 11/5/2016]).

[13] Even though our lower-bound estimates suggest lower valuations than previous studies.

Play Store offered approximately 400,000 apps, and in 2015 the number increased to 1.5 million. Google distinguishes thirty categories of apps, such as Communication, Education, Productivity, Shopping or Weather apps as well as Games. The latter category, consisting of Arcade & Action, Brain & Puzzle, Cards & Casino, Casual, Racing, and Sports Games apps, make up around 17 percent of our estimation sample.

A central feature of the Android app ecosystem is its permission system. This system is specific to the Android OS and provides a setting in which the money-for-privacy trade-off can be meaningfully studied. First, developers can choose among standardized blocks of information, so-called permissions, in which some enable access to a user's location, communication, browsing behavior, and so on. Before their installation, apps must declare which permissions they require and must request these from the users. More precisely, the system provides a list of permission names alongside a short explanation for each permission requested. Users must accept this list and explicitly acknowledge that they are granting these permissions to proceed with the installation. Alternatively, if they feel uncomfortable about the set of permissions requested, they can cancel the installation. Note that such explicit consumer consent for the set of permissions does not exist in Apple's iOS, in which this information remains implicit before installation. In essence, this procedure has been in place since 2012, despite the rapid growth of the Android Market.

In 2012, developers could choose among 136 predefined permissions.[14] This large number illustrates the quantity and diversity of information app developers can potentially collect about app users. Appendix Figure A2 illustrates the way in which permissions were displayed in the Android Market in 2012.[15] Since then, Google has introduced several small modifications in how permissions are displayed to the user. Before 2014, the list of permissions showed permission names next to short explanations of the permissions. Since 2014 the system shows only the names of aggregated permission *groups* (though users can open a more detailed dialogue for each permission group). Still, users must approve of the permission list before proceeding with installation.[16]

3.2 Monetization and Trading of App Data

Developers can monetize their apps via four main channels. According to AppBrain (2016), around 20 percent of the apps in the Google Play Store are paid apps, whereas the remaining 80 percent

[14] As of March 2017, the count stands at 148 permissions (although the precise contents of some permissions changed; see `http://developer.android.com/reference/android/Manifest.permission.html`).

[15] Although requested permissions were displayed slightly differently on smartphones, the text of the permissions' descriptions is *identical* in both versions (mobile and browser) of the Play Store. The layout is almost identical.

[16] Since 2016, Google allows users to withdraw individual permissions from an app after installation. However, this only affects recent versions of the Android OS (starting with version 6.0.

of apps are free.[17] Alternative revenue channels are in-app advertising, in-app purchases, and data trade. The size of these alternative revenue channels has been relatively stable since 2012 except for in-app purchases, which were introduced shortly before our period of observation.[18] In 2012, when we collected our data, the "freemium" model based on in-app purchases barely existed. Since then, a marked increase has occurred in the use of this model, in which users may install the apps without paying but must pay to unlock usefull additional functions. The two other channels, in-app advertising and data trade, were already common.

Data sharing for monetization purposes is common in the mobile app industry, as in many online markets (see, e.g., Woodcock, 2017 and references therein). Christl and Spiekermann (2016) and OECD (2013) survey several studies that show that apps commonly shared data with third parties.[19] They also provide a brief discussion of the data-sharing business model and the three main channels that developers have to exploit app usage data to generate revenues: (1) The most important channel is using information about users to sell targeted advertising.[20] This channel has arguably been the powerhouse of the market for free apps, representing a sizable industry.[21] The value in this industry is generated from collecting and aggregating individuals' digital traces from various sources (e.g., an app), in order to profile users' location (Dubé, Fang, Fong, and Luo, 2017; Fong, Fang, and Luo, 2015) and behavior for targeted advertising.[22] Yet, although this type of information brokerage potentially offers greater efficiency in matching the ads, sharing the data implies that the app users' information is passed on, and their safety depends solely on the integrity of the developer's advertising partner. (2) Another prominent channel of data flows arises when

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

[18] 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`).

[19] The pieces of information most common transmitted about the user are identifying information (name, e-mail address, phone ID, gender, age, or birthdate), location data, contacts, and usage data. Sometimes "data input" (e.g. search terms) is transmitted as well. This information is more sensitive in some contexts, such as health, than in others (Appthority, 2014; Schmidt, ten Venne, and Eikenberg, 2012; Zang, Dummit, Graves, Lisker, and Sweeney, 2015). Most of these studies are based on very small samples. For example, Seneviratne, Kolamunna, and Seneviratne (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 percent of these users were exposed to more than 25 trackers.

[20] Successful apps with sufficient traffic are able to sell their advertising space directly on the market. However, doing so is costly for less successful apps, which can achieve better targeting by sharing their data with a third-party broker. Such data brokers can provide advertisers with access to users from a bigger pool of multiple apps (and developers to more advertisers).

[21] Peer-reviewed estimates are not available. Existing estimates range from several hundred million to several billion US dollars (OECD (2013) based on Beales (2010)).

[22] To use ads (even without targeting), developers need access to a phone's network state to simplify communication between ad library and the ad server. For targeting consumers, developers frequently access a phone's fine location and permission which allows identifying the app's user (READ_PHONE_STATE; see e.g. Book et al., 2013).

developers trade their data for valuable third-party services. A good example of such services is app analytics. App analytics help developers gain insight on who uses their app when and where, along with which other processes on the phone are simultaneously being used. 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 advertising purposes. As with third-party advertising, the secondary usage of the user data depends on the analytics site's own policy and integrity. (3) Finally, app developers can trade their data with third parties in several other common ways. For example, they can exchange their data for direct monetary benefit. Such data can be purchased by known industry analysts such as *Alexa.com* or by lesser-known data aggregators. As with third-party libraries, the secondary usage of the user data depends on the analytics site's policy and integrity.

4 Data and Descriptive Evidence

We first describe our main dataset and then turn to our two alternative datasets (Section 4.2). The descriptive analysis in Section 4.3 illustrates two key findings about the money-for-privacy trade-off in the market for mobile apps.

4.1 Main Data and Variables

For our main dataset, we extracted all publicly available information on nearly all the products available in the Google Play Store in 2012 (around 300,000 apps). We collected the data on a monthly basis from April to September 2012. Appendix Figure A1 shows the design of Google's app store in 2012, which corresponds to the information we were able to collect. Using this information, we created two main datasets: (1) a cross section of all apps available in April 2012 and (2) a balanced monthly panel of apps that were available from April 2012 until September 2012 and had within-variation in the number of privacy-sensitive permissions. In addition, we collected a cross section of data in 2014 to analyze long-term outcomes. Also, for some of the robustness checks, we used additional data sources, such as Alexa.com, Amazon Mechanical Turk, PrivacyGrade.org, and AppAnnie.com.

Given our research questions, we need three main types of information: a price measure, a demand measure, and a measure of an app's ability to collect private information about users. In the following, we introduce and discuss each of these measures as well as the core control variables.

Main Outcome Variables: Our first main outcome is the **price** of apps. We directly observe price (in euros) in the data. We measure (1) the decision to provide a free app or a paid app as a dummy ($D_{Paid} = 1$, if the app is paid) and (2) the price if a paid app is offered ($\ln price$). Our second main outcome is **demand**, which we observe with two measures. Our first measure of app demand is the monthly change in the number of ratings (ΔRatings). Alternatively, we use the monthly change in the number of installations (ΔInstallations). Both proxies measure the number of new customers who consider the benefits of installing an app greater than the cost of paying the product's price or the associated loss of privacy. The two demand measures are highly correlated (see Figure 1 and Appendix A.2). The high correlation is not surprising given that users have to install an app in order to rate it. Our preferred demand variable is the change in the number of ratings, since we observe the exact number of ratings, whereas the number of installations is observed only in discrete step-size form with 17 levels (e.g. 11-50, 51-100, 101-500 installations), which leads to low intertemporal variation in a panel specification.[23] It is well known that both ratings and installations are not without drawbacks, for example, because both can be purchased by developers to promote their app (Li, Bresnahan, and Yin, 2016) and only a small fraction of users who install an app might actually use or rate it (Hu, Zhang, and Pavlou, 2009).[24] Hence, we show the robustness of our findings with respect to the choice and definition of our demand measure in Section 6.4. There and in our main demand specifications, we use several alternative demand measures based on ratings, installations, rankings, and alternative time windows.[25] The industry's most relevant measure for an app's demand or usage would be daily active users (DAU), but this measure is not available, because it is only known by an app's developer team.

FIGURE 1

[23] The discrete measure introduces attenuation bias through noise in the dependent variable.

[24] Purchased ratings or installations could be problematic for our empirical strategy for two reasons: (1) Widespread fraudulent reviews or installations could lead to considerable noise in our dependent variable, which might result in attenuation bias, and small estimates; (2) If apps consistently differ in their strategies to encourage ratings and in their data collection behavior, our results might be biased. Specifically, if apps that use more permissions more actively recruit raters, this would bias our coefficients toward zero. To address the first concern, we compared installations and ratings and found little evidence for fake ratings as a widespread phenomenon. "Hiring" ratings is most effective for young (Muchnik, Aral, and Taylor, 2013) and sometimes intermediate apps, which begin to have imitators (copycats). However, it is less likely to have an effect on monthly changes in later periods. We address the second issue, which is of critical importance, in our panel fixed-effects specifications, but it may affect the cross-section estimation with attenuation bias. We also do not see many such sudden peaks in ratings over time.

[25] In Appendix A.2 we discuss the close relationship between our main demand measure and our measures based on app rankings. However, app rankings are available only for the most successful apps, which limits the scope of this comparison to a small and arguably selected set of apps.

Identifying the Privacy Sensitivity of Apps: To measure an app's ability to collect private information, we take advantage of the fact that, as described earlier, Google provides precise details on the permissions an app uses. This feature allows us to understand exactly which private information an app can collect about the app user. Among the 136 permissions available to app developers, some can be considered innocuous with respect to the privacy of the user, while others grant an app access to sensitive information. Based on an extensive literature review (Dini, Martinelli, Matteucci, Petrocchi, Saracino, and Sgandurra, 2012; Jeon, Micinski, Vaughan, Fogel, Reddy, Foster, and Millstein, 2012; Mylonas, Theoharidou, and Gritzalis, 2013; Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy, 2012; Taylor and Martinovic, 2016), we were able to identify a total of 25 privacy-sensitive permissions, which allow apps to access information about users' ID, location, profile, or communication (see Appendix Table A1). These potentially problematic permissions include "fine (GPS) location," which allows access to the user's current location, "read browser data," which grants access to a user's browsing history and their bookmarks, or "read contact data," which allows an app to access the user's contact data. 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 25 privacy-sensitive permissions and zero otherwise. To capture the intensity of an app's ability to collect private information, we additionally use the number of privacy-sensitive permissions in an app ($\#Privacy$). In addition, we evaluate our findings when using other, alternative privacy measures in section 6.4. These measures include (1) a category-specific modification of our baseline measure, (2) a classification based on Google's assessment of problematic permissions, (3) a classification derived by hiring 400 classifiers at Amazon Mechanical Turk, and (4) a classification to account for the fact that apps require internet access to transmit user information. Appendix Tables A1 and A2 summarize all classifications applied and describe all privacy-sensitive permission. In addition, these tables show the details of how we grouped the privacy-sensitive permissions into four subgroups: location-, profile-, communication-, and ID-specific permissions.

Control Variables: In addition to our main variables, we also observe a rich set of app-specific characteristics relevant for explaining app supply and demand: the app category, the average rating, the code size, the required version of Android OS, developer-specific information (name of developer, number of its other apps, top developer status, etc.), and the app's description (length, number of screenshots, video) and permissions related to internet access ($D_{Internet}$) and advertising (D_{Ads}).[26]

[26] D_{Ads} equals one if an app uses the ACCESS_NETWORK_STATE permission (see Table A1). Around 80 percent of ad libraries in 2013 used this permission, which simplifies communication between ad library and the ad

Also, we use the section "users who viewed this also viewed" to identify app-specific competitors and to construct three additional control variables: the average price of competing apps, competitors' average number of installations, and the average rating of competing apps. Finally, to control for an app's functionality unrelated to the ability to collect private information, we include in all specifications the number of unproblematic permissions to which an app has access ($\#CleanPerm$). For a list and definition of our key variables, see Appendix Table A2.

4.2 Two Alternative Datasets

In addition to our two main datasets (cross section and balanced panel from 2012), we additionally use a set of app pairs and a set of apps for which we could retrieve app download rankings from 2012 in both the Google Play Store and Apple's App Store. We use these two alternative datasets for robustness checks that control for the unobserved heterogeneity of apps. We also use the second one to show the close relationship between app download rankings and our demand measure.

App Pairs: App developers often offer two versions of the same app, a free version and a paid version. The paid version typically offers some advantages over the free version: It may offer additional functionality, contain less advertising, or require fewer (privacy-sensitive) permissions.[27] We identified pairs within our cross-section dataset, which consist of app siblings with the same functionality.[28] To do so, we used a two-to-three-step procedure. First, we used a word-processing algorithm that identified app pairs with the same name except for the addition of one of the following: "free," "paid," "lite," "full," "demo," "pro," "premium," "donate," "trial," "plus" (which results in 7,211 app pairs). Second, within this broad set of pairs, we identified those for which the code size of the paid version is not larger than that of the free version and for which the description of both apps is more or less the same length.[29] This gives us a sample of 1,999 pairs. Third, to identify an even more rigorously matched subset of app pairs we *manually* checked the description of the remaining

server (see e.g. Book et al., 2013). For the provision of targeted advertisement, especially READ_PHONE_STATE (which allows identifying app users) and ACCESS_FINE_LOCATION seem to be of high relevance.

[27]Introducing sensitive permissions in the free but not the paid version is the most explicit instance of a money-for-privacy trade-off. A typical, though not ideal, pair consists of a free "lite" version and the full (paid) version of the same app, in which the free version has more permissions but less functionality. In our research, we can use pairs to find unnecessary permissions, because the paid version is the proof that the same app can run without that permission. This inference is valid even if the paid version of the app offers *more* features than the free version, if only the paid app does not offer less functionality.

[28]By "functionality," we mean the utility-generating service to the user. Ads, for example, might imply code and functionality, but their purpose is generating (data-driven) revenue for the developer (not increasing user satisfaction).

[29]According to our procedure, a pair's free app is of the same code size or a little larger (on average, 300 KB larger). In the full cross-section, the free apps are smaller (on average $> 3MB$ smaller), so we attribute the difference in the matched pairs to the developer's monetization strategy. A pair's length of description differs by less than 30 percent.

app pairs and kept those whose description indicated no difference in functionality. The versions in this last set of pairs should thus differ only in price, permissions, and possibly the existence of ads. This most restrictively matched sample consists of 354 pairs and excludes app siblings with identical descriptions if the text mentioned any differences between free and paid versions.

Android and iOS App Ranking: Many apps are offered for both Google's Android OS and Apple's iOS. This fact allows us to compare the relative success of apps in both systems and the extent to which it depends on their use of privacy-sensitive permissions. We used AppAnnie.com to retrospectively collect (in 2016) the apps' rankings from 2012 for both OS. Unfortunately, in 2016, we found only a small number of 192 (mostly successful) apps for which the app ranking from 2012 was available for both OS. Nevertheless, the dataset allows a rigorous comparison across platforms. Unlike in our main datasets, which contain global success measures, such as the worldwide number of installations, rankings are available only at a country-specific level. Thus, we collected and aggregated rankings for seven large markets: Germany, India, Japan, South Korea, Russia, the UK, and the US. For these countries, we have rankings for April and September 2012. For free apps, we use the overall free app rank, and for paid apps we use the overall paid ranking.[30] Based on the collected country-, time-, and OS-specific ranks we compute average OS-specific values for each app. We use the order of these average values to construct our own OS-specific in-sample rankings, which are designed to ensure that the values are comparable across OS and limit the potential effect of outliers. Using these standardized in-sample rankings, we construct two measures of relative app success on the two operating systems ($\mathit{\Delta Rank}^{iOS-And}$): We compute (1) the difference between rankings for Android OS and iOS based on the average OS-specific rankings that we were able to collect for the two versions of the app, and (2) the same difference but based only on the ranking information of the largest market, i.e., the US.[31] This dataset contains up to 192 apps observed in both OS.

4.3 Descriptive Evidence

Before providing descriptive evidence on two key results of this paper, we discuss the summary statistics of the variables in our main datasets.

[30] In addition to overall rankings we include category-specific rankings, e.g., for games, weather, or education apps.

[31] Specifically, we compute average rankings by app and OS based on time- and country-specific ranks (up to 7 countries and 2 points in time per OS, i.e., up to a total of 14 ranks). Using these average values, we create a new in-sample ranking that ranks apps by the order of their average rank values and ranges from 1 to 192 for each OS (because we have 192 apps available). In the estimation, we use the difference between the standardized rank we generated for the iOS version and the one from the Android version (i.e., iOS rank minus Android rank).

Summary Statistics: Table 1 provides an overview of the most relevant variables in our main datasets. For each dataset, the left column shows averages for free apps, and the right column shows those for paid apps. The first set of columns (cols. 1-2) shows permission use in the entire cross section of apps, consisting of 177,193 observations.[32] Free apps are installed more often and have a lower average rating. Crucially, an average free app is more likely to use potentially privacy-sensitive permissions and will require more such permissions than an average paid app. For example, free apps use on average 1.19 privacy-sensitive permissions, whereas paid apps use only 0.52 such permissions. Similarly, 50 percent of the free apps have at least one privacy-sensitive permission, while only 28 percent of the paid apps use such permissions. This relation does not change when we compare alternative groups of privacy-sensitive permissions, and this pattern also holds on our three other datasets: the panel dataset of apps that have within variation in the number of privacy-sensitive permissions over time (cols. 3-4), one of our sets of matched app pairs (cols. 5-6), and the set of apps for which we were able to retrieve app rankings from 2012 for both Android OS and iOS. Despite covering very different subsets of apps, all four datasets show consistent patterns with respect to the differences in the permission use by free and paid apps: No matter how we look at the data, free apps always request more privacy-sensitive permissions than paid apps. This holds for essentially all criteria we analyzed, and this pattern is found even when we match pairs of apps that come from the same developer and have the same functionality. In the appendix (available online), we provide additional information (Appendix Table A2) and summary statistics for the cross-sectional data (Appendix Table A3) and pairs (Appendix Table B1).

Below, we document two key results about the supply side. First, some mainly free apps request permissions that are not necessary for their functionality. We conjecture that these unnecessary permissions are used for monetization. Second, privacy-sensitive permissions are more likely to appear in free apps than in paid apps. This indicates a negative relationship between price and permission use. Both results suggest a trade-off between price and privacy. Consumers can choose between apps that are cheaper but request more privacy sensitive information and apps that are more expensive but more privacy respecting.

Unnecessary Permissions: We use both the full sample of app pairs and the sample of selected app pairs to study whether apps request permissions without needing them for their functionality. In both datasets, the paid version serves as a technological reference, because the paid version can

[32]The discrepancy between our sample and the full population of around 300,000 apps is mainly due to excluding apps that were unavailable in some of the monthly waves in our panel analysis, lacked key variables, had zero installations or ratings, or are outliers with respect to our main demand measure.

be safely assumed to provide the same or even more functionality than the free one.[33] Hence, any permission that is present in the free version but not in the paid sister is unnecessary in terms of functionality, and can be expected to be related to monetization.

Figure 2 compares the number of privacy-sensitive permissions in a pair's free and paid versions. Panel (a) shows how frequently we observed a given difference in permissions for the full pair sample. A frequency of 280 for a value of 2 means that we find 280 pairs for which the free version requested two permissions more than the paid version. Panel (b) shows the same comparison for the selected sample of app pairs, for which we ensured that functionality does not differ between apps. Across both samples, we find more pairs in which the free version requires more privacy-sensitive permissions than the paid one. Particularly in panel (b), the paid version hardly ever requests more permissions than its free counterpart. We conclude that several of the free versions request unnecessary, privacy-sensitive permissions, which are not required for their functionality.

FIGURE 2

Price and Permissions: When we analyze all apps, we see that developers, who charge zero prices, are more likely to request more privacy-sensitive permissions. Figure 3 shows the 10 most frequently used privacy-sensitive permissions and contrasts the share of free and paid apps that request them. All privacy-sensitive permissions are more common in free apps. Developers choose between either offering their apps free with additional privacy-sensitive permissions or charging a fee but including fewer privacy-sensitive permissions.

FIGURE 3

The goal of the subsequent sections is to provide rigorous econometric evidence of the trade-off that is implied by the key findings we just presented.

5 Empirical Framework

This section discusses our empirical framework. The first subsection presents an overview of our estimation approach for the supply side. Subsequently, we discuss estimation of the demand side.

[33] Because a free sibling is available, it is hard to imagine that users would be willing to pay a positive amount for less functionality, such that if two siblings have any difference in functionality, we assume that the paid app sibling provides more functionality.

5.1 Supply-Side Estimation

We run two types of correlational supply-side regressions to provide insights on the role of private information for app developers. First, we examine the choice of the business model. Specifically, we analyze how the developer's decision to offer a free app or a paid app is related to the app's use of privacy-sensitive permissions. Our estimation equation is:

$$D_i^{Paid} = \alpha + \beta D_i^{Privacy} + \theta X_i + \varepsilon_i, \tag{1}$$

where D_i^{Paid} is a dummy equal to zero if the app i is free, and ε_i is the error term. The dummy $D_i^{Privacy}$ is equal to one if an app requires at least one privacy-sensitive permission. A negative β indicates that free apps are more likely to request privacy-sensitive permissions.[34] The control variables X_i are: the number of unproblematic permissions, a dummy for ad-related permissions, a dummy for internet access, dummies for the app category and the maturity level (e.g., "recommended for users aged 13+"), the average rating (in logs), the code size (in logs), the app version, the length of the app description (in logs), the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the number of apps by the developer (in logs), the average number of installations of the developer's other apps (in logs), the minimum and the maximum compatible Android OS version, and information about all of the competing apps's characteristics.

In a second specification we restrict our sample to paid apps and study how price levels are related to the use of privacy-sensitive permissions. The estimation equation is:

$$\ln Price_i = \alpha + \beta D_i^{Privacy} + \theta X_i + \varepsilon_i, \tag{2}$$

where the dependent variable is the log-standardized price of an app. A negative coefficient β indicates a lower price for apps that require privacy-sensitive permissions. We stress that these cross-sectional regressions do not account for unobserved heterogeneity.

Panel Data Analysis: We apply two main strategies to address the concerns over unobserved heterogeneity and to validate our results on alternative data structures. In our first approach to control for unobserved heterogeneity, we use a panel specification. We include an app fixed effect (α_i) and estimate our model based on within-app variation. We combined the cross-sections from April and September 2012 to form a panel covering our observation period of five months. Analyzing the within-app variation focuses on changes in price or the app's business model and allows us to

[34] Apps with privacy-sensitive permissions are $\beta \times 100$ percent less likely to be paid.

control for unobserved heterogeneity. However, this approach is limited by the small number of apps that changed price or switched business model during our sample period. For the business model, we estimate the following fixed-effects model:

$$D_{it}^{Paid} = \alpha_i + \beta D_{it}^{Privacy} + \theta X_{it} + \varepsilon_{it}, \tag{3}$$

where α_i captures unobserved heterogeneity of app i and t represents the month. We estimate the analogous relationship for the price of apps (provided they request payment) and use the waves from April and September 2012 for both specifications.

App Pairs: Our second main strategy to deal with unobserved heterogeneity exploits variation within pairs of apps, in which the free and the paid version of an app differ mainly in price and the number of permissions they use. These app pairs shed light on the money-for-privacy trade-off, as perceived by developers. As explained in Section 4.2, this is because the paid version serves as a technological reference that allows us to identify permissions that likely serve purposes other than facilitating the app's core services. The regression equation is:

$$D_{pi}^{Paid} = \alpha_p + \beta D_{pi}^{Privacy} + \theta X_{pi} + \varepsilon_{pi},$$

where α_p controls for unobserved heterogeneity of app pair p. This framework is valid if we can ensure that the two siblings have identical functionality. Hence the need to identify app pairs without any discernible difference in functionality, which differ only in permissions and price.[35] We can run this regression only for the app's business model choice, because one of the siblings is free by definition.

5.2 Demand-Side Estimation

Our baseline specification to analyze the relationship between app demand and permissions is based on the cross-section sample. We model demand as a function of its permissions, its price $Price_i$, and the same set of control variables X_i as in the supply side specifications. We estimate the following baseline model:

$$\ln Demand_i = \alpha + \beta D_i^{Privacy} + \gamma \ln Price_i + \theta X_i + \varepsilon_i. \tag{4}$$

In the main specification, we approximate $Demand_i$ for app i by its change in the number of ratings in that month (ΔRatings), and in our robustness checks we use installation-based measures. Again β is the coefficient of interest, indicating that the use of a certain permission comes with a $\beta \times 100$

[35] For any remaining differences in functionality, our results would continue to be biased toward zero if greater functionality is in the paid version and is positively correlated with price or more (privacy-sensitive) permissions.

percent change in demand. Prices are in logs, so that γ can be interpreted as own-price elasticity.[36]

Since OLS is sensitive to unobserved heterogeneity we test our results using (1) a 2SLS model, (2) a panel approach (3) a difference-in-differences style analysis between Android OS and iOS, and (4) an extensive series of robustness checks, which we present in the following section.

Endogeneity of Prices, Permissions, and Exit: Both permissions and price are strategic choices for the developer. Hence, the estimated demand coefficients might suffer from endogeneity bias. For monetary prices, endogeneity is well understood and usually leads to an upward bias in the estimated price coefficients (Wooldridge, 2010). We address this issue using standard instruments for price, such as cost shifters including the code size and the number of apps a developer offers (Berry, Levinsohn, and Pakes, 1995; Hausman, 1996; Nevo, 2000).[37] Using 2SLS techniques, we test whether our coefficients of interest (privacy-sensitive permissions) are affected by the potential endogeneity in monetary prices. We show these results, which confirm our baseline findings, in Appendix Table A10.

The second source of potential endogeneity is permissions. If developers charge a higher "permission price" for unobserved high quality, this could introduce a positive correlation between permissions and the error term. The resulting estimation bias would be in the same upward direction as for the monetary price. We attempt to reduce the effect of unobserved quality using panel estimation techniques and our difference-in-differences-style approach.[38] Any remaining correlation would lead to an upward bias in the regression coefficients, and a truly negative coefficient would be biased toward zero. Without an instrument, we can only provide a lower-bound estimate of the effect of permissions on demand. While it is hard to find the ideal instrument for this source of endogeneity, we can use developer's and competitors' behavior as an instrument. Specifically, we use the same developer's share of apps with privacy-sensitive permissions (excluding the focal app) as our first set of instrumental variables (IV). Alternatively, we use the share of the competing apps that use privacy-sensitive permissions as a second IV. While both approaches suffer from limitations, they are good predictors of the app's permission choices and arguably do not directly depend on the app's success. Again, the results are shown in Appendix Table A10 of Appendix A.1.

[36] A 1 percent price change leads to a γ percent change in expected demand.

[37] Given product quality, neither code length nor the number of products by a developer should generate utility for the user, but reflect the production cost and developer experience as well as potential for code sharing. As an alternative, "Hausman instruments" could be used in the panel, but the variation in prices is too small. Furthermore, note that free apps have a constant (zero) price, such that no instrumentation is required in their case.

[38] However, if permissions are introduced at a later point, even a panel approach could fail. In our favor, the endogeneity problem is somewhat mediated by the fact that functionality and offering pairs of apps are strategic choices. Developers must make these choices and introduce the implied permissions before demand can be observed.

A third source of endogeneity could be a survivor bias that could affect our cross section specifications. This is because our cross-section consists only of survivors (until 2012) and the likelihood of app survival might depend on an app's use of permissions. If app survival is negatively related to permission use, e.g., through app demand, we would only observe apps that are relatively successful such that a potentially negative permission effect on demand would again be underestimated. To study whether the use of privacy-sensitive permissions is related to apps' likelihood of survival, we correct for the potential bias from this selection process. We provide results based on Heckman selection models (Heckman, 1979), which confirm our baseline findings, in Appendix Table A10(Appendix A.1).

Our remaining strategies to deal with the threats to identification are based on using different data structures that aim to eliminate unobserved heterogeneity.

Panel Data Analysis: In our first approach to tackling unobserved heterogeneity, we use an app fixed effect (α_i) to estimate a panel regression model. We thus only exploit within-app variation to address the concern that we cannot observe all heterogeneity between apps in the cross-sectional analysis, despite our rich set of control variables. In particular, unmeasured quality and functionality of apps could be positively correlated with both app demand and permission use (resulting in an upward bias). To address this concern, we estimate the following fixed-effects specification:

$$\ln Demand_{it} = \alpha_i + \beta D_{it}^{Privacy} + \gamma \ln Price_{it} + \theta X_{it} + \varepsilon_{it}. \tag{5}$$

where $Demand_{it}$ is measured as the monthly change in the number of ratings, and the interpretation of the coefficient of interest, β is the same as for the cross-sectional results. A negative coefficient would indicate that increased permission use comes with a subsequent decrease in demand.

Difference-in-Differences-Style Analysis: Our second strategy to tackle the challenge of unobserved heterogeneity between apps is a difference-in-differences-style analysis. Specifically, we compare the relative performance of apps with and without privacy-sensitive permissions that are available both on Android OS and iOS and are successful enough that we could obtain their 2012 download ranking retrospectively. We run the following regression:

$$\Delta Rank_i^{iOS-And} = \alpha + \beta D_i^{Privacy} + \gamma \# CleanPerm_i + \varepsilon_i,$$

with $\Delta Rank_i^{iOS-And}$ being defined as the difference between app i's Apple and Android standardized in-sample download ranks ($iOSRank_i - AndroidRank_i$).[39] A negative β-coefficient indicates that apps that use privacy-sensitive permissions have a lower download rank in Apple's OS compared to Android's OS (i.e., they are downloaded relatively more often in iOS than in Android OS). This approach exploits the fact that the Apple platform is less explicit about app permissions when users download the apps than the Android counterpart. Therefore, Apple users may be less aware than Android users about the private information they are sharing through the app. We thus obtain two points of observations from each platform and can difference out the innate app characteristics in this comparison. This allows us to measure the effect of app permissions more precisely.[40] The downside of this analysis is the small and selective sample, which consists of only 192 apps in the top segment. However, while this analysis can only complement our main analysis of the demand side, it can do so with a relevant set of widely used apps and an especially clean design.

6 Results

We first analyze whether cheaper apps request more privacy-sensitive permissions. Second, we present our results on the demand side. Third, we analyze circumstantial factors (such as the reputation of app developers), robustness, and the implied valuations of private data.

6.1 Money vs. Privacy on the Supply Side

[Table 2 (MAIN SUPPLY RESULTS) HERE]

Table 2 shows descriptive regressions that relate the supply side's pricing choices to the use of privacy-sensitive permissions. The two outcomes of interest are the app's business model (cols. 1-6) and the price charged, given that it was positive (cols. 7-10). All these regressions control for a large set of variables that could drive app supply. These controls include the number of unproblematic permissions and permissions that are related to ads and internet access. Columns 1-6 analyze the developer's decision to offer a free app or a paid app. The dependent variable is a dummy, which equals 1 if users have to pay a positive price for downloading the app. Columns 1 and 2 analyze the cross section, columns 3 and 4 use the panel of apps that change their business model at least once over the five-month sample period. Hence, we use only the first and the last wave of our

[39] As described in section 4, we provide two versions of this measure which are based on average OS-specific download ranks (1) according to the average rank in seven countries and (2) exclusively according US ranks.

[40] We thank an anonymous referee for suggesting this specification.

panel in these regressions. Columns 5 and 6 use the app pairs. In the cross section, we find that apps that require privacy-sensitive permissions are 3.3 percent more likely to be free. Also when we use the number of permissions to measure the presence of privacy sensitive permissions, we find that apps are more likely to be free when they have one or more sensitive permissions.[41] Thus, the correlational results confirm the descriptive evidence showing that "a price comes with fewer privacy-sensitive permissions." This applies to both the presence of any privacy-sensitive permissions (col. 1) and the number of sensitive permissions (col. 2). The panel regressions in columns 3 and 4 highlight two patterns. First, less than 1 in 1000 apps switches its business model and the number of permissions it uses over a period of five months. This low percentage suggests that the choice of the pricing model is a strategic decision, which is not easily adjusted. Second, when using the Dummy that indicates a switch between models, we find that moving from paid to free coincides with an insignificant increase in the likelihood that the app requests privacy-sensitive and ad-related permissions. However, analyzing the number of privacy-sensitive permissions (col. 4), we see that more such permissions come with a weakly significantly lower likelihood that an app is paid. Our analysis of the app pairs (cols. 5 and 6) confirms the cross sectional patterns even more strongly. If one of the siblings uses sensitive permissions, it is 15.6 percent more likely to be the free version, which again is confirmed if we use the number of permissions rather than the dummy variable.

In columns 7-10, we analyze the determinants of paid apps' price level. Columns 7 and 8 show cross section results, while columns 9 and 10 use the panel data. First, note that only around 47,000 apps (26.75%) have a positive price, and of those, only 211 apps (less than 1%) had any variation in price and the number of permissions over the five-month period of observation. As in the business model, the price of an app is hardly ever adjusted. In the cross section, in both specifications, we see an insignificant correlation between using privacy-sensitive permissions and prices. As discussed in section 5, the cross section coefficient for permissions could be confounded by functionality and other factors. We can control for such unobserved heterogeneity in panel estimations, and indeed we see a much stronger and weakly significant negative relationship in the panel. Although the scope of our panel estimation is limited by the small number of apps that can be included, this negative relationship is in line with the notion that apps might be trading privacy-sensitive permissions for a lower price (and vice-versa).

[41] We find a coefficient equal to 0.024 which suggests that each additional permission increases the probability that the app is free by 2.4 percent. When using the dummy $D_{Privacy}$ and the number of permissions $\#Privacy$ together, we find the dummy to be insignificant and the coefficient of #permissions still equals 0.024. We further investigate the marginal effect in Appendix Table A4.

Further supply-side results are shown in Appendix Table A4. In columns 1-4, we vary the specification in the supply-side regressions to highlight further findings. Column 1 shows that apps are more likely to be free when the number of privacy-sensitive permissions increases. In column 2, we show that developers strategically use privacy-sensitive permissions in free apps only together with internet access, which highlights that data collection (via permissions) is more valuable for developers if they can easily transfer the data from the app (via internet access). Columns 3 and 4 use our category-specific privacy measure (see Appendix A.3 for more details on this measure). In columns 5-10, we contrast games and non-games, analyze moderating factors and show the results for even more restrictively matched app pairs.

Taking all supply specifications together, we conclude that developers are trading access to privacy-sensitive information for money. Pay strategies are particularly associated with more privacy for the users, because paid apps request fewer sensitive permissions.

6.2 Demand-Side Analysis

[Table 3 (MAIN DEMAND RESULTS) HERE]

Cross Section: Table 3 shows our main results for the relationship between app demand and the presence of privacy-sensitive permissions. We show (1) descriptive regressions for the full cross section (cols. 1-3), (2) fixed-effects panel regressions (cols. 4 and 5), and (3) a difference-in-differences-style (DiD) analysis (cols. 6 and 7). The demand measure (dependent variable) is the monthly change in the number of ratings (in logs). Column 1 shows the raw correlation between permissions and demand. Absent any control variables, the coefficients for both the privacy-sensitive permission dummy and the number of nonsensitive permissions have positive signs, which are presumably related to confounding factors. More permissions could be related to greater functionality, which leads to more demand. After we add control variables and controls for permissions that allow internet access and to show ads (col. 2), the positive coefficient of the dummy capturing privacy-sensitive permissions becomes negative (-6.5%) and significant. In contrast, the coefficient of unproblematic permissions remains positive and significant throughout all cross-section-based specifications. Also the coefficient for ad-related permissions is positive, which points to the need for more rigorous estimation strategies, such as panel techniques. In column 3 we use the number of privacy-sensitive permissions as an alternative measure of potential intrusion. Also with this measure we find a significant and negative coefficient for privacy-sensitive permissions, whereas we find a positive one for unproblematic permissions.[42] Both columns (2 and 3) document that privacy-sensitive permissions are associated with lower demand.

Panel Data Analysis: Columns 4-7 in Table 3 probe the robustness of our cross section results by accounting for the unobserved heterogeneity of apps. Columns 4 and 5 show the results of our fixed-effects panel regression analysis. We run a fixed-effects panel regression for apps that changed the number of privacy-sensitive permissions during our period of observation. The results show that privacy-sensitive permissions are associated with lower demand, independently of whether we look at the presence of any (dummy) or the number of sensitive permissions. Similarly, the introduction of ads is no longer positive (unlike in the cross section, where ads are associated with greater demand, due to potential endogeneity).

Difference-in-Differences-Style Analysis: Our second approach to tackle unobserved heterogeneity is a DiD-style comparison of the same app's success in both the Google Play Store and in Apple's App Store. In columns 6 and 7 we compare the difference in apps' download rankings in Google's Play Store and Apple's App Store for apps that request sensitive permissions to the difference in the rankings for apps that do not request such permissions. Column 6 uses the aggregated rankings of seven countries whereas column 7 uses only US rankings. In both estimations, the value of the within-app rank difference between the two OS (iOS minus Android) is weakly significantly smaller for apps that require privacy-sensitive permissions, which indicates that apps with privacy-sensitive permissions perform relatively poorly on Android's platform.[43] Specifically, our results suggest a difference in the rank difference of the two groups of apps equal to 12.6 ranks for the results based on the broader set of countries and a difference of 15 ranks for the US-based results. Thus, both specifications confirm the cross section-based results, indicating a negative relationship between privacy-sensitive permissions and demand.

Above, we have presented our baseline result. Consumers have lower demand for apps that request privacy-sensitive permissions. We now move on to analyze moderators that amplify or reduce

[42] The coefficient for $\#_{Privacy}$ of 0.021 suggests a 2.1 percent lower demand for each additional permission. However, when estimating the dummy $D_{Privacy}$ and the number of permissions $\#_{Privacy}$ jointly, we find a significant coefficient of -0.049 for $D_{Privacy}$ and the coefficient of $\#_{Privacy}$ is significant at -0.011 which suggests a non-linear effect, which we further investigate in column 1 of Appendix Table A9.

[43] Already the raw mean values of the OS-specific rankings confirm this finding: The mean rank of apps with(out) privacy-sensitive permissions in the iOS is 50.2(58.5) whereas in the Android OS these mean values equal 52.4 (with such permissions) and 52 (without such permissions). Thus, in the iOS apps with privacy-sensitive permissions have a lower mean rank, i.e., they are more successful than apps without such permissions. In contrast, in the Android OS, if one group has a higher mean rank, it is the group of apps with privacy-sensitive permissions, i.e., apps using such permissions in the Android OS are less successful than apps that do not use such permissions.

the size of this relationship (Table 4) in subsection 6.3. Section 6.4 then provides an extensive series of robustness checks that aim to reduce the role of unobserved heterogeneity in the cross section. Also, we use alternative measures of demand and of privacy, apply alternative estimation specifications, and split our estimation sample into various subsamples.

6.3 Moderating Factors on the Demand Side

Previous literature (cf. Acquisti, Taylor, and Wagman, 2016) has shown that the role of privacy concerns depends on an app's context. In this subsection we analyze how privacy concerns for mobile apps depend on their context. We take inspiration from a framework by Chellappa and Sin (2005) which highlights the trade-offs that consumers face when choosing how much personal information to share in exchange for better services. Chellappa, Sambamurthy, and Saraf (2010) embed these trade-offs in a formal mechanism design problem that captures three key moderators of aggregate privacy concerns, that arise in our empirical context: (i) the composition of consumers, (ii) trust-inspiring measures by firms, and (iii) the nature and amount of the requested private information.[44] Three predictions can be derived from this model (and the broader privacy literature in general). First, if developers can implement trust-inspiring measures, privacy concerns are reduced, that is, users are more willing to adopt products that request sensitive information. Second, we expect differences in the willingness to share sensitive information among subgroups of the total population of users - specifically, we expect mature users to be more aware or cautious than younger users in adopting privacy-sensitive apps. Third, and lastly, users might more hesitantly adopt apps with the ability to collect personal information if they operate in sensitive contexts, in which the shared information might be abused or used to discriminate against the consumer. Examples of such contexts are health-related or financial apps.

In Table 4 we highlight examples in which trust (cols. 1-4), the targeted user group (cols. 5-6), and the sensitivity of the shared information (cols. 7-8) indeed moderate the relationship between privacy sensitive permissions and app demand: For each of these factors, we separately estimate apps of a specific type (e.g., apps that have a privacy policy) and apps that do not satisfy the criterion. The coefficient of interest is $D_{Privacy}$ and, specifically, we compare the coefficient for two groups of apps and analyze whether it is equal for both groups or not. This analysis serves a double function.

[44] In Chellappa, Sambamurthy, and Saraf (2010) consumers have heterogeneous privacy concerns, and a monopolist uses technology to reduce the disutility from requesting personal information. The monopolist offers a menu and decides on optimal personalization with and without subsidization. The model's technology parameter allows accommodating trust-inspiring measures or extending personalization disutility beyond nuisance from targeting to accommodating worries about very sensitive information, or other sources of concern, such as identity theft or embarrassment.

First, if privacy concerns drive our results, the relationship of interest should vary by context as theory predicts. Second, we can generate additional insights into how and in which contexts privacy concerns matter in the market for mobile apps.

[Table 4 (MODERATING FACTORS) HERE]

In columns 1-4 of Table 4, we analyze the role of apps' past success and trust-enhancing measures such as privacy policies. Columns 1 and 2 study how the effects vary for top apps with a very large user base (i.e. apps with more than 75,000 installations). Among apps with a very large user base, we observe no negative correlation between the presence of privacy-sensitive permissions and demand, whereas for less successful apps the negative effect prevails. Columns 3 and 4 estimate how the relationship of interest differs for apps that are transparent about their privacy policy. Apps that publish a directly accessible privacy policy in the Google Play Store would be expected to inspire additional trust from app users. Indeed, among apps with an easily accessible privacy policy, sensitive permissions have no statistically significant effect.[45] Columns 5-6 in Table 4 analyze the role of privacy concerns for different user groups. Columns 5-6 look at how the coefficient estimates differ for apps that are not suitable for children or young adults. Privacy might be a greater concern in such apps, because they target a different group of users, who are, on average, more privacy sensitive.[46] The significantly more negative coefficient for $D_{Privacy}$ in column 5 highlights that mature users (or no indication) avoid privacy sensitive permissions more actively, while apps that are suitable for children are less punished for using privacy-sensitive permissions.[47] In columns 7-8, we separately analyze the most sensitive category of apps, medical and health-related apps, because we would expect the relevance of privacy concerns to differ for this category.[48] Indeed, we see that users of medical and health-related apps seem to avoid privacy-sensitive permissions more than users of other genres.

This subsection highlights that reputation, the app's target group, and the sensitivity of the app's context moderate the negative relationship between using privacy-sensitive permissions and

[45] Similar findings emerge for apps from "top developers" (which receive a visible badge from the Play Store administration) or when we separately analyze corporate apps or apps from large and well-known websites (Facebook, Amtrak, banks, Starbucks, etc.; cf. Appendix Table A7). Note that we do not claim that these coefficients measure the causal relationship. We point out, however, that we control for the mere fact that apps with privacy policies might be published by more experienced suppliers (and possibly are better apps).

[46] The results could also be explained by differences in the type of content.

[47] Similar results emerge for games (cf. Appendix Table A7). In this group of apps, users reveal the least sensitive information, and many users are children or young adults.

[48] Besides a general reluctance to digitize health-related information, users might suspect developers to have incentives for sharing their information with insurance companies, for which health-related information about users and their health-related behaviors is very valuable.

app demand. We found the strongest positive moderating effect for well known apps and for apps that adopt a privacy policy. The strongest negative moderation was found for health-related apps. In Appendix Tables A7 and B5 we show the robustness of these results to considering alternative moderating factors or different specifications.

6.4 Robustness

We begin this section by discussing the results of our alternative supply-side estimations (Appendix Table A4). Specifically, we show that the main results do not change if we use more rigorous matching or subsamples and that the moderating factors from the demand-side analysis are also present on the supply side. We then move on to the robustness of our demand-side results. We show additional panel specifications (Appendix Table A5), and more DiD results (Appendix Table A6). In addition, we analyze alternative demand measures (e.g., installation-based demand measures; Appendix Table A8), different ways of measuring privacy-sensitive permissions (Appendix Tables A9 and A14), employing alternative estimation strategies (Appendix Tables A10), and considering/omitting subsegments of the market (Appendix Tables A7, and A11).

6.4.1 Supply Side

The supply side is robust to using alternative definitions of privacy and subsamples: In Appendix Table A4 we test the robustness of our supply-side results that analyze the business model choice, that is, the dependent variable is again the dummy equal to one if an app is a paid version. In columns 1-6, we varied the specification in the supply-side regressions to verify both that our results do not depend on our privacy variable and that they hold within subsamples of our data such as games and normal apps. The main cross section results do not depend on the choice of the specific privacy measure. For increasing numbers of privacy-sensitive permissions, apps are more likely to be free (column 1). Column 2 shows that developers use privacy-sensitive permissions in free apps only together with internet access, highlighting that data collection (via permissions) requires internet access to transfer the data from the app. In columns 3-4, we use our category-specific privacy measure and show that our results are robust to this context-specific definition of privacy-intrusiveness (see Appendix A.3 for more details on this measure). Columns 5 and 6 show that our baseline finding holds for both non-game apps (column 5) and games (column 6). For both groups, we find a negative significant effect.

Supply-side results are robust to including moderating factors and to using more rigorously matched app pairs: In columns 7-10 in Appendix Table A4, we analyze the role of moderating factors on the supply side and restrict the pairs dataset even further to rule out unobserved heterogeneity as an explanation of our results. Columns 7 and 8 show that the role of reputation as a moderating factor is just as influential for the supply side as for the demand side. In column 7 we separately analyze apps with a large user base (more than 10,000 installations), and in column 8 we analyze apps that are associated with a popular website (low traffic rank on Alexa.com). Such apps are generally less likely to be paid versions, but if they are paid, they are more likely to require privacy-sensitive permissions. Columns 9 and 10 analyze the robustness of our supply-side results from the app pairs data by applying more restrictive matching for the pairs data: here we only consider pairs with no difference in description and code length, which was verified by human coders. These results show again that privacy-sensitive permissions are more likely in free apps, independent of how restrictive we are in our matching of the app pairs.

Taken together, these results confirm our baseline findings for the supply side in Table 2.

6.4.2 Demand Side

Panel estimations are robust to using alternative samples and specifications: Next, we run alternative panel fixed-effect regressions, which use an alternative set of controls, employ alternative datasets and vary the explanatory variable of interest. This check ensures that our demand-side results do not critically depend on our choice of controls, the sampling, or the variable of interest. In Appendix Table A5 we show these additional panel fixed-effect regressions. The estimated coefficient of interest remains essentially unchanged for the specification in column 1, where we include only a reduced set of controls (we include only alternative sets of permissions) or when the data are limited to apps that introduced permissions, but no major update to functionality, during our period of observation (column 3). Moreover, we see similar results when analyzing the number of sensitive permissions, rather than an indicator of their presence (cols. 4-6). Finally, the panel results are also robust to using a category-specific definition of privacy-sensitive permissions, as shown in Appendix Table A14. Altogether, the main results are confirmed, and the estimated effect of introducing permissions is even a bit larger when we restrict the sample to apps without any improvements in functionality.

DiD-style comparison is robust to using alternative measures of the ranking difference: In Appendix Table A6 we test the robustness of our DiD-style setup to alternative methods of computing the difference in iOS and Android rankings as well as to splitting the sample into game and non-game apps. As before, the dependent variable captures differences in the download rankings of an app on the two platforms (iOS App Store rankings minus Google Play Store rankings). Columns 1, 2, 5, and 6 compare the apps' average rankings in the seven countries for which we collected rankings, whereas columns 3 and 4 use only US ranks. The dependent variable in columns 1 and 3 is difference between the simple ranking averages, whereas in columns 2 and 4 we use the difference between the newly created in-sample ranks within the OS (as in the baseline demand table), which is based on the average download rankings we observed on AppAnnie.com. Columns 5 and 6 also use the difference between the in-sample rankings based on the average download ranks but contrast games (col. 6) with other apps (col. 5). All specifications except specification 6, which covers only games, show a significant negative effect for privacy-sensitive permissions.[49] This corroborates that apps with privacy-sensitive permissions are on average more successful in the Apple iOS than in the Android OS where these permissions are visible to the user before the app is installed. Thus, the results reaffirm our baseline demand results despite using a completely different sample of apps and a completely different identification approach.

Demand-side results are robust to using alternative demand and app popularity measures: In Appendix Table A8 we verify that our main findings for the demand side do not depend on our preferred demand measure. Specifically, we estimate our main specification with eight ten alternative demand measures based on ratings (cols. 1 and 2), the direct measure of installations (cols. 3-5), and three measures of predicted monthly new installations (cols. 6-8). Our main demand-side results remain the same, independent of whether we use measures based on installations or measures based on ratings or of how we vary the time window of ratings- or installation growth (cols. 1-5).

Similarly, when we use predicted changes in the number of installations (cols. 6-8), our results are also confirmed. In columns 6-8 we apply three measures of predicted installation numbers. For each of the measures, we exploit the cross-section information on changes in ratings to predict changes in installation numbers. In column 6, we predict monthly installation changes in April 2012 based on the observed change in the number of ratings. In column 7, we predict installation changes

[49] The insignificant effect for games is in line with additional alternative results, which show that a weaker or even an insignificant relationship between privacy-sensitive permissions and demand exists for games (see Appendix Tables A7 and B5).

between April and September 2012 based on the observed change in ratings during this period. For both specifications, we use a parametric log-log-specification to predict the change in the number of installations (cf. cols. 2 and 4 of Appendix Table A12). In column 8 of Appendix Table A8 we use a measure of non-parametrically predicted monthly installations in April 2012.

Finally, in columns 9 and 10 we study how the existence of privacy-sensitive permissions affects apps' popularity as measured by their average ratings. In column 9 the dependent variable is the average rating in April 2012 (which is between one and five stars), whereas in column 10 it is the average rating between April and September 2012. Both specifications indicate that apps receive worse ratings if they use privacy-sensitive permissions.

Demand-side results are robust to using alternative privacy measures: We verify that our results do not depend on our workhorse definition of what is considered "privacy sensitive" by estimating our main specification with multiple alternative definitions. Specifically, in Appendix Table A9 we show our main regression for eight alternative privacy measures. These measures include Google's own classification, the definition by Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy (2012), a category-specific measure, and one derived via classifiers from Amazon Mechanical Turk. The results show that the negative relationship between privacy-sensitive permissions and demand holds across six of the eight measures that we consider, and in column 7 only one out of four permission groups has a positive coefficient.[50] The only measure with a positive relationship is the *privacy grade* provided by Lin, Hong, and Sadeh. (2014), which grades apps' privacy intrusiveness in 2014 for a subsample of our data. The positive coefficient indicates that demand is higher for apps with a bad privacy grade. This finding was unexpected but could be explained by the fact that the privacy grades are based on analyzing app's software libraries, which is possible only if apps use standard software libraries. Using such libraries might require a certain level of experience and professionalism and thus might pick up unobserved developer quality.

As in our baseline results, D_{Ads} and the number of unproblematic permissions are positively associated with demand in all specifications, which indicates that unproblematic permissions, unlike privacy-sensitive permissions, do not lead to lower demand but might capture functionality and therefore might have a positive coefficient.[51] We also would like to highlight the results in column 7, where we distinguish between permissions by type. Two out of four permission groups have a

[50]That D_{ID} would have a positive coefficient is surprising at first sight, but the permission was hard to recognize for users as intrusive in 2012, as it was called "READ_PHONE_STATE."

[51]Recall, though, that D_{Ads} has a negative effect when using within variation in our panel data in Table 3.

negative, significant coefficient ($D_{Location}, D_{Profile}$), one coefficient ($D_{Communication}$) is insignificant, and one (D_{ID}) seems to be positively associated to demand. Even though they are based on an analysis of the cross section, these findings might indicate that users have different degrees of willingness to share certain types of information and warrant further research using individual user-level data. Moreover, we analyze category-specific privacy-sensitive permissions in additional depth. The "category-specific" definition takes into account that privacy is context dependent, and considers that apps in a certain category might need access to certain data to perform their services.[52] In Appendix A.3 we discuss the measure, and especially in Appendix Table A14, we show the main demand-side results for the cross section and panel also hold if we use the category-specific definition.

Demand-side results are robust to accounting for censoring, selection, network effects, and endogeneity: We also run the main regressions with alternative estimation methods to verify that our demand-side findings are not merely due to our OLS-specification. In Appendix Table A10 we use Tobit estimation to account for the possibility of censoring in the dependent variable (cols. 1-2). In addition, in columns 3 and 4 we provide results from Heckman selection models, where we model apps' survival between April and September 2012 (col. 3) as well as apps' survival between 2012 and 2014 (col. 4). We also include controls that account for potential network effects (col. 5) and introduce instruments for potential endogeneity in prices and developers' use of permissions (cols. 6-9). None of these alternative specifications drastically changes our baseline findings.

Sample splits and success: In Appendix Table A7 we show that alternative moderating factors highlight similar patterns as documented in Table 4. The table analyzes the moderating role of the following factors: (1) top developer-status (cols. 1-2), (2) apps connected or not connected to a well known website (using Alexa ranks) (cols. 3-4), (3) games vs. non-games (cols. 5-6), and (4) pricing strategy (cols. 7-8). Also for these moderating factors, the strength of the effect varies along similar lines as in our main results.

Next, we test whether our main result is robust to excluding the most and the least successful apps.[53] This robustness check allows us to exclude the possibility that only a few apps drive our results and provides additional insights into the anatomy of our main demand-side results. Appendix Table A11 shows that excluding the most successful apps (5% of apps with the highest demand; cols. 1 and 2) and excluding both the most and least successful apps (those without any new rating

[52]E.g., a running app needs access to its user's location.

[53]We thank one of the anonymous referees for suggesting this additional test.

in April 2012; cols. 5 and 6) does not change our findings. We found an insignificant permission coefficient only by disregarding the least successful apps (cols. 3 and 4). However, this finding is not robust to including a dummy that equals one if an app already has a user base of more than 10,000 installations as well as an interaction term with privacy-sensitive permissions. Doing so, we also find a negative significant effect for privacy-sensitive permissions for that sample. The dummy and the interaction term together capture the influence of an app's reputation (as in the analysis of moderating factors). This finding corroborates the role of reputation as a moderating factor and suggests that large and established apps may have easier access to sensitive user data than other players. Unknown apps face a penalty for requesting privacy-sensitive information.[54]

6.5 Implied Valuations

We now turn to discussing the value of privacy-sensitive permissions, as implied by our demand and supply estimates. We do so by comparing the estimated demand coefficient for privacy-sensitive permissions with the own-price elasticity of apps. Both higher prices and the use of sensitive permissions are associated with lower demand. From the cross-section results the coefficient for apps that use sensitive permissions is about 0.9 times the size of that corresponding to a 1 percent increase in price. The panel estimates indicate that ratio is 1.5, suggesting that the penalty for using sensitive permissions is equal to the penalty corresponding to a 1.5 percent higher price. The average price of paid apps of 2 euros implies a lower bound "willingness to avoid" of 2-3 euro cents. We reiterate, however, that our work suffers from limitations, and we consider our coefficients to be lower-bound estimates of the effect of introducing such permissions, because of the endogeneity concerns that result from imprecise demand measures and unobserved app quality. Hence, the implied valuations should be used carefully, and any policy implications require validation with individual-level or experimental data.

To derive a valuation for data on the developer side, a first meaningful way of quantifying this value can be based on columns 1-2 in Table 2, where we observe that a privacy-sensitive permission is associated with a 15 percent higher probability that an app is free. Multiplying this by the average price of a paid app (2 euros) gives an estimate of 30 euro cents. We find a similar value if we use the results from the price choice model (Table 2, col. 9). The coefficient of privacy-sensitive permissions (-0.118) indicates that developers are willing to reduce the app price by around 12 percent if the app has a privacy-sensitive permission. This implies a price reduction of 24 euro cents for the average

[54] Alternatively, we split the sample by apps' past success and verified that the negative relationship between demand and privacy-sensitive permissions is not driven by a specific segment of apps with large or small user bases.

paid app. Again, these estimates suffer from severe limitations, and they are likely confounded by other sources of income that the free apps can generate (building a user base, etc.). We leave it to future research to evaluate the developers' valuations more precisely.

7 Discussion

We document that private information plays a critical role in the market for mobile apps, because it resembles a second "means of payment." On the supply side, app developers ask for more privacy-sensitive permissions in exchange for a free app, or they offer a less privacy-sensitive app for a higher price. On the demand side, we observe fewer installations for apps that request more privacy-sensitive permissions. Together these results highlight a money-for-privacy trade-off in the market for mobile apps.

We use various subsets of data to obtain our results: a full cross section, panel data, carefully selected app siblings, and a DiD-style strategy that compares the same app on the two biggest app platforms. We can show the relationship for both the supply and demand for apps, and our extensive set of robustness checks highlights the robustness of our findings. Our conclusions persist across (1) different ways of quantifying intrusiveness, (2) several alternative classifications to measure the sensitivity of permissions (Google's own classification, the definition by Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy (2012), a category-specific measure, and one derived via classifiers from Amazon Mechanical Turk), (3) several conceptualizations of demand (ratings, installations, first-differences, etc.), and (4) several alternative estimation approaches.

Two patterns emerge from our analysis. First, we find economically significant demand-side effects.[55] To be precise, our (lower-bound) estimate of consumers' willingness to avoid sharing personal information is between 2 and 3 euro cents. These estimates are between the extremes of the privacy valuations found in experimental work and other contexts (Beresford, Kübler, and Preibusch, 2012; Carrascal, Riederer, Erramilli, Cherubini, and de Oliveira, 2013; Grossklags and Acquisti, 2007). However, our estimates are considerably smaller than previous estimates for apps based on surveys (Savage and Waldman, 2014), even though the Android OS makes the information very explicit. The potential explanations include: (1) Users might simply show behavioral biases and overstate their preference for privacy in surveys or experiments. (2) We provide only lower-bound estimates of the users' true reaction to privacy sensitive permissions, because of data limitations in our demand measure and the challenge of fully isolating the effect of privacy sensitivity from that of

[55] Note that our data predate increased public awareness of data collection by the US National Security Agency.

functionality. (3) Having too much unstructured information about permissions might be like having no information at all. While the first explanation stresses the importance of using observational data for appropriately estimating consumers' valuations, the third hypothesis could align our small coefficients with the findings of Savage and Waldman (2014), who found that consumers valued an app's access to private information much more negatively, but only *after* receiving information about the permissions' implications. To test whether information overflow is the underlying reason, simple pieces of summary information should suffice to reduce the gap between earlier findings and ours.[56]

The second key pattern in our analysis emerges from our analysis of moderating factors. We confirm users' generally negative attitude toward privacy-sensitive permissions but also show that this finding does not always hold. The strength of the relationship depends on the context as well as app and user characteristics, which confirms previous research (Acquisti, John, and Loewenstein, 2013; Acquisti, Taylor, and Wagman, 2016; Chellappa and Sin, 2005; Goldfarb and Tucker, 2012). For example, the negative effect of permissions is weaker if privacy-sensitive permissions are common in an app category or if it is likely that a consumer is relatively young. However, the negative relationship between demand and permissions can also become stronger. If an app is in more sensitive categories (such as health-related apps), intrusive permissions are associated with even lower demand. This finding suggests that users are aware of the value of specific types of personal information and share it more cautiously. Trust appears to be the most prominent factor, as is documented by the insignificant effect of sensitive permissions on apps that adopt a privacy policy and on widely adopted apps that already have a large user base. Supplementary results show similar patterns for apps by top developers or apps that can be associated to widely used websites (via Alexa.com). Consumers who care but do not fully understand all permissions might create this pattern, because they might rationally prefer apps that are well known to those that are unknown. On the positive side, such behavior allows established firms to improve their products based on consumer data, which can increase quality. On the negative side, users' reliance on outside reputation might create a barrier to entry: well established apps can gather user information more easily than newcomers, reinforcing the arguments in Campbell, Goldfarb, and Tucker (2015) and potentially reducing innovation in the app market.

Our work suffers from several limitations, which offer avenues for further research. Although

[56]Suggestions include a traffic light scale of an app's intrusiveness (Tsai, Egelman, Cranor, and Acquisti, 2011) or local storage of the user's information on their device (Sutanto, Palme, Tan, and Phang, 2013). Our supporting analysis (Online Appendix Table B5) indeed indicates that Google's "maliciousness warning" is associated with greater avoidance. However, caution is needed, since such schemes might make it harder for developers to monetize the private information on free apps which could reduce both supply and the adoption of apps.

we do our best to scrutinize the robustness of our main results, we do not have ideal data. We do not have access to the exact numbers of installations and even more precise information on an app's characteristics, quality, or use of ads and have to approximate these variables. Approximation carries the risk of picking up confounding influences from ads or unmeasurable aspects of quality. We address these concerns by intensely testing the robustness of our findings, and we further validate our main findings on more rigorously matched subsamples. However, these strategies involve their own compromises, such as smaller datasets, the risk of attenuation bias, and lower significance levels. Future work could improve our research by using the exact installations numbers or daily active users, which is developers' most preferred measure of usage. Moreover, we cannot completely eliminate two problems: (1) Time-variant unobserved heterogeneity cannot be controlled by a panel fixed effect.[57] (2) When conditioning on apps without functionality improvement, we fail to observe developers who introduce sensitive and unnecessary permissions alongside a new feature. This caveat reduces the power of our specification. Similarly, the analyzed moderating factors do not change over time, which disallows a panel analysis. Hence, these additional results are based on conditional correlations in the cross section. For example, the weaker negative relationship between permissions and demand for younger consumers should be validated with user-level data. Such information is needed to decide whether young users install apps with too little reflection or whether older users are too cautious. Hence, great care is warranted before any policy changes can be implemented in this potentially sensitive two-sided market, and further research must carefully evaluate the effects of policy changes that aim to highlight the potential intrusiveness of an app. Lastly, further research should analyze developers' dynamic strategies to gain access to data in the market for mobile apps.

In sum, we see our findings as a first step in understanding the role of privacy in app markets. Any policy implications that we suggest should be validated with individual-level or experimental data. We believe that such a careful evaluation would be a fruitful avenue for further research. Hence, significant unleveraged potentials might exist to further improve the performance of the market, especially in sensitive categories.

8 Conclusion

The so-called American perspective on privacy views it as a property right (Camp, 1999). This perspective is interesting to managers, as it suggests that private information can be negotiated

[57] E.g., if many apps introduce permissions at exactly the same stage of their development, and if this were positively correlated with installation or review growth, then we would overestimate the effect;

between firms and consumers. Our paper speaks to this view and documents trade-offs with respect to privacy-sensitive information for both firms and consumers in the market for mobile apps. The app market is economically relevant and offers previously unknown potential for collecting information about consumers. A money-for-privacy trade-off in this market matters, because it suggests a cost for greater privacy protection, independent of whether such protection is desired.

We base our evidence on large observational data that reflect revealed preferences. We analyzed information on nearly all the apps in Google's Android Market in 2012 (around 300,000 apps). We augmented the data with additional information from Alexa.com, privacygrade.org, and AppAnnie.com. Using these data, we study the role of privacy in both the supply and demand for mobile apps. Specifically, we study (1) whether developers use app permissions to collect private information about users (e.g., information about their communication behavior, location and profile), and (2) how consumers' installation behavior is related to privacy-sensitive permissions.

Using several datasets, with alternative measures of intrusiveness and multiple specifications, we document that private information plays a prominent role on both sides of this market: App developers offer either a lower price for a more privacy-sensitive app or a higher price for more privacy. Consumers choose between these two options (privacy and money), and we indeed find a small but robust and economically significant negative relationship between permissions and demand.

In addition, the sensitivity of consumers toward apps' permissions was found to depend on the app's context: Demand is lower for apps with content that is rated as appropriate for older users, or in more sensitive categories, such as health-related apps. The negative correlation between demand and permissions disappears for known apps and apps that adopt a privacy policy. This differential behavior is in line with consumers who find it difficult to evaluate whether they are installing a privacy-endangering product and react by favoring brands and apps that have been widely adopted. Yet, we also found more cautious behavior in sensitive contexts, such as health, which suggests that consumers are considering costs and benefits when sharing their information.

Our results suggest that the app market is a favorable environment for established apps who seek to collect information about their consumers. The flip-side interpretation of this result is that the resulting lack of trust of unknown app developers might constitute a significant barrier to entry. Such a barrier could point to suboptimal innovative performance of the market for mobile apps especially in "sensitive" categories, such as health-related apps. If consumers lack sufficient knowledge to rely on their judgment regarding privacy-sensitive permissions, a simple certificate or label by a third party (e.g., a traffic light symbol) might help to reduce this entry barrier.

References

Acquisti, A., L. Brandimarte, and G. Loewenstein (2015): "Privacy and Human Behavior in the Age of Information," *Science*, 347(6221), 509–514.

Acquisti, A., L. K. John, and G. Loewenstein (2013): "What is Privacy Worth?," *The Journal of Legal Studies*, 42(2), 249–274.

Acquisti, A., C. R. Taylor, and L. Wagman (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54(2), 442–492.

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

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

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

Appthority (2014): "App Reputation Report," Available at: `https://www.onegini.com/hubfs/White_Papers/Mobile_App_Reputation_Report-Summer-2014.pdf?t=1454942139543` [2/18/2017].

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

Askalidis, G. (2015): "The Impact of Large Scale Promotions on the Sales and Ratings of Mobile Apps: Evidence from Apple's App Store," *Cornell University Working Paper*.

Aziz, A., and R. Telang (2015): "What is a Cookie Worth?," *Heinz College Working Paper*.

Beales, H. (2010): "The Value of Behavioral Targeting," *Network Advertising Initiative*, 1.

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

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bertschek, I., and R. Kesler (2017): "Let the User Speak: Is Feedback on Facebook a Source of Firms' Innovation?," *ZEW Discussion Paper*.

Book, T., A. Pridgen, and D. S. Wallach (2013): "Longitudinal Analysis of Android Ad Library Permissions," *arXiv preprint arXiv:1303.0857*.

Bresnahan, T., J. Orsini, and P.-L. Yin (2014): "Platform Choice by Mobile App Developers," *NBER Working Paper*.

Camp, L. J. (1999): "Web Security and Privacy: An American Perspective," *The Information Society*, 15(4), 249–256.

Campbell, J., A. Goldfarb, and C. Tucker (2015): "Privacy Regulation and Market Structure," *Journal of Economics & Management Strategy*, 24(1), 47–73.

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

Carrascal, J. P., C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira (2013): "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," in *Proceedings of the 22nd international conference on World Wide Web*, pp. 189–200. International World Wide Web Conferences Steering Committee.

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

Chaudhari, H., and J. Byers (2017): "The Impact of Zero-Price Promotions on Sales: A Case Study of Amazon Appstore," *Boston University Working Paper*.

CHELLAPPA, R. K., V. SAMBAMURTHY, AND N. SARAF (2010): "Competing in crowded markets: Multimarket contact and the nature of competition in the enterprise systems software industry," *Information Systems Research*, 21(3), 614–630.

CHELLAPPA, R. K., AND R. G. SIN (2005): "Personalization versus privacy: An empirical examination of the online consumer's dilemma," *Information technology and management*, 6(2), 181–202.

CHIA, P. H., Y. YAMAMOTO, AND N. ASOKAN (2012): "Is this App Safe? A Large Scale Study on Application Permissions and Risk Signals," pp. 311–320. ACM.

CHRISTL, W., AND S. SPIEKERMANN (2016): *Networks of Control - A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*. Facultas, Vienna.

CONITZER, V., C. R. TAYLOR, AND L. WAGMAN (2012): "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases," *Marketing Science*, 31(2), 277–292.

DAVIS, J. P., Y. MUZYRYA, AND P.-L. YIN (2014): "Experimentation Strategies and Entrepreneurial Innovation: Inherited Market Differences in the iPhone Ecosystem," *INSEAD Working Paper*.

DINI, G., F. MARTINELLI, I. MATTEUCCI, M. PETROCCHI, A. SARACINO, AND D. SGANDURRA (2012): "A Multi-Criteria-Based Evaluation of Android Applications," in *International Conference on Trusted Systems*, pp. 67–82. Springer.

DUBÉ, J.-P., Z. FANG, N. FONG, AND X. LUO (2017): "Competitive Price Targeting with Smartphone Coupons," *Marketing Science*, 36(6), 944–975.

EGELMAN, S., A. P. FELT, AND D. WAGNER (2013): "Choice Architecture and Smartphone Privacy: There is a Price for That," in *The Economics of Information Security and Privacy*, pp. 211–236. Springer.

ERSHOV, D. (2016): "The Effect of Consumer Search Costs on Entry and Quality in the Mobile App Market," *University of Toronto Working Paper*.

FAHL, S., M. HARBACH, T. MUDERS, L. BAUMGÄRTNER, B. FREISLEBEN, AND M. SMITH (2012): "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*, pp. 50–61. ACM.

FONG, N. M., Z. FANG, AND X. LUO (2015): "Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions," *Journal of Marketing Research*, 52(5), 726–735.

GHOSE, A., AND S. P. HAN (2014): "Estimating Demand for Mobile Applications in the New Economy," *Management Science*, 60(6), 1470–1488.

GOLDFARB, A., AND C. TUCKER (2011): "Privacy Regulation and Online Advertising," *Management Science*, 57(1), 57–71.

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

GROSS, R., AND A. ACQUISTI (2005): "Information Revelation and Privacy in Online Social Networks," in *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, pp. 71–80. ACM.

GROSSKLAGS, J., AND A. ACQUISTI (2007): "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)*.

HANN, I.-H., K.-L. HUI, S.-Y. T. LEE, AND I. P. PNG (2007): "Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach," *Journal of Management Information Systems*, 24(2), 13–42.

HANN, I.-H., K.-L. HUI, S.-Y. T. LEE, AND I. P. PNG (2008): "Consumer Privacy and Marketing Avoidance: A Static Model," *Management Science*, 54(6), 1094–1103.

HAUSMAN, J. A. (1996): "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, pp. 207–248. University of Chicago Press.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Hu, J., H.-J. Zeng, H. Li, C. Niu, and Z. Chen (2007): "Demographic prediction based on user's browsing behavior," in *Proceedings of the 16th international conference on World Wide Web*, pp. 151–160. ACM.

Hu, N., J. Zhang, and P. A. Pavlou (2009): "Overcoming the J-Shaped Distribution of Product Reviews," *Communications of the ACM*, 52(10), 144–147.

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

Jeon, J., K. K. Micinski, J. A. Vaughan, A. Fogel, N. Reddy, J. S. Foster, and T. Millstein (2012): "Dr. Android and Mr. Hide: Fine-Grained Permissions in Android Applications," in *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pp. 3–14. ACM.

Johnson, G. A. (2013a): "The Impact of Privacy Policy on the Auction Market for Online Display Advertising," *Simon School Working Paper*.

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

Li, X., T. Bresnahan, and P.-L. Yin (2016): "Paying Incumbents and Customers to Enter an Industry: Buying Downloads," *SSRN*.

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

Lin, J., J. Hong, and N. Sadeh. (2014): "Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings.," *Symposium on Usable Privacy and Security (SOUPS 2014)*.

Liu, Y. (2017): "Mobile App Platform Choice," *SSRN*.

Liu, Y., D. Nekipelov, and M. Park (2014): "Timely versus quality innovation: The case of mobile applications on iTunes and Google Play," *NBER Working Paper*.

Marthews, A., and C. Tucker (2017): "Government Surveillance and Internet Search Behavior," *SSRN Working Paper*.

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

Mobasher, B., R. Cooley, and J. Srivastava (2000): "Automatic personalization based on web usage mining," *Communications of the ACM*, 43(8), 142–151.

Muchnik, L., S. Aral, and S. J. Taylor (2013): "Social Influence Bias: A Randomized Experiment," *Science*, 341(6146), 647–651.

Mylonas, A., M. Theoharidou, and D. Gritzalis (2013): "Assessing Privacy Risks in Android: A User-Centric Approach," in *Proceedings of the 1st international workshop on risk assessment and risk-driven testing (RISK-2013), Springer, Turkey (November 2013)*.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513–548.

Norberg, P. A., D. R. Horne, and D. A. Horne (2007): "The privacy paradox: Personal information disclosure intentions versus behaviors," *Journal of Consumer Affairs*, 41(1), 100–126.

OECD (2013): "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value," *OECD Digital Economy Papers*, (220).

Preibusch, S., and J. Bonneau (2013): "The Privacy Landscape: Product Differentiation on Data Collection," in *Economics of Information Security and Privacy III*, pp. 263–283. Springer.

Pu, Y., and J. Grossklags (2016): "Towards a Model on the Factors Influencing Social App Users' Valuation of Interdependent Privacy," *Proceedings on Privacy Enhancing Technologies*, (2), 61–81.

Racherla, P., J. S. Babb, and M. J. Keith (2011): "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, pp. 1–13.

Rainie, L., S. Kiesler, R. Kang, M. Madden, M. Duggan, S. Brown, and L. Dabbish (2013): "Anonymity, Privacy, and Security Online," Discussion paper.

Ransbotham, S., and S. Mitra (2009): "Choice and Chance: A Conceptual Model of Paths to Information Security Compromise," *Information Systems Research*, 20(1), 121–139.

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

Sarma, B. P., N. Li, C. Gates, R. Potharaju, C. Nita-Rotaru, and I. Molloy (2012): "Android Permissions: A Perspective Combining Risks and Benefits," in *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, pp. 13–22. ACM.

Savage, S. J., and D. M. Waldman (2014): "The Value of Online Privacy: Evidence from Smartphone Applications," .

Schmidt, J., P. K. ten Venne, and R. Eikenberg (2012): "Selbstbedienungsladen Smartphone," *c't*, Heft 7, 114ff, https://shop.heise.de/katalog/selbstbedienungsladen-smartphone.

Seneviratne, S., H. Kolamunna, and A. Seneviratne (2015): "A measurement study of tracking in paid mobile applications," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, p. 7. ACM.

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

Sutanto, J., E. Palme, C.-H. Tan, and C. W. Phang (2013): "Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users.," *Mis Quarterly*, 37(4), 1141–1164.

Tambe, P., L. M. Hitt, and E. Brynjolfsson (2012): "The extroverted firm: How external information practices affect innovation and productivity," *Management Science*, 58(5), 843–859.

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

Taylor, C. R., V. Conitzer, and L. Wagman (2010): "Online Privacy and Price Discrimination," Discussion paper.

Taylor, V. F., and I. Martinovic (2016): "Quantifying Permission-Creep in the Google Play Store," *arXiv preprint arXiv:1606.01708*.

Tsai, J. Y., S. Egelman, L. Cranor, and A. Acquisti (2011): "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research*, 22(2), 254–268.

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

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

Turow, J., J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy (2009): "Americans Reject Tailored Advertising and Three Activities that Enable It," Discussion paper.

Wathieu, L. (2002): "Privacy, Exposure and Price Discrimination," Discussion paper.

Woodcock, R. A. (2017): "Big Data, Price Discrimination, and Antitrust," *Hastings Law Journal*, 68(6), 1371.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2nd edn.

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

Zang, J., K. Dummit, J. Graves, P. Lisker, and L. Sweeney (2015): "Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps.," *Technology Science.*, 30.

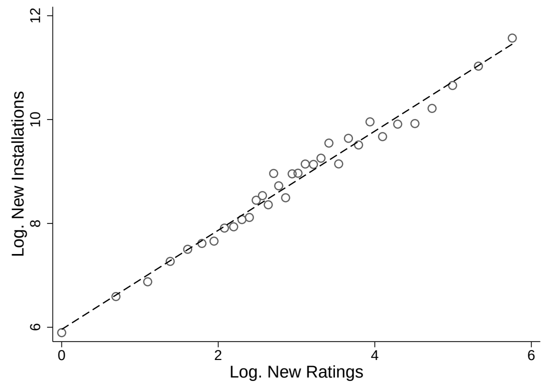
Tables and Graphs

Table 1: Summary Statistics of the Main Datasets

	Cross Section		Panel		Pairs		Android-iOS	
	Free	Paid	Free	Paid	Free	Paid	Free	Paid
Outcome Measures:								
Δ Ratings	9.23	2.31	89.07	34.19	67.33	6.08	27107.85	807.74
Δ Installations	3178	105	16559	743	8213	93	1.5e+06	61333
Ratings	158.53	44.50	1083.07	372.03	623.53	81.05	1.7e+05	5096.88
Installations	43374	1936	1.9e+05	11719	1.3e+05	2740	2.7e+07	2.7e+05
Price	0.00	2.01	0.00	4.41	0.00	1.37	0.00	3.04
Permissions:								
$\#_{TotalPerm}$	4.10	2.32	8.08	6.67	3.78	3.10	11.17	5.71
$D_{Privacy}$	0.50	0.28	0.87	0.76	0.49	0.40	0.95	0.60
$\#_{Privacy}$	1.19	0.52	2.82	1.88	0.93	0.70	3.40	1.03
$\#_{CleanPerm}$	2.92	1.80	5.26	4.79	2.85	2.40	7.77	4.68
$D_{PrivCatSpec}$	0.22	0.09	0.52	0.27	0.14	0.08	0.38	0.12
$D_{MTurkEP2}$	0.12	0.08	0.22	0.27	0.12	0.12	0.39	0.11
D_{Google}	0.35	0.16	0.66	0.49	0.25	0.21	0.64	0.20
$D_{Sarmaetal}$	0.47	0.25	0.85	0.70	0.46	0.35	0.90	0.52
D_{ID}	0.32	0.14	0.69	0.47	0.30	0.22	0.76	0.42
$D_{Location}$	0.29	0.10	0.61	0.31	0.21	0.12	0.46	0.17
$D_{Communication}$	0.06	0.05	0.14	0.18	0.06	0.06	0.21	0.06
$D_{Profile}$	0.21	0.11	0.45	0.41	0.19	0.18	0.57	0.22
$D_{Internet}$	0.81	0.44	0.98	0.85	0.82	0.56	1.00	0.86
D_{Ads}	0.55	0.21	0.89	0.63	0.56	0.31	0.91	0.60
Control Variables:								
Average Rating	3.91	3.99	3.98	4.07	3.79	4.14	4.38	4.21
Size (in KB)	2192	3928	3684	8886	3207	3039	10324	13034
Length Description	730	996	1004	1612	877	874	1717	2019
Number Screenshots	3.36	3.69	4.08	5.25	3.73	3.87	5.56	5.17
Dummy: Video	0.10	0.12	0.12	0.24	0.17	0.17	0.47	0.59
Dummy: Top-Developer	0.00	0.01	0.01	0.04	0.01	0.01	0.61	0.43
Apps by Developer	92	127	63	34	14	14	10	8
Average Installations of Developer	74231	58691	1.4e+05	1.6e+05	57843	91749	1.1e+07	2.5e+06
Observations	128921	48272	30131	2964	1999	1999	94	98

Notes: The table provides an overview over the most important variables, and shows the corresponding descriptive statistics for the four main datasets in this paper. For each dataset we show two columns, where the left column shows averages for free apps and the second column for paid apps. The first two columns show the summary statistics for the entire cross section. The second pair of columns (Col. 3-4) shows those for our panel data set consisting of apps having varied their use of privacy-sensitive permissions at least once between April and September 2012. Columns 5 and 6 use the data set consisting of app-pairs where the paid version of the app has the same or a smaller code size and where both apps have more or less the same description length. Columns 7 and 8 show statistics for those apps for which we were able to collect for 2012 information on app ranks both in the Play Store and in the iOS store.

Figure 1: Relationship between New Installations and New Ratings

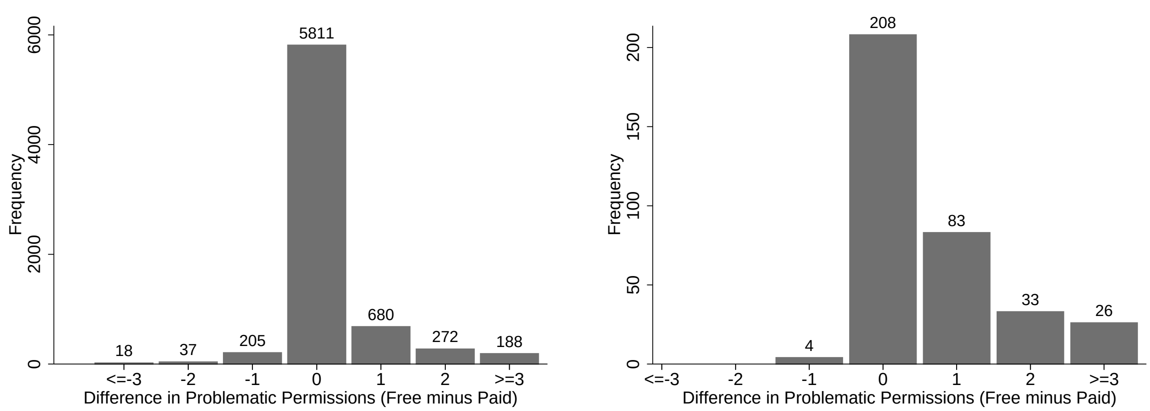


Notes: The figure displays the relationship between apps' log number of new installations and their log number of new ratings. A linear trend is added. The data used is an aggregated version of the cross-section data set used in the estimation sections.

Figure 2: Difference in Privacy-Sensitive Permissions between Free and Paid Twin-Apps

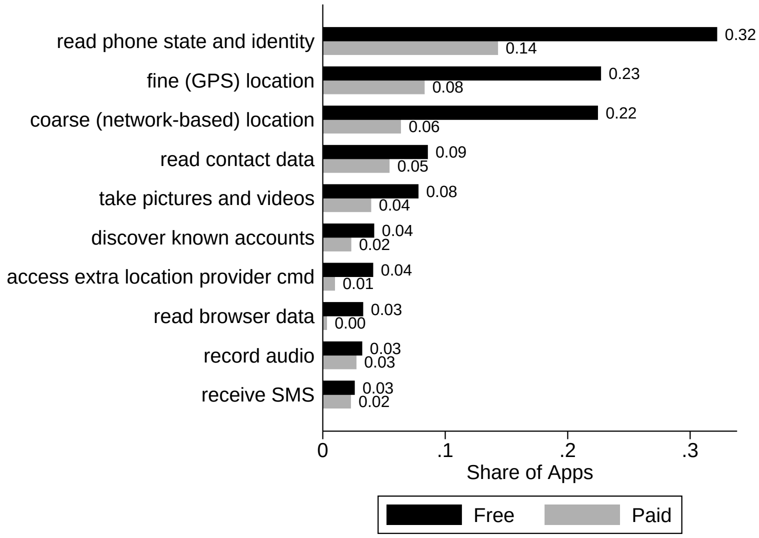
(a) Full Pairs Sample ($N = 7211$)

(b) Selected Pairs Sample ($N = 354$)



Notes: Both figures compare free and paid versions of an app-pair. On the x-axis we show the differences in the number of privacy-sensitive permissions when comparing the pair's free and paid version (= number of privacy-sensitive permissions of the free app minus the number of privacy-sensitive permissions of the paid twin). The y-axis shows the frequency, with which each difference occurs. Panel (a) shows those frequencies for the full pairs sample, whereas panel (b) shows them for the selected sample of app pairs where we ensure that the functionality between apps does not differ. In most cases both versions request the same number of permissions (80% of all pairs, and 59% of the pairs with matched functionality). While the difference in average permission usage is less pronounced when analyzing all pairs, the free app is much more likely to request sensitive permissions once we condition on observing a difference. This pattern emerges for both pair samples, and is strongest on the pairs with no discernable difference in functionality (1151 vs. 267 (or 80% vs. 20%) overall and 142 vs. 4 (or 97% vs. 3%) on pairs with matched functionality).

Figure 3: Frequency of Top 10 Privacy-Sensitive Permissions - Free & Paid



Notes: This diagram shows how often privacy-sensitive permissions are used, both by free and paid apps. The majority of apps do not use these permissions. However, given such permissions are used, free apps use them more frequently than paid apps. The data used is equal to the cross-section data set used in the estimation sections.

Table 2: Main Supply Side Results

	Business Model Choice (D_{Paid})						Price Choice (Log. Price)			
	Cross-Section		Panel		Pairs		Cross-Section		Panel	
$D_{Privacy}$	-0.033***		-0.104		-0.156***		0.012		-0.119*	
	(0.002)		(0.261)		(0.042)		(0.008)		(0.070)	
$\#_{Privacy}$		-0.024***		-0.150*		-0.137***		-0.005		-0.060**
		(0.001)		(0.086)		(0.025)		(0.004)		(0.025)
$\#_{CleanPerm}$	0.005***	0.011***	-0.003	0.035	0.204***	0.226***	0.018***	0.020***	0.048**	0.068***
	(0.001)	(0.001)	(0.024)	(0.030)	(0.023)	(0.025)	(0.002)	(0.002)	(0.020)	(0.023)
$D_{Internet}$	-0.219***	-0.225***	-0.627**	-0.582**	-0.496***	-0.517***	0.067***	0.068***	0.004	-0.015
	(0.003)	(0.003)	(0.290)	(0.246)	(0.037)	(0.038)	(0.007)	(0.007)	(0.085)	(0.083)
D_{Ads}	-0.119***	-0.124***	-0.347	-0.165	-0.535***	-0.547***	0.014	0.014	-0.003	-0.039
	(0.002)	(0.002)	(0.301)	(0.250)	(0.036)	(0.036)	(0.009)	(0.009)	(0.119)	(0.112)
Constant	0.094***	0.090**	10.787***	9.913***	3.495**	3.291**	0.060	0.065	-2.943***	-3.558***
	(0.035)	(0.035)	(2.840)	(2.754)	(1.533)	(1.520)	(0.106)	(0.106)	(0.870)	(0.835)
Category	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176000	176000	96	96	3998	3998	47079	47079	422	422
Num. of Groups			48	48	1999	1999			211	211
Mean of dep. Var.	0.27	0.27	0.50	0.50	0.50	0.50	0.20	0.20	0.53	0.53
SD of dep. Var.	0.44	0.44	0.50	0.50	0.50	0.50	0.60	0.60	0.62	0.62
Adjusted R^2	0.320	0.322	0.963	0.971	0.618	0.622	0.223	0.223	0.177	0.183

NOTES: The table shows the relationship between privacy-sensitive permissions and the strategic choices of app developers: the choice of the business model in Columns 1-6, and the price choice in Columns 7-10. In Columns 1-6 the dependent variable D_{Paid} measures the developer's decision to offer their app for money or for free. It takes the value 1 if users have to pay to download the app. Columns 1 and 2 show descriptive regressions based on the cross-section of data, where the independent variable of interest is (1) an indicator for one or more privacy-sensitive permissions ($D_{Privacy}$ in Column 1) or (2) the number of privacy-sensitive permissions ($\#Privacy$ in Column 2). Columns 3 and 4 show panel fixed effects regressions where we restrict the sample to such apps which changed both the number of privacy-sensitive permissions and the business model at least once between April and September 2012. Here, we use only the first and the last wave of our data to maximize the within variation. Columns 5 and 6 use data on app-pairs where the paid version of the app has the same or a smaller code size and where both apps have more or less the same description length. Columns 7-10 show the results for price-level choices (of paid apps). The dependent variable is the app's price (in logs). Columns 7 and 8 show cross sectional regressions, and Columns 9 and 10 show panel fixed effects regressions based on the first and last wave of our data, where we restrict the sample to such apps which change both the number of privacy-sensitive permissions and their price at least once. In all specifications we drop outliers with respect to app prices, i.e. apps with prices above 8 Euros. All of these regressions control for the number of clean permissions and permissions that are needed to show ads. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Main Demand Side Results

	Cross-Section (Log.$\Delta Ratings$)			Panel (Log.$\Delta Ratings$)		Difference-in-Difference	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$D_{Privacy}$	0.012	-0.065***		-0.059***		-12.660*	-14.988**
	(0.011)	(0.009)		(0.020)		(6.634)	(7.215)
$\#_{Privacy}$			-0.021***		-0.014**		
			(0.003)		(0.007)		
$\#_{CleanPerm}$	0.111***	0.037***	0.041***	-0.010**	-0.005	0.849	1.011*
	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.643)	(0.605)
$D_{Internet}$		-0.200***	-0.211***	0.003	-0.014		
		(0.009)	(0.009)	(0.054)	(0.053)		
D_{Ads}		0.236***	0.230***	0.002	-0.014		
		(0.009)	(0.009)	(0.033)	(0.033)		
Log. Price		-0.071***	-0.071***	-0.044***	-0.043***		
		(0.001)	(0.001)	(0.011)	(0.011)		
Constant	-0.289***	-3.871***	-3.872***	1.198**	1.201**	4.262	2.428
	(0.007)	(0.122)	(0.122)	(0.561)	(0.563)	(5.735)	(6.327)
Category	No	Yes	Yes	Yes	Yes	No	No
Controls	No	Yes	Yes	Yes	Yes	No	No
Month	No	No	No	Yes	Yes	No	No
Observations	177193	177193	177193	33095	33095	192	162
Num. of Groups				6619	6619		
Mean of dep. Var.	0.10	0.10	0.10	1.63	1.63	-0.24	-3.22
SD of dep. Var.	1.55	1.55	1.55	2.23	2.23	31.81	29.27
Adjusted R^2	0.047	0.294	0.294	0.033	0.033	0.013	0.025

NOTES: The table shows the relationship between the presence of privacy-sensitive permissions and app demand on three different data sets: Columns 1-5 exploit our cross-section and panel data; Columns 6-7 show a difference-in-differences style setup between Google's Play Store and the iOS App Store. In Columns 1-5 the dependent variable is demand proxied by the log. number of monthly new ratings of an app. Columns 1-3 contain cross-section results. Column 1 shows the raw correlation between the use of privacy-sensitive permissions and demand. Column 2 adds in controls for the app's observed characteristics. Column 3 uses the number of privacy-sensitive permissions as privacy indicator. Columns 4-5 show panel fixed effects regressions for those apps within our data set that varied their use of privacy-sensitive permissions at least once between April and September 2012. We show the results for the presence of privacy-sensitive permissions (Column 4) and their number (Column 5). In Columns 6 and 7 the dependent variable is the difference between the app's download ranks on the iOS App Store vs. Google Play Store. In Column 6 this difference is based on the average download ranks of seven countries, whereas in Column 7 it is based only on US ranks. All specifications control for the number of unproblematic permissions ($\#CleanPerm.$). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Moderating Factors of the Demand Side Relationship

	Top Apps		Privacy Policy		Maturity Level		Med.&Health Apps	
Log. *ΔRatings*	No	Yes	No	Yes	High	Low	Yes	No
$D_{Privacy}$	-0.068***	0.008	-0.070***	0.038	-0.109***	-0.054***	-0.151***	-0.060***
	(0.008)	(0.029)	(0.009)	(0.057)	(0.019)	(0.010)	(0.043)	(0.009)
$\#_{CleanPerm}$	0.029***	0.014***	0.033***	0.054***	0.028***	0.038***	0.029**	0.036***
	(0.002)	(0.005)	(0.003)	(0.010)	(0.008)	(0.003)	(0.014)	(0.003)
$D_{Internet}$	-0.122***	-0.320***	-0.192***	-0.254***	-0.108***	-0.211***	-0.106**	-0.203***
	(0.008)	(0.040)	(0.009)	(0.063)	(0.019)	(0.010)	(0.041)	(0.009)
D_{Ads}	0.188***	0.364***	0.229***	0.225***	0.292***	0.222***	0.180***	0.238***
	(0.008)	(0.033)	(0.009)	(0.056)	(0.025)	(0.010)	(0.042)	(0.009)
Log. Price	-0.048***	0.003	-0.070***	-0.114***	-0.050***	-0.076***	-0.056***	-0.072***
	(0.001)	(0.007)	(0.001)	(0.005)	(0.002)	(0.001)	(0.003)	(0.001)
Constant	-3.470***	-4.694***	-3.753***	-4.497***	-3.467***	-3.980***	-3.811***	-3.865***
	(0.106)	(0.499)	(0.122)	(0.976)	(0.247)	(0.140)	(0.669)	(0.125)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	164157	13036	170658	6535	27843	149350	6218	170975
Mean of dep. Var.	-0.14	3.16	0.06	1.11	-0.18	0.15	-0.08	0.11
SD of dep. Var.	1.27	1.54	1.52	2.02	1.38	1.58	1.38	1.56
Adjusted R^2	0.221	0.293	0.285	0.371	0.305	0.291	0.270	0.295

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data. App demand is measured by the log. number of monthly new ratings of an app. Columns 1 and 2 split the sample into apps which have a high or a low stock of installations (more or less than 75000 installations). Columns 3 and 4 split the sample into apps with and without a privacy policy. Columns 5 and 6 split them into apps which require a high (Column 5) or low (Column 6) maturity of the user (apps are defined as appropriate for low maturity if they classified as being recommended for 'everyone' or for 'low maturity'-users). Columns 7 and 8 split the sample into medical and health-related apps as well as into other apps. All specifications control for the number of unproblematic permissions ($\#CleanPerm.$). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

ONLINE APPENDIX for: When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications

Contents

A Online Appendix A **A1**

A.1 Additional Tables and Graphs A1

A.2 Demand Measures A13

A.3 Context-Dependent Privacy-Sensitive Permissions A14

B Online Appendix B **B1**

B.1 Additional Descriptive Statistics B1

B.2 Additional Robustness Checks B2

B.2.1 Privacy and Demand: The Role of Apps' Previous Success B2

B.2.2 User Assessment and Survival B2

B.2.3 Moderating Factors (Alternative Specifications) B4

B.3 Main Estimation Tables - Detailed Version B8

B.4 Robustness Checks - Detailed Version B11

B.4.1 Supply-side Analysis: Privacy Definition and Including Moderating Factors . B11

B.4.2 Panel estimations are robust to using alternative samples and specifications: . B12

B.4.3 Alternative Difference-in-Differences-Style Estimation Results: B15

B.4.4 Robustness of Demand-side Results to Using Alternative Demand Measures . B17

B.4.5 Robustness of Demand-side results to Using Alternative Privacy Measures . . B20

B.4.6 Robustness of Demand-side Results to using Alternative Estimation TechniquesB23

B.4.7 Robustness of Results to Splitting the Sample B24

B.4.8 Excluding the Most and Least Successfull Apps from Estimation B30

B.5 AppAnnie.com B30

A Online Appendix A - Supporting Materials

A.1 Additional Tables and Graphs

Figure A1: App Information in the Android Market 2012

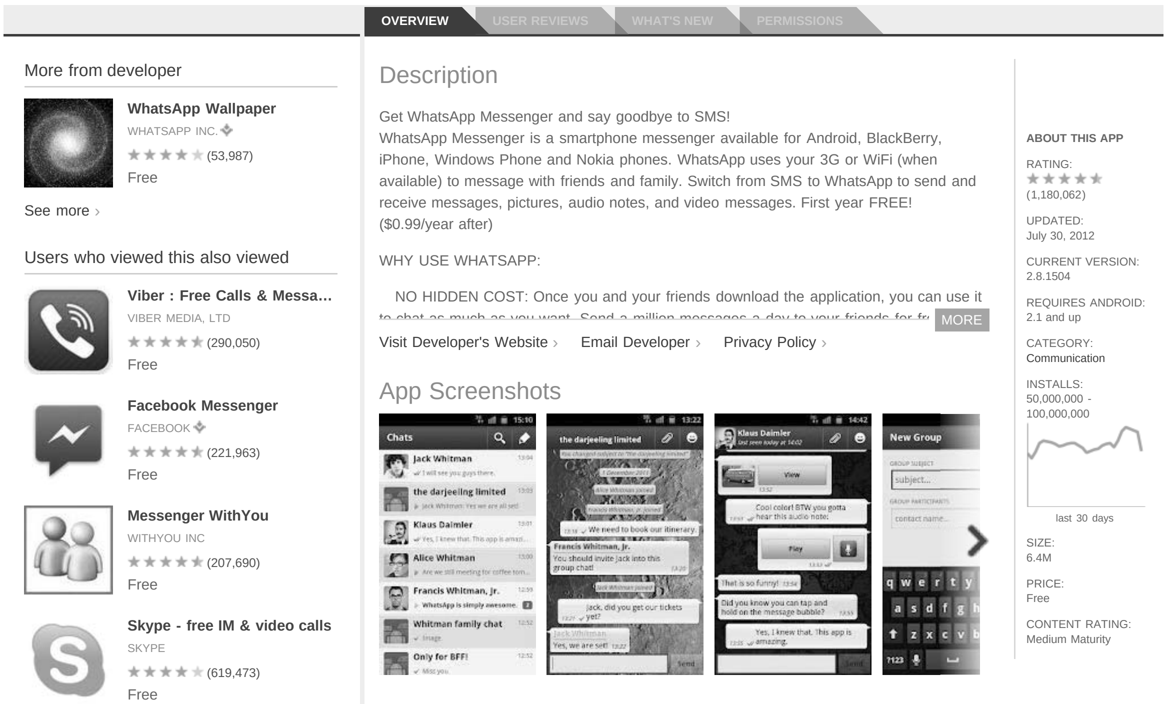


Figure A2: Permission Information in the Android Market 2012

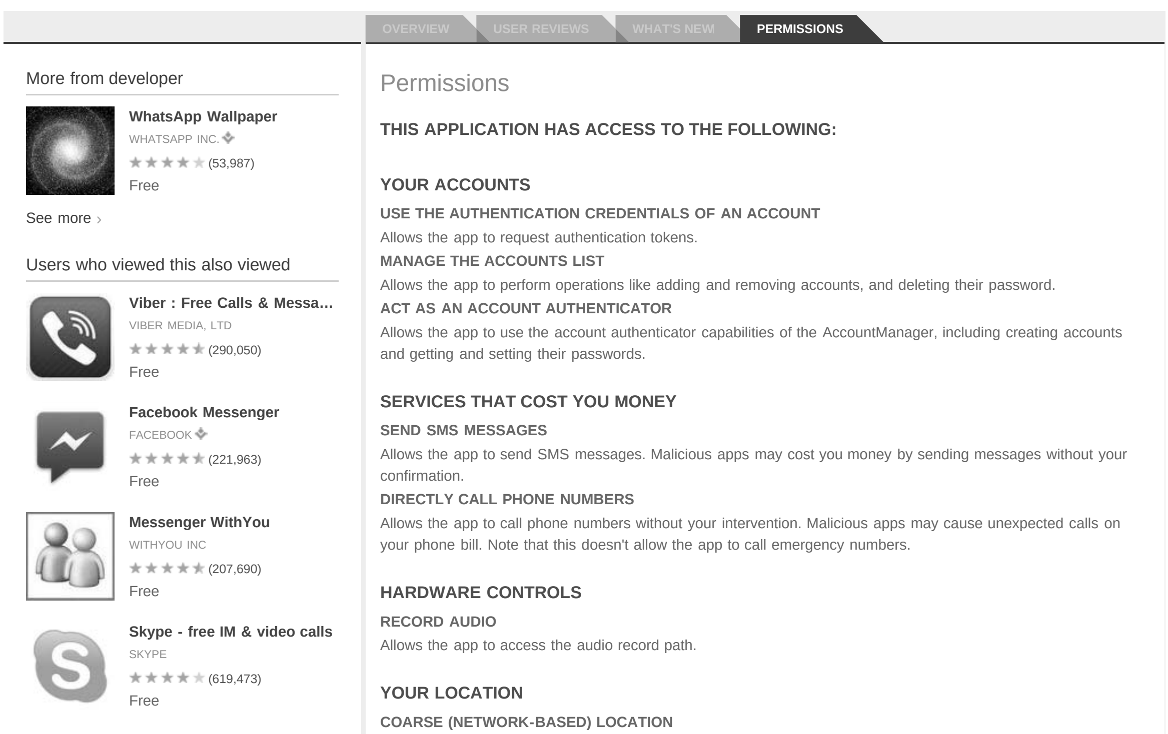


Table A1: Permission Definitions

Permissions (Group)	Description (provided by Google)
$D_{Privacy}$	
D_{ID}	
READ_PHONE_STATE	Allows an app to the read phone status and identity.
$D_{Location}$	
ACCESS_COARSE_LOCATION	Allows an app to access approximate location derived from network location sources such as cell towers and Wi-Fi.
ACCESS_FINE_LOCATION	Allows an app to access precise location from location sources such as GPS, cell towers, and Wi-Fi.
ACCESS_LOCATION_EXTRA_COMMANDS	Allows an apps to access extra location provider commands.
NFC	Allows apps to control Near Field Communication.
$D_{Communication}$	
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	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_PUSH	Allows an app to monitor incoming WAP push messages.
$D_{Profile}$	
READ_CONTACTS	Allows an app to read the user's contacts data.
READ_HISTORY_BOOKMARKS	Allows an app to read (but not write) the user's browsing history and bookmarks.
READ_LOGS	Allows an app to read the low-level system log files.
GOOGLE_AUTH	Allows apps to see the usernames (email addresses) of the Google account(s) you have configured.
ACCOUNT_MANAGER	Allows an app to act as an AccountAuthenticator for the AccountManager.
MANAGE_ACCOUNTS	Allows an app to manage the list of accounts in the AccountManager.
GET_ACCOUNTS	Allows access to the list of accounts in the Accounts Service.
USE_CREDENTIALS	Allows an app to request auth tokens from the AccountManager.
READ_SYNC_STATS	Allows applications to read the sync stats.
SUBSCRIBED_FEEDS_READ	Allows an app to allow access the subscribed feeds ContentProvider.
CAMERA	Allows an app to take pictures and videos.
ACCESS_DOWNLOAD_MANAGER	Allows an app to access the download manager and to use it to download files.
READ_INPUT_STATE	Allows an app to record what you type and actions that you take.
MOUNT_UNMOUNT_FILESYSTEMS	Allows mounting and unmounting file systems for removable storage.
$D_{Internet}$	
INTERNET	Allows apps to open network sockets.
D_{Ads}	
ACCESS_NETWORK_STATE	Allows apps to access information about networks.

Source: `http://developer.android.com/reference/android/Manifest.permission.html` and `https://android.izzysoft.de/applists/perms?lang=en`.

Table A2: List of Variables

Variable	Description
ΔRatings	Monthly change in the number of ratings
ΔInstallations	Monthly change in the number of installations
Ratings	Number of ratings
Installations	Number of installations
ΔR_{AprSep}	Change in the number of ratings between April and September 2012
ΔR_{1214}	Change in the number of ratings between 2012 and 2014
ΔI_{Apr}	Change in the number of installations in April 2012
ΔI_{AprSep}	Change in the number of installations between April and September 2012
ΔI_{1214}	Change in the number of installations between 2012 and 2014
$\Delta Rank^{iOS-And}$	Aggregated difference of app i's iOS and Android rankings ($iOSAppStoreRank_i - AndroidRank_i$)
Price	Price of apps (in Euro)
D_{Paid}	Dummy equal to one if price > 0
$\#_{TotalPerm}$	Number of total permissions
$D_{Privacy}$	Dummy equal to one if an app uses at least one of the 25 privacy-sensitive permissions from the groups D_{ID}, $D_{Location}$, $D_{Communication}$ or $D_{Profile}$ as defined by Table A1
$\#Privacy$	Number of privacy-sensitive permissions
$\#CleanPerm$	Number of unproblematic permissions (i.e. $\#_{TotalPerm} - \#Privacy$).
$D_{Sarmaetal}$	Dummy equal to one if the app uses at least one of the permissions classified as privacy-sensitive by Sarma et. al (2012), i.e. if it uses at least one of ACCESS_COARSE_LOCATION, ACCESS_FINE_LOCATION, INTERCEPT_OUTGOING_CALLS, READ_CONTACTS, READ_HISTORY_BOOKMARKS, READ_SMS, RECEIVE_SMS, RECEIVE_MMS, RECORD_AUDIO, RECEIVE_WAP_PUSH, READ_LOGS or READ_PHONE_STATE
D_{Google}	Dummy equal to one if the app uses at least one 'potentially malicious' permission (as classified and indicated by Google), i.e. if it uses at least one of the following permissions: sendsmsmessages, accessdownloadmanager, interceptoutgoingcalls, addormodifycalendareventsandsend, sendsmsreceivedbroadcast, accessextralocationprovidercomma, deleteapplications, sendwappushreceivedbroadcast, enableapplicationdebugging, retrievesysteminternalstate, readsmsormms, directlycallanyphonenumbers, permissiontoinstallalocationprov, writebrowser39shistoryandbookmar, finegpslocation, receivemms, mocklocationsourcesfortesting, receivewap, editsmsormms, monitorandcontrolallapplicationl, receivedatafrominternet, directlyinstallapplications, presskeysandcontrolbuttons, sendstickybroadcast, readcontactdata, displaysystemlevelalerts, modifyglobalsystemsettings, retrieverunningapplications, readcalendarevents, enableordisableapplicationcompon, setpreferredapplications, reorderrunningapplications, receivesms, directlycallphonenumbers, writecontactdata, writesubscribedfeeds
$D_{PrivCatSpec}$	Dummy equal to one if an app uses at least one category-specific privacy-sensitive permission. We flag a privacy-sensitive permission as problematic within a app category if paid apps of this category use this permission on average less often than the overall average paid app (for more details see section A.3)

Continued on next page

Table A2 – continued from previous page

Variable	Description
$D_{MTurkEP2}$	Dummy equal to one if the app uses at least one of the privacy-sensitive permissions classified as extremely problematic by Amazon MTurk survey participants, i.e. if it uses at least one of READ_SMS, RECORD_AUDIO, INTERCEPT_OUTGOING_CALLS, READ_LOGS
D_{PGrade}	Dummy equal to one if the app got a bad rating equal to 'B', 'C' or 'D' from Lin et al. (2014.) (published on privacygrade.org in 2014), i.e. if it got a rating indicating the app is privacy-intrusive
Average rating	Average of the ratings the app has received so far (between 1 and 5 stars)
AR_{Apr}	Average of the ratings the app has received in April 2012
AR_{AprSep}	Average of the ratings the app has received between April and September 2012
Size	Code size of the app (in KB)
App description	Length of the app description (in number of characters)
Number of screenshots	Number of screenshots available in the app description
Video	Dummy equal to one if a video is available in the app description
Top developer (D_{TopDev})	Dummy equal to one if the app is provided by a top developer (measured by a badge, awarded by Google)
Apps by developer ($AppByDev$)	Number of available apps of the developer
Average installations of developer	Average number of installations of the developer's other apps
$ShareDev_{Privacy}$	Share of other apps by developer which use at least on privacy-sensitive permission
Average price of competitors	Average price of competitor apps
Average installations of competitors	Average installations of competitor apps
Average rating of competitors	Average rating of competitor apps
$ShareComp_{Privacy}$	Share of competitor apps which use at least one privacy-sensitive permission
App category	Categorical variable indicating apps' category
Maturity level	Categorical variable indicating the recommended necessary maturity level of app users ('everyone', 'low maturity', 'medium maturity', 'high maturity', 'not rated')
$D_{Maturity}$	Dummy equal to one if the app requires medium or high maturity or is not rated
App version	Version number of app
Minimum Android version	Minimum compatible Android OS version
Maximum Android version	Maximum compatible Android OS version
$D_{NumInst}$	Dummy equal to one if the app has accumulated 10000 or more installations since its market entry
D_{Transp}	Dummy equal to one if the app has published a privacy policy, email and website address that are directly accessible from the Play Store
$D_{AlexaRank}$	Dummy equal to one if the app has a high ranking on Alexa.com, i.e. its website traffic rank is lower than 10000
D_{Game}	Dummy equal to one if the app is a game
$D_{MedHealth}$	Dummy equal to one if the app is a health or medical app

Table A3: Summary Statistics - Cross-Section April 2012

	mean	sd	min	p10	p50	p90	max
Outcome Measures:							
ΔRatings	7.35	29.78	0	0	0	12	403
ΔInstallations	2340	64892	0	0	0	0	22500000
Ratings	127.47	1026.34	1	1	7	142	160404
Installations	32085	2.1e+05	3	30	3000	30000	30000000
Price	0.55	2.41	0	0	0	1	136
Permissions:							
$\#_{TotalPerm}$	3.62	3.77	0	0	3	9	114
$D_{Privacy}$	0.44	0.50	0	0	0	1	1
$\#_{Privacy}$	1.00	1.56	0	0	0	3	23
$\#_{CleanPerm}$	2.61	2.57	0	0	2	6	92
$D_{PrivCatSpec}$	0.19	0.39	0	0	0	1	1
$D_{MTurkEP2}$	0.11	0.31	0	0	0	1	1
D_{Google}	0.30	0.46	0	0	0	1	1
$D_{Sarmaetal}$	0.41	0.49	0	0	0	1	1
D_{ID}	0.27	0.45	0	0	0	1	1
$D_{Location}$	0.24	0.43	0	0	0	1	1
$D_{Communication}$	0.06	0.24	0	0	0	0	1
$D_{Profile}$	0.18	0.39	0	0	0	1	1
$D_{Internet}$	0.71	0.45	0	0	1	1	1
D_{Ads}	0.46	0.50	0	0	0	1	1
Control Variables:							
Average Rating	3.93	0.97	1	3	4	5	5
Size (in KB)	2665	5809	4	76	852	6500	461000
Length Description	803	791	7	172	509	1837	12285
Number Screenshots	3.45	1.89	0	2	3	6	8
Dummy: Video	0.10	0.31	0	0	0	1	1
Dummy: Top-Developer	0.01	0.07	0	0	0	0	1
Apps by Developer	101	388	1	1	6	186	3548
Average Installations of Developer	69998	6.8e+05	0	210	8157	130178	75000000
Observations	177193						

Notes: The table provides summary statistics for our estimation sample based on the cross-section from April 2012. The discrepancy between our sample and the full app population of around 300,000 apps is mainly due to excluding apps which (a) were not available in some of the subsequent monthly waves we use for our panel analysis, (b) apps which lack relevant variables, (c) apps which have a stock of zero installations or ratings as well as (d) outliers with respect to our main demand measure.

Table A4: Alternative Supply Side Results

	Privacy Measures				Non-Games vs Games		Moderating Factors		App Pairs	
D_{Paid}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\#_{Privacy=1}$	-0.010***									
	(0.003)									
$\#_{Privacy=2}$	-0.058***									
	(0.003)									
$\#_{Privacy\geq 3}$	-0.111***									
	(0.004)									
$D_{Privacy}$		0.012*			-0.012***	-0.057***	-0.062***	-0.037***	-0.246***	
		(0.006)			(0.003)	(0.005)	(0.003)	(0.003)	(0.054)	
$D_{Privacy} \times D_{Internet}$		-0.055***								
		(0.007)								
$D_{PrivCatSpec}$			-0.102***							
			(0.003)							
$\#_{PrivCatSpec}$				-0.045***						
				(0.001)						
$D_{Privacy} \times D_{NumInst}$							0.144***			
							(0.003)			
$D_{NumInst}$							-0.322***			
							(0.003)			
$D_{Privacy} \times D_{AlexaRank}$								0.046**		
								(0.018)		
$D_{AlexaRank}$								-0.109***		
								(0.016)		
$\#_{Privacy}$										-0.199***
										(0.029)
$\#_{CleanPerm}$	0.009***	0.005***	0.006***	0.008***	0.003***	0.012***	0.005***	0.004***	0.031	0.051**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.024)	(0.025)
$D_{Internet}$	-0.223***	-0.209***	-0.216***	-0.220***	-0.206***	-0.221***	-0.209***	-0.217***	-0.392***	-0.422***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.003)	(0.003)	(0.047)	(0.047)
D_{Ads}	-0.122***	-0.117***	-0.120***	-0.120***	-0.127***	-0.139***	-0.106***	-0.114***	-0.522***	-0.531***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.051)	(0.048)
Constant	0.081**	0.085**	0.080**	0.087**	-0.067*	0.087	0.127***	0.133***	-0.411	-0.167
	(0.035)	(0.035)	(0.035)	(0.035)	(0.040)	(0.088)	(0.034)	(0.039)	(0.687)	(0.667)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176000	176000	176000	176000	145972	30028	176000	145126	708	708
Num. of Groups									354	354
Mean of dep. Var.	0.27	0.27	0.27	0.27	0.28	0.23	0.27	0.27	0.50	0.50
SD of dep. Var.	0.44	0.44	0.44	0.44	0.45	0.42	0.44	0.44	0.50	0.50
Adjusted R^2	0.323	0.320	0.324	0.324	0.348	0.260	0.367	0.319	0.848	0.862

NOTES: The table shows additional supply-side results for the developer's choice of their business model. The dependent variable is the developer's decision to offer their app for money or for free (D_{Paid}). Column 1 uses as privacy measure an individual dummy for each number of privacy-sensitive permissions (1, 2 or 3 and more permissions). Column 2 uses a cross term equal to one for apps which simultaneously use sensitive permissions and have access to the internet. Columns 3 and 4 use our category-specfic privacy measures. Within a category we flag a privacy-sensitive permission as problematic only if paid apps of this category use this permission on average less often than the overall average paid app. Columns 5 and 6 split the sample into normal apps (Col. 5) and games (Col. 6). Column 7 adds a cross term for apps with a very large total number of installations (10000 or more). Column 8 adds a cross term for apps that could be associated with a top-ranked website (on Alexa.com), i.e. for a website with rank lower than 10000. Columns 9 and 10 analyze the most restrictive set of matched pairs (no difference in description or only difference describes existence of ads, verified by human coders). In all specifications we drop outliers with respect to app prices, i.e. apps with prices above 8 Euros. All of these regressions control for the number of clean permissions and permissions that are needed to show ads. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Alternative Panel Demand Side Estimation Results

Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-0.070***	-0.059***	-0.063**			
	(0.020)	(0.020)	(0.027)			
$\#_{Privacy}$				-0.023***	-0.014**	-0.010
				(0.006)	(0.007)	(0.009)
$\#_{CleanPerm}$	-0.010**	-0.010**	-0.003	-0.000	-0.005	-0.001
	(0.004)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
$D_{Internet}$	-0.003	0.003	0.033	-0.021	-0.014	0.012
	(0.054)	(0.054)	(0.072)	(0.054)	(0.053)	(0.071)
D_{Ads}	0.007	0.002	-0.041	-0.011	-0.014	-0.058
	(0.033)	(0.033)	(0.043)	(0.032)	(0.033)	(0.043)
Log. Price	-0.044***	-0.044***	-0.037***	-0.044***	-0.043***	-0.037***
	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.013)
Constant	1.158***	1.198**	0.749***	1.156***	1.201**	0.741***
	(0.127)	(0.561)	(0.154)	(0.128)	(0.563)	(0.155)
Category	No	Yes	No	No	Yes	No
Month	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No	Yes	No
Observations	33095	33095	15220	33095	33095	15220
Num. of Groups	6619	6619	3044	6619	6619	3044
Mean of dep. Var.	1.63	1.63	1.17	1.63	1.63	1.17
SD of dep. Var.	2.23	2.23	2.10	2.23	2.23	2.10
Adjusted R^2	0.026	0.033	0.020	0.026	0.033	0.020

NOTES: This table shows the results from fixed-effect panel regressions. The dependent variable is the log. number of monthly new ratings of an app. We restrict our sample to apps within our data set that varied their use of privacy-sensitive permissions at least once between April and September 2012. Columns 1-3 analyze the effect of introducing any privacy-sensitive permissions (measured by the indicator $D_{Privacy}$), whereas Columns 4-6 use the number of privacy-sensitive permissions as the variable of interest. Columns 1 and 4 show the raw fixed effects regressions without controls. In Columns 2 and 5 we add control variables and dummies to control for the apps' categories. Columns 3 and 6 restrict the analysis to apps that introduced new permissions without changing the app's description (no change in the length of description). In all specifications we include monthly fixed-effects and control for the number of unproblematic permissions. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A6: Alternative Difference-in-Differences-Style Estimation Results:

	Global Ranks		US Ranks		Non-Games vs Games	
	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-116.468***	-12.660**	-182.504***	-14.988**	-25.368**	1.533
	(35.327)	(6.088)	(58.130)	(6.321)	(10.316)	(8.488)
$\#_{CleanPerm}$	0.266	0.849	0.344	1.011*	1.498**	-1.310
	(3.473)	(0.598)	(5.476)	(0.595)	(0.694)	(1.263)
Constant	264.070***	4.262	320.040***	2.428	10.214	6.258
	(29.120)	(5.018)	(48.089)	(5.229)	(9.514)	(6.209)
Observations	192	192	162	162	96	96
Mean of dep. Var.	175.94	-0.24	176.85	-3.22	-1.07	0.59
SD of dep. Var.	188.76	31.81	274.05	29.27	32.96	30.78
Adjusted R^2	0.056	0.013	0.059	0.025	0.062	-0.006

Notes: This table shows further demand side results from comparing the download ranks of an app in the iOS Appstore and Google's Play Store depending on whether it's Android version uses privacy-sensitive permissions or not. The dependent variable captures differences in the ranks on the two platforms (iOS App Store ranks minus Google Play Store ranks). Colums 1, 2, 5 and 6 compare the apps' average ranks of the seven countries we collected ranks for, whereas Columns 3 & 4 use only the US ranks. Columns 1 & 3 use as dependent variable the difference between the simple rank averages whereas Columns 2 & 4 use as the dependent variable the difference between the newly created ranks within the operating system, which are based on the average download ranks we observed on AppAnnie.com. Columns 5 & 6 also use the difference between the newly created ranks based on the average download ranks but contrast games (Column 6) with other apps (Column 5). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Alternative Moderating Factors

	Top-Developer		Alexa Rank		Games		Pricing Model	
Log. *ΔRatings*	No	Yes	No	Yes	No	Yes	Free	Paid
$D_{Privacy}$	-0.065***	0.088	-0.055***	0.010	-0.082***	-0.020	-0.026**	-0.121***
	(0.009)	(0.151)	(0.010)	(0.172)	(0.010)	(0.022)	(0.011)	(0.014)
$\#_{CleanPerm}$	0.036***	0.097***	0.035***	0.051*	0.037***	0.098***	0.037***	0.054***
	(0.003)	(0.031)	(0.003)	(0.028)	(0.003)	(0.009)	(0.003)	(0.006)
$D_{Internet}$	-0.195***	0.023	-0.187***	0.152	-0.211***	-0.201***	-0.256***	-0.075***
	(0.009)	(0.231)	(0.010)	(0.264)	(0.010)	(0.024)	(0.012)	(0.013)
D_{Ads}	0.232***	0.360**	0.226***	-0.251	0.228***	0.157***	0.241***	0.039**
	(0.009)	(0.164)	(0.010)	(0.173)	(0.010)	(0.024)	(0.010)	(0.017)
Log. Price	-0.071***	-0.108***	-0.071***	-0.107***	-0.068***	-0.095***	0.000	0.099***
	(0.001)	(0.011)	(0.001)	(0.024)	(0.001)	(0.002)	(.)	(0.007)
Constant	-3.852***	-3.587*	-3.832***	-6.650***	-3.398***	-5.120***	-3.557***	-3.416***
	(0.123)	(2.124)	(0.136)	(2.355)	(0.140)	(0.306)	(0.152)	(0.172)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176224	969	145367	771	147143	30050	128921	48272
Mean of dep. Var.	0.09	2.15	0.11	2.19	0.07	0.26	0.33	-0.50
SD of dep. Var.	1.54	2.25	1.55	2.09	1.52	1.70	1.65	1.06
Adjusted R^2	0.286	0.538	0.298	0.348	0.280	0.370	0.285	0.262

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data. App demand is measured by the log. number of monthly new ratings of an app. Columns 1 and 2 split the sample into apps which are from a top developer (according to Google's classification) or not. Columns 3 and 4 split the sample into normal apps and games. Columns 5 and 6 split them into normal apps and games. Columns 7 and 8 split the sample into free and paid apps. All specifications control for the number of unproblematic permissions (*#CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Alternative Demand Measures

	Ratings-Based		Installations-Based			Predicted New Installations			Average Rating	
	ΔR_{AprSep}	ΔR_{1214}	ΔI_{Apr}	ΔI_{AprSep}	ΔI_{1214}	param.	param. 1214	non-param.	AR_{Apr}	AR_{AprSep}
$D_{Privacy}$	-0.093***	-0.070***	-0.054***	-0.197***	-0.388***	-0.030***	-0.053***	-0.183***	-0.022***	-0.018***
	(0.011)	(0.015)	(0.016)	(0.025)	(0.036)	(0.004)	(0.007)	(0.024)	(0.003)	(0.003)
$\#_{CleanPerm}$	0.042***	0.035***	0.010***	0.014**	-0.010	0.017***	0.030***	0.069***	0.000	0.000
	(0.003)	(0.004)	(0.004)	(0.006)	(0.008)	(0.001)	(0.002)	(0.006)	(0.001)	(0.001)
$D_{Internet}$	-0.220***	-0.250***	-0.078***	-0.206***	-0.317***	-0.092***	-0.163***	-0.416***	-0.054***	-0.056***
	(0.012)	(0.016)	(0.015)	(0.025)	(0.037)	(0.004)	(0.007)	(0.026)	(0.004)	(0.004)
D_{Ads}	0.326***	0.356***	0.248***	0.618***	0.925***	0.109***	0.192***	0.613***	0.009**	0.014***
	(0.011)	(0.016)	(0.015)	(0.024)	(0.037)	(0.004)	(0.007)	(0.025)	(0.004)	(0.003)
Log. Price	-0.114***	-0.173***	-0.038***	-0.113***	-0.307***	-0.033***	-0.058***	-0.201***	-0.002***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
Constant	-4.583***	-5.540***	-3.557***	-6.936***	-13.206***	-2.248***	-3.083***	-6.608***	0.354***	0.360***
	(0.156)	(0.229)	(0.196)	(0.324)	(0.505)	(0.056)	(0.100)	(0.345)	(0.055)	(0.046)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176212	124482	177193	177193	126987	177193	177193	177193	68248	102594
Mean of dep. Var.	0.97	2.40	-0.42	0.86	4.63	-0.42	0.16	2.76	1.29	1.29
SD of dep. Var.	2.03	2.52	2.34	3.89	5.16	0.72	1.27	4.25	0.35	0.35
Adjusted R^2	0.360	0.388	0.037	0.086	0.201	0.294	0.294	0.272	0.043	0.052

NOTES: The table shows descriptive regressions analyzing the relationship between privacy-sensitive permissions and app demand by using various different demand and app popularity measures. Columns 1 and 2 use alternative demand measures based on ratings, Columns 3-5 use demand measures based on installations, whereas Columns 6-8 use three measures of predicted new installations which we estimated based on the information about the number of new ratings. Columns 9 and 10 use the average ratings of apps as a measure of app popularity. In Column 1 (2) the dependent variable is the log. number of new ratings between April and September 2012 (log. number of new ratings between 2012 and 2014). In Column 3 the dependent variable is the log. number of new installations in April 2012. Columns 4 and 5 use the log. number of new installations between April and September 2012 (Column 4) and between 2012 and 2014 (Column 5) as demand measures. In Columns 6-8 we apply three measures of predicted download numbers. For each of the measures we exploit the cross-section information on changes in ratings to predict changes in installation numbers. In Column 6, we use a measure of predicted monthly installation changes in April 2012 which is based on the observed change in the number of ratings in this month (see Column 2 of Table A12). In Column 7, we use a measure of predicted installation changes between April and September 2012 which is based on the observed change in the number of ratings in this period (see Column 4 of Table A12). In Column 8, we again use a measure of predicted monthly installation changes in April 2012 which is based on the observed change in the number of ratings in this month (see Column 2 of Table A12), but instead of employing a parametric log-log-specification to the data, we employ a non-parametric approach to it. In Columns 9 the dependent variable is the log. average of the ratings the app has received in April 2012, whereas in Column 10 it is the log. average of the ratings the app has received between April and September 2012. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A9: Alternative Privacy Measures

Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#_{Privacy=1}$	-0.054***							
	(0.010)							
$\#_{Privacy=2}$	-0.087***							
	(0.014)							
$\#_{Privacy\geq 3}$	-0.089***							
	(0.014)							
$D_{Privacy}$		-0.060***	-0.048***	-0.040***		-0.086***		
		(0.009)	(0.010)	(0.014)		(0.018)		
$D_{MTurkEP2}$		-0.021*						
		(0.013)						
$D_{PrivCatSpec}$			-0.048***					
			(0.012)					
D_{PGrade}				0.148***				
				(0.020)				
$D_{Privacy} \times D_{Internet}$					-0.057***	0.025		
					(0.010)	(0.019)		
D_{ID}							0.046***	
							(0.009)	
$D_{Location}$							-0.200***	
							(0.014)	
$D_{Communication}$							-0.015	
							(0.016)	
$D_{Profile}$							-0.034***	
							(0.011)	
$D_{Sarmaetal}$								-0.048***
								(0.009)
$\#_{CleanPerm}$	0.038***	0.037***	0.037***	0.048***	0.036***	0.037***	0.038***	0.036***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
$D_{Internet}$	-0.202***	-0.201***	-0.199***	-0.288***	-0.192***	-0.204***	-0.201***	-0.202***
	(0.009)	(0.009)	(0.009)	(0.015)	(0.009)	(0.010)	(0.009)	(0.009)
D_{Ads}	0.234***	0.234***	0.235***	0.260***	0.236***	0.235***	0.219***	0.235***
	(0.009)	(0.009)	(0.009)	(0.014)	(0.009)	(0.009)	(0.009)	(0.009)
Log. Price	-0.071***	-0.071***	-0.071***	-0.094***	-0.071***	-0.071***	-0.072***	-0.071***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-3.878***	-3.872***	-3.880***	-3.785***	-3.878***	-3.867***	-3.914***	-3.870***
	(0.122)	(0.122)	(0.123)	(0.322)	(0.123)	(0.123)	(0.122)	(0.122)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177193	177193	177193	86834	177193	177193	177193	177193
Mean of dep. Var.	0.10	0.10	0.10	0.51	0.10	0.10	0.10	0.10
SD of dep. Var.	1.55	1.55	1.55	1.71	1.55	1.55	1.55	1.55
Adjusted R^2	0.294	0.294	0.294	0.288	0.294	0.294	0.295	0.294

NOTES: The table shows descriptive regressions analyzing the relationship between the presence of privacy sensitive permissions and app demand. The dependent variable is demand for the app measured by log. number of monthly new ratings of an app. The coefficient of interest analyzes the relationship between an app's demand and our privacy measures. Each column presents the results obtained when using an alternative privacy measure. Column 1 introduces an indicator for each number of permission (having 1, 2, and 3 or more permissions). Column 2 uses a dummy variable which is equal to one if an app uses at least one privacy-sensitive permission which was classified as very problematic by 450 microworkers we surveyed on Amazon's mechanical turk. In Column 3 we look at privacy sensitive permissions that are unusual for the app's category. Within a category we flag a privacy-sensitive permission as problematic only if paid apps of this category use this permission on average less often than the overall average paid app. Column 4 uses the 'privacygrade' by Lin, Hong, and Sadeh. (2014), that was made available on *privacygrade.org* in 2014. The dummy is equal to one if the the app got a rating equal to 'B', 'C' or 'D', i.e. a rating indicating the app being privacy-intrusive. In Columns 5 & 6 we introduce a crossterm that is equal to one if an app uses both at least one privacy-sensitive permission and has internet access. Column 7 disaggregates the privacy sensitive permissions into functionality-related types of permissions. Column 8 uses an alternative definition of privacy-sensitive permissions from previous research by Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy (2012), which defines only 12 permissions as privacy-sensitive. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A10: Alternative Estimation Specifications

	Tobit		Heckman		Netw.Eff.	IV-Privacy		IV-Price	
Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$D_{Privacy}$	-0.094***	-0.095***	-0.063***	-0.164***	-0.069***	-0.918***	-0.372***	-0.091***	-0.063***
	(0.017)	(0.017)	(0.009)	(0.024)	(0.007)	(0.049)	(0.017)	(0.009)	(0.017)
$\#_{CleanPerm}$	0.055***	0.057***	0.038***	0.019***	0.026***	0.065***	0.044***	0.043***	0.082***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.002)	(0.004)	(0.003)	(0.003)	(0.007)
Log. Installations					0.417***				
					(0.002)				
$D_{Internet}$	-0.325***	-0.333***	-0.184***	-0.423***	-0.168***	-0.099***	-0.177***	-0.289***	-0.005
	(0.020)	(0.021)	(0.010)	(0.030)	(0.007)	(0.012)	(0.011)	(0.014)	(0.017)
D_{Ads}	0.436***	0.443***	0.234***	0.291***	0.203***	0.243***	0.253***	0.198***	0.043**
	(0.018)	(0.018)	(0.010)	(0.023)	(0.007)	(0.011)	(0.011)	(0.011)	(0.020)
Log. Price	-0.172***	-0.174***	-0.068***	-0.157***	0.040***	-0.074***	-0.074***	-0.108***	-0.687***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.094)
Constant	-8.802***	-8.975***	-3.533***	-5.781***	-6.472***	-4.141***	-4.305***	-3.970***	-3.300***
	(0.266)	(0.271)	(0.127)	(0.319)	(0.100)	(0.136)	(0.141)	(0.128)	(0.197)
Category	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177193	177193	185533	177193	177193	136040	137146	177193	48272
Mean of dep. Var.	0.10	0.10	0.10	2.40	0.10	0.10	0.10	0.10	-0.50
SD of dep. Var.	1.55	1.55	1.55	2.52	1.55	1.55	1.56	1.55	1.06
Adjusted R^2					0.534	0.256	0.324	0.269	0.047

NOTES: The table analyzes the robustness of our main demand-side results to using alternative estimation strategies. The dependent variable is app demand measured by the log. number of monthly new ratings, and the main variable of interest is a dummy that indicates the presence of privacy-sensitive permissions. All columns show cross section results. Columns 1 & 2 show Tobit-regressions that account for the fact that the dependent variable might be censored, especially might be left-censored at demand equal to 0. Column 1 sets the left-censoring limit to 0 new ratings, whereas in Column 2 in addition a right-censoring limit equal to 5 is set. Columns 3 & 4 contain results from Heckman selection models, where the regression equation is identical to our baseline cross-section demand specification, i.e. the dependent variable is the log. number of monthly or biannual new ratings, and the selection equation models app survival. In column 3 survival is modeled by comparing apps which are observed throughout the period April to September 2012 to those which are observed in April 2012 but which cannot be observed in later months. In column 4 survival is modeled by comparing apps within our baseline cross-section from April 2012 which survive until 2014 to those which are observed in April 2012 but are not observed in 2014. In both selection models we apply Heckman's two-step consistent estimator and use the information on code size as the selection variable, i.e. we include the code size only in the selection equation but not in the regression equation. In Column 5 we control for the existing user-base by including a control for the stock of existing installations (log. number of installations). In Columns 6 and 7 we estimate a 2SLS model and instrument the variable of interest to account for the endogeneity of the developers' privacy model choice. In Column 6 we instrument the privacy-dummy by the share of competing apps which use privacy-sensitve permissions ($ShareComp_{Privacy}$). In Column 7 we instrument the privacy-dummy by the share of the apps of the developer which use privacy-sensitive permissions ($ShareDev_{Privacy}$). In Columns 8 and 9 we instrument the app price by using in both specifications two potential cost shifters: the log. code size and the log. number of apps a developer offers in the Google Play Store. In Column 8 we use the full cross-section, whereas in Column 9 we use only the sample of paid apps. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A11: Excluding the Most and Least Successful Apps

	W/o Top-Apps		W/o Flop-Apps		W/o Top- and Flop-Apps	
Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-0.064***	-0.079***	-0.011	-0.086***	-0.030***	-0.050***
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.010)
$D_{NumInst}$		1.545***		1.303***		0.946***
		(0.011)		(0.012)		(0.011)
$D_{Privacy} \times D_{NumInst}$		0.036**		0.126***		0.009
		(0.015)		(0.017)		(0.014)
$\#_{CleanPerm}$	0.019***	0.019***	0.036***	0.032***	0.017***	0.018***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
$D_{Internet}$	-0.114***	-0.093***	-0.197***	-0.172***	-0.092***	-0.090***
	(0.008)	(0.007)	(0.014)	(0.012)	(0.011)	(0.010)
D_{Ads}	0.170***	0.154***	0.182***	0.188***	0.102***	0.116***
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.009)
Log. Price	-0.055***	-0.026***	-0.071***	-0.023***	-0.046***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-2.777***	-2.924***	-3.228***	-3.462***	-1.535***	-1.906***
	(0.103)	(0.090)	(0.192)	(0.167)	(0.151)	(0.135)
Category	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	168767	168767	79847	79847	71421	71421
Mean of dep. Var.	-0.12	-0.12	1.44	1.44	1.09	1.09
SD of dep. Var.	1.23	1.23	1.44	1.44	1.04	1.04
Adjusted R^2	0.231	0.416	0.240	0.409	0.149	0.303

NOTES: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data where we exclude the tails of the distribution with respect to our dependent variable, i.e. the most and the least successful apps. App demand is measured by the log. number of monthly new ratings of an app. In Columns 1 and 2 we exclude the most succesful apps, i.e. the upper 5 percentiles with respect to the number of new ratings in April 2012. In Columns 3 and 4 we exclude the least successful apps, i.e. those without any new ratings in April 2012. In Columns 5 and 6 we exclude both groups, i.e. the upper 5 percent of most successful apps and those having no new rating in April 2012. In Columns 2, 4 and 6 we add a dummy which is equal to one if the app in the past had accumulated a stock of at least 10000 or more installations and also add an interaction of this dummy with the privacy-dummy. All specifications control for the number of unproblematic permissions ($\#CleanPerm.$). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

A.2 Demand Measures

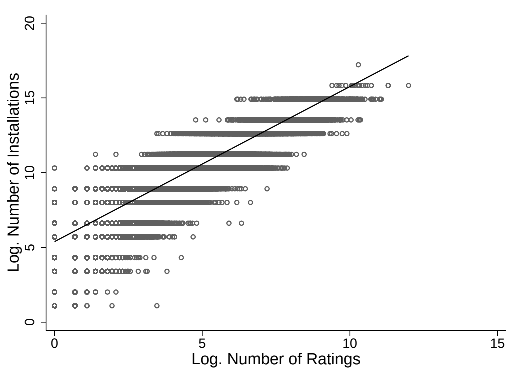
In this section we illustrate the relationship between apps' installations and ratings as well as the relationship between apps' number of ratings and apps' success measured by app ranks.

Relationship between Installations and Ratings: Our main measure of app demand is the log. number of new ratings an app has received within a month. We use it as a demand proxy since the optimal measure of app demand, the number of new installations, is a very noisy measure due to the fact that the total number of installations is published in a discrete form (17 steps). More explicitly, changes in the number of installations for most of the apps are observed only very rarely, i.e. in most cases the monthly number of new installations is equal to zero, and in cases where we observe a change in the number of installations this observed change most likely overstates the true change in the monthly number of installations. However, since we have exact and continuous information about the number of ratings of an app, we can use this as an alternative demand measure. We can do so, since, as we show in the following, (changes in) the number of installations and (changes in) the number of ratings are on average extremely closely related to each other.

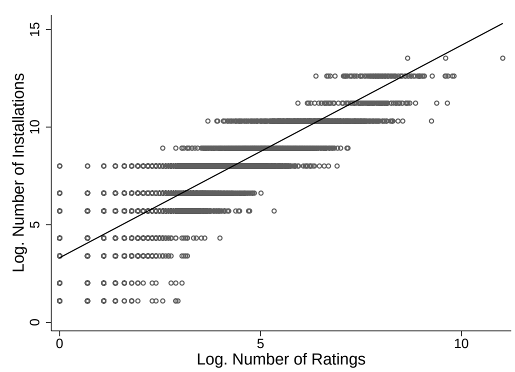
To illustrate the relationship between both measures visually, Figure A3 shows for the April 2012 cross-section of free and paid apps the relationship between the log. number of installations and the log. number of ratings. As can be seen, for both types of apps a strong positive relationship between both measures exists.

Figure A3: Relationship between Level of Installations and Ratings for Free and Paid Apps

(a) Free Apps ($N = 128921$)



(b) Paid Apps ($N = 48272$)



Notes: Both figures display the relationship between apps' log. number of installations and their log. number of ratings. The left plot shows the relationship for free apps, whereas the right one contains that of paid apps. In both cases a linear trend is added. The data used is equal to the cross-section data set from April 2012 used in the estimation sections.

In addition, Table A12 shows more descriptive, econometric, evidence on the relationship between installations and ratings. The first column shows evidence on the relationship between log. installations and log. ratings, where we find an estimation coefficient slightly above one and a R^2 of around 0.7. This indicates that an increase in the number of ratings by one percent comes with an increase in the number of installations by around one percent. Columns 2 and 3 use the cross-section from April 2012 and analyze the relationship based on changes in both variables within one month. Column 2 uses all available observations, whereas Column 3 drops observations with no change in the number of installations in April. As can be seen, even in these cases where our measure of monthly changes in the number of installations is extremely noisy (due to the short time period of one month), we can observe for both versions a strong and positive relationship between installations and ratings. In Column 4 and 5 we also use the cross-section from April 2012 but analyze changes between April and September. Column 4 uses all observations, whereas Column 5 drops those which exhibit no change in installations between April and September. Doing so, the previous finding of a strong and positive relationship is even more pronounced, which is not too surprising since measurement error looses relative weight compared to the version based on monthly changes. Columns 6 and 7 use the sample of observations for which we have information both in 2012 and in 2014 and analyzes changes in installations and ratings within those two years. Column 6 uses all observations, whereas Column 7 uses only those with changes in the number of installations between 2012 and 2014. Going one step further and exploiting changes within two years of observation, we again find, as for the specification in levels, coefficients of around one and R^2 values of 0.4 and 0.8. To estimate the relationship in Column 8, which is based on aggregated data, we apply a two-step approach: (1) First, we aggregate single data points into average ones. For doing so, we split the sample, which is ordered by the number of new ratings, into 100 quartiles and then compute for each quartile the average number of new ratings and new installations.[58] (2) Second, based on the log-values of the average data points we estimate a simple linear log-log model of the relationship between new installations and new ratings, for which the results are given in Column 8. As can be seen, an extremely close relationship between the log. number of new installations and the log. number of new ratings exists, with an R^2 close to one. Visually, this relationship is illustrated by Figure 1.

Relationship between Ratings and App Ranks: In a second step we aim at illustrating the close relationship between our demand measure and a demand measure based on app ranks. The Google Play store, as well as Apple's iOS store, provide app ranks for the most successful apps by category. In table A13 we show the correlation between (a) app ranks which we compute based on changes in apps' number of ratings and (b) their official app ranks in the Google Play Store (taken from App Annie). The correlation is given for apps for which we have an overall rank (Columns 1 to 3) and for apps for which we have a rank within the games category (Columns 4 to 6). Column 1 and 4 contain all apps, whereas Columns 2 and 5 use only apps which are for free, wherease Columns 3 and 6 analyze paid apps. The dependent variable is the official app rank, whereas the independent variable is its rank based on the change in the number of ratings which we observe. In all regression we do not include a constant.

A.3 Context-Dependent Privacy-Sensitive Permissions

In this robustness check, we test whether the main demand-side results are robust to using a context-specific definition of privacy-sensitive permissions. We consider this relevant, since such a definition allows to consider category-specific functionalities. For example, 'running'-apps can provide better service if they are able to geo-locate their users, while other apps (e.g., a mail client) do not need this ability. Thus, the use of certain permissions might allow an app to access private information

[58] Due to the skewed distribution of the data, STATA only generates 34 quartiles.

Table A12: Relationship between Installations and Ratings

	Level	Growth		Growth Apr-Sep		Growth 12-14		Means
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log. Ratings	1.178***							
	(0.002)							
Log. Δ Ratings		0.461***	0.938***					0.955***
		(0.006)	(0.010)					(0.024)
Log. Δ_{AprSep} Ratings				0.815***	0.952***			
				(0.005)	(0.004)			
Log. Δ_{1214} Ratings						1.132***	0.929***	
						(0.005)	(0.002)	
Constant	4.518***	-0.463***	6.752***	0.073***	6.158***	1.966***	5.815***	5.959***
	(0.006)	(0.005)	(0.026)	(0.006)	(0.014)	(0.014)	(0.009)	(0.076)
Observations	177193	177193	10988	176212	34620	125213	74040	34
Adjusted R^2	0.720	0.094	0.427	0.181	0.609	0.318	0.788	0.980

NOTES: The table shows descriptive econometric evidence on the relationship between installations and ratings. The first Column shows the relationship in levels, i.e. the relationship between the log. number of installations and log. the number ratings. Columns 2 to 8 contain estimates of the relationship in changes, i.e. the relationship between the log. number of new installations and the log. number of new ratings. Columns 2 and 3 use the cross-section from April 2012 and analyze the relationship based on changes in both variables within one month. Column 2 uses all available observations, whereas Column 3 drops observations with no change in the number of installations in April. In Column 4 and 5 we also use the cross-section from April 2012 but analyze changes between April and September. Column 4 uses all observations, whereas Column 5 drops those which exhibit no change in installations between April and September. Columns 6 and 7 use the sample of observations for which we have information both in 2012 and in 2014 and analyzes changes in installations and ratings within those two years. Column 6 uses all observations, whereas Column 7 uses only those with changes in the number of installations between 2012 and 2014. For the estimation in Column 8 we aggregate single data points into average ones by splitting the sample, which is ordered by the number of new ratings, into 100 quartiles and then compute for each quartile the average number of new ratings and new installations and then, based on the log-values of the average data points, we estimate a simple linear log-log model of the relationship between new installations and new ratings. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A13: Relationship between Ranks based on Ratings and Ranks from App-Annie

	All Apps			Games		
	Free&Paid	Free	Paid	Free&Paid	Free	Paid
$RANK_{AllApps}$	0.880***	0.855***	0.893***			
	(0.023)	(0.036)	(0.029)			
$RANK_{GAMES}$				0.889***	0.851***	0.909***
				(0.020)	(0.033)	(0.026)
Observations	407	180	227	465	207	258
Adjusted R^2	0.774	0.729	0.796	0.787	0.729	0.818

NOTES: The table shows correlations between (a) app ranks which we compute based on changes in apps' number of ratings and (b) their official app ranks in the Google Play Store (taken from App Annie). The correlation is given for apps for which we have an overall rank (Columns 1 to 3) and for apps for which we have a rank within the games category (Columns 4 to 6). Column 1 and 4 contain all apps, whereas Columns 2 and 5 use only apps which are for free, wherease Columns 3 and 6 analyze paid apps. The dependent variable is the official app rank, whereas the independent variable is its rank based on the change in the number of ratings which we observe. In all regression we do not include a constant. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

might be related to its functionality and might not only exist for collecting information about users. In Table A14 we consider category-specific definitions of privacy-sensitive permissions, and tolerate such permissions that are common and necessary for a certain category of apps.

Data and Variables: To identify category-specific permission requirements, we exploit the Play Store's category tree and analyze which permissions are common in each category and thus might be required for the functionality of typical apps within this category. Based on this information we generate an alternative measure of category-specific privacy-sensitive permissions, which takes the categories usual functionality and permission-needs into account. We generate a dummy variable $D_{PrivcCatSpec}$ which is equal to one if an app uses at least one privacy-sensitive permission which is relatively unusual within the app's category. Based on the cross-section data set from April 2012, we define permissions as relatively uncommon within an app category if paid apps of this category use this permission on average less often than the overall paid app.[59] Thus, a privacy-sensitive permission is classified as problematic within a category if it is relatively uncommon in this category, i.e. if a higher share of apps in other categories use this permission (for an overview about which permissions are classified as category-specific privacy-sensitive in which category, see Table A15).

Results: Table A14 shows the results which we obtained using the category-specific measure of privacy-sensitive permissions. The dependent variable is as in the baseline specifications the monthly number of new reviews, and $D_{PrivcCatSpec}$ indicates the presence of privacy-sensitive permissions that are unusual for a given app category. We show four specifications using the cross-section (cols. 1-4) and three different panel specifications (cols. 5-7). The results show that the negative demand effect of category-specific privacy-sensitive permissions is statistically significant for the cross-section (cols. 1-4) and in two out of the three panel specifications (cols. 5-7). Only in specification 6, where we control for time-varying factors but do not ensure that apps do not change their description over time and thus might have changed their functionality, we find a negative but insignificant effect.

In Columns 1 and 2 we use the full sample and vary the number of controls. Column three shows the results only for free apps and column 4 shows the results exclusively for paid apps. The coefficient is largest for paid apps, suggesting that paying users avoid apps who request sensitive *and unusual* permissions (which cannot be related to a functionality). Among the panel results, Columns 5 & 6 use the large panel, while Column 7 focuses on apps that did not update their description of functionality (but changed their permissions). We see the strongest negative effect when apps update their permission use without changing functionality in a visible way.

[59] We use paid apps for this definition, since we expect them to use to a higher degree only such permissions which are necessary for their functionality, whereas free apps, according to our findings, do also use permissions which are not necessary for the apps' functionality but which are related to monetisation.

Table A14: Using a Category-Specific Privacy Measure

	Cross-Section				Panel		
Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$D_{PrivCatSpec}$	-0.064***	-0.072***	-0.025**	-0.156***	-0.039*	-0.015	-0.062**
	(0.012)	(0.011)	(0.012)	(0.020)	(0.020)	(0.022)	(0.028)
$\#_{CleanPerm}$	0.115***	0.037***	0.037***	0.053***	-0.010**	-0.011**	-0.002
	(0.003)	(0.003)	(0.003)	(0.006)	(0.004)	(0.005)	(0.006)
$D_{Internet}$		-0.204***	-0.258***	-0.080***	-0.021	-0.015	0.014
		(0.009)	(0.012)	(0.013)	(0.054)	(0.053)	(0.071)
D_{Ads}		0.233***	0.240***	0.037**	-0.008	-0.011	-0.048
		(0.009)	(0.010)	(0.017)	(0.032)	(0.033)	(0.043)
Log. Price		-0.071***	0.000	0.100***	-0.043***	-0.043***	-0.037***
		(0.001)	(.)	(0.007)	(0.011)	(0.011)	(0.013)
D_{Other}		0.047***	0.014	0.012		0.017	
		(0.009)	(0.011)	(0.014)		(0.022)	
Log. Length Desc.		0.261***	0.279***	0.133***		0.000	
		(0.004)	(0.005)	(0.006)		(0.023)	
Log. Size (in KB)		0.062***	0.079***	0.010***		0.041**	
		(0.002)	(0.003)	(0.003)		(0.017)	
Number Screenshots		0.085***	0.106***	0.057***		0.029**	
		(0.002)	(0.003)	(0.003)		(0.011)	
Dummy: Video		0.204***	0.271***	0.224***		-0.139**	
		(0.012)	(0.016)	(0.018)		(0.058)	
Log. Average Rating		0.200***	0.372***	0.083***		0.272***	
		(0.007)	(0.010)	(0.007)		(0.098)	
Dummy: Top-Dev.		1.130***	1.150***	1.142***		-0.178***	
		(0.058)	(0.088)	(0.082)		(0.014)	
App Version		0.028***	0.038***	0.035***		0.030*	
		(0.004)	(0.005)	(0.006)		(0.015)	
Log. AppsByDev		-0.113***	-0.168***	-0.046***		-0.161***	
		(0.002)	(0.003)	(0.002)		(0.029)	
Log. InstByDev		0.183***	0.236***	0.100***		-0.011	
		(0.002)	(0.002)	(0.002)		(0.014)	
Log. InstByComp		-0.007***	-0.010***	-0.008***		0.004	
		(0.002)	(0.002)	(0.002)		(0.004)	
Log. PriceOfComp		0.011***	0.012***	0.003***		-0.001	
		(0.001)	(0.001)	(0.001)		(0.001)	
Log. RatByComp		-0.698***	-0.819***	0.159**		-0.028	
		(0.061)	(0.078)	(0.078)		(0.129)	
Min. Android Vers.		-0.015	-0.029**	0.122***		0.059	
		(0.010)	(0.013)	(0.015)		(0.044)	
Max. Android Vers.		0.170***	0.171***	0.144***		-0.145	
		(0.020)	(0.024)	(0.029)		(0.096)	
Constant	-0.283***	-3.879***	-3.557***	-3.432***	1.162***	1.251**	0.740***
	(0.008)	(0.123)	(0.152)	(0.172)	(0.128)	(0.561)	(0.154)
Category	No	Yes	Yes	Yes	No	Yes	No
Month	No	No	No	No	Yes	Yes	Yes
Observations	177193	177193	128921	48272	33095	33095	15220
Num. of Groups					6619	6619	3044
Mean of dep. Var.	0.10	0.10	0.33	-0.50	1.63	1.63	1.17
SD of dep. Var.	1.55	1.55	1.65	1.06	2.23	2.23	2.10
Adjusted R^2	0.048	0.294	0.285	0.262	0.026	0.033	0.020

NOTES: The table shows descriptive regressions analyzing the relationship between the presence of category-specific privacy-sensitive permissions and app demand. We look at privacy-sensitive permissions that are unusual for the app's category. Within a category we flag a privacy-sensitive permission as problematic only if paid apps of this category use this permission on average less often than the overall average paid app. In Columns 1 to 4 we provide cross-sectional results, whereas Columns 5 to 7 contain panel results. Column 3 restricts the sample to free apps, whereas Column 4 uses paid apps. Columns 5 shows the raw fixed effects regressions without controls, whereas in Column 6 we add control variables and dummies to control for the apps' categories and in Column 7 we restrict the analysis to apps that introduced new permissions without changing the app's description (no change in the length of description). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A15: Overview of Category-Specific Privacy-Sensitive Permissions

	INTERCEPT_OUTGOING_CALLS	GOOGLE_AUTH	READ_HISTORY_BOOKMARKS	READ_SMS	ACCESS_FINE_LOCATION	GET_ACCOUNTS	READ_CONTACTS	READ_SYNC_STATS	READ_PHONE_STATE	USE_CREDENTIALS	READ_LOGS	SUBSCRIBED_FEEDS_READ	RECORD_AUDIO	CAMERA	ACCESS_COARSE_LOCATION	RECEIVE_SMS	RECEIVE_MMS	RECEIVE_WAP_PUSH	ACCESS_DOWNLOAD_MANAGER	ACCESS_LOCATION_EXTRA_COMMANDS	MANAGE_ACCOUNTS	NFC	READ_INPUT_STATE	MOUNT_UNMOUNT_FILESYSTEMS	ACCOUNT_MANAGER
Arcade & Action	●	○	●	●	●	○	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	○	○	●	●
Books & Reference	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Brain & Puzzle	●	○	●	●	●	○	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●
Business	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	●	○	●	○	○	○	●	●	○
Cards & Casino	●	○	●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Casual	●	●	●	●	●	○	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Comics	●	●	●	●	●	○	●	●	●	○	●	●	●	●	●	●	●	●	○	●	●	●	●	○	●
Communication	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	●	○
Education	●	●	○	●	●	●	●	●	●	●	●	○	○	●	●	●	●	●	○	●	●	●	●	●	●
Entertainment	●	●	●	○	●	○	●	●	○	●	●	●	○	○	●	●	●	●	●	●	●	●	●	●	●
Finance	●	●	●	●	●	○	●	●	●	○	●	●	●	○	●	●	●	●	●	●	○	●	●	●	○
Health & Fitness	●	●	●	●	○	●	●	●	○	●	●	●	●	○	○	●	●	●	●	○	●	●	●	●	●
Libraries & Demo	○	●	○	●	●	●	●	○	●	○	●	○	●	○	○	●	○	○	●	○	○	●	○	○	○
Lifestyle	●	●	●	●	○	●	●	○	●	●	●	●	●	○	○	●	●	●	●	○	●	●	○	○	●
Media & Video	○	○	○	●	●	○	●	●	○	○	○	●	○	○	○	●	●	●	●	●	○	●	●	○	●
Medical	●	●	●	●	●	○	●	●	○	●	●	●	●	●	●	●	●	●	●	●	○	●	●	○	●
Music & Audio	○	○	●	●	●	●	●	●	○	○	○	●	○	●	●	●	○	○	○	○	●	●	●	●	○
News & Magazines	●	○	○	●	○	○	●	●	○	○	○	●	●	○	○	○	●	●	●	○	○	○	●	○	●
Personalization	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●
Photography	●	●	●	●	○	●	●	●	○	○	●	○	○	○	○	●	●	●	●	○	○	○	●	●	●
Productivity	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○
Racing	●	○	●	●	●	○	●	●	○	○	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●
Shopping	●	●	○	●	○	●	○	●	●	●	○	●	●	○	○	○	○	○	●	○	●	●	●	○	○
Social	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	●	○	●
Sports	●	●	●	●	○	○	○	●	○	○	●	●	○	○	○	●	●	●	●	○	●	●	●	○	○
Sports Games	●	●	●	●	●	○	●	●	○	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●
Tools	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Transportation	○	●	○	○	○	●	○	●	○	●	●	●	○	○	○	○	●	●	○	○	●	○	●	●	●
Travel & Local	●	●	●	●	○	●	○	●	○	●	○	●	○	○	○	○	●	●	●	○	●	●	●	○	●
Weather	●	●	●	●	○	○	●	●	●	●	○	●	●	●	○	●	●	●	●	○	●	●	●	●	●

Notes: A '●' indicates that a permission is relatively uncommon in a category and is thus classified as category-specific privacy-sensitive. A '○' indicates that a permission is relatively common in a category and is thus classified as unproblematic. A permission is defined to be relatively common within an app category if paid apps of this category use this permission on average more often than the overall paid app.

B Online Appendix B - Additional Evidence Detailed Regressions

B.1 Additional Descriptive Statistics

Table B1: Summary Statistics of the Pairs Datasets

	Pairs I		Pairs II		Pairs III	
	Free	Paid	Free	Paid	Free	Paid
Δ Ratings	73.81	8.80	67.33	6.08	81.16	2.65
Δ Installations	23895.46	456.34	8213.48	92.71	3833.90	101.76
Ratings	980.13	119.90	623.53	81.05	1420.45	62.58
Installations	2.1e+05	4428.21	1.3e+05	2739.82	4.1e+05	4167.34
$\#_{TotalPerm}$	3.63	2.99	3.78	3.10	4.39	2.18
$D_{Privacy}$	0.46	0.37	0.49	0.40	0.57	0.29
$\#_{Privacy}$	0.88	0.66	0.93	0.70	1.12	0.47
$\#_{CleanPerm}$	2.75	2.33	2.85	2.40	3.27	1.71
$D_{PrivCatSpec}$	0.15	0.10	0.14	0.08	0.23	0.05
$D_{MTurkEP2}$	0.11	0.11	0.12	0.12	0.11	0.09
D_{Google}	0.24	0.19	0.25	0.21	0.27	0.14
$D_{Sarmaetal}$	0.42	0.32	0.46	0.35	0.53	0.24
D_{ID}	0.27	0.19	0.30	0.22	0.40	0.15
$D_{Location}$	0.19	0.11	0.21	0.12	0.27	0.05
$D_{Communication}$	0.07	0.07	0.06	0.06	0.05	0.05
$D_{Profile}$	0.18	0.16	0.19	0.18	0.19	0.13
$D_{Internet}$	0.76	0.52	0.82	0.56	0.99	0.30
D_{Ads}	0.53	0.28	0.56	0.31	0.84	0.11
Price	0.00	1.46	0.00	1.37	0.00	1.19
Average Rating	3.85	4.20	3.79	4.14	3.93	4.21
Size (in KB)	3043.97	3734.91	3206.59	3039.23	2102.13	1848.41
Length Description	1052.04	1027.52	876.84	873.62	982.41	853.95
Number Screenshots	3.92	4.15	3.73	3.87	4.01	3.99
Dummy: Video	0.19	0.19	0.17	0.17	0.14	0.13
Dummy: Top-Dev.	0.01	0.01	0.01	0.01	0.02	0.02
Apps by Developer	33.55	33.55	14.47	14.47	10.60	10.60
Average Installations of Developer	58752.29	1.3e+05	57842.83	91748.60	65842.46	2.1e+05
Observations	14422		3998		708	

Notes: The table provides an overview over the most important variables, and shows the corresponding descriptive statistics for the three pairs datasets in this paper. For each dataset we show two columns, where the left column shows averages for free apps and the second column for paid apps.

B.2 Additional Robustness Checks

Note: This section is not intended for publication in the journal, instead we will make it available online. This section presents additional results from robustness checks that had to be excluded from the paper for reasons of space as well as more extensive descriptions of data sets and estimation results presented in the paper.

B.2.1 Privacy and Demand: The Role of Apps' Previous Success

We further analyze how the relationship between privacy-sensitive permissions and app demand varies with an app's past success. This robustness check shows that our main results are not driven by either very successful or unsuccessful apps only. In Table B2, we divide the estimation sample into four groups according to apps previous success.

Data Preparation and Additional Variables: To analyze how the relationship of interest varies with previous app success, we split our cross-section sample into 20 quantiles according to the total number of ratings an app has received so far. For example, those apps which are in the first two quantiles have the greatest past success, and the highest stock of ratings. We then separately estimated the relationship of interest for the apps in quantiles 1-2 (Column 1), quantiles 3-4 (Column 2), quantliles 5-10 (Column 3) and the remaining apps (quantiles 11 to 20) Column 4.

Results: As before, app demand in Table B2 is measured by the log. number of monthly new ratings of an app. The coefficient of interest analyzes the relationship between an app's downloads and our measures of privacy-sensitive permissions. The dependent variable is app demand and the variable of interest is an indicator for presence of privacy-sensitive permissions ($D_{Privacy}$). Column 1 analyzes the top 10% most successful apps. Column 2 analyzes the top 10-20% most successful apps. Column 3 focuses on the top 20-50% most successful apps, and column 4 analyzes the remaining 50% of least successful apps. All specifications control for the number of unproblematic permissions ($CleanPerm.$), and for an app's other observed characteristics on the Play Store (the app's price, description, ratings, categorical dummies, etc.). We also control for internet access, and ad-specific permissions.

While the coefficient of interest is insignificant for the top 10% of the apps, it is largest in columns 2 and 3, i.e. for intermediate apps, and somewhat weaker for unsuccessful apps. These findings highlight, that our main results are driven by the large mass of "average" apps, and that they actually apply to top apps to a far lesser extent. Moreover, the findings in this robustness check confirm the insights from the analysis in Table B5, and especially the finding, that successful apps almost do not have to worry about using privacy-intrusive permissions. However, apps that are not in the top-segment are faced with a negative relationship of permissions and demand, suggesting that they are at a competitive disadvantage compared to bestselling apps.

B.2.2 User Assessment and Survival

In Table B3 we explore whether apps with fewer intrusive permissions induce higher user satisfaction and are more likely to survive. Given our results for demand effects of privacy-sensitive permissions,

Table B2: Samples based on Past App Success

	1st-10th pct	11th-20th pct	21st-50th pct	51st-100th pct
Log. $\Delta Ratings$	(1)	(2)	(3)	(4)
$D_{Privacy}$	-0.016	-0.133***	-0.090***	-0.026***
	(0.026)	(0.024)	(0.013)	(0.006)
$\#_{CleanPerm}$	0.011***	0.013**	0.013***	0.004***
	(0.004)	(0.006)	(0.003)	(0.001)
$D_{Internet}$	-0.204***	-0.121***	-0.092***	-0.021***
	(0.034)	(0.029)	(0.014)	(0.005)
D_{Ads}	0.420***	0.394***	0.264***	0.088***
	(0.029)	(0.026)	(0.013)	(0.006)
Log. Price	-0.085***	-0.065***	-0.049***	-0.014***
	(0.003)	(0.003)	(0.001)	(0.000)
D_{Other}	0.083***	0.083***	0.065***	0.023***
	(0.025)	(0.025)	(0.013)	(0.006)
Log. Length Desc.	0.321***	0.201***	0.137***	0.064***
	(0.014)	(0.013)	(0.006)	(0.002)
Log. Size (in KB)	0.120***	0.100***	0.056***	0.013***
	(0.008)	(0.007)	(0.003)	(0.001)
Number Screenshots	0.082***	0.084***	0.059***	0.022***
	(0.006)	(0.006)	(0.003)	(0.001)
Dummy: Video	0.032	-0.074**	0.019	0.014*
	(0.028)	(0.032)	(0.018)	(0.008)
Log. Average Rating	2.183***	1.226***	0.557***	0.019***
	(0.089)	(0.061)	(0.022)	(0.004)
Dummy: Top-Dev.	0.387***	-0.029	-0.086	-0.008
	(0.062)	(0.128)	(0.092)	(0.045)
App Version	-0.051***	-0.080***	-0.023***	-0.009***
	(0.013)	(0.012)	(0.006)	(0.002)
Log. AppsByDev	-0.025***	-0.003	-0.024***	-0.031***
	(0.010)	(0.008)	(0.004)	(0.001)
Log. InstByDev	0.115***	0.048***	0.058***	0.025***
	(0.005)	(0.005)	(0.003)	(0.001)
Log. InstByComp	0.041***	0.012**	-0.006**	-0.008***
	(0.007)	(0.006)	(0.003)	(0.001)
Log. PriceOfComp	0.024***	0.001	-0.003***	-0.004***
	(0.003)	(0.002)	(0.001)	(0.000)
Log. RatByComp	-0.671***	-0.422**	-0.247***	-0.223***
	(0.187)	(0.167)	(0.091)	(0.040)
Min. Android Vers.	0.681***	0.786***	0.413***	0.074***
	(0.036)	(0.033)	(0.017)	(0.006)
Max. Android Vers.	0.274***	0.322***	0.234***	0.060***
	(0.082)	(0.064)	(0.026)	(0.010)
Constant	-7.316***	-5.818***	-4.139***	-1.517***
	(0.447)	(0.369)	(0.173)	(0.073)
Category	Yes	Yes	Yes	Yes
Observations	17711	17396	51021	91065
Mean of dep. Var.	2.91	1.37	0.20	-0.74
SD of dep. Var.	1.67	1.47	1.21	0.60
Adjusted R^2	0.340	0.308	0.193	0.080

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data which are defined according to apps past success. App demand is measured by the log. number of monthly new ratings of an app. We split our cross-section sample into 20 quantiles according to the total number of ratings an app has received so far. In Column 1 we estimate our model for those apps which are in the first two quantiles, i.e. which have the highest stock of ratings. In Columns 2 apps from the 3rd and 4th quantile are used. In Column 3 apps from the 5th to the 10th quantile are analyzed. In Column 4 the remaining apps, i.e. those from the 11th to the 20th quantile are used. All specifications control for the number of unproblematic permissions (*CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

we expect average apps that use privacy-sensitive permissions also to face lower user-satisfaction and, ultimately, lower survival rates.

Data Preparation and Additional Variables: Table B3 explores the relationship between apps' use of privacy-sensitive permissions and two alternative success measures: apps' average ratings (Columns 1-4) and their long-run survival rates (Columns 5-6). All results are based on the cross-section from April 2012. In Columns 1 and 2 the dependent variable is the log. average value of the ratings the app has received in April 2012. In Columns 3 and 4 the dependent variable is the log. average value of the ratings the app has received between April and September 2012. In Columns 5 and 6 we analyze app survival over two years. The dependent variable is a dummy which equals one, if an app from April 2012 was still available in 2014 (i.e. if we observe it in the data set we collected in 2014). All specifications control for the number of unproblematic permissions ($CleanPerm.$).

Results: The first four columns show that apps which do not use privacy-sensitive permissions or do use less privacy-sensitive permissions are rated better by users. Moreover, apps which use privacy-sensitive permissions are also less likely to survive until 2014, as can be seen in Columns 5 and 6. In short, the demand side results carry over to user satisfaction and survival.

B.2.3 Moderating Factors (Alternative Specifications)

Table B3: User Assessment and Survival

	User Assessment (Avg. Rating)				Survival ($D_{Surviver}$)	
	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-0.022***		-0.018***		-0.028***	
	(0.003)		(0.003)		(0.003)	
$\#_{Privacy}$		-0.011***		-0.009***		-0.017***
		(0.001)		(0.001)		(0.001)
$\#_{CleanPerm}$	0.000	0.003***	0.000	0.003***	-0.005***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$D_{Internet}$	-0.054***	-0.058***	-0.056***	-0.060***	-0.041***	-0.046***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
D_{Ads}	0.009**	0.006	0.014***	0.012***	-0.025***	-0.030***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Log. Price	-0.002***	-0.003***	-0.005***	-0.005***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
D_{Other}	0.006*	0.005	-0.003	-0.004	-0.005*	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log. Length Desc.	0.006***	0.006***	0.006***	0.006***	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Log. Size (in KB)	0.009***	0.009***	0.012***	0.012***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number Screenshots	0.012***	0.012***	0.013***	0.013***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Dummy: Video	0.014***	0.014***	0.012***	0.012***	0.007**	0.008**
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Dummy: Top-Dev.	0.007	0.006	-0.013	-0.014	-0.041***	-0.042***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.015)	(0.015)
App Version	0.004**	0.004**	0.004***	0.004***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Log. AppsByDev	-0.007***	-0.006***	-0.001	-0.000	-0.061***	-0.061***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log. InstByDev	0.015***	0.015***	0.015***	0.015***	-0.001	-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Log. InstByComp	-0.007***	-0.007***	-0.005***	-0.005***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log. PriceOfComp	-0.001***	-0.001***	-0.001***	-0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log. RatByComp	0.449***	0.446***	0.436***	0.434***	-0.051***	-0.051***
	(0.024)	(0.024)	(0.020)	(0.020)	(0.020)	(0.020)
Min. Android Vers.	-0.000	0.001	0.004	0.004	-0.044***	-0.044***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
Max. Android Vers.	0.020**	0.020**	0.006	0.006	0.011	0.011
	(0.010)	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)
Log. Average Rating					0.017***	0.016***
					(0.003)	(0.003)
Constant	0.354***	0.353***	0.360***	0.361***	1.130***	1.126***
	(0.055)	(0.055)	(0.046)	(0.045)	(0.041)	(0.041)
Category	Yes	Yes	Yes	Yes	Yes	Yes
Observations	68248	68248	102594	102594	177193	177193
Mean of dep. Var.	1.29	1.29	1.29	1.29	0.73	0.73
SD of dep. Var.	0.35	0.35	0.35	0.35	0.45	0.45
Adjusted R^2	0.043	0.044	0.052	0.052	0.118	0.119

NOTES: This table shows the relationship between apps' use of privacy-sensitive permissions and two alternative success measures: apps' average rating (Columns 1-4) and their long-run survival (Columns 5-6). All results are based on the cross-section from April 2012. In Columns 1 and 2 the dependent variable is the log. average of the ratings the app has received in April 2012. In Columns 3 and 4 the dependent variable is the log. average of the ratings the app has received between April and September 2012. In Columns 5 and 6 the dependent variable is a dummy which is equal to one, if an app from April 2012 was still available in our data set collected in 2014. All specifications control for the number of unproblematic permissions (*CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B4: Alternative Moderating Factors

	Top-Developer		Alexa Rank		Games		Pricing Model	
Log. $\Delta Ratings$	No	Yes	No	Yes	No	Yes	Free	Paid
$D_{Privacy}$	-0.065***	0.088	-0.055***	0.010	-0.082***	-0.020	-0.026**	-0.121***
	(0.009)	(0.151)	(0.010)	(0.172)	(0.010)	(0.022)	(0.011)	(0.014)
$\#_{CleanPerm}$	0.036***	0.097***	0.035***	0.051*	0.037***	0.098***	0.037***	0.054***
	(0.003)	(0.031)	(0.003)	(0.028)	(0.003)	(0.009)	(0.003)	(0.006)
$D_{Internet}$	-0.195***	0.023	-0.187***	0.152	-0.211***	-0.201***	-0.256***	-0.075***
	(0.009)	(0.231)	(0.010)	(0.264)	(0.010)	(0.024)	(0.012)	(0.013)
D_{Ads}	0.232***	0.360**	0.226***	-0.251	0.228***	0.157***	0.241***	0.039**
	(0.009)	(0.164)	(0.010)	(0.173)	(0.010)	(0.024)	(0.010)	(0.017)
Log. Price	-0.071***	-0.108***	-0.071***	-0.107***	-0.068***	-0.095***	0.000	0.099***
	(0.001)	(0.011)	(0.001)	(0.024)	(0.001)	(0.002)	(.)	(0.007)
D_{Other}	0.053***	-0.293**	0.070***	0.137	0.023**	0.144***	0.016	0.022
	(0.009)	(0.129)	(0.010)	(0.164)	(0.010)	(0.027)	(0.011)	(0.015)
Log. Length Desc.	0.262***	0.492***	0.277***	0.372***	0.255***	0.291***	0.280***	0.135***
	(0.004)	(0.084)	(0.004)	(0.087)	(0.004)	(0.011)	(0.005)	(0.006)
Log. Size (in KB)	0.063***	-0.008	0.062***	0.200***	0.050***	0.117***	0.080***	0.011***
	(0.002)	(0.065)	(0.002)	(0.054)	(0.002)	(0.006)	(0.003)	(0.003)
Number Screenshots	0.085***	0.050**	0.083***	0.221***	0.086***	0.085***	0.105***	0.057***
	(0.002)	(0.026)	(0.002)	(0.040)	(0.002)	(0.006)	(0.003)	(0.003)
Dummy: Video	0.199***	0.347***	0.191***	0.910***	0.204***	0.177***	0.271***	0.225***
	(0.012)	(0.134)	(0.013)	(0.224)	(0.015)	(0.024)	(0.016)	(0.018)
Log. Average Rating	0.198***	0.747***	0.204***	-0.216	0.167***	0.355***	0.371***	0.083***
	(0.007)	(0.228)	(0.008)	(0.300)	(0.008)	(0.020)	(0.010)	(0.007)
Dummy: Top-Dev.	0.000	0.000	0.949***	0.638**	1.026***	1.295***	1.151***	1.169***
	(.)	(.)	(0.065)	(0.292)	(0.074)	(0.092)	(0.088)	(0.082)
App Version	0.029***	-0.161*	0.032***	-0.027	0.029***	0.017*	0.039***	0.036***
	(0.004)	(0.083)	(0.004)	(0.153)	(0.004)	(0.009)	(0.005)	(0.006)
Log. AppsByDev	-0.113***	-0.387***	-0.110***	0.148	-0.105***	-0.180***	-0.169***	-0.047***
	(0.002)	(0.072)	(0.002)	(0.126)	(0.002)	(0.005)	(0.003)	(0.002)
Log. InstByDev	0.183***	0.192***	0.188***	0.154***	0.179***	0.203***	0.236***	0.100***
	(0.002)	(0.040)	(0.002)	(0.032)	(0.002)	(0.004)	(0.002)	(0.002)
Log. InstByComp	-0.008***	0.041	-0.008***	-0.037	-0.007***	0.003	-0.010***	-0.007***
	(0.002)	(0.035)	(0.002)	(0.047)	(0.002)	(0.006)	(0.002)	(0.002)
Log. PriceOfComp	0.010***	0.045**	0.011***	-0.003	0.008***	0.019***	0.012***	0.004***
	(0.001)	(0.018)	(0.001)	(0.015)	(0.001)	(0.002)	(0.001)	(0.001)
Log. RatByComp	-0.721***	0.399	-0.776***	-0.043	-0.696***	-1.085***	-0.823***	0.148*
	(0.061)	(0.967)	(0.067)	(1.341)	(0.067)	(0.147)	(0.078)	(0.078)
Min. Android Vers.	-0.012	-0.580***	-0.011	-0.302	-0.025**	0.039	-0.029**	0.125***
	(0.010)	(0.193)	(0.012)	(0.195)	(0.011)	(0.028)	(0.013)	(0.015)
Constant	-3.852***	-3.587*	-3.832***	-6.650***	-3.398***	-5.120***	-3.557***	-3.416***
	(0.123)	(2.124)	(0.136)	(2.355)	(0.140)	(0.306)	(0.152)	(0.172)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176224	969	145367	771	147143	30050	128921	48272
Mean of dep. Var.	0.09	2.15	0.11	2.19	0.07	0.26	0.33	-0.50
SD of dep. Var.	1.54	2.25	1.55	2.09	1.52	1.70	1.65	1.06
Adjusted R^2	0.286	0.538	0.298	0.348	0.280	0.370	0.285	0.262

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data. App demand is measured by the log. number of monthly new ratings of an app. Columns 1 and 2 split the sample into apps which are from a top developer (according to Google's classification) or not. Columns 3 and 4 split the sample into normal apps and games. Columns 5 and 6 split them into normal apps and games. Columns 7 and 8 split the sample into free and paid apps. All specifications control for the number of unproblematic permissions (*#CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B5: Moderating Factors (Alternative Specifications)

	Reputation				User Groups			Visibility
Log. $\Delta Ratings$	TopDev	Installs.	Priv. Policy	Alexa	Maturity	Games	Med&Health	Warning
$D_{Privacy}$	-0.067***	-0.077***	-0.079***	-0.058***	-0.051***	-0.131***	-0.058***	-0.032***
	(0.009)	(0.008)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
$D_{Privacy} \times D_{TopDev}$	0.303***							
	(0.116)							
$D_{Privacy} \times D_{NumInst}$		0.101***						
		(0.025)						
$D_{NumInst}$		2.530***						
		(0.019)						
$D_{Privacy} \times D_{Transp}$			0.397***					
			(0.041)					
D_{Transp}			0.291***					
			(0.030)					
$D_{Privacy} \times D_{AlexaRank}$				0.613***				
				(0.133)				
$D_{AlexaRank}$				0.926***				
				(0.099)				
$D_{Privacy} \times D_{Maturity}$					-0.070***			
					(0.018)			
$D_{Maturity}$					0.105***			
					(0.017)			
$D_{Privacy} \times D_{Game}$						0.318***		
						(0.019)		
D_{Game}						-0.635***		
						(0.055)		
$D_{Privacy} \times D_{MedHealth}$							-0.181***	
							(0.032)	
$D_{MedHealth}$							-0.055**	
							(0.028)	
$D_{Privacy} \times D_{Google}$								-0.084***
								(0.012)
$\#_{CleanPerm}$	0.037***	0.029***	0.035***	0.035***	0.036***	0.039***	0.036***	0.040***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$D_{Internet}$	-0.200***	-0.141***	-0.198***	-0.186***	-0.199***	-0.206***	-0.200***	-0.205***
	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
D_{Ads}	0.235***	0.204***	0.232***	0.224***	0.235***	0.233***	0.236***	0.229***
	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Log. Price	-0.071***	-0.049***	-0.071***	-0.071***	-0.071***	-0.071***	-0.071***	-0.071***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_{TopDev}	0.962***	0.550***	1.047***	0.904***	1.139***	1.139***	1.151***	1.138***
	(0.091)	(0.047)	(0.058)	(0.064)	(0.058)	(0.058)	(0.058)	(0.058)
D_{Other}	0.052***	0.057***	0.052***	0.071***	0.051***	0.055***	0.052***	0.052***
	(0.009)	(0.008)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Log. Length Desc.	0.263***	0.209***	0.256***	0.278***	0.262***	0.261***	0.262***	0.262***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log. Size (in KB)	0.063***	0.047***	0.061***	0.062***	0.062***	0.063***	0.063***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number Screenshots	0.085***	0.076***	0.082***	0.084***	0.085***	0.085***	0.085***	0.085***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Dummy: Video	0.204***	0.155***	0.186***	0.193***	0.205***	0.191***	0.204***	0.203***
	(0.012)	(0.011)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)
Log. Average Rating	0.199***	0.173***	0.199***	0.203***	0.199***	0.194***	0.199***	0.199***
	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
App Version	0.029***	0.004	0.027***	0.032***	0.028***	0.029***	0.029***	0.028***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log. AppsByDev	-0.114***	-0.096***	-0.113***	-0.110***	-0.113***	-0.118***	-0.114***	-0.113***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByDev	0.184***	0.121***	0.181***	0.187***	0.184***	0.183***	0.184***	0.183***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByComp	-0.007***	-0.020***	-0.006***	-0.008***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log. PriceOfComp	0.011***	0.000	0.011***	0.011***	0.011***	0.010***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log. RatByComp	-0.703***	-0.198***	-0.705***	-0.776***	-0.702***	-0.752***	-0.707***	-0.706***
	(0.061)	(0.053)	(0.061)	(0.067)	(0.061)	(0.061)	(0.061)	(0.061)
Min. Android Vers.	-0.015	0.106***	-0.013	-0.015	-0.015	-0.010	-0.015	-0.014
	(0.010)	(0.009)	(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)
Max. Android Vers.	0.170***	0.158***	0.167***	0.168***	0.170***	0.170***	0.169***	0.170***
	(0.020)	(0.017)	(0.020)	(0.022)	(0.020)	(0.020)	(0.020)	(0.020)
Constant	-3.879***	-3.686***	-3.791***	-3.844***	-3.882***	-3.257***	-3.872***	-3.884***
	(0.122)	(0.104)	(0.122)	(0.136)	(0.122)	(0.132)	(0.122)	(0.122)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177193	177193	177193	146138	177193	177193	177193	177193
Interaction	0.236	0.024	0.317	0.555	0.054	0.187	-0.239	-0.116
SE of Interaction	0.116	0.025	0.042	0.133	0.021	0.018	0.032	0.012
Mean of dep. Var.	0.10	0.10	0.10	0.12	0.10	0.10	0.10	0.10
SD of dep. Var.	1.55	1.55	1.55	1.56	1.55	1.55	1.55	1.55
Adjusted R^2	0.294	0.459	0.298	0.304	0.294	0.295	0.294	0.294

NOTES: This table analyzes factors that moderate the relationship between privacy-sensitive permissions and app demand. In each specification we add interaction-terms to analyze the moderating role of reputation and user groups. Columns 1-4 analyze measures of developer or app reputation/success, whereas columns 5-7 analyze the role of user groups. The dependent variable is demand measured by the log. number of monthly new ratings of an app. Column 1 adds an interaction of privacy-sensitiveness with a dummy for 'top developers' (measured by a badge, awarded by Google). Column 2 adds an interaction and a dummy which is equal to one if the app had accumulated 75000 or more installations since its market entry. Column 3 adds an interaction and a dummy which is equal to one if the app is transparant about its origin and privacy policy, i.e. if it has published a privacy policy, email and website address that are directly accessible from the Play Store. Column 4 adds an interaction and dummy which is equal to one if the app has a high ranking on Alexa.com, i.e. its website traffic rank is lower than 10000. Column 5 uses a dummy which is equal to one if the app is a high maturity app, i.e. if it requires medium or high maturity or is not rated at all. In Column 6 it is equal to one if the app is a game. Column 7 contains a dummy which is equal to one if the app is a health or medical app. In Column 8 we use a interaction term which is equal to one if the app uses at least one privacy-sensitive permission and one permission for which Google provides in its permission description a warning, which states that a malicious app with this permission could create harm to the user. All specifications control for the app's observed characteristics on the Play Store (the app's price, description, ratings, categorical dummies, etc.), and also control for internet access, and ad-specific permissions as well as unproblematic permissions. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.3 Main Estimation Tables - Detailed Version

Table B6: Main Supply Side Results

	Business Model Choice (D_{Paid})						Price Choice (Log. Price)			
	Cross-Section		Panel		Pairs		Cross-Section		Panel	
$D_{Privacy}$	-0.033***		-0.104		-0.156***		0.012		-0.119*	
	(0.002)		(0.261)		(0.042)		(0.008)		(0.070)	
$\#_{Privacy}$		-0.024***		-0.150*		-0.137***		-0.005		-0.060**
		(0.001)		(0.086)		(0.025)		(0.004)		(0.025)
$\#_{CleanPerm}$	0.005***	0.011***	-0.003	0.035	0.204***	0.226***	0.018***	0.020***	0.048**	0.068***
	(0.001)	(0.001)	(0.024)	(0.030)	(0.023)	(0.025)	(0.002)	(0.002)	(0.020)	(0.023)
$D_{Internet}$	-0.219***	-0.225***	-0.627**	-0.582**	-0.496***	-0.517***	0.067***	0.068***	0.004	-0.015
	(0.003)	(0.003)	(0.290)	(0.246)	(0.037)	(0.038)	(0.007)	(0.007)	(0.085)	(0.083)
D_{Ads}	-0.119***	-0.124***	-0.347	-0.165	-0.535***	-0.547***	0.014	0.014	-0.003	-0.039
	(0.002)	(0.002)	(0.301)	(0.250)	(0.036)	(0.036)	(0.009)	(0.009)	(0.119)	(0.112)
D_{Other}	0.010***	0.009***	0.363***	0.546***	-0.153**	-0.155**	0.119***	0.121***	-0.015	-0.009
	(0.002)	(0.002)	(0.110)	(0.139)	(0.068)	(0.067)	(0.008)	(0.007)	(0.076)	(0.077)
Log. Length Desc.	0.054***	0.054***	0.133	0.124	-0.607***	-0.601***	0.079***	0.079***	0.160**	0.133**
	(0.001)	(0.001)	(0.092)	(0.111)	(0.203)	(0.201)	(0.004)	(0.004)	(0.069)	(0.064)
Log. Size (in KB)	0.032***	0.031***	-0.591***	-0.605***	-0.186***	-0.177***	0.017***	0.017***	0.040	0.045
	(0.001)	(0.001)	(0.187)	(0.154)	(0.037)	(0.037)	(0.002)	(0.002)	(0.033)	(0.033)
Number Screenshots	0.011***	0.011***	0.111***	0.131***	0.063***	0.062***	0.008***	0.008***	0.040*	0.038*
	(0.001)	(0.001)	(0.035)	(0.041)	(0.015)	(0.015)	(0.002)	(0.002)	(0.021)	(0.023)
Dummy: Video	-0.003	-0.002	-0.432	-0.208	0.055	0.040	0.045***	0.045***	-0.123	-0.162
	(0.003)	(0.003)	(0.282)	(0.283)	(0.120)	(0.124)	(0.008)	(0.008)	(0.112)	(0.108)
Log. Average Rating	-0.038***	-0.039***	-0.063	0.295	0.370***	0.372***	-0.021***	-0.021***	-0.036	-0.069
	(0.003)	(0.003)	(0.111)	(0.241)	(0.051)	(0.050)	(0.007)	(0.007)	(0.138)	(0.137)
Dummy: Top-Dev.	0.186***	0.184***	0.000	0.000	0.000	0.000	0.097***	0.097***	0.000	0.000
	(0.015)	(0.014)	(.)	(.)	(.)	(.)	(0.030)	(0.030)	(.)	(.)
App Version	-0.048***	-0.048***	-0.142	-0.194*	-0.192***	-0.191***	-0.014***	-0.014***	0.069	0.083*
	(0.001)	(0.001)	(0.150)	(0.109)	(0.014)	(0.014)	(0.003)	(0.003)	(0.053)	(0.049)
Log. AppsByDev	0.021***	0.022***	-1.619**	-1.236*	0.000	0.000	0.019***	0.019***	0.021	-0.008
	(0.001)	(0.001)	(0.793)	(0.693)	(.)	(.)	(0.002)	(0.002)	(0.094)	(0.094)
Log. InstByDev	-0.018***	-0.019***	-0.488***	-0.407***	0.158***	0.158***	0.010***	0.010***	0.017	0.011
	(0.000)	(0.000)	(0.120)	(0.119)	(0.005)	(0.005)	(0.001)	(0.001)	(0.022)	(0.019)
Log. InstByComp	-0.018***	-0.018***	-0.052*	-0.048	-0.023***	-0.022***	-0.016***	-0.016***	0.003	0.002
	(0.001)	(0.001)	(0.030)	(0.035)	(0.008)	(0.008)	(0.001)	(0.001)	(0.012)	(0.012)
Log. PriceOfComp	0.017***	0.017***	-0.001	0.002	0.007**	0.007***	0.008***	0.008***	0.003	0.002
	(0.000)	(0.000)	(0.007)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)	(0.012)	(0.011)
Log. RatByComp	0.189***	0.188***	-0.147	-0.450	1.344***	1.351***	-0.526***	-0.531***	0.331	0.566
	(0.017)	(0.017)	(0.409)	(0.390)	(0.228)	(0.227)	(0.048)	(0.048)	(0.448)	(0.430)
Min. Android Vers.	-0.045***	-0.045***	-1.124***	-1.262***	-0.229	-0.191	-0.061***	-0.061***	0.057	0.126
	(0.003)	(0.003)	(0.238)	(0.229)	(0.160)	(0.147)	(0.008)	(0.008)	(0.153)	(0.138)
Max. Android Vers.	0.026***	0.026***	1.371***	1.135***	-0.256	-0.260	0.040**	0.039**	0.264***	0.291***
	(0.006)	(0.006)	(0.380)	(0.319)	(0.176)	(0.177)	(0.019)	(0.019)	(0.053)	(0.052)
Constant	0.094***	0.090**	10.787***	9.913***	3.495**	3.291**	0.060	0.065	-2.943***	-3.558***
	(0.035)	(0.035)	(2.840)	(2.754)	(1.533)	(1.520)	(0.106)	(0.106)	(0.870)	(0.835)
Category	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	176000	176000	96	96	3998	3998	47079	47079	422	422
Num. of Groups			48	48	1999	1999			211	211
Mean of dep. Var.	0.27	0.27	0.50	0.50	0.50	0.50	0.20	0.20	0.53	0.53
SD of dep. Var.	0.44	0.44	0.50	0.50	0.50	0.50	0.60	0.60	0.62	0.62
Adjusted R^2	0.320	0.322	0.963	0.971	0.618	0.622	0.223	0.223	0.177	0.183

NOTES: The table shows the relationship between privacy-sensitive permissions and the strategic choices of app developers: the choice of the business model in Columns 1-6, and the price choice in Columns 7-10. In Columns 1-6 the dependent variable D_{Paid} measures the developer's decision to offer their app for money or for free. It takes the value 1 if users have to pay to download the app. Columns 1 and 2 show descriptive regressions based on the cross-section of data, where the independent variable of interest is (1) an indicator for one or more privacy-sensitive permissions ($D_{Privacy}$ in Column 1) or (2) the number of privacy-sensitive permissions ($\#Privacy$ in Column 2). Columns 3 and 4 show panel fixed effects regressions where we restrict the sample to such apps which changed both the number of privacy-sensitive permissions and the business model at least once between April and September 2012. Columns 5 and 6 use data on app-pairs where the paid version of the app has the same or a smaller code size and where both apps have more or less the same description length. Columns 7-10 show the results for price-level choices (of paid apps). The dependent variable is the app's price (in logs). Columns 7 and 8 show cross sectional regressions, and Columns 9 and 10 show panel fixed effects regressions where we restrict the sample to such apps which change both the number of privacy-sensitive permissions and their price at least once. In all specifications we drop outliers with respect to app prices, i.e. apps with prices above 8 Euros. All of these regressions control for the number of clean permissions and permissions that are needed to show ads. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B7: Main Demand Side Results

	Cross-Section (Log.$\Delta Ratings$)			Panel (Log.$\Delta Ratings$)		Difference-in-Difference	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$D_{Privacy}$	0.012	-0.065***		-0.059***		-12.660*	-14.988**
	(0.011)	(0.009)		(0.020)		(6.634)	(7.215)
$\#_{Privacy}$			-0.021***		-0.014**		
			(0.003)		(0.007)		
$\#_{CleanPerm}$	0.111***	0.037***	0.041***	-0.010**	-0.005	0.849	1.011*
	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.643)	(0.605)
$D_{Internet}$		-0.200***	-0.211***	0.003	-0.014		
		(0.009)	(0.009)	(0.054)	(0.053)		
D_{Ads}		0.236***	0.230***	0.002	-0.014		
		(0.009)	(0.009)	(0.033)	(0.033)		
Log. Price		-0.071***	-0.071***	-0.044***	-0.043***		
		(0.001)	(0.001)	(0.011)	(0.011)		
D_{Other}		0.052***	0.045***	0.022	0.024		
		(0.009)	(0.009)	(0.022)	(0.022)		
Log. Length Desc.		0.263***	0.262***	0.003	0.002		
		(0.004)	(0.004)	(0.023)	(0.023)		
Log. Size (in KB)		0.063***	0.062***	0.044***	0.043***		
		(0.002)	(0.002)	(0.017)	(0.017)		
Number Screenshots		0.085***	0.085***	0.029**	0.027**		
		(0.002)	(0.002)	(0.011)	(0.011)		
Dummy: Video		0.204***	0.204***	-0.138**	-0.139**		
		(0.012)	(0.012)	(0.058)	(0.059)		
Log. Average Rating		0.198***	0.199***	0.267***	0.272***		
		(0.007)	(0.007)	(0.098)	(0.098)		
Dummy: Top-Dev.		1.139***	1.136***	-0.206***	-0.179***		
		(0.058)	(0.058)	(0.016)	(0.014)		
App Version		0.029***	0.028***	0.029*	0.032**		
		(0.004)	(0.004)	(0.015)	(0.015)		
Log. AppsByDev		-0.114***	-0.114***	-0.161***	-0.157***		
		(0.002)	(0.002)	(0.029)	(0.029)		
Log. InstByDev		0.184***	0.183***	-0.011	-0.011		
		(0.002)	(0.002)	(0.014)	(0.014)		
Log. InstByComp		-0.007***	-0.007***	0.004	0.004		
		(0.002)	(0.002)	(0.004)	(0.004)		
Log. PriceOfComp		0.011***	0.011***	-0.001	-0.001		
		(0.001)	(0.001)	(0.001)	(0.001)		
Log. RatByComp		-0.707***	-0.703***	-0.030	-0.031		
		(0.061)	(0.061)	(0.129)	(0.129)		
Min. Android Vers.		-0.015	-0.014	0.058	0.056		
		(0.010)	(0.010)	(0.044)	(0.044)		
Max. Android Vers.		0.170***	0.170***	-0.143	-0.137		
		(0.020)	(0.020)	(0.096)	(0.097)		
Constant	-0.289***	-3.871***	-3.872***	1.198**	1.201**	4.262	2.428
	(0.007)	(0.122)	(0.122)	(0.561)	(0.563)	(5.735)	(6.327)
Category	No	Yes	Yes	Yes	Yes	No	No
Month	No	No	No	Yes	Yes	No	No
Observations	177193	177193	177193	33095	33095	192	162
Num. of Groups				6619	6619		
Mean of dep. Var.	0.10	0.10	0.10	1.63	1.63	-0.24	-3.22
SD of dep. Var.	1.55	1.55	1.55	2.23	2.23	31.81	29.27
Adjusted R^2	0.047	0.294	0.294	0.033	0.033	0.013	0.025

NOTES: The table shows the relationship between the presence of privacy-sensitive permissions and app demand on three different data sets: Columns 1-5 exploit our cross-section and panel data; Columns 6-7 show a difference-in-differences style setup between Google's Play Store and the iOS App Store. In Columns 1-5 the dependent variable is demand proxied by the log. number of monthly new ratings of an app. Columns 1-3 contain cross-section results. Column 1 shows the raw correlation between the use of privacy-sensitive permissions and demand. Column 2 adds in controls for the app's observed characteristics. Column 3 uses the number of privacy-sensitive permissions as privacy indicator. Columns 4-5 show panel fixed effects regressions for those apps within our data set that varied their use of privacy-sensitive permissions at least once between April and September 2012. We show the results for the presence of privacy-sensitive permissions (Column 4) and their number (Column 5). In Columns 6 and 7 the dependent variable is the difference between the app's download ranks on the iOS App Store vs. Google Play Store. In Column 6 this difference is based on the average download ranks of seven countries, whereas in Column 7 it is based only on US ranks. All specifications control for the number of unproblematic permissions ($\#CleanPerm.$). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B8: Moderating Factors of the Demand Side Relationship

Log. $\Delta Ratings$	Top Apps Low	Top Apps High	Privacy Policy No	Privacy Policy Yes	Maturity Level High	Maturity Level Low	Med.&Health Apps Yes	Med.&Health Apps No
$D_{Privacy}$	-0.068***	0.008	-0.070***	0.038	-0.109***	-0.054***	-0.151***	-0.060***
	(0.008)	(0.029)	(0.009)	(0.057)	(0.019)	(0.010)	(0.043)	(0.009)
$\#_{CleanPerm}$	0.029***	0.014***	0.033***	0.054***	0.028***	0.038***	0.029**	0.036***
	(0.002)	(0.005)	(0.003)	(0.010)	(0.008)	(0.003)	(0.014)	(0.003)
$D_{Internet}$	-0.122***	-0.320***	-0.192***	-0.254***	-0.108***	-0.211***	-0.106**	-0.203***
	(0.008)	(0.040)	(0.009)	(0.063)	(0.019)	(0.010)	(0.041)	(0.009)
D_{Ads}	0.188***	0.364***	0.229***	0.225***	0.292***	0.222***	0.180***	0.238***
	(0.008)	(0.033)	(0.009)	(0.056)	(0.025)	(0.010)	(0.042)	(0.009)
Log. Price	-0.048***	0.003	-0.070***	-0.114***	-0.050***	-0.076***	-0.056***	-0.072***
	(0.001)	(0.007)	(0.001)	(0.005)	(0.002)	(0.001)	(0.003)	(0.001)
D_{Other}	0.047***	0.122***	0.050***	0.067	0.116***	0.040***	0.057	0.053***
	(0.008)	(0.028)	(0.009)	(0.052)	(0.025)	(0.010)	(0.044)	(0.009)
Log. Length Desc.	0.202***	0.275***	0.252***	0.404***	0.237***	0.268***	0.257***	0.262***
	(0.003)	(0.016)	(0.004)	(0.027)	(0.011)	(0.004)	(0.018)	(0.004)
Log. Size (in KB)	0.044***	0.093***	0.061***	0.065***	0.044***	0.065***	0.064***	0.062***
	(0.002)	(0.009)	(0.002)	(0.015)	(0.005)	(0.002)	(0.010)	(0.002)
Number Screenshots	0.073***	0.071***	0.083***	0.076***	0.086***	0.084***	0.075***	0.085***
	(0.002)	(0.006)	(0.002)	(0.012)	(0.007)	(0.002)	(0.010)	(0.002)
Dummy: Video	0.151***	0.083**	0.181***	0.250***	0.072*	0.218***	0.024	0.208***
	(0.011)	(0.033)	(0.012)	(0.050)	(0.039)	(0.013)	(0.068)	(0.012)
Log. Average Rating	0.155***	2.242***	0.200***	0.192***	0.207***	0.204***	0.257***	0.196***
	(0.006)	(0.102)	(0.007)	(0.072)	(0.014)	(0.008)	(0.029)	(0.007)
Dummy: Top-Dev.	0.588***	0.333***	1.143***	0.614***	1.505***	1.104***	0.155	1.272***
	(0.063)	(0.070)	(0.066)	(0.130)	(0.212)	(0.060)	(0.126)	(0.063)
App Version	0.010***	-0.066***	0.028***	-0.012	0.018*	0.032***	0.069***	0.027***
	(0.003)	(0.015)	(0.004)	(0.026)	(0.010)	(0.004)	(0.020)	(0.004)
Log. AppsByDev	-0.097***	-0.085***	-0.112***	-0.173***	-0.077***	-0.122***	-0.100***	-0.114***
	(0.002)	(0.010)	(0.002)	(0.018)	(0.004)	(0.002)	(0.010)	(0.002)
Log. InstByDev	0.124***	0.099***	0.178***	0.289***	0.140***	0.192***	0.164***	0.184***
	(0.001)	(0.006)	(0.002)	(0.010)	(0.004)	(0.002)	(0.008)	(0.002)
Log. InstByComp	-0.022***	0.039***	-0.009***	0.013	-0.007*	-0.007***	0.009	-0.007***
	(0.002)	(0.008)	(0.002)	(0.011)	(0.004)	(0.002)	(0.011)	(0.002)
Log. PriceOfComp	-0.002**	0.024***	0.010***	0.031***	0.007***	0.011***	0.013***	0.011***
	(0.001)	(0.003)	(0.001)	(0.005)	(0.002)	(0.001)	(0.004)	(0.001)
Log. RatByComp	-0.273***	-0.181	-0.685***	-1.048***	-0.424***	-0.761***	-0.765**	-0.716***
	(0.055)	(0.199)	(0.061)	(0.358)	(0.134)	(0.068)	(0.304)	(0.062)
Min. Android Vers.	0.094***	0.371***	-0.012	-0.026	-0.022	-0.011	-0.076	-0.014
	(0.009)	(0.040)	(0.010)	(0.069)	(0.026)	(0.011)	(0.051)	(0.011)
Max. Android Vers.	0.159***	0.112	0.172***	0.028	0.187***	0.171***	0.146	0.171***
	(0.017)	(0.095)	(0.019)	(0.196)	(0.032)	(0.023)	(0.127)	(0.020)
Constant	-3.470***	-4.694***	-3.753***	-4.497***	-3.467***	-3.980***	-3.811***	-3.865***
	(0.106)	(0.499)	(0.122)	(0.976)	(0.247)	(0.140)	(0.669)	(0.125)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	164157	13036	170658	6535	27843	149350	6218	170975
Mean of dep. Var.	-0.14	3.16	0.06	1.11	-0.18	0.15	-0.08	0.11
SD of dep. Var.	1.27	1.54	1.52	2.02	1.38	1.58	1.38	1.56
Adjusted R^2	0.221	0.293	0.285	0.371	0.305	0.291	0.270	0.295

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data. App demand is measured by the log. number of monthly new ratings of an app. Columns 1 and 2 split the sample into apps which have a high or a low stock of installations (more or less than 75000 installations). Columns 3 and 4 split the sample into apps with and without a privacy policy. Columns 5 and 6 split them into apps which require a high (Column 5) or low (Column 6) maturity of the user (apps are defined as appropriate for low maturity if they classified as being recommended for 'everyone' or for 'low maturity'-users). Columns 7 and 8 split the sample into medical and health-related apps as well as into other apps. All specifications control for the number of unproblematic permissions (*#CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.4 Robustness Checks - Detailed Version

This section provides more detailed descriptions of our robustness checks. It fills in the descriptive details that we had to omit for the brief presentation in the paper's main body in order to save space.

B.4.1 Supply-side Analysis: Privacy Definition and Including Moderating Factors

In Table B9, we analyze the robustness of our supply-side findings to using alternative definitions of privacy, to including moderating factors and the use of alternative samples.

The table shows additional supply-side results for the developer's choice of their business model. The dependent variable is the developer's decision to offer their app for money or for free (D_{Paid}). All specifications, except those using the pairs data set, use the cross section data from April 2012. Column 1 uses as privacy measure an individual dummy for each number of privacy-sensitive permissions (1, 2 or 3 and more permissions). Column 2 uses a cross term equal to one for apps which simultaneously use sensitive permissions and have access to the internet. Columns 3 and 4 split the sample into normal apps (Col. 3) and games (Col. 4). Column 5 adds a cross term for apps with a very large total number of installations (10000 or more). Column 6 adds a cross term for apps that could be associated with a top-ranked website (on Alexa.com), i.e. for a website with rank lower than 10000. Columns 7 and 8 analyze the most restrictive set of matched pairs (no difference in description or only difference describes existence of ads, verified by human coders). In all specifications we drop outliers with respect to app prices, i.e. apps with prices above 8 Euros. All of these regressions control for the number of clean permissions and permissions that are needed to show ads.

In columns 1-4, we varied the specification in the supply-side regressions to verify (i) that our results do not depend on our privacy variable, and (ii) that they hold within subsamples of our data, such as games and normal apps. The results show that our main cross section results do not depend on the choice of the specific privacy measure. Apps that request more privacy-sensitive permissions are more likely to be free (column 1). Column 2 shows that developers strategically use privacy-sensitive permissions in free apps only together with internet access. This highlights that data collection (via permissions) is more valuable for developers if they can easily transfer the data from the app (via internet access). Columns 3 and 4 show that our baseline finding holds for both non-game apps (column 3) and games (column 4). For both groups, we find a negative significant effect.

Also, our supply-side results are robust to including moderating factors and to the use of a more rigorous matched set of app pairs: In columns 5-8, we analyze the role of moderating factors on the supply side and restrict the pairs data set even further to rule out unobserved heterogeneity as an explanation of our results. Columns 5 and 6 show that the role of reputation as a moderating factor is equally important for the supply side as for the demand side. In column 5 we separately analyze apps with a large user base, and in column 6 we analyze apps that are associated with a popular website (low traffic rank on Alexa.com). Such apps are generally less likely to be paid versions, but if they are paid, they are more likely to require privacy-sensitive permissions. Columns 7 and 8 analyze the robustness of our supply-side results, which we obtained based on the app pairs data set by applying a more restrictive matching for the pairs data: here we only consider pairs with no difference in description and code length, which was verified by human coders. These results show again that privacy-sensitive permissions are more likely in free apps, independently of how restrictive we are in our matching of the app pairs. Taken together, these results confirm our baseline findings for the supply side of the market.

B.4.2 Panel estimations are robust to using alternative samples and specifications:

Beyond the panel specifications in Table B7 we run additional fixed-effect panel regressions. We thus verfiy the robustness of our panel results to using alternative specifications. The results are shown in Table B10.

Our demand side panel data set consists of apps which we observed in each of the five waves between April and September 2012 and that varied their use of privacy-sensitive permissions at least once in this period. In columns 3 and 6 we restrict our sample in addition to apps which did not change the length of the app's description in the Google Play Store during this period. We interpret such a change in permission without a change in the app's description as an indication of a change in permissions which came without a change in functionality.

The dependent variable is the log. number of monthly new ratings of an app. Columns 1-3 analyze the effect of introducing any privacy-sensitive permissions (measured by the indicator, whereas Columns 4-6 use the number of privacy-sensitive permissions as the variable of interest. In all specifications we include monthly fixed-effects and control for the number of unproblematic permissions. We first analyzed the raw fixed effects regressions without controls (Columns 1 and 4), and then we added control variables and dummies to control for the apps' categories. (Columns 2 and 5). Finally we analyze our reduced sample of apps that introduced new permissions without changing their description (Columns 3 and 6).

The estimated coefficient of interest remains essentially unchanged when looking at the specification in column 1 where we include only a reduced set of controls, or when we restrict the data to apps that introduced permissions but no major update to functionality (column 3). Moreover, we see similar results when analyzing the number of sensitive permissions, rather than an indicator for their presence (cols. 4-6).

Table B9: Alternative Supply Side Results

D_{Paid}	Privacy Measures (1)	(2)	(3)	(4)	Non-Games vs Games (5)	(6)	Moderating Factors (7)	(8)	App Pairs (9)	(10)
$\#_{Privacy=1}$	-0.010***									
	(0.003)									
$\#_{Privacy=2}$	-0.058***									
	(0.003)									
$\#_{Privacy\geq 3}$	-0.111***									
	(0.004)									
$D_{Privacy}$		0.012*			-0.012***	-0.057***	-0.062***	-0.037***	-0.246***	
		(0.006)			(0.003)	(0.005)	(0.003)	(0.003)	(0.054)	
$D_{Privacy} \times D_{Internet}$		-0.055***								
		(0.007)								
$D_{PrivCatSpec}$			-0.102***							
			(0.003)							
$\#_{PrivCatSpec}$				-0.045***						
				(0.001)						
$D_{Privacy} \times D_{NumInst}$							0.144***			
							(0.003)			
$D_{NumInst}$							-0.322***			
							(0.003)			
$D_{Privacy} \times D_{AlexaRank}$								0.046**		
								(0.018)		
$D_{AlexaRank}$								-0.109***		
								(0.016)		
$\#_{Privacy}$										-0.199***
										(0.029)
$\#_{CleanPerm}$	0.009***	0.005***	0.006***	0.008***	0.003***	0.012***	0.005***	0.004***	0.031	0.051**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.024)	(0.025)
$D_{Internet}$	-0.223***	-0.209***	-0.216***	-0.220***	-0.206***	-0.221***	-0.209***	-0.217***	-0.392***	-0.422***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.003)	(0.003)	(0.047)	(0.047)
D_{Ads}	-0.122***	-0.117***	-0.120***	-0.120***	-0.127***	-0.139***	-0.106***	-0.114***	-0.522***	-0.531***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.051)	(0.048)
D_{Other}	0.013***	0.009***	0.013***	0.012***	0.004	-0.004	0.011***	0.014***	-0.023	-0.015
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.007)	(0.002)	(0.003)	(0.091)	(0.088)
Log. Length Desc.	0.054***	0.054***	0.053***	0.053***	0.058***	0.033***	0.061***	0.054***	0.006	0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.050)	(0.051)
Log. Size (in KB)	0.032***	0.032***	0.031***	0.030***	0.026***	0.045***	0.032***	0.030***	-0.014	-0.026
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.029)	(0.030)
Number Screenshots	0.011***	0.011***	0.011***	0.011***	0.012***	0.011***	0.012***	0.012***	0.029	0.028
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.020)	(0.019)
Dummy: Video	0.000	-0.003	-0.001	-0.003	0.027***	0.014**	0.006**	-0.001	-0.422**	-0.401**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.007)	(0.003)	(0.004)	(0.192)	(0.184)
Log. Average Rating	-0.038***	-0.037***	-0.036***	-0.037***	-0.050***	0.065***	-0.033***	-0.036***	0.122	0.118
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.003)	(0.003)	(0.079)	(0.073)
Dummy: Top-Dev.	0.184***	0.186***	0.177***	0.178***	0.118***	0.217***	0.223***	0.201***	0.000	0.000
	(0.014)	(0.015)	(0.014)	(0.014)	(0.016)	(0.025)	(0.014)	(0.015)	(.)	(.)
App Version	-0.048***	-0.048***	-0.049***	-0.049***	-0.048***	-0.065***	-0.040***	-0.044***	-0.041*	-0.022
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.024)	(0.025)
Log. AppsByDev	0.022***	0.022***	0.023***	0.023***	0.034***	-0.020***	0.016***	0.019***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(.)	(.)
Log. InstByDev	-0.019***	-0.018***	-0.018***	-0.019***	-0.025***	0.010***	-0.005***	-0.018***	0.062***	0.059***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.009)	(0.009)
Log. InstByComp	-0.017***	-0.018***	-0.017***	-0.018***	-0.016***	-0.021***	-0.015***	-0.018***	-0.003	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.013)	(0.012)
Log. PriceOfComp	0.017***	0.017***	0.017***	0.017***	0.017***	0.012***	0.017***	0.017***	-0.005	-0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.004)
Log. RatByComp	0.190***	0.190***	0.196***	0.195***	0.291***	0.004	0.103***	0.199***	0.628*	0.498
	(0.017)	(0.017)	(0.017)	(0.017)	(0.019)	(0.039)	(0.016)	(0.019)	(0.337)	(0.324)
Min. Android Vers.	-0.044***	-0.046***	-0.045***	-0.043***	-0.052***	-0.035***	-0.065***	-0.042***	-0.034	-0.074
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.003)	(0.003)	(0.167)	(0.163)
Max. Android Vers.	0.025***	0.027***	0.025***	0.027***	0.025***	0.023	0.025***	0.016**	0.000	0.000
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.017)	(0.006)	(0.007)	(.)	(.)
Constant	0.081**	0.085**	0.080**	0.087**	-0.067*	0.087	0.127***	0.133***	-0.411	-0.167
	(0.035)	(0.035)	(0.035)	(0.035)	(0.040)	(0.088)	(0.034)	(0.039)	(0.687)	(0.667)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176000	176000	176000	176000	145972	30028	176000	145126	708	708
Num. of Groups	0.27	0.27	0.27	0.27	0.28	0.23	0.27	0.27	0.50	0.50
Mean of dep. Var.	0.44	0.44	0.44	0.44	0.45	0.42	0.44	0.44	0.50	0.50
SD of dep. Var.	0.323	0.320	0.324	0.324	0.348	0.260	0.367	0.319	0.848	0.862

NOTES: The table shows additional supply-side results for the developer's choice of their business model. The dependent variable is the developer's decision to offer their app for money or for free (D_{Paid}). Column 1 uses as privacy measure an individual dummy for each number of privacy-sensitive permissions (1, 2 or 3 and more permissions). Column 2 uses a cross term equal to one for apps which simultaneously use sensitive permissions and have access to the internet. Columns 3 and 4 use our category-specfic privacy measures. Within a category we flag a privacy-sensitive permission as problematic only if paid apps of this category use this permission on average less often than the overall average paid app. Columns 5 and 6 split the sample into normal apps (Col. 5) and games (Col. 6). Column 7 adds a cross term for apps with a very large total number of installations (10000 or more). Column 8 adds a cross term for apps that could be associated with a top-ranked website (on Alexa.com), i.e. for a website with rank lower than 10000. Columns 9 and 10 analyze the most restrictive set of matched pairs (no difference in description or only difference describes existence of ads, verified by human coders). In all specifications we drop outliers with respect to app prices, i.e. apps with prices above 8 Euros. All of these regressions control for the number of clean permissions and permissions that are needed to show ads. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B10: Alternative Panel Demand Side Estimation Results

Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-0.070***	-0.059***	-0.063**			
	(0.020)	(0.020)	(0.027)			
$\#_{Privacy}$				-0.023***	-0.014**	-0.010
				(0.006)	(0.007)	(0.009)
$\#_{CleanPerm}$	-0.010**	-0.010**	-0.003	-0.000	-0.005	-0.001
	(0.004)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
$D_{Internet}$	-0.003	0.003	0.033	-0.021	-0.014	0.012
	(0.054)	(0.054)	(0.072)	(0.054)	(0.053)	(0.071)
D_{Ads}	0.007	0.002	-0.041	-0.011	-0.014	-0.058
	(0.033)	(0.033)	(0.043)	(0.032)	(0.033)	(0.043)
Log. Price	-0.044***	-0.044***	-0.037***	-0.044***	-0.043***	-0.037***
	(0.011)	(0.011)	(0.013)	(0.011)	(0.011)	(0.013)
D_{Other}		0.022			0.024	
		(0.022)			(0.022)	
Log. Length Desc.		0.003			0.002	
		(0.023)			(0.023)	
Log. Size (in KB)		0.044***			0.043***	
		(0.017)			(0.017)	
Number Screenshots		0.029**			0.027**	
		(0.011)			(0.011)	
Dummy: Video		-0.138**			-0.139**	
		(0.058)			(0.059)	
Log. Average Rating		0.267***			0.272***	
		(0.098)			(0.098)	
Dummy: Top-Dev.		-0.206***			-0.179***	
		(0.016)			(0.014)	
App Version		0.029*			0.032**	
		(0.015)			(0.015)	
Log. AppsByDev		-0.161***			-0.157***	
		(0.029)			(0.029)	
Log. InstByDev		-0.011			-0.011	
		(0.014)			(0.014)	
Log. InstByComp		0.004			0.004	
		(0.004)			(0.004)	
Log. PriceOfComp		-0.001			-0.001	
		(0.001)			(0.001)	
Log. RatByComp		-0.030			-0.031	
		(0.129)			(0.129)	
Min. Android Vers.		0.058			0.056	
		(0.044)			(0.044)	
Max. Android Vers.		-0.143			-0.137	
		(0.096)			(0.097)	
Constant	1.158***	1.198**	0.749***	1.156***	1.201**	0.741***
	(0.127)	(0.561)	(0.154)	(0.128)	(0.563)	(0.155)
Category	No	Yes	No	No	Yes	No
Month	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33095	33095	15220	33095	33095	15220
Num. of Groups	6619	6619	3044	6619	6619	3044
Mean of dep. Var.	1.63	1.63	1.17	1.63	1.63	1.17
SD of dep. Var.	2.23	2.23	2.10	2.23	2.23	2.10
Adjusted R^2	0.026	0.033	0.020	0.026	0.033	0.020

NOTES: This table shows the results from fixed-effect panel regressions. The dependent variable is the log. number of monthly new ratings of an app. We restrict our sample to apps within our data set that varied their use of privacy-sensitive permissions at least once between April and September 2012. Columns 1-3 analyze the effect of introducing any privacy-sensitive permissions (measured by the indicator $D_{Privacy}$), whereas Columns 4-6 use the number of privacy-sensitive permissions as the variable of interest. Columns 1 and 4 show the raw fixed effects regressions without controls. In Columns 2 and 5 we add control variables and dummies to control for the apps' categories. Columns 3 and 6 restrict the analysis to apps that introduced new permissions without changing the app's description (no change in the length of description). In all specifications we include monthly fixed-effects and control for the number of unproblematic permissions. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.4.3 Alternative Difference-in-Differences-Style Estimation Results:

Table A6 shows demand-side results from comparing the download ranks of an app in the iOS Appstore and Google's Play Store depending on whether its Android version uses privacy-sensitive permissions or not. It provides results which illustrate the robustness of our DiD-style results to the use of alternative rank measures and subsamples of the data.

Data Preparation and Additional Variables: To compare apps' relative success in the Android OS and iOS in dependence of their use of privacy-sensitive permissions, we collected in 2016 app rankings from 2012 for both OS from AppAnnie.com. We found only a small number of (mostly successful) apps for which the app ranking from 2012 was available for both OS. Unlike in our main data sets, which contain global success measures such as the world wide number of installations, rankings are only available at a country-specific basis. Thus, we collected rankings for April and September 2012 and for seven important markets: Germany, India, Japan, Korea, Russia, the UK and the US April and September 2012. For free apps we use the 'overall free app rank' and for paid apps we use the 'overall paid ranking'.[60] For the estimation sample we kept those apps for which we were able to collect at least four rankings per OS (out of the up to 14 rankings per OS).

Based on that information we construct four measures of the relative app success on the two operating systems ($\Delta Rank^{iOS-And}$):

- In column 1 of Table A6 the dependent variable is the simple difference between the iOS and the Android country ranks we were able to collect for each app, i.e. it is the difference between the average iOS rank (which is based on up to 14 country- and time-specific ranks) and the Android OS average rank.

- In column 2 the dependent variable is also a difference based on the average country ranks. However, here we construct for each OS based on the observed average ranks an in-sample ranking which ranks apps within the OS according to its average OS-specific rank. The difference between the iOS and the Android is then computed as the difference between the two self-generated in-sample ranks.[61] This procedure guarantees that the observed apps' ranks have a similar distribution in Apple's iOS and Google's Android OS.

- In column 3 the dependent variable is again a simple difference as in column 1 but is based only on the US ranks.

- In column 4 the dependent variable is also based only on US ranks but again uses the difference between self-generated rankings which we created as in column 2 for each OS.

[60] In addition to overall rankings there are also category-specific rankings such as rankings for games, weather apps, and education apps.

[61] Specifically: we compute average rankings by app and operating system based on the time- and country-specific ranks (up to seven countries and two points in time per OS, i.e. based on up to 14 ranks). Using this information we create a new in-sample ranking which ranks apps by the order of their average rank values and ranges from 1 to 192 for each operating system. The measure we use in the estimation is then the difference between the rank we have computed for the iOS version and the one we have computed for the Android version (i.e. iOS rank minus Android rank).

Columns 5 & 6 finally also use, as in column 2, the difference between the newly created ranks based on the average download ranks but split the sample into games (Column 6) and other apps (Column 5).

Results: The results show that the DiD-style comparison is robust to using alternative measures of the ranking difference: All specifications, except specification 6 which covers only games, show a significant negative effect for privacy-sensitive permissions. This corroborates the conclusion that apps that request privacy-sensitive permissions are on average less successful in the Android OS than in Apple's iOS, which could be due to the fact that only in the Android OS are these permissions visible to the user before installation of the app. These significant findings are all the more impressive given the low number of observations and the fact that our previous results suggest a lower effect for privacy-sensitive permissions of well-known apps.[62] Also, the insignificant effect for games is in line with our main results which show that a weaker or even an insignificant relationship between privacy-sensitive permissions and demand exists for games. Thus, the results reaffirm our baseline demand results as well as the results shown in our analysis of moderating factors despite using a completely different sample of apps and using a completely different identification approach.

Table B11: Alternative Difference-in-Differences-Style Estimation Results:

	Global Ranks		US Ranks		Non-Games vs Games	
	(1)	(2)	(3)	(4)	(5)	(6)
$D_{Privacy}$	-116.468***	-12.660**	-182.504***	-14.988**	-25.368**	1.533
	(35.327)	(6.088)	(58.130)	(6.321)	(10.316)	(8.488)
$\#_{CleanPerm}$	0.266	0.849	0.344	1.011*	1.498**	-1.310
	(3.473)	(0.598)	(5.476)	(0.595)	(0.694)	(1.263)
Constant	264.070***	4.262	320.040***	2.428	10.214	6.258
	(29.120)	(5.018)	(48.089)	(5.229)	(9.514)	(6.209)
Observations	192	192	162	162	96	96
Mean of dep. Var.	175.94	-0.24	176.85	-3.22	-1.07	0.59
SD of dep. Var.	188.76	31.81	274.05	29.27	32.96	30.78
Adjusted R^2	0.056	0.013	0.059	0.025	0.062	-0.006

NOTES: This table shows further demand side results from comparing the download ranks of an app in the iOS Appstore and Google's Play Store depending on whether it's Android version uses privacy-sensitive permissions or not. The dependent variable captures differences in the ranks on the two platforms (iOS App Store ranks minus Google Play Store ranks). Colums 1, 2, 5 and 6 compare the apps' average ranks of the seven countries we collected ranks for, whereas Columns 3 & 4 use only the US ranks. Columns 1 & 3 use as dependent variable the difference between the simple rank averages whereas Columns 2 & 4 use as the dependent variable the difference between the newly created ranks within the operating system, which are based on the average download ranks we observed on AppAnnie.com. Columns 5 & 6 also use the difference between the newly created ranks based on the average download ranks but contrast games (Column 6) with other apps (Column 5). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

[62] Such apps are clearly overrepresented in the small sample of apps for which we were able to retrieve ranks in 2016.

B.4.4 Robustness of Demand-side Results to Using Alternative Demand Measures

It is crucial for the validity of our research to analyze whether our main findings for the demand-side depend on the definition of our demand-measure. To verify that this is not the case, we varied the dependent variable, and run our main specification with eight alternative demand measures and two alternative measures of app popularity. The results are shown in Table B12, where we show our main specification with alternative dependent variables based on actual installations, alternative measures based on ratings, and when using predicted installations that we calibrated using the available information on ratings and installations. Verifying our results for predicted installations highlights that our findings are in line with the previous literature that has used predicted installations.

Data Preparation and Additional Variables: We estimate our main specification with eight alternative demand measures based on ratings (cols. 1&2), the direct measure of installations (cols. 3-5), and three measures of predicted monthly new installations (cols. 6-8). In addition we provide results with respect to apps' average ratings (cols. 9&10). All specifications use the cross section data from April 2012. In the following we describe the various different demand measures we use in Table B12.

- In column 1 the dependent variable is ΔR_{AprSep} which equals the change in the number of ratings between April and September 2012.
- In column 2 the dependent variable is ΔR_{1214} which equals the change in the number of ratings between 2012 and 2014.
- In column 3 the dependent variable is ΔI_{Apr} which equals the change in the number of installations in April 2012.
- In column 4 the dependent variable is ΔI_{AprSep} which equals the change in the number of installations between April and September 2012.
- In column 5 the dependent variable is ΔI_{1214} which equals the change in the number of installations between 2012 and 2014.
- In column 6 the dependent variable equals the predicted change in the number of installations in April 2012. We predict this change based on a log-log-specified estimation specification where the dependent variable is the log. number of new installations and the explanatory variable is the log. number of new ratings an app has received in April 2012 (see Column 2 of Table A12).
- In column 7 the dependent variable equals the predicted change in the number of installations between April and September 2012. We predict this change based on a log-log estimation specification where the dependent variable is the log. number of new installations and the explanatory variable is the log. number of new ratings an app has received between April and September 2012 (see Column 4 of Table A12).
- In column 8 the dependent variable equals the predicted change in the number of installations in April 2012. We predict this change based on a non-parametric estimation specification

where the dependent variable is the log. number of new installations and the explanatory variable is the log. number of new ratings an app has received in April 2012 (see Column 8 of Table A12).

- In column 9 the dependent variable equals AR_{Apr} which is the average of the ratings the app has received in April 2012.

- In column 10 the dependent variable equals AR_{AprSep} which is the average of the ratings the app has received between April and September 2012.

Results: Our main demand-side results remain the same, independently of whether we use measures based on installations or measures based on ratings as well as whether we consider a longer time window of ratings- or installation growth (cols. 1-5). Similarly, when we use predicted changes in the number of installations (cols. 6-8), our results are also confirmed. Finally, the results in Columns 9 and 10 indicate that apps receive worse ratings if they use privacy-sensitive permissions and thus seem to be less popular among users.

Table B12: Alternative Demand Measures

	Ratings-Based		Installations-Based			Predicted New Installations			Average Rating	
	ΔR_{AprSep}	ΔR_{1214}	ΔI_{Apr}	ΔI_{AprSep}	ΔI_{1214}	param.	param. 1214	non-param.	AR_{Apr}	AR_{AprSep}
$D_{Privacy}$	-0.093***	-0.070***	-0.054***	-0.197***	-0.388***	-0.030***	-0.053***	-0.183***	-0.022***	-0.018***
	(0.011)	(0.015)	(0.016)	(0.025)	(0.036)	(0.004)	(0.007)	(0.024)	(0.003)	(0.003)
$\#_{CleanPerm}$	0.042***	0.035***	0.010***	0.014**	-0.010	0.017***	0.030***	0.069***	0.000	0.000
	(0.003)	(0.004)	(0.004)	(0.006)	(0.008)	(0.001)	(0.002)	(0.006)	(0.001)	(0.001)
$D_{Internet}$	-0.220***	-0.250***	-0.078***	-0.206***	-0.317***	-0.092***	-0.163***	-0.416***	-0.054***	-0.056***
	(0.012)	(0.016)	(0.015)	(0.025)	(0.037)	(0.004)	(0.007)	(0.026)	(0.004)	(0.004)
D_{Ads}	0.326***	0.356***	0.248***	0.618***	0.925***	0.109***	0.192***	0.613***	0.009**	0.014***
	(0.011)	(0.016)	(0.015)	(0.024)	(0.037)	(0.004)	(0.007)	(0.025)	(0.004)	(0.003)
Log. Price	-0.114***	-0.173***	-0.038***	-0.113***	-0.307***	-0.033***	-0.058***	-0.201***	-0.002***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
D_{Other}	0.071***	0.058***	0.062***	0.165***	0.168***	0.024***	0.042***	0.155***	0.006*	-0.003
	(0.011)	(0.016)	(0.015)	(0.025)	(0.035)	(0.004)	(0.008)	(0.024)	(0.003)	(0.003)
Log. Length Desc.	0.373***	0.443***	0.157***	0.377***	0.564***	0.121***	0.214***	0.690***	0.006***	0.006***
	(0.005)	(0.007)	(0.007)	(0.011)	(0.016)	(0.002)	(0.003)	(0.011)	(0.002)	(0.001)
Log. Size (in KB)	0.092***	0.104***	0.029***	0.077***	0.160***	0.029***	0.051***	0.146***	0.009***	0.012***
	(0.003)	(0.004)	(0.004)	(0.006)	(0.009)	(0.001)	(0.002)	(0.006)	(0.001)	(0.001)
Number Screenshots	0.113***	0.142***	0.058***	0.142***	0.260***	0.039***	0.069***	0.204***	0.012***	0.013***
	(0.003)	(0.004)	(0.004)	(0.006)	(0.008)	(0.001)	(0.002)	(0.006)	(0.001)	(0.001)
Dummy: Video	0.212***	0.302***	0.019	0.007	0.131***	0.094***	0.166***	0.408***	0.014***	0.012***
	(0.014)	(0.020)	(0.022)	(0.033)	(0.047)	(0.006)	(0.010)	(0.030)	(0.004)	(0.003)
Log. Average Rating	0.272***	0.421***	0.168***	0.410***	0.886***	0.092***	0.162***	0.511***		
	(0.010)	(0.016)	(0.011)	(0.020)	(0.035)	(0.003)	(0.006)	(0.024)		
Dummy: Top-Dev.	1.263***	1.668***	0.124	-0.066	0.262	0.525***	0.929***	1.733***	0.007	-0.013
	(0.062)	(0.084)	(0.102)	(0.151)	(0.235)	(0.027)	(0.047)	(0.110)	(0.009)	(0.009)
App Version	0.045***	0.062***	-0.036***	-0.050***	-0.030*	0.013***	0.023***	0.075***	0.004**	0.004***
	(0.005)	(0.007)	(0.006)	(0.010)	(0.016)	(0.002)	(0.003)	(0.010)	(0.002)	(0.001)
Log. AppsByDev	-0.182***	-0.211***	-0.066***	-0.154***	-0.173***	-0.052***	-0.093***	-0.346***	-0.007***	-0.001
	(0.002)	(0.004)	(0.003)	(0.005)	(0.010)	(0.001)	(0.002)	(0.005)	(0.001)	(0.001)
Log. InstByDev	0.258***	0.311***	0.043***	0.111***	0.182***	0.085***	0.150***	0.476***	0.015***	0.015***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.006)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)
Log. InstByComp	-0.025***	-0.041***	-0.007**	-0.046***	-0.073***	-0.003***	-0.006***	-0.059***	-0.007***	-0.005***
	(0.002)	(0.003)	(0.003)	(0.005)	(0.008)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)
Log. PriceOfComp	0.007***	0.003**	-0.006***	-0.012***	-0.020***	0.005***	0.009***	0.006***	-0.001***	-0.001***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
Log. RatByComp	-1.118***	-1.451***	-0.106	-0.375**	-0.632**	-0.326***	-0.577***	-2.445***	0.449***	0.436***
	(0.076)	(0.108)	(0.103)	(0.172)	(0.258)	(0.028)	(0.050)	(0.169)	(0.024)	(0.020)
Min. Android Vers.	-0.022*	-0.022	0.294***	0.751***	1.316***	-0.007	-0.012	0.018	-0.000	0.004
	(0.013)	(0.019)	(0.018)	(0.030)	(0.043)	(0.005)	(0.009)	(0.029)	(0.005)	(0.004)
Max. Android Vers.	0.302***	0.636***	0.165***	0.508***	1.629***	0.078***	0.139***	0.529***	0.020**	0.006
	(0.026)	(0.039)	(0.029)	(0.048)	(0.080)	(0.009)	(0.016)	(0.057)	(0.010)	(0.008)
Constant	-4.583***	-5.540***	-3.557***	-6.936***	-13.206***	-2.248***	-3.083***	-6.608***	0.354***	0.360***
	(0.156)	(0.229)	(0.196)	(0.324)	(0.505)	(0.056)	(0.100)	(0.345)	(0.055)	(0.046)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	176212	124482	177193	177193	126987	177193	177193	177193	68248	102594
Mean of dep. Var.	0.97	2.40	-0.42	0.86	4.63	-0.42	0.16	2.76	1.29	1.29
SD of dep. Var.	2.03	2.52	2.34	3.89	5.16	0.72	1.27	4.25	0.35	0.35
Adjusted R^2	0.360	0.388	0.037	0.086	0.201	0.294	0.294	0.272	0.043	0.052

Notes: The table shows descriptive regressions analyzing the relationship between privacy-sensitive permissions and app demand by using various different demand and app popularity measures. Columns 1 and 2 use alternative demand measures based on ratings, Columns 3-5 use demand measures based on installations, whereas Columns 6-8 use three measures of predicted new installations which we estimated based on the information about the number of new ratings. Columns 9 and 10 use the average ratings of apps as a measure of app popularity. In Column 1 (2) the dependent variable is the log. number of new ratings between April and September 2012 (log. number of new ratings between 2012 and 2014). In Column 3 the dependent variable is the log. number of new installations in April 2012. Columns 4 and 5 use the log. number of new installations between April and September 2012 (Column 4) and between 2012 and 2014 (Column 5) as demand measures. In Columns 6-8 we apply three measures of predicted download numbers. For each of the measures we exploit the cross-section information on changes in ratings to predict changes in installation numbers. In Column 6, we use a measure of predicted monthly installation changes in April 2012 which is based on the observed change in the number of ratings in this month (see Column 2 of Table A12). In Column 7, we use a measure of predicted installation changes between April and September 2012 which is based on the observed change in the number of ratings in this period (see Column 4 of Table A12). In Column 8, we again use a measure of predicted monthly installation changes in April 2012 which is based on the observed change in the number of ratings in this month (see Column 2 of Table A12), but instead of employing a parametric log-log-specification to the data, we employ a non-parametric approach to it. In Columns 9 the dependent variable is the log. average of the ratings the app has received in April 2012, whereas in Column 10 it is the log. average of the ratings the app has received between April and September 2012. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.4.5 Robustness of Demand-side results to Using Alternative Privacy Measures

We verified that our results do not depend on our definition of privacy-sensitive permissions. Clearly, such a dependency would cast doubt on the generalizeability of our findings. To test the robustness of our findings to using alternative implementations, we estimate our main specification with seven alternative privacy measures.

Data Preparation and Additional Variables: All specifications use the cross section data from April 2012. In the following we describe the various different privacy measures we use in Table B13.

1. We generated indicators for each number of permissions. This results in three dummies for 1, 2, and 3 or more permissions respectively (Column 1).

2. We asked 450 microworkers on Amazon's mechanical turk to classify permissions as to whether they thought they were neutral, problematic or very problematic. From these classifications we generated a dummy variable which is equal to one if an app uses at least one privacy-sensitive permission which was classified as problematic by the microworkers (Column 2).

3. We identified privacy-sensitive permissions that are unusual for the app's category. We flag a privacy-sensitive permission as category-specific problematic if paid apps of the category use this permission less frequently than the average paid app across all categories (Column 3).

4. We used additional data from *privacygrade.org* by Lin, Hong, and Sadeh. (2014), which evaluated apps' intrusiveness in 2014.[63] We created a dummy which equals 1 if an app got a rating that indicated the app was privacy-intrusive in 2014 (ratings equal to 'B', 'C' or 'D'). (Column 4).

5. We introduce and estimate a crossterm that indicates apps, which used both at least one privacy-sensitive permission and internet access. We use this specification to test if users distinguish sensitive permissions that come with internet access from sensitive permissions that come without the ability to transmit the sensitive data (Columns 5 & 6).

6. We disaggregated the privacy sensitive permissions into functionality related types of permissions, distinguishing location-, communication-, user ID-, or profile-specific permissions (Column 7).

7. We used an alternative definition of privacy-sensitive permissions from previous research by Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy (2012). This definition classified only 12 permissions as privacy-sensitive, and is thus more restrictive (Column 8).[64]

We then proceeded to run our main specification with these alternative measures/definitions of a "privacy-sensitive" app.

[63]To provide this measure a group of researchers evaluated all apps with respect to how they privacy-sensitive data Lin, Hong, and Sadeh. (2014).

[64]These permissions include: read phone state and ID, coarse location, fine gps location, intercept outgoing calls, read sms or mms, receive sms, receive mms, record audio, receive wap, read contact data, read browser data, read sensitive log data.

Results: The results are shown in Table B13. The dependent variable is demand for the app measured by log. number of monthly new ratings of an app. The coefficient of interest analyzes the relationship between an app's demand and our measures of privacy-sensitive permissions. Each column presents the results obtained when using an alternative measure for the presence of privacy-sensitive permissions. The results show that the weakly negative relationship between permissions and downloads holds across almost all definitions we consider. To be precise, the coefficient that measures the relationship between app-demand and the use privacy-sensitive permissions is negative for almost all definitions. The only exception is the *privacygrade* measure, which was published in 2014 on *privacygrade.org*. In contrast, the number of clean/unproblematic permissions is positively associated with demand in most specifications, which indicates that such permissions do not face lower demand.

Table B13: Alternative Privacy Measures

Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#_{Privacy=1}$	-0.054***							
	(0.010)							
$\#_{Privacy=2}$	-0.087***							
	(0.014)							
$\#_{Privacy\geq3}$	-0.089***							
	(0.014)							
$D_{Privacy}$		-0.060***	-0.048***	-0.040***		-0.086***		
		(0.009)	(0.010)	(0.014)		(0.018)		
$D_{MTurkEP2}$		-0.021*						
		(0.013)						
$D_{PrivCatSpec}$			-0.048***					
			(0.012)					
D_{PGrade}				0.148***				
				(0.020)				
$D_{Privacy} \times D_{Internet}$					-0.057***	0.025		
					(0.010)	(0.019)		
D_{ID}							0.046***	
							(0.009)	
$D_{Location}$							-0.200***	
							(0.014)	
$D_{Communication}$							-0.015	
							(0.016)	
$D_{Profile}$							-0.034***	
							(0.011)	
$D_{Sarmaetal}$								-0.048***
								(0.009)
$\#_{CleanPerm}$	0.038***	0.037***	0.037***	0.048***	0.036***	0.037***	0.038***	0.036***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
$D_{Internet}$	-0.202***	-0.201***	-0.199***	-0.288***	-0.192***	-0.204***	-0.201***	-0.202***
	(0.009)	(0.009)	(0.009)	(0.015)	(0.009)	(0.010)	(0.009)	(0.009)
D_{Ads}	0.234***	0.234***	0.235***	0.260***	0.236***	0.235***	0.219***	0.235***
	(0.009)	(0.009)	(0.009)	(0.014)	(0.009)	(0.009)	(0.009)	(0.009)
Log. Price	-0.071***	-0.071***	-0.071***	-0.094***	-0.071***	-0.071***	-0.072***	-0.071***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
D_{Other}	0.053***	0.052***	0.053***	0.017	0.048***	0.052***	0.048***	0.047***
	(0.009)	(0.009)	(0.009)	(0.014)	(0.009)	(0.009)	(0.009)	(0.009)
Log. Length Desc.	0.262***	0.263***	0.262***	0.287***	0.262***	0.263***	0.261***	0.262***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Log. Size (in KB)	0.063***	0.063***	0.062***	0.087***	0.063***	0.062***	0.061***	0.062***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Number Screenshots	0.085***	0.085***	0.085***	0.105***	0.085***	0.085***	0.086***	0.085***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Dummy: Video	0.205***	0.205***	0.205***	0.256***	0.203***	0.204***	0.208***	0.203***
	(0.012)	(0.012)	(0.012)	(0.020)	(0.012)	(0.012)	(0.012)	(0.012)
Log. Average Rating	0.199***	0.198***	0.199***	0.456***	0.199***	0.198***	0.201***	0.199***
	(0.007)	(0.007)	(0.007)	(0.015)	(0.007)	(0.007)	(0.007)	(0.007)
Dummy: Top-Dev.	1.140***	1.139***	1.135***	1.121***	1.139***	1.139***	1.128***	1.138***
	(0.058)	(0.058)	(0.058)	(0.099)	(0.058)	(0.058)	(0.058)	(0.058)
App Version	0.029***	0.028***	0.028***	0.050***	0.028***	0.029***	0.027***	0.028***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Log. AppsByDev	-0.113***	-0.114***	-0.113***	-0.145***	-0.114***	-0.114***	-0.113***	-0.114***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByDev	0.183***	0.184***	0.184***	0.245***	0.184***	0.184***	0.182***	0.184***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByComp	-0.007***	-0.007***	-0.007***	-0.001	-0.007***	-0.007***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Log. PriceOfComp	0.011***	0.011***	0.011***	0.016***	0.011***	0.011***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log. RatByComp	-0.707***	-0.707***	-0.704***	-0.833***	-0.705***	-0.708***	-0.697***	-0.704***
	(0.061)	(0.061)	(0.061)	(0.097)	(0.061)	(0.061)	(0.061)	(0.061)
Min. Android Vers.	-0.014	-0.015	-0.015	-0.036**	-0.016	-0.014	-0.013	-0.015
	(0.010)	(0.010)	(0.010)	(0.016)	(0.010)	(0.010)	(0.010)	(0.010)
Max. Android Vers.	0.170***	0.170***	0.170***	-0.127*	0.171***	0.170***	0.174***	0.170***
	(0.020)	(0.020)	(0.020)	(0.071)	(0.020)	(0.020)	(0.020)	(0.020)
Constant	-3.878***	-3.872***	-3.880***	-3.785***	-3.878***	-3.867***	-3.914***	-3.870***
	(0.122)	(0.122)	(0.123)	(0.322)	(0.123)	(0.123)	(0.122)	(0.122)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177193	177193	177193	86834	177193	177193	177193	177193
Mean of dep. Var.	0.10	0.10	0.10	0.51	0.10	0.10	0.10	0.10
SD of dep. Var.	1.55	1.55	1.55	1.71	1.55	1.55	1.55	1.55
Adjusted R^2	0.294	0.294	0.294	0.288	0.294	0.294	0.295	0.294

Notes: The table shows descriptive regressions analyzing the relationship between the presence of privacy sensitive permissions and app demand. The dependent variable is demand for the app measured by log. number of monthly new ratings of an app. The coefficient of interest analyzes the relationship between an app's demand and our privacy measures. Each column presents the results obtained when using an alternative privacy measure. Column 1 introduces an indicator for each number of permission (having 1, 2, and 3 or more permissions). Column 2 uses a dummy variable which is equal to one if an app uses at least one privacy-sensitive permission which was classified as very problematic by 450 microworkers we surveyed on Amazon's mechanical turk. In Column 3 we look at privacy sensitive permissions that are unusual for the app's category. Within a category we flag a privacy-sensitive permission as problematic only if paid apps of this category use this permission on average less often than the overall average paid app. Column 4 uses the 'privacygrade' by Lin, Hong, and Sadeh. (2014), that was made available on *privacygrade.org* in 2014. The dummy is equal to one if the the app got a rating equal to 'B', 'C' or 'D', i.e. a rating indicating the app being privacy-intrusive. In Columns 5 & 6 we introduce a crossterm that is equal to one if an app uses both at least one privacy-sensitive permission and has internet access. Column 7 disaggregates the privacy sensitive permissions into functionality-related types of permissions. Column 8 uses an alternative definition of privacy-sensitive permissions from previous research by Sarma, Li, Gates, Potharaju, Nita-Rotaru, and Molloy (2012), which defines only 12 permissions as privacy-sensitive. Heteroscedasticity-consistent standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.4.6 Robustness of Demand-side Results to using Alternative Estimation Techniques

Demand side results are robust to accounting for censoring, selection, network effects, or endogeneity: In Table B14 we analyze the robustness of our main demand-side results to the use of alternative estimation strategies that account for several potential endogeneity concerns. Specifically we check the sensitivity of our results with respect to censoring in the dependent variable, survivor bias, network effects, and strategic behavior of the developers when choosing requested permissions and prices. All of these checks require different estimation strategies that are discussed in the following.

Data Preparation, Additional Variables, and Estimation: In all specifications, the dependent variable is our main measure of app demand, i.e. the log. number of monthly new ratings, and the main variable of interest is a dummy that indicates the presence of privacy-sensitive permissions. All columns are based on the cross section from April 2012.

Columns 1 & 2 show Tobit-regressions that account for the fact that the dependent variable might be censored, especially might be left-censored at demand equal to zero. Column 1 sets the left-censoring limit to zero new ratings, whereas in Column 2 in addition a right-censoring limit equal to five is added.

Columns 3 & 4 contain results from Heckman selection models which aim to control for survivor bias, i.e. for the fact that apps using privacy-sensitive permissions might have lower demand and because of that might also have lower survival rates which would result in biased OLS estimates (underestimating the true effect of permissions) if these are based on a sample of surviving apps (as it is the case in our baseline demand estimates). In both heckman selection specifications (cols. 3 & 4) the regression equation is identical to our baseline cross-section demand specification and the selection equation models app survival. In column 3 survival is modeled by comparing apps which are observed throughout the period April to September 2012 to those which are observed in April, May and June, but which are not observed in the last two monthly waves (which we consider as an indication of their drop out). In column 4, survival is modeled by comparing apps within our baseline cross-section from April 2012 which survive until 2014 to those which are observed in April 2012 but are not observed in 2014. In both selection models we apply Heckman's (1979) two-step consistent estimator and use the information on code size as a selection variable, i.e. we include the code size only in the selection equation but not in the regression equation. In Table B14 we show only the results for the regression equation and not those for the selection equation but show both results together in Table B15. In addition, in Table B15 we also provide a detailed comparison of the reference OLS estimates for both specifications. Columns 2 &4 in this table shows the full Heckman (1979) specification, and tables 1&3 show the OLS regression that mirrors the second step of the Heckman's procedure.

In Column 5 we control for the existing user-base, which could affect demand through the existence of network effects, by including a control for the stock of existing installations (i.e. the log. number of installations).

In the following columns we estimate 2SLS models to instrument the variables of interest and to account for the endogeneity of the developers' privacy model choices. In Column 6 we instrument the privacy-dummy by the share of competing apps which use privacy-sensitve permissions

($ShareComp_{Privacy}$). Competing apps are those which are identified by Google as those which "users who viewed this [app] also viewed". In Column 7 we instrument the privacy-dummy by the share of the developer's other apps which use privacy-sensitive permissions (excluding the focal app) ($ShareDev_{Privacy}$). In Columns 8 and 9 we instrument the app price by using in both specifications two potential cost shifters: the log. code size and the log. number of apps a developer offers in the Google Play Store. In Column 8 we use the full cross-section, whereas in Column 9 we use only the sample of paid apps. Again, in Table B14 we only show the main results of these specifications, i.e. the second stage results, but provide for all four 2SLS specifications in Table B16 also the related first stage results.

Results: The results of all specifications support our baseline findings. The results in the first four columns suggest that neither accounting for censoring in the dependent variable (Col. 1-2), nor accounting for survivor bias (Col. 3-4) result in drastic changes of our main estimates. If anything, these specifications suggest a higher effect size, that is, these results suggest that privacy-sensitive permissions come with a stronger demand reduction than the baseline specifications. The Heckman selection models indeed support the idea that privacy-sensitive permissions come with a lower survival rate and thus with a bias of the OLS estimates towards zero (an underestimation of the true effect). Similar, controlling for past success and potential network effects (Col. 5) does not change our baseline conclusions drastically. The IV-estimations in Columns 6-9 show that our attempts to account for endogeneity in prices and permissions result in larger coefficient estimates, as would be expected if developers of better apps were to charge higher prices or ask for more permissions. In all four IV-specifications, in the first stage, our IV variables are highly significant and can thus be considered relevant variables.

B.4.7 Robustness of Results to Splitting the Sample

In this robustness check we verify that the negative relationship between demand and privacy-sensitive permissions is not driven by a specific type of app. We do so by splitting the sample into groups of apps where privacy could matter differently. This rules out an important alternative explanation, and highlights that the phenomenon we highlight affects the entire market.

Data Preparation and Additional Variables: To analyze the robustness of the demand-side results across different subsamples, we split our cross section data from April 2012 into groups of apps where the role of privacy could be of varying importance and estimate the main specification from Table B7 for each group separately. We divide the data along four dimensions: (i) pricing strategy (Col. 1-2), (ii) Game or normal app (Col. 3-4), (iii) user groups by maturity requirements (Col. 5-6), and (iv) by the presence or absence of health-relevant content (Col. 7-8). More precisely, in column 1 we consider free apps whereas in column 2 we consider paid apps. In column 3 we analyze non-game apps whereas in column 4 we consider games. In column 5 we restrict the sample to apps which are for users of higher maturity (age), whereas column 6 is based on apps for users recommended for 'everyone' or for 'low maturity'-users. Finally, in column 7 we consider apps from the categories 'health'- and 'medical'-apps whereas in column 8 we use the remaining apps.

Table B14: Alternative Estimation Specifications

	Tobit		Heckman		Netw.Eff.	IV-Privacy		IV-Price	
Log. $\Delta Ratings$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$D_{Privacy}$	-0.094***	-0.095***	-0.063***	-0.164***	-0.069***	-0.918***	-0.372***	-0.091***	-0.063***
	(0.017)	(0.017)	(0.009)	(0.024)	(0.007)	(0.049)	(0.017)	(0.009)	(0.017)
Log. Installations					0.417***				
					(0.002)				
$\#_{CleanPerm}$	0.055***	0.057***	0.038***	0.019***	0.026***	0.065***	0.044***	0.043***	0.082***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.002)	(0.004)	(0.003)	(0.003)	(0.007)
$D_{Internet}$	-0.325***	-0.333***	-0.184***	-0.423***	-0.168***	-0.099***	-0.177***	-0.289***	-0.005
	(0.020)	(0.021)	(0.010)	(0.030)	(0.007)	(0.012)	(0.011)	(0.014)	(0.017)
D_{Ads}	0.436***	0.443***	0.234***	0.291***	0.203***	0.243***	0.253***	0.198***	0.043**
	(0.018)	(0.018)	(0.010)	(0.023)	(0.007)	(0.011)	(0.011)	(0.011)	(0.020)
Log. Price	-0.172***	-0.174***	-0.068***	-0.157***	0.040***	-0.074***	-0.074***	-0.108***	-0.687***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.094)
D_{Other}	0.098***	0.098***	0.063***	0.058***	0.046***	0.182***	0.115***	0.079***	0.117***
	(0.017)	(0.017)	(0.009)	(0.020)	(0.007)	(0.014)	(0.011)	(0.010)	(0.020)
Log. Length Desc.	0.521***	0.529***	0.272***	0.503***	0.094***	0.290***	0.253***	0.298***	0.219***
	(0.008)	(0.008)	(0.004)	(0.010)	(0.003)	(0.005)	(0.004)	(0.005)	(0.012)
Log. Size (in KB)	0.120***	0.122***			0.044***	0.069***	0.064***		
	(0.005)	(0.005)			(0.002)	(0.003)	(0.002)		
Number Screenshots	0.152***	0.155***	0.095***	0.183***	0.068***	0.077***	0.075***	0.106***	0.070***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Dummy: Video	0.291***	0.299***	0.230***	0.361***	0.108***	0.235***	0.211***	0.221***	0.283***
	(0.021)	(0.021)	(0.011)	(0.027)	(0.010)	(0.014)	(0.013)	(0.012)	(0.020)
Log. Average Rating	0.656***	0.667***	0.215***	0.444***	0.289***	0.188***	0.204***	0.211***	0.079***
	(0.023)	(0.023)	(0.010)	(0.024)	(0.007)	(0.009)	(0.008)	(0.008)	(0.009)
Dummy: Top-Dev.	1.186***	1.291***	1.192***	1.544***	0.499***	1.201***	1.038***	1.242***	1.490***
	(0.075)	(0.082)	(0.044)	(0.108)	(0.043)	(0.064)	(0.058)	(0.060)	(0.094)
App Version	0.069***	0.070***	0.027***	0.038***	-0.026***	0.038***	0.023***	-0.004	0.016**
	(0.008)	(0.008)	(0.004)	(0.009)	(0.003)	(0.004)	(0.004)	(0.004)	(0.007)
Log. AppsByDev	-0.332***	-0.337***	-0.115***	-0.459***	-0.025***	-0.099***	-0.093***		
	(0.005)	(0.005)	(0.003)	(0.025)	(0.002)	(0.002)	(0.002)		
Log. InstByDev	0.406***	0.413***	0.181***	0.317***	0.029***	0.190***	0.206***	0.178***	0.105***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByComp	-0.050***	-0.051***	-0.008***	-0.055***	0.015***	-0.000	-0.003	-0.012***	-0.019***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Log. PriceOfComp	0.008***	0.009***	0.011***	0.007***	0.010***	0.011***	0.015***	0.018***	0.012***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Log. RatByComp	-2.002***	-2.017***	-0.743***	-1.684***	0.560***	-0.799***	-0.613***	-0.750***	-0.380***
	(0.122)	(0.124)	(0.063)	(0.151)	(0.050)	(0.063)	(0.067)	(0.063)	(0.111)
Min. Android Vers.	0.047**	0.045**	0.007	-0.145***	0.313***	-0.022*	0.014	0.039***	0.067***
	(0.022)	(0.022)	(0.010)	(0.030)	(0.009)	(0.012)	(0.012)	(0.011)	(0.018)
Max. Android Vers.	0.409***	0.416***	0.175***	0.830***	0.158***	0.187***	0.182***	0.152***	0.177***
	(0.046)	(0.047)	(0.022)	(0.056)	(0.016)	(0.024)	(0.024)	(0.020)	(0.034)
Constant	-8.802***	-8.975***	-3.533***	-5.781***	-6.472***	-4.141***	-4.305***	-3.970***	-3.300***
	(0.266)	(0.271)	(0.127)	(0.319)	(0.100)	(0.136)	(0.141)	(0.128)	(0.197)
Category	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Observations	177193	177193	185533	177193	177193	136040	137146	177193	48272
Mean of dep. Var.	0.10	0.10	0.10	2.40	0.10	0.10	0.10	0.10	-0.50
SD of dep. Var.	1.55	1.55	1.55	2.52	1.55	1.55	1.56	1.55	1.06
Adjusted R^2					0.534	0.256	0.324	0.269	0.047

NOTES: The table analyzes the robustness of our main demand-side results to using alternative estimation strategies. The dependent variable is app demand measured by the log. number of monthly new ratings, and the main variable of interest is a dummy that indicates the presence of privacy-sensitive permissions. All columns show cross section results. Columns 1 & 2 show Tobit-regressions that account for the fact that the dependent variable might be censored, especially might be left-censored at demand equal to 0. Column 1 sets the left-censoring limit to 0 new ratings, whereas in Column 2 in addition a right-censoring limit equal to 5 is set. Columns 3 & 4 contain results from Heckman selection models, where the regression equation is identical to our baseline cross-section demand specification, i.e. the dependent variable is the log. number of monthly or biannual new ratings, and the selection equation models app survival. In column 3 survival is modeled by comparing apps which are observed throughout the period April to September 2012 to those which are observed in April 2012 but which cannot be observed in later months. In column 4 survival is modeled by comparing apps within our baseline cross-section from April 2012 which survive until 2014 to those which are observed in April 2012 but are not observed in 2014. In both selection models we apply Heckman's two-step consistent estimator and use the information on code size as the selection variable, i.e. we include the code size only in the selection equation but not in the regression equation. In Column 5 we control for the existing user-base by including a control for the stock of existing installations (log. number of installations). In Columns 6 and 7 we estimate a 2SLS model and instrument the variable of interest to account for the endogeneity of the developers' privacy model choice. In Column 6 we instrument the privacy-dummy by the share of competing apps which use privacy-sensitve permissions ($ShareComp_{Privacy}$). In Column 7 we instrument the privacy-dummy by the share of the apps of the developer which use privacy-sensitive permissions ($ShareDev_{Privacy}$). In Columns 8 and 9 we instrument the app price by using in both specifications two potential cost shifters: the log. code size and the log. number of apps a developer offers in the Google Play Store. In Column 8 we use the full cross-section, whereas in Column 9 we use only the sample of paid apps. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B15: Baseline OLS and Full Heckman Selection Models

	Heckman (1)		Heckman (2)	
	OLS	Heckman	OLS	Heckman
main				
$D_{Privacy}$	-0.055***	-0.063***	-0.048***	-0.164***
	(0.009)	(0.009)	(0.016)	(0.024)
$\#_{CleanPerm}$	0.041***	0.038***	0.043***	0.019***
	(0.003)	(0.002)	(0.005)	(0.005)
$D_{Internet}$	-0.187***	-0.184***	-0.228***	-0.423***
	(0.009)	(0.010)	(0.016)	(0.030)
D_{Ads}	0.243***	0.234***	0.366***	0.291***
	(0.009)	(0.010)	(0.016)	(0.023)
Log. Price	-0.069***	-0.068***	-0.169***	-0.157***
	(0.001)	(0.001)	(0.001)	(0.002)
Log. AppsByDev	-0.112***	-0.115***	-0.207***	-0.459***
	(0.002)	(0.003)	(0.004)	(0.025)
Constant	-3.520***	-3.533***	-4.958***	-5.781***
	(0.122)	(0.127)	(0.229)	(0.319)
select				
$D_{Privacy}$		-0.099***		-0.081***
		(0.014)		(0.009)
$\#_{CleanPerm}$		-0.026***		-0.013***
		(0.002)		(0.002)
$D_{Internet}$		0.026		-0.158***
		(0.018)		(0.010)
D_{Ads}		-0.131***		-0.054***
		(0.015)		(0.009)
Log. Price		0.003**		0.008***
		(0.001)		(0.001)
Log. AppsByDev		-0.042***		-0.165***
		(0.003)		(0.002)
Log. Size (in KB)		-0.019***		-0.012***
		(0.004)		(0.002)
Constant		2.383***		1.066***
		(0.208)		(0.129)
mills				
lambda		0.486**		2.978***
		(0.194)		(0.281)
Category	Yes	No	Yes	No
Controls	Yes	Yes	Yes	Yes
Observations	177193	185533	124482	177193
Mean of dep. Var.	0.10	0.10	2.40	2.40
SD of dep. Var.	1.55	1.55	2.52	2.52
Adjusted R^2	0.291		0.384	

NOTES: The table analyzes the robustness of our main demand-side results to using alternative estimation strategies. The dependent variable is app demand measured by the log. number of monthly new ratings, and the main variable of interest is a dummy that indicates the presence of privacy-sensitive permissions. All columns show cross section results. Column 1 shows OLS estimates for April 2012 (like in our baseline demand equation but without the code size variable). Column 2 & 4 contain results from Heckman selection models, where the regression equation is identical to our baseline cross-section demand specification, i.e. the dependent variable is the log. number of monthly or biannual new ratings, and the selection equation models app survival. In column 2 survival is modeled by comparing apps which are observed throughout the period April to September 2012 to those which are observed in April 2012 but which cannot be observed in later months. Column 3 contains OLS estimates for those apps which are available both in 2012 and 2014. The dependent variable as in column 4 is the log. change in the number of ratings between 2012 and 2014. In column 4 survival is modeled by comparing apps within our baseline cross-section from April 2012 which survive until 2014 to those which are observed in April 2012 but are not observed in 2014. In both selection models we apply Heckman's two-step consistent estimator and use the information on code size as the selection variable, i.e. we include the code size only in the selection equation but not in the regression equation. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table B16: IV-Specifications (1st and 2nd Stage)

	IV-Privacy (1)		IV-Privacy (2)		IV-Price (1)		IV-Price (2)	
	2nd	1st	2nd	1st	2nd	1st	2nd	1st
$D_{Privacy}$	-0.918***		-0.372***		-0.091***	-0.337***	-0.063***	0.072***
	(0.049)		(0.017)		(0.009)	(0.029)	(0.017)	(0.009)
$\#_{CleanPerm}$	0.065***	0.027***	0.044***	0.016***	0.043***	0.069***	0.082***	0.032***
	(0.004)	(0.002)	(0.003)	(0.001)	(0.003)	(0.007)	(0.007)	(0.003)
$D_{Internet}$	-0.099***	0.108***	-0.177***	0.086***	-0.289***	-2.483***	-0.005	0.085***
	(0.012)	(0.003)	(0.011)	(0.003)	(0.014)	(0.035)	(0.017)	(0.009)
D_{Ads}	0.243***	0.032***	0.253***	0.014***	0.198***	-1.442***	0.043**	0.000
	(0.011)	(0.004)	(0.011)	(0.003)	(0.011)	(0.029)	(0.020)	(0.011)
Log. Price	-0.074***	-0.003***	-0.074***	-0.001***	-0.108***		-0.687***	
	(0.001)	(0.000)	(0.001)	(0.000)	(0.004)		(0.094)	
$ShareComp_{Privacy}$		0.279***						
		(0.004)						
$ShareDev_{Privacy}$				0.594***				
				(0.003)				
Log. Size (in KB)	0.069***	0.011***	0.064***	0.002***		0.381***		0.022***
	(0.003)	(0.001)	(0.002)	(0.001)		(0.008)		(0.002)
Log. AppsByDev	-0.099***	0.017***	-0.093***	0.005***		0.255***		0.022***
	(0.002)	(0.001)	(0.002)	(0.000)		(0.007)		(0.002)
D_{Other}	0.182***	0.144***	0.115***	0.090***	0.079***	0.170***	0.117***	0.122***
	(0.014)	(0.004)	(0.011)	(0.003)	(0.010)	(0.029)	(0.020)	(0.009)
Log. Length Desc.	0.290***	0.016***	0.253***	0.015***	0.298***	0.688***	0.219***	0.104***
	(0.005)	(0.001)	(0.004)	(0.001)	(0.005)	(0.012)	(0.012)	(0.004)
Number Screenshots	0.077***	-0.003***	0.075***	-0.001	0.106***	0.146***	0.070***	0.009***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.007)	(0.003)	(0.002)
Dummy: Video	0.235***	0.041***	0.211***	0.017***	0.221***	-0.002	0.283***	0.045***
	(0.014)	(0.003)	(0.013)	(0.003)	(0.012)	(0.040)	(0.020)	(0.010)
Log. Average Rating	0.188***	-0.011***	0.204***	-0.014***	0.211***	-0.477***	0.079***	-0.015**
	(0.009)	(0.003)	(0.008)	(0.002)	(0.008)	(0.034)	(0.009)	(0.008)
Dummy: Top-Dev.	1.201***	0.064***	1.038***	0.032***	1.242***	2.787***	1.490***	0.425***
	(0.064)	(0.014)	(0.058)	(0.010)	(0.060)	(0.169)	(0.094)	(0.045)
App Version	0.038***	0.008***	0.023***	0.005***	-0.004	-0.618***	0.016**	-0.022***
	(0.004)	(0.001)	(0.004)	(0.001)	(0.004)	(0.014)	(0.007)	(0.004)
Log. InstByDev	0.190***	0.008***	0.206***	-0.001***	0.178***	-0.219***	0.105***	0.003**
	(0.002)	(0.000)	(0.002)	(0.000)	(0.002)	(0.005)	(0.002)	(0.001)
Log. InstByComp	-0.000	-0.001***	-0.003	0.003***	-0.012***	-0.220***	-0.019***	-0.017***
	(0.002)	(0.001)	(0.002)	(0.000)	(0.002)	(0.007)	(0.003)	(0.002)
Log. PriceOfComp	0.011***	0.001***	0.015***	0.001***	0.018***	0.204***	0.012***	0.010***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)	(0.001)
Log. RatByComp	-0.799***	-0.044***	-0.613***	-0.061***	-0.750***	1.918***	-0.380***	-0.692***
	(0.063)	(0.016)	(0.067)	(0.015)	(0.063)	(0.204)	(0.111)	(0.056)
Min. Android Vers.	-0.022*	-0.015***	0.014	-0.005**	0.039***	-0.601***	0.067***	-0.100***
	(0.012)	(0.003)	(0.012)	(0.003)	(0.011)	(0.035)	(0.018)	(0.010)
Max. Android Vers.	0.187***	-0.023***	0.182***	-0.008	0.152***	0.328***	0.177***	0.062***
	(0.024)	(0.007)	(0.024)	(0.007)	(0.020)	(0.069)	(0.034)	(0.020)
Constant	-4.141***	-0.195***	-4.305***	-0.069*	-3.970***	-10.267***	-3.300***	0.087
	(0.136)	(0.038)	(0.141)	(0.035)	(0.128)	(0.415)	(0.197)	(0.117)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	136040	136040	137146	137146	177193	177193	48272	48272
Mean of dep. Var.	0.10	0.44	0.10	0.43	0.10	-8.30	-0.50	0.26
SD of dep. Var.	1.55	0.50	1.56	0.50	1.55	5.26	1.06	0.73
Adjusted R^2	0.256	0.475	0.324	0.635	0.269	0.320	0.047	0.277

NOTES: The dependent variable is app demand measured by the log. number of monthly new ratings, and the main variable of interest is a dummy that indicates the presence of privacy-sensitive permissions. All columns show cross section 2SLS results (as in Table A10). Columns 1, 3, 5, and 6 contain 2nd-stage results, whereas columns 2, 4, 6, and 8 contain 1st-stage results. In Columns 1 to 4 we instrument the privacy-dummy to account for the endogeneity of the developers' privacy model choice. In Columns 1 and 2 we instrument the privacy-dummy by the share of competing apps which use privacy-sensitve permissions ($ShareComp_{Privacy}$). In Columns 3 and 4 we instrument the privacy-dummy by the share of the apps of the developer which use privacy-sensitive permissions ($ShareDev_{Privacy}$). In Columns 5 to 8 we instrumtent the app price by using in both specifications two potential cost shifters: the log. code size and the log. number of apps a developer offers in the Google Play Store. In Columns 5 and 6 we use the full cross-section, whereas in Columns 7 and 8 we use only the sample of paid apps. Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Results: Table B17 shows the results. In all specifications, the dependent variable is app demand and the variable of interest is an indicator for presence of privacy sensitive permissions ($D_{Privacy}$). App demand is measured by the log. number of monthly new ratings of an app. Column 1 shows the results for free apps only, and column 2 for paid apps only. The estimated coefficient is much more negative in paid apps. Columns 3 and 4 contrast non-game apps (col.3) with games (col. 4), and shows that the association between permissions and usage is negative only for non-games. Column 5 considers only apps for mature users, and column 6 all others. The estimations suggest that privacy-sensitive permissions matter more in apps for mature users. Column 7 focuses on non-health apps and column 8 on health related apps. All specifications control for the app's observed characteristics on the Play Store (the app's price, description, ratings, categorical dummies, etc.), and also control for internet access, and ad-specific permissions. Moreover, all specifications control for the number of unproblematic permissions ($CleanPerm$. The results show that our main demand-side result, the negative relationship between demand and permissions, can be found for almost all types of apps, except for games. The varying size of the coefficients indicates that the app's type matters for the strength of the relationship between demand and permissions. Users of free apps and games seem more willing to share data, and apps for mature users or health-related apps are less demanded when using privacy-sensitive permissions.

Table B17: Alternative Estimation Samples

Log. $\Delta Ratings$	Free vs. Paid (1)	(2)	Non-Games vs. Games (3)	(4)	High vs. Low Maturity (5)	(6)	Med.&Health vs. Oth. (7)	(8)
$D_{Privacy}$	-0.026**	-0.121***	-0.082***	-0.020	-0.109***	-0.054***	-0.151***	-0.060***
	(0.011)	(0.014)	(0.010)	(0.022)	(0.019)	(0.010)	(0.043)	(0.009)
$\#_{CleanPerm}$	0.037***	0.054***	0.037***	0.098***	0.028***	0.038***	0.029**	0.036***
	(0.003)	(0.006)	(0.003)	(0.009)	(0.008)	(0.003)	(0.014)	(0.003)
$D_{Internet}$	-0.256***	-0.075***	-0.211***	-0.201***	-0.108***	-0.211***	-0.106**	-0.203***
	(0.012)	(0.013)	(0.010)	(0.024)	(0.019)	(0.010)	(0.041)	(0.009)
D_{Ads}	0.241***	0.039**	0.228***	0.157***	0.292***	0.222***	0.180***	0.238***
	(0.010)	(0.017)	(0.010)	(0.024)	(0.025)	(0.010)	(0.042)	(0.009)
Log. Price	0.000	0.099***	-0.068***	-0.095***	-0.050***	-0.076***	-0.056***	-0.072***
	(.)	(0.007)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)
D_{Other}	0.016	0.022	0.023**	0.144***	0.116***	0.040***	0.057	0.053***
	(0.011)	(0.015)	(0.010)	(0.027)	(0.025)	(0.010)	(0.044)	(0.009)
Log. Length Desc.	0.280***	0.135***	0.255***	0.291***	0.237***	0.268***	0.257***	0.262***
	(0.005)	(0.006)	(0.004)	(0.011)	(0.011)	(0.004)	(0.018)	(0.004)
Log. Size (in KB)	0.080***	0.011***	0.050***	0.117***	0.044***	0.065***	0.064***	0.062***
	(0.003)	(0.003)	(0.002)	(0.006)	(0.005)	(0.002)	(0.010)	(0.002)
Number Screenshots	0.105***	0.057***	0.086***	0.085***	0.086***	0.084***	0.075***	0.085***
	(0.003)	(0.003)	(0.002)	(0.006)	(0.007)	(0.002)	(0.010)	(0.002)
Dummy: Video	0.271***	0.225***	0.204***	0.177***	0.072*	0.218***	0.024	0.208***
	(0.016)	(0.018)	(0.015)	(0.024)	(0.039)	(0.013)	(0.068)	(0.012)
Log. Average Rating	0.371***	0.083***	0.167***	0.355***	0.207***	0.204***	0.257***	0.196***
	(0.010)	(0.007)	(0.008)	(0.020)	(0.014)	(0.008)	(0.029)	(0.007)
Dummy: Top-Dev.	1.151***	1.169***	1.026***	1.295***	1.505***	1.104***	0.155	1.272***
	(0.088)	(0.082)	(0.074)	(0.092)	(0.212)	(0.060)	(0.126)	(0.063)
App Version	0.039***	0.036***	0.029***	0.017*	0.018*	0.032***	0.069***	0.027***
	(0.005)	(0.006)	(0.004)	(0.009)	(0.010)	(0.004)	(0.020)	(0.004)
Log. AppsByDev	-0.169***	-0.047***	-0.105***	-0.180***	-0.077***	-0.122***	-0.100***	-0.114***
	(0.003)	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.010)	(0.002)
Log. InstByDev	0.236***	0.100***	0.179***	0.203***	0.140***	0.192***	0.164***	0.184***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.008)	(0.002)
Log. InstByComp	-0.010***	-0.007***	-0.007***	0.003	-0.007*	-0.007***	0.009	-0.007***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.004)	(0.002)	(0.011)	(0.002)
Log. PriceOfComp	0.012***	0.004***	0.008***	0.019***	0.007***	0.011***	0.013***	0.011***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)
Log. RatByComp	-0.823***	0.148*	-0.696***	-1.085***	-0.424***	-0.761***	-0.765**	-0.716***
	(0.078)	(0.078)	(0.067)	(0.147)	(0.134)	(0.068)	(0.304)	(0.062)
Min. Android Vers.	-0.029**	0.125***	-0.025**	0.039	-0.022	-0.011	-0.076	-0.014
	(0.013)	(0.015)	(0.011)	(0.028)	(0.026)	(0.011)	(0.051)	(0.011)
Max. Android Vers.	0.171***	0.140***	0.155***	0.250***	0.187***	0.171***	0.146	0.171***
	(0.024)	(0.030)	(0.021)	(0.053)	(0.032)	(0.023)	(0.127)	(0.020)
Constant	-3.557***	-3.416***	-3.398***	-5.120***	-3.467***	-3.980***	-3.811***	-3.865***
	(0.152)	(0.172)	(0.140)	(0.306)	(0.247)	(0.140)	(0.669)	(0.125)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	128921	48272	147143	30050	27843	149350	6218	170975
Mean of dep. Var.	0.33	-0.50	0.07	0.26	-0.18	0.15	-0.08	0.11
SD of dep. Var.	1.65	1.06	1.52	1.70	1.38	1.58	1.38	1.56
Adjusted R^2	0.285	0.262	0.280	0.370	0.305	0.291	0.270	0.295

NOTES: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data. App demand is measured by the log. number of monthly new ratings of an app. Columns 1 and 2 split the sample into apps which are for free or for paid. Columns 3 and 4 split the sample into normal apps and games. Columns 5 and 6 split them into apps which require a high (Column 5) or low (Column 6) maturity of the user (apps are defined as appropriate for low maturity if they classified as being recommended for 'everyone' or for 'low maturity'-users). Columns 7 and 8 split the sample into medical and health-related apps as well as into other apps. All specifications control for the number of unproblematic permissions (*#CleanPerm.*). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.4.8 Excluding the Most and Least Successfull Apps from Estimation

In Table B18 we further analyze the robustness of the main demand-side results to excluding the most or least successful apps from the analysis. This robustness check guarantees, that our results are not driven by the apps in the tails of the distribution in our dependent variable.

Data Preparation and Additional Variables: For this robustness check, we excluded the most and least successful apps. First we excluded the most succesful apps, i.e. the upper 5 percentiles with respect to the number of new ratings in April 2012 in Columns 1 and 2. Analogously, we exclude the least successful apps, without any new ratings in April 2012 (the least successful ones) in Columns 3 and 4. In Columns 5 and 6 we exclude both groups. Moreover we generated a dummy which equals 1 for apps which had accumulated more than 10000 installations (Col. 2, 4 & 6). As before app demand is measured by the log. number of monthly new ratings of an app. All specification use the cross section from April 2012.

Results: The results in columns 1, 3, and 5 of Table B18 are shown without accounting for apps with a very high stock of installations. In Columns 2, 4 and 6 we account for such highly successful apps, by adding a dummy which equals one if the app had accumulated a stock of at least 10000 installations in the past, and also add an interaction of this dummy with the privacy-dummy. All specifications control for the number of unproblematic permissions ($CleanPerm.$), and for the app's observed characteristics on the Play Store (the app's price, description, ratings, categorical dummies, etc.). We also control for internet access, and ad-specific permissions. The coefficient of interest analyzes the relationship between an app's downloads and our measures of privacy sensitive permissions. While the coefficient is negative for all specifications, it becomes smaller and statistically insignificant when excluding only the least successful ones (column 3). Thus, the table highlights that the negative relationship between privacy-sensitive permission and demand holds across a wide range of apps and does not depend on the inclusion of the distribution tails. In addition, it indicates that the stock of installations affects the sensitivity of users with respect to the existence of privacy-sensitive permissions.

Table B18: Excluding the Most and Least Successful Apps

Log. $\Delta Ratings$	W/o Top-Apps (1)	W/o Top-Apps (2)	W/o Flop-Apps (3)	W/o Flop-Apps (4)	W/o Top- and Flop-Apps (5)	W/o Top- and Flop-Apps (6)
$D_{Privacy}$	-0.064***	-0.079***	-0.011	-0.086***	-0.030***	-0.050***
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.010)
$D_{NumInst}$		1.545***		1.303***		0.946***
		(0.011)		(0.012)		(0.011)
$D_{Privacy} \times D_{NumInst}$		0.036**		0.126***		0.009
		(0.015)		(0.017)		(0.014)
$\#_{CleanPerm}$	0.019***	0.019***	0.036***	0.032***	0.017***	0.018***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
$D_{Internet}$	-0.114***	-0.093***	-0.197***	-0.172***	-0.092***	-0.090***
	(0.008)	(0.007)	(0.014)	(0.012)	(0.011)	(0.010)
D_{Ads}	0.170***	0.154***	0.182***	0.188***	0.102***	0.116***
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.009)
Log. Price	-0.055***	-0.026***	-0.071***	-0.023***	-0.046***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_{Other}	0.041***	0.043***	0.028**	0.038***	0.014	0.020**
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.009)
Log. Length Desc.	0.193***	0.138***	0.214***	0.151***	0.132***	0.099***
	(0.003)	(0.003)	(0.006)	(0.005)	(0.004)	(0.004)
Log. Size (in KB)	0.039***	0.029***	0.069***	0.057***	0.035***	0.030***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
Number Screenshots	0.058***	0.057***	0.070***	0.065***	0.040***	0.041***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Dummy: Video	0.131***	0.101***	0.167***	0.110***	0.099***	0.068***
	(0.010)	(0.009)	(0.015)	(0.014)	(0.012)	(0.011)
Log. Average Rating	0.137***	0.140***	0.558***	0.576***	0.316***	0.352***
	(0.006)	(0.005)	(0.017)	(0.015)	(0.014)	(0.012)
Dummy: Top-Dev.	0.558***	0.345***	0.859***	0.644***	0.414***	0.296***
	(0.052)	(0.044)	(0.055)	(0.048)	(0.049)	(0.042)
App Version	0.020***	-0.007***	0.032***	-0.004	0.020***	-0.004
	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)
Log. AppsByDev	-0.091***	-0.068***	-0.128***	-0.094***	-0.083***	-0.064***
	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Log. InstByDev	0.134***	0.073***	0.177***	0.094***	0.109***	0.057***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Log. InstByComp	-0.016***	-0.018***	0.010***	0.002	-0.003	-0.008***
	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
Log. PriceOfComp	0.002***	-0.001**	0.020***	0.012***	0.008***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log. RatByComp	-0.611***	-0.128***	-0.518***	0.091	-0.354***	0.043
	(0.052)	(0.045)	(0.083)	(0.073)	(0.068)	(0.061)
Min. Android Vers.	0.008	0.145***	-0.001	0.205***	0.024**	0.168***
	(0.009)	(0.008)	(0.015)	(0.013)	(0.012)	(0.011)
Max. Android Vers.	0.138***	0.144***	0.113***	0.149***	0.072***	0.103***
	(0.016)	(0.014)	(0.035)	(0.030)	(0.027)	(0.024)
Constant	-2.777***	-2.924***	-3.228***	-3.462***	-1.535***	-1.906***
	(0.103)	(0.090)	(0.192)	(0.167)	(0.151)	(0.135)
Category	Yes	Yes	Yes	Yes	Yes	Yes
Observations	168767	168767	79847	79847	71421	71421
Mean of dep. Var.	-0.12	-0.12	1.44	1.44	1.09	1.09
SD of dep. Var.	1.23	1.23	1.44	1.44	1.04	1.04
Adjusted R^2	0.231	0.416	0.240	0.409	0.149	0.303

Notes: The table shows the relationship between the presence of privacy-sensitive permissions and app demand for subsamples of our data where we exclude the tails of the distribution with respect to our dependent variable, i.e. the most and the least successful apps. App demand is measured by the log. number of monthly new ratings of an app. In Columns 1 and 2 we exclude the most succesful apps, i.e. the upper 5 percentiles with respect to the number of new ratings in April 2012. In Columns 3 and 4 we exclude the least successful apps, i.e. those without any new ratings in April 2012. In Columns 5 and 6 we exclude both groups, i.e. the upper 5 percent of most successful apps and those having no new rating in April 2012. In Columns 2, 4 and 6 we add a dummy which is equal to one if the app in the past had accumulated a stock of at least 10000 or more installations and also add an interaction of this dummy with the privacy-dummy. All specifications control for the number of unproblematic permissions ($\#CleanPerm.$). Heteroscedasticity-consistent standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.5 AppAnnie.com

Figure B1: Ranking Information on AppAnnie.com in 2016