Spillover Effects and Freemium Strategy in the Mobile App Market

Yiting Deng   Anja Lambrecht   Yongdong Liu*

*The authors are listed alphabetically. Yiting Deng is an Assistant Professor at the UCL School of Management, University College London (email: yiting.deng@ucl.ac.uk). Anja Lambrecht is a Professor at London Business School (email: alambrecht@london.edu). Yongdong Liu is an Assistant Professor at the UCL School of Management, University College London (email: yongdong.liu@ucl.ac.uk). The authors would like to thank AppMonsta for providing data on App Store and Google Play, and thank Dave Verwer for providing data on now-defunct website AppReviewTimes.com. We would also like to thank Catherine Tucker, Prasad Vana, Rob Waiser, Wei Wang, participants at the 2017 Marketing Science Conference, the 2017 Marketing Dynamics Conference, the 2018 Berlin IO Day, and the Workshop on Perception and Public Policy at Bar Ilan University, as well as seminar participants at Cornell, ESSEC Business School, HKUST, Johns Hopkins University, Peking University, SHUFE, Tsinghua University, Sun Yat-sen University, UIBE, University of California, Irvine, University of Southern California, University of Colorado Boulder, University College London, University of Cambridge, University of Mannheim, University of Toulouse, Western University, Wharton and Yale for useful comments. Elias Djurfeldt, Julian Hohlweg and Lucas Weidenholzer provided excellent research assistance. The usual disclaimer applies.

Abstract

"Freemium" whereby a basic service level is provided free of charge but consumers are charged for more advanced features has become a popular business model for firms selling digital goods. However, it is not clear whether the launch of a free version helps or hurts the demand of an existing paid version. The free version may allow consumers to sample the product before making a purchase decision and subsequently increase demand of the paid version, but it may also cannibalize demand of the paid version. We use a comprehensive data set on game apps from Apple's App Store that tracks the launch of both the paid and the free versions of individual apps on a daily level, to identify whether a freemium strategy stimulates or hurts demand of an existing paid version. We estimate the spillover effects between the free version and the paid version of the same app under a difference-in-difference framework, relying on the fact that app developers cannot predict the exact launch date of the free version of the app due to Apple's review and approval of apps prior to release and accounting for app-level product heterogeneity. We find that the launch of a free version increases demand of the paid version of the same app. Under the main specification, if the daily number of ratings before the free version's launch is at the mean, then all else equal, the launch of the free version leads to an 8.9% increase in the daily number of ratings. We then describe multiple robustness checks. Finally, we present evidence that the results are driven by consumers sampling the free version as well as enhanced app discovery, and explore the relative importance of the two mechanisms.

Keywords: Freemium, mobile app, spillover effect, cannibalization, sampling, discovery

JEL Classification Codes: C21, D12, L15, M31

1 Introduction

For firms selling digital products, "freemium" – providing a basic service level for free but charging consumers for using their "premium" version – has in recent years become an increasingly popular business model. Well-known examples include Dropbox, Evernote, LinkedIn and Spotify. Freemium is a prevalent pricing model for mobile apps, a market that is predicted to total almost 935 billion U.S. dollars in worldwide revenues by 2023 (Statista, 2019). In 2014, about half of the apps downloaded from the Google Play store were freemium, and these apps garnered 98% of global Google Play revenue (AppAnnie, 2015). Similarly, 92% of revenue on Apple's App Store ("App Store" hereafter) is generated by freemium apps (Kosner, 2015).[1] For example, the game Fruit Ninja is available for free with a separate paid version for a price of $1.99.

Understanding whether and when offering a free version alongside a paid version is an effective strategy is important: the free version may allow consumers to sample the product before making a purchase decision and so ultimately increase demand of the paid version. But the free version may also cannibalize demand of the paid version if consumers who otherwise would have paid instead opt for the free version. As of yet, there is contradictory evidence on whether the launch of a free version helps or hurts the demand of the existing paid version. While some findings suggest that offering feature-limited free versions in the mobile app market negatively affects the paid app's adoption speed (Arora et al., 2017), others suggest that a free version increases sales of the paid app (Liu et al., 2014) and that the relationship between the free and paid versions may change along the app's life cycle (Lee et al., 2021).

The objective of the present paper is threefold. First, we aim to identify the effect of introducing a free version of an app on the demand of its existing paid counterpart by focusing on a short time window around the free version's launch. Thus, we circumvent identification issues encountered by prior research that analyzed an extended time period where broader changes in the marketplace, in patterns of consumer demand or in the app itself may have affected the result. Second, we aim to provide evidence that the patterns we observe are indeed driven by the free version's launch and not by correlated marketing actions. Third, we explore the mechanisms behind the results in order to make recommendations to firms on when

[1] During the period we study, apps with a freemium offering typically offered both a free and a paid app version of the same app. More recently, firms have started to adopt another variant of freemium: they offer free apps with in-app purchases. The data by AppAnnie (2015) and Kosner (2015) cover both types of freemium apps.

offering a free version can be an effective strategy.

We use daily data on individual game apps from App Store. We focus the analysis on apps that have both a paid version and a free counterpart, and construct a group of control apps for each freemium app. We estimate the model in a difference-in-difference framework within a tight four-week window around the launch of the free version. Using the number of ratings an app received as a proxy for demand, we find that the launch of the free version increases the demand of the paid version. Under the main specification, if the daily number of ratings before the free version's launch is at the mean, then all else equal, after the launch of the free version, the daily number of ratings would increase by 8.9%. The interpretation that this effect is causal holds under two assumptions that are both supported by our data: first, the assumption that the firm cannot time precisely the date for the free version to be approved and launched by the platform and, second, the assumption that within this brief four-week time period around the launch, no other actions correlated with the launch affect demand. We then conduct a battery of robustness checks, and demonstrate the effect holds under alternative specifications. Further, we provide evidence suggesting that the effect is not driven by other events correlated with the introduction of the free version, such as changes to the paid version's design or price, or advertising and promotion.

We then explore two potential mechanisms. First, we explore the possibility that the free version allows consumers who are uncertain about the quality of the product to sample before purchasing the full version (e.g., Bawa and Shoemaker, 2004; Lee and Tan, 2013; Shoemaker and Shoaf, 1975; Wang and Zhang, 2009). We conceptualize the process of consumers sampling the free version and upgrading to the paid version. We then propose how observable characteristics of an app such as the average star rating contribute to a consumer's likelihood to sample and upgrade. Perhaps most importantly, we suggest that sampling should be most effective for apps that have a moderate star rating. This is because consumers who expect a product might be of very low quality will find the expected benefits too low given the cost associated with sampling. At the same time, if the product's quality is guaranteed to be above a high threshold, consumers may purchase the paid version without sampling, so introducing the free version provides little added benefit. Further empirical evidence related to the number of ratings likewise supports sampling as a mechanism. In addition, we propose that the effect of sampling should be more pronounced when the paid version provides significantly greater utility for consumers to upgrade. Again, we find evidence consistent with these predictions, supporting further that sampling is one mechanism behind the findings.

Second, we examine whether the free version may have enhanced the paid version's visibility thus facilitating app discovery for consumers in a market where tens of thousands of options are available (Li et al., 2016).[2] In such a setting, having two versions of the same app can make the app more visible to consumers. We provide two pieces of evidence supporting that enhanced visibility can increase demand in the market for mobile apps. First, we document that even for apps that initially launched a free version and later added a paid version, the introduction of the paid version increases the initial free version's demand. Second, we demonstrate that the effect of launching a second version decreases as the size of the category increases, suggesting that the effect on the marginal probability to discover an app becomes less pronounced as a category becomes more crowded. Both patterns are consistent with an app's additional version enhancing visibility of the original version but cannot directly be attributed to sampling. We also demonstrate the relative importance of the sampling and discovery mechanisms.

Our work relates to three streams of academic research. First, it relates to a nascent stream of empirical research on freemium pricing when firms offer both a free and a paid version of a digital product. Existing research has demonstrated the effect of providing a free version in various contexts such as software (Chen et al., 2017; Lee and Tan, 2013; Runge et al., 2016), cloud storage (Lee et al., 2017), digital TV service (Foubert and Gijsbrechts, 2016) and on-demand dramas (Hoang and Kauffman, 2018). For apps on Google Play, Liu et al. (2014) find that paid apps with a free counterpart receive more sales than those that do not have a free counterpart. By contrast, in the same empirical context, Arora et al. (2017) find that offering free versions reduces the paid app's adoption speed. Lee et al. (2021) find contemporaneous cannibalization effects as well as positive inter-temporal effects between the free and paid versions of the same app, and model app developers' versioning decision. However, this existing literature has focused on an overall effect of a free version, without being able to tie it directly to the actual timing of the introduction of the free version. Our work adds to this literature in three ways. First, in order to identify the effect of the free version's launch, we identify the precise launch date of the free version and focus our analysis on a tight time window around the launch. To control for unobservable time-varying factors that can potentially influence app demand, we construct a group of control apps for each freemium app and estimate the model in a difference-in-difference framework. We demonstrate that launching a free version increases the demand

[2] See also `http://www.tbray.org/ongoing/When/201x/2010/07/18/How-to-run-an-App-store` and `https://www.cnet.com/news/the-mobile-app-discovery-problem/`.

for the app's paid version. Second, we discuss in detail that other marketing actions that could in theory be correlated with the launch are unlikely to drive this effect of interest. Third, to our knowledge we are the first to empirically identify mechanisms that lead to a spillover effect of the presence of a free version on an existing paid version. Specifically, we provide evidence for both sampling and app discovery as the mechanisms.

The second stream of research we contribute to is a broader literature on the pricing of digital goods. Pauwels and Weiss (2008) demonstrate for an online content provider for marketing professionals that moving from free to fee can be profitable, despite loss of advertising revenue. Chiou and Tucker (2013) find that paywalls decrease viewership of an online newspaper and, as a result, depress advertising revenues. Lambrecht and Misra (2016) show that an online content provider that also generates advertising revenues can benefit from offering more free content in periods of high demand than in periods of low demand. Kannan et al. (2009) explore how a firm that offers a print and a PDF version of books should set prices. We contribute to this literature by demonstrating that offering a free version of a product can generate additional demand for the paid version, potentially making this a profitable pricing strategy.

Third, our research relates to prior research on sampling and versioning. The earlier literature focused on analytical models for the sampling of physical goods (Bawa and Shoemaker, 2004; Heiman et al., 2001; Jain et al., 1995). More recently, researchers have investigated free trials of digital goods that limit available features (Chellappa and Mehra, 2017; Cheng and Liu, 2011; Dey and Lahiri, 2016; Halbheer et al., 2014; Lahiri and Dey, 2018; Niculescu and Wu, 2014; Wang and Zhang, 2009) or the time of use (Cheng and Liu, 2011; Dey et al., 2013). For example, Halbheer et al. (2014) model an online content publisher's choice between offering paid content, sampling and free content. They find that it can be optimal for the publisher to generate advertising revenue by offering free samples even when sampling reduces prior quality expectations and content demand. Appel et al. (2019) develop an analytical framework for app monetization, and demonstrate that even if advertisers do not pay for ads, an app developer can still profit from offering a free version of the app. Empirical research on sampling and versioning is also growing. For example, Gu et al. (2018) and Li et al. (2019) conduct field experiments at a book publisher to investigate how to design free samples for books to increase revenue. Sunada (2020) models consumer learning in the context of a video game and explores the optimal design of free trial. Brecko (2020) models consumer choice of product versions, renewal and upgrades in the context of software. We add to this literature by providing empirical evidence that consumers' sampling of digital products can indeed drive consumers' decision to purchase the paid version and so potentially increase firm revenue, and demonstrating conditions under which the effect is more prominent.

At the same time, our findings are relevant to managers. We demonstrate, at least in the context of mobile apps, that a freemium strategy can indeed increase demand for the paid version of a product. In addition, our findings also suggest that firms may not want to indiscriminately offer free versions, but should instead assess the likely benefits and cost given the specific product under consideration. First, firms should consider investing into a free version particularly when the paid product is of medium quality. If a product is of low quality, opportunity cost associated with the free version may prevent consumers from sampling; while for high quality products, consumers may forgo sampling and instead directly purchase the paid version. So in either case a free version may not be as effective in driving demand. Second, firms need to ensure that the paid version offers substantial additional benefits relative to the free version as only then will it be attractive for consumers to upgrade. Third, especially in crowded marketplaces such as the market for mobile apps – at the time of our data, App Store offered close to 200,000 game apps – a free version may be useful in enhancing product visibility.

2 Industry Background and Data

2.1 Industry Background

Since Apple and Google launched their respective mobile application stores in 2008, the number of apps has grown exponentially: App Store has 3.42 million available apps as of July 2020 (Statista, 2020b) and Google Play offers about 3.04 million apps as of September 2020 (Statista, 2020a). Each platform offers a wide range of different apps developed by a large number of independently operating developers, the vast majority of which are small. Because the iOS and Android operating systems are incompatible, apps are not portable between the two platforms without a substantial change in the source code. In this paper, we focus on App Store, which, during the sample period (2011-2013), was the largest app platform.

During our sample period, a developer who adopted a freemium strategy typically offered both a free and a paid versions of an app. The free version may have differed from the paid version along a number of dimensions but generally had a reduced set of capabilities. To access the full version, users were required to separately download the paid app. In recent years, a different variant of freemium pricing, in-app purchases, has become popular. While App Store started allowing for in-app purchase microtransactions in 2009, these were not yet common during the period of our data.

2.2 Data Collection

We use a comprehensive data set that was collected from App Store between September 2011 and October 2013. The data set was provided to us by AppMonsta, a company specialized in app data analysis and consulting. The company scraped the U.S. App Store on a daily basis and each day collected all data that were publicly available.

We focus on game apps as it is the most popular category in App Store and purchases tend to be for personal use instead of business purposes. In addition, game apps are highly heterogeneous in terms of functionality and quality, thus, before using it, consumers are likely to have limited knowledge about how well an app fits their needs. In total, the data cover 197,884 mobile game apps, of which 95,990 (48.5%) are paid apps and the remaining are free apps. Among these paid apps, 55,320 were released during our data period. There are 18 genres of game apps: Action, Adventure, Role Play, Arcade, Racing, Sports, Trivia, Word, Education, Family, Dice, Board, Card, Puzzle, Casino, Simulation, Strategy, and Music. The data include the name of the developer, app characteristics such as price, version, size, genre, and release date, as well as user ratings, reviews and app rankings on various top charts.[3] In the data, the paid version and the free version of the same app are marked as different products. Table 1 summarizes the relevant variables in the raw data, some of which such as app price, the number of ratings, the star rating (on a scale of 1 to 5) and ranking can change on a daily basis.

[3] App Store provides three types of top charts: Top Free, Top Paid, and Top Grossing. The Top Free chart ranks free apps. The Top Paid chart ranks paid apps. The Top Grossing chart ranks all apps based on total revenue. For each genre (e.g., Action, Board, Trivia), there is a top free chart, a top paid chart, and a top grossing chart.

Table 1: Variables in Raw Data

Variable Category	Item	Description
App Identifier	App name App ID	App name shown on mobile devices A string of numbers and/or characters that uniquely identifies an app
Developer Identifier	Developer name Developer ID	Developer name shown on mobile devices A string of numbers and/or characters that uniquely identifies a developer
App Characteristics	App price App size Release date Genre App description Version	App price in U.S. Dollars App size in bytes App launch date App category One or several paragraphs outlining app functions App version number
Ratings	Number of ratings Star ratings	Accumulative number of ratings received Daily and overall ratings
Ranking	Ranking on top charts	Rankings on each related genre's top charts

2.3 Freemium Apps

We define freemium apps as paid apps that have a free counterpart. We take three steps to identify among all available apps those apps that are freemium apps (see Online Appendix A for details). First, we identify apps launched by the same developer. Second, within these apps, we compare similarities in app names for all possible app pairs. The pairs that are associated with high similarity measures are identified as potential freemium pairs. Third, we manually check all pairs to make sure they are indeed different versions of the same app. The free version and the paid version can usually be distinguished by words such as "lite", "free", "pro", and "premium".

We identify 9,806 app pairs that adopted the freemium strategy. For 8,284 (84.5%) apps, the paid version was launched first. In only 1,522 (15.5%) cases, the free version was launched first. There are no instances when the two versions were launched at the same time. We focus on the vast majority of cases where the paid version was launched first. Furthermore, for identification purposes, we drop from our sample 5,586 app pairs where the free version was launched before the start of our sample period. For 167 app pairs, data are missing for an extensive period of time after the release and are dropped from the sample. Thus, we retain a total of 2,531 app pairs which form the focus of our empirical analysis. For these app pairs, on average, the gap between the two versions' launch dates is 115 days (s.d. = 200 days). We will later use the

86,094 paid apps that did not have a free counterpart during our sample period to identify suitable control apps.

For each app, App Store displays a description of the app on its download page. We use the descriptions of the two versions displayed on the day when the free version was launched to identify differences between the paid and the free versions. Research assistants manually coded the differences into six categories. Each app was coded by two coders independently. We manually checked inconsistent codings to make final decisions. Table 2 summarizes the six dimensions along which the two versions may differ. In 34.6% of cases, the paid version allows the user to progress to more game levels than the free version. For example, the description of the game Phase 10's free version indicates "Phase 10 Free lets you play 3 random phases. Update to the full version for $.99 and play all 10 phases." Typically, this means a more complete, advanced or challenging game experience. 17.9% of paid versions offer more modes or themes than the free versions. For instance, Bean's Quest's paid version includes five themes (worlds) whereas the free version includes three themes. This means that while the ability to progress or the difficulty of the game does not change, the user experience can be customized. In 16.7% of cases, the paid version offers more functions or features (e.g., more powerful weapons, record keeping) than the free version. Further, in 11.7% of cases, the paid version allows for social interactions while the free version does not. For example, the paid version may be integrated with Game Center, Apple's social gaming network, which enables users to track their best scores on a leaderboard, compare their achievements, invite friends to play a game, and start a multiplayer game through auto-matching. The paid version may also allow the user to link to Facebook to post scores and share progress with friends. Lastly, in 11.0% of cases, the paid version is ad-free while the free version is ad-supported and in 1.5% of cases, the paid version provides better user support, for example an email contact to address user questions. These differences are not mutually exclusive, and there can be multiple differences between the paid version and the free version. For 43.3% of apps, there is no difference in the descriptions of the two versions.[4]

[4]Note that for these apps, the two versions may still differ along one or more dimensions, but we were unable to identify the differences from our data. Note also that while app versions are designed endogenously, app designs may not be optimal for many apps, especially those designed by inexperienced independent developers. Some evidence for this is that there are many discussions on Quora about what features should be added into a mobile app and whether to introduce a lite version in addition to the paid version, and suggestions are not always aligned (e.g., `https://www.quora.com/How-many-levels-should-I-start-with-in-a-mobile-game-I-am-about-to-create`; `https://www.quora.com/What-features-make-mobile-app-development-unique`).

Table 2: Version Differences based on App Description

Difference	Percent
More levels	34.6%
More modes/themes	17.9%
More functions/features	16.7%
Social interactions	11.7%
Ad free	11.0%
Better support	1.5%
No difference	43.3%

2.4 Descriptive Statistics

For our analysis, we focus on App Store and retain two types of data: paid apps that also have a free version (2,531 apps), and paid apps that did not have a free version in our data period (87,706 apps). Table 3 provides descriptive statistics for these apps, measured on the last day of the sample period (October 31, 2013).[5] Columns (1) and (2) respectively show the descriptive statistics for the paid version and the free version of the freemium apps, and Column (3) shows the descriptive statistics for apps that only had a paid version. The data suggest that the paid versions of freemium apps are broadly similar to apps that only exist as a paid version, but tend to have a somewhat lower price, somewhat higher star rating, and appear somewhat more likely to be ranked highly. In addition, compared with its paid counterpart, the free version tends to receive somewhat fewer ratings and a somewhat lower star rating.

Table 3: Descriptive Statistics of Apps

	(1) Paid version of freemium apps	(2) Free version of freemium apps	(3) Only paid
Price ($)	1.2 (1.0)	0.0 (0.0)	1.8 (6.5)
If rated	39.5% (48.9%)	27.1% (44.5%)	37.0% (48.27%)
ln (No. ratings+1)	1.4 (2.0)	0.9 (1.6)	1.4 (2.2)
Average star rating	4.2(0.7)	4.0 (0.7)	3.6 (0.9)
Ranked on top 10	0.04% (2.02%)	0.00% (0.00%)	0.02% (1.43%)
Ranked on top 11-20	0.04% (2.02%)	0.04% (2.02%)	0.01% (1.22%)
Ranked on top 21-50	0.00% (0.00%)	0.12% (3.49%)	0.04% (2.00%)
Ranked on top 51-100	0.24% (4.94%)	0.28% (5.33%)	0.09% (2.94%)
Ranked on top 101-	1.71% (12.96%)	5.62% (23.03%)	2.87% (16.70%)
Age (Days)	515 (292)	398 (220)	691 (459)

Note: Each entry is the average value of the corresponding variable for the group, with standard deviations in parentheses

[5]Table A.1 in the Online Appendix B reports summary statistics for the paid version of freemium apps in the panel setting, which shows variations both across apps and within an app over time.

2.5 Google Play as an Alternative Data Source

Our main analysis focuses on data from App Store. We collect similar data for the same data period (September 2011 to October 2013) from Google Play. In total, the data cover 1,169,351 apps, including 150,822 game apps of which 124,004 are free. The sample of freemium apps in Google Play is significantly smaller, with a total of 405 cases where a free version of a pre-existing paid version was launched during our data period. To allow for a meaningful analysis, we need to observe the paid version's number of ratings for two weeks before the launch of the free version as well as for two weeks after the launch of the free version, reducing the data to 217 pairs of freemium apps.[6] The significantly smaller sample size means that we focus the majority of our analysis on App Store but demonstrate the robustness of results using the data from Google Play.

2.6 Proxying App Demand

Ideally, we would want to explore the effect of launching a free version on the number of downloads of the existing paid version. As download data are not publicly available,[7] we use the number of ratings an app has accumulated as a proxy for demand. Using ratings as a proxy for demand is possible because the free and paid versions are treated as separate products in App Store, meaning that each has a separate measure for the number and valence of ratings. In this section, we first explain why the number of ratings of an app is a good proxy for demand. We then discuss alternatives proxies for demand.

Number of ratings: Our approach to proxy demand by an app's number of ratings builds on previous literature. Kummer and Schulte (2019) use the number of ratings as the preferred demand variable for Google Play apps, and demonstrate that the number of incremental ratings correlates with download ranges.[8]

[6]This means we exclude apps where the gap between the release dates of the free and paid versions was less than two weeks as well as those where the free version was launched close to the start or the end of the data period. More generally, two reasons for the smaller number of freemium apps are that at the time of our data, Google Play was still smaller than App Store and that it had a greater focus on free apps.

[7]Download data are confidential and only held by the platforms (App Store or Google Play) and individual developers, typically small firms. Third-party firms offer so-called "download data" for sale but these firms do not access true download data at scale. Instead, they typically build models to estimate downloads based on apps' rankings, sometimes supplemented with information on release date and ratings and calibrate the models based on downloads of a selection of apps. They then estimate downloads only for days on which an app was ranked. Thus, these firms only offer a proxy for downloads, the data are skewed to more successful apps on days when they performed particularly well, and the scope of such data would be too narrow as very few apps are ranked consistently across a 28- or 14-day time period, which we will require for identification. As a result, such data are not useful for our purpose.

[8]They also use change in the number of installations as an alternative measure but point out that this variable is less preferred than changes in the number of ratings, because the number of installations is observed only in discrete step-size form which leads to low intertemporal variation.

Boudreau (2018) reports that for apps sold in App Store, sales are highly correlated with the number of user ratings. He then approximates the rank order of app demand by the rank order of the numbers of user ratings.

We test whether the number of ratings and the number of downloads are indeed highly correlated as we assume. Such a test is difficult to implement using data from App Store as no download information is provided. However, Google Play reports for each app and day a range for the cumulative number of downloads the app has reached as of yet.[9] These download ranges allow us to broadly test the relationship between the number of ratings and downloads. To do so, we use all apps available in the Google Play data (i.e., not limited to freemium apps) and test the relationship between the number of ratings and the upper, respectively lower, bound of the download range, controlling for app fixed effects. Table 4 displays the estimation results. The dependent variables in Column (1) and Column (2) are respectively the logarithm of the upper bound and the logarithm of the lower bound of the download range, and the independent variable is the logarithm of the number of ratings.[10] In both cases, $R^2 = 0.97$, indicating a strong relationship between the logarithm of the bounds of download ranges and the logarithm of the number of ratings. To rule out the possibility that the results are driven by highly popular apps, Columns (3) and (4) replicate Columns (1) and (2), excluding app-day observations that have more than 492 ratings (95th percentile). The results are robust and together give us confidence that the number of ratings is indeed a good proxy for demand.

Table 4: Relationship between the Number of Ratings and Download Range Bounds on Google Play

Variables	All observations		Exclude observations with number of ratings over 95th percentile	
	(1) ln(Upper bound+1)	(2) ln(Lower bound+1)	(3) ln(Upper bound+1)	(4) ln(Lower bound+1)
ln (No. Ratings+1)	1.245***	1.240***	1.298***	1.294***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	4.583***	3.282***	4.507***	3.204**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
App FE	Yes	Yes	Yes	Yes
Number of Observations	279,547,860	279,547,860	265,555,690	265,555,690
Number of Apps	1,169,351	1,169,351	1,158,507	1,158,507
Adj. R-squared	0.97	0.97	0.96	0.96

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

[9]Google Play reports data for 18 download ranges: [1, 5], [5, 10], [10, 50], [50, 100], [100, 500], [500 1,000], [1,000, 5,000], [5,000, 10,000], [10,000, 50,000], [50,000, 100,000], [100,000, 500,000], [500,000, 1,000,000], [1,000,000, 5,000,000], [5,000,000, 10,000,000], [10,000,000, 50,000,000], [50,000,000, 100,000,000], [100,000,000, 500,000,000], and [500,000,000, 1,000,000,000].

[10]We use the logarithm of these variables as both the download range bounds and the number of ratings are skewed.

Despite the strong relationship between download range bounds and number of ratings, relying on any proxy for sales could introduce bias. In our context, we expect that if such a bias existed, it would lead us to underestimate, not overestimate, the effect of the launch of the free version on the sales of the paid version. Prior research suggests that the number of new consumer reviews decreases in the total number of reviews available (Wu and Huberman, 2008; Godes and Silva, 2012). By implication, the fraction of paid app downloads that lead to app ratings should have a downward trend. Thus, if the free version's launch does not affect the paid version's downloads, we would expect the paid version to accumulate fewer daily ratings after the free version has been launched. If instead we observe an increase in the paid version's daily incremental number of ratings after the free version's launch, this would strongly suggest an increase in the paid version's downloads. We further acknowledge the possibility that consumers might delay rating an app after downloading, which would again lead to an underestimation of the effect immediately following adoption.

Download ranges: As an alternative to the number of ratings as a proxy for demand, previous research has used download ranges reported by Google Play to proxy for demand (Arora et al., 2017; Liu et al., 2014). As these ranges vary in size and can be very broad (the lowest download range is 0 to 5, the highest is 500,000,000 to 1,000,000,000; see footnote 9), this approach requires apps to have a significant variation in downloads (for example because apps quickly become very popular) or have data covering a long time period.

Proxying demand using download ranges is difficult for two reasons. First, App Store does not report download ranges. Second, there are significantly fewer freemium apps on Google Play, and in the Google Play data, apps rarely change download ranges, likely owing to the large width of most ranges. For an individual app, we can proxy the total number of downloads by the upper bound of the previous interval or the lower bound of the new interval at the time of a download range change. Observing two changes would thus allow us to infer the speed of download accumulation. However, our identification approach requires us to observe such changes within a short time window around the launch of a free version. Out of the 217 freemium apps, 54 apps have at least one change in download ranges during the four-week window around the time of the launch of the free version, 16 apps have two or more download range changes, and no app has two download range changes both during the two weeks before and during the two weeks after the launch of the free version. Thus, such data would not provide sufficient variation to identify the effect of launching a free version.[11]

Ranking: Since Chevalier and Goolsbee (2003) demonstrated for books a relationship between actual sales and sales rank at Amazon and Barnes & Noble, researchers have tried using ranking to proxy for sales or downloads.[12] In our context, however, there are multiple challenges with using ranking as a proxy for sales. First, ranks are only available for a small fraction (15%) of freemium apps during the period of two weeks before and two weeks after the launch of the free version.[13] Second, relying on ranks would require excluding unranked apps and so bias the sample towards more successful apps. Third, ranking is a relative measure as a change in rank of one app could be caused by a change in performance of another app and not the focal app. Fourth, app rankings not only reflect the number of downloads, but also account for other factors such as the average rating, freshness (e.g., recency, updates) of the app or app retention (Liu et al., 2014).

Notwithstanding, we will report a robustness check in §3.3.3 where we use ranking in App Store as an alternative outcome variable.

3 Empirical Strategy and Estimation

3.1 Model-Free Evidence

To identify the effect of launching a free version of a mobile game app on the demand for the existing paid version, proxied for by its number of ratings, we focus the analysis on a four-week window around the free version's launch.[14] The assumption is that within this brief time period, changes to the incremental number of ratings can be attributed only to the launch of the new version. Figure 1 shows the average of the logarithms of the number of ratings received by the paid version in the two-week windows before

[11] Liu et al. (2014) and Arora et al. (2017) circumvent these constraints because of three characteristics of their empirical approaches. First, their approaches do not require them to observe the free version's launch. Second, they extend their observation period until the end of their data period, even if this lies many weeks or months after the launch of the free version. As a result, they observe apps over a longer time period and thus have more variation in download ranges. As we discuss in more detail in §3, in our case, we are concerned that such an approach would make identification of a causal effect difficult. Third, Liu et al. (2014) focus only on apps that were ranked at least once. These are obviously popular apps, suggesting that if they are followed over a long time period, one would observe a significantly greater variation in downloads than for the majority of apps that have never been ranked. Instead, we want to ensure that our results apply to a broad range of freemium apps.

[12] Garg and Telang (2013) build on this approach and calibrate the relationship between app ranks and sales. Ghose and Han (2014) calibrate the relationship between app ranks and sales using additional data that contain rank and download information from an app store.

[13] Kummer and Schulte (2019) demonstrate a close relationship between the number of ratings and the app-ranking-based demand measures, and suggest that using app rankings as the demand measure limits the scope of the analysis because rankings are available only for the most successful apps.

[14] In the remainder of the paper we simplify the language for ease of exposition. When referring to the demand for the paid version, we refer to demand as proxied for by the number of ratings, even if we do not explicitly spell this out.

and after the free version's launch. Clearly, the paid version received more incremental ratings after the introduction of the free version, suggesting a relationship between the free version's launch and the paid version's demand. The figure also indicates that there is significant variation in the number of ratings across apps.

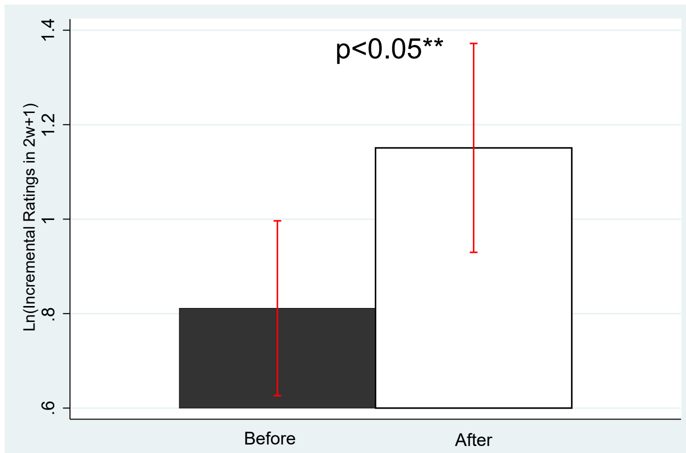

Figure 1: ln (Number of Incremental Ratings +1) in a Two-week Window

This analysis does, of course, not control for other potentially confounding factors. It is possible that the relationship in Figure 1 simply reflects a time trend. To address this concern, we select control apps for each freemium app and estimate the effect of the free version's introduction in a difference-in-difference framework, controlling for a set of fixed effects as well as a set of covariates. Further, we present a battery of robustness checks which illustrate that the results are robust to alternative definitions of the time window and other marketing actions related to product, price, and promotion.

One might also be concerned that a developer would introduce a free version strategically at the time when they expect a rise in demand. There are two reasons why we believe this selection effect is limited. First, many developers in the data are small developers, who are likely to have little knowledge on whether and when it would be effective to introduce a free version for the paid app. Second, even if the developers had enough knowledge to strategically decide whether to introduce a free version, this would influence only the selection of apps but not the precise timing of launch, because Apple first reviews and approves each individual app before making it available. In the data period, there was considerable uncertainty for developers over an app's review time and in turn the precise launch date. While app review times are not usually publicly available, a now-defunct website `https://appreviewtimes.com/` tracked App Store review times crowdsourced from developers from 2011 to 2018, because the founder of the website noted that "from the very start of the App Store, until May 2016 the length of an app review was often very lengthy, causing problems for developers trying to schedule releases or set the expectations of their clients". We obtained from the owner all the data he collected when the website was active. 9,617 review times were reported in our data period, ranging between 0 and 49 days with a mean of 7.05 days and a standard deviation of 4.05 days (see Figure 2 for boxplots of the review times by month). This high variation makes it unlikely that a developer could determine with precision the day or week of a launch to coincide with a time when demand is expected to be higher.

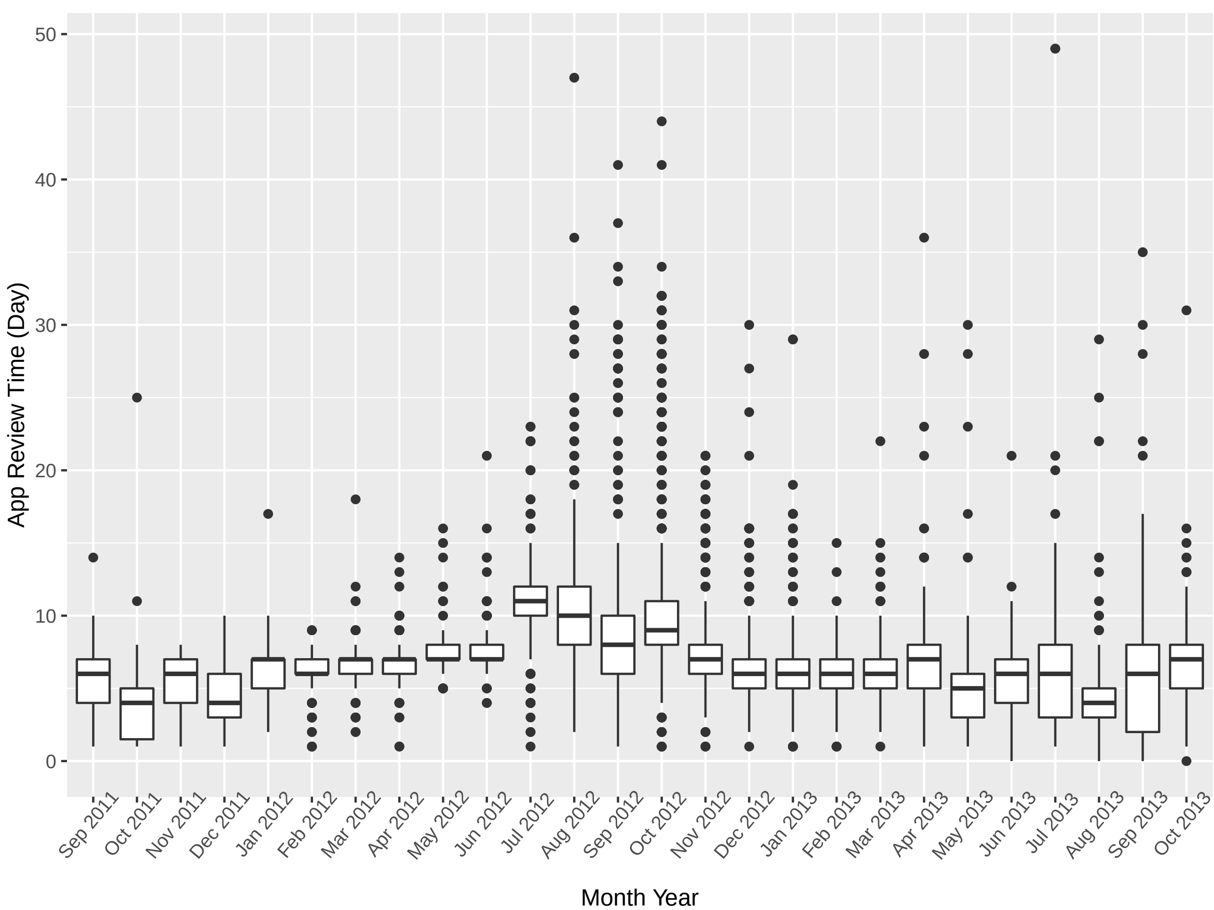

Figure 2: Distribution of App Review Times by Month. Each box is drawn from the 25th percentile to the 75th percentile and the horizontal line in the middle denotes the median.

3.2 Empirical Strategy

Ideally, to estimate the effect of the launch of a free version on the existing paid version's number of ratings, we would compare the paid version's number of ratings in the presence of a free version to its number of ratings in the absence of a free version, holding everything else constant. However, we cannot simultaneously observe for the same paid version of an app the number of ratings in the presence and in the absence of a free version. As a result, we focus our analysis on a tight time window around the launch and compare the paid version's daily incremental number of ratings before the launch of the free version to its daily incremental number of ratings after the launch. We include app fixed effects to control for time-invariant app features that can potentially influence rating accumulation.

To control for unobservable time-varying factors that can potentially influence rating accumulation, we construct a set of control apps for each freemium app based on most of the observable app characteristics and estimate the model in a difference-in-difference framework. Specifically, for each freemium app, we look for candidate control apps among paid apps that meet the following four criteria: (1) The control app is in the same genre as the freemium app; (2) the control app is not developed by the same developer as the freemium app (to ensure no strategic interactions between the control and freemium apps); (3) the time difference between the launches of the control app and the freemium app is less than six months; and (4) if the freemium app was updated during the two weeks before the free version's launch, the control app was likewise updated, alternatively if the freemium app was not updated, the control app needs to not have had an update during that time period. Within these candidate apps, we conduct a propensity score matching based on the values of six variables at both two weeks and one week before the free version's launch, that is, we evaluate control apps based on whether they are similar to the freemium app in their characteristics on both of these dates. The variables include the app's price, and the number of 1-, 2-, 3-, 4- and 5-star ratings. Among candidate apps whose propensity score distance to the freemium app is less than 0.25, we select the top ten apps (or all apps if fewer than ten apps satisfy the criterion) as control apps for the freemium app. We identify suitable control apps for the vast majority of freemium apps (94%). Table A.2 in Online Appendix B shows the similarity between the control apps and freemium apps.

For the purpose of our estimation, we denote a freemium app and all its control apps as an *app group*. A control app may appear in multiple app groups if it is matched to multiple freemium apps, but because the free version's launch date varies across freemium apps, the time windows that are matched to freemium apps and so contributed to the data, vary across app groups.[15]

Figure 3 shows for freemium and control apps, the average daily number of ratings around the free version's launch date. For control apps, we first calculate the average for all control apps within an app group, and then calculate the average value across all app groups. Clearly, following the free version's launch, there is an increase in the daily number of ratings for its paid counterpart, relative to the number of ratings of the control apps. The difference in trends prior to the free version's launch is insignificant (p=0.55).

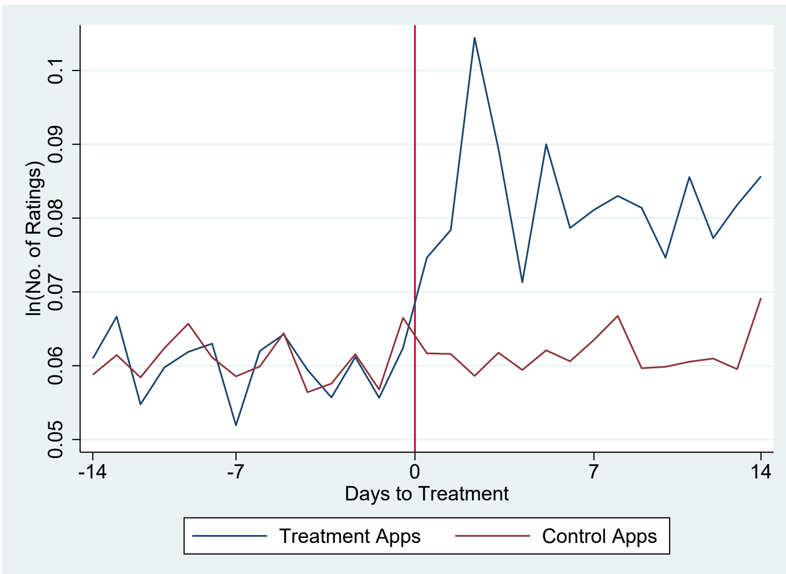

Figure 3: Daily Average Number of Ratings for Freemium and Control Apps

Based on the control apps selected, we specify a stacked difference-in-difference model as follows:

$$ln\left(N_{ijt}+1\right)=\beta F_i T_{jt}+X_{i,t-1}\gamma+\delta_{ij}+\eta_{jt}+\varepsilon_{ijt}, \tag{1}$$

where i denotes a paid app, j denotes the group of freemium-control apps, t denotes day. N_{ijt} denotes the number of ratings accumulated by app i of app group j on day t which we use to proxy for demand. We specify our dependent variable in log-form as the data are skewed. F_i is a binary variable which takes the value of 1 if app i is a freemium app, and takes the value of 0 otherwise. T_{jt} is a binary variable which takes

[15]Because of this variation in time windows across freemium apps, we allow each control app to be matched with multiple freemium apps, that is we use matching with replacement. We restrict, however, each candidate control app to be used as a control app no more than ten times to ensure that the results are not driven by a small group of apps.

the value of 1 if t is in the window after the launch of the free version of the freemium app in app group j. $X_{i,t-1}$ denotes the vector of control variables for app i measured on day $t-1$, including ranking on top charts, price, the total number of ratings app i had accumulated, and the average star rating of the app. δ_{ij} is a fixed effect that is specific to app i and group j (in the results tables, we refer to this as *App - App group FE*), which captures unobservable characteristics to app i and app group j. η_{jt} is a fixed effect that is specific to app group j and day t (referred to as *App group - Day FE*), which captures factors such as favorable demand periods that are app group specific. ε_{ijt} is an idiosyncratic error term. Because the total number of ratings accumulated up to day $t-1$ is included in $X_{i,t-1}$ and it is a function of the number of new ratings on day $t-1$, $N_{i,t-1}$, error terms might be serially correlated. We therefore cluster standard errors at app level.

3.3 Main Results

As discussed in §3.2, our identification relies on the difference-in-difference specification with both freemium and control apps. Because the OLS model cannot control for unobservable time-varying factors that influence rating accumulation, this difference-in-difference specification forms the body of the empirical analysis. Since, however, there is a small fraction of freemium apps for which we were unable to find suitable control apps, we first present in §3.3.1 the results based on a first-difference approach, abstracting from our control group, for the full sample of freemium apps. In §3.3.2, we then present the difference-in-difference results. Finally, in §3.3.3 we discuss the difference-in-difference results with alternative dependent variables.

3.3.1 First-Difference Results

Before presenting the difference-in-difference estimation results, we first present in Table 5 the first-difference estimates for only freemium apps, where we control for app fixed effects and day fixed effects. Columns (1) and (2) report OLS estimates for a time window of two weeks before the launch of the free version and two weeks after the launch. We include in this analysis 1,696 apps as we need to ensure the paid version's number of ratings is observed in the entire time window before and after the free version's launch. This means that the gap between the release dates of the two versions must be at least two weeks, and the free app must be launched at least two weeks before the end of our data period.

Table 5: First-Difference Results

Variables	ln (Number of Ratings of Paid Version+1)			
	OLS, 2-week window		OLS, 1-week window	Type I Tobit, 2-week window
	(1)	(2)	(3)	(4)
T (After new version launch)	0.0185***	0.0291***	0.0315***	0.211***
	(0.00426)	(0.00362)	(0.00371)	(0.0522)
Lag (Ranked on top 10)		1.479***	1.967***	2.873***
		(0.282)	(0.527)	(0.376)
Lag (Ranked on top 11-20)		1.592***	1.980***	2.698***
		(0.273)	(0.584)	(0.528)
Lag (Ranked on top 21-50)		1.337***	0.645	3.105***
		(0.337)	(0.412)	(0.299)
Lag (Ranked on top 51-100)		0.770***	0.650***	2.431***
		(0.168)	(0.207)	(0.296)
Lag (Ranked on top 101-)		0.111***	0.0863***	1.270***
		(0.0272)	(0.0280)	(0.0973)
Lag (Price ($))		–0.00272	–0.00135	–0.0133
		(0.00259)	(0.00153)	(0.0340)
Lag (ln (Total # ratings+1))		0.0139	–0.0745	0.733***
		(0.0500)	(0.0696)	(0.0405)
Lag (If rated)		–0.629***	–0.388	–2.126***
		(0.225)	(0.326)	(0.455)
Lag (Average star rating)		0.0941**	0.0638	0.236**
		(0.0388)	(0.0653)	(0.0941)
App FE	Yes	Yes	Yes	
App RE				Yes
Day FE	Yes	Yes	Yes	Yes
Number of Observations	48,785	48,785	31,644	48,785
Number of Apps	1,696	1,696	2,143	1,696
Adj. R-squared	0.5005	0.5319	0.5292	-

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Column (1) controls only for app and day fixed effects. It indicates that the paid version receives more ratings in the window after the launch of the free version. Column (2) shows that the result is robust to including the full set of controls. Not surprisingly, the number of ratings increases with the average star rating.

We next assess the magnitude of the effects. Denote the daily number of ratings before and after the free version's launch as N_0 and N_1 respectively. With β being the estimated coefficient on treatment (i.e., the free version's launch), we obtain $ln(N_1+1) - ln(N_0+1) = \beta$, hence $ln\left(\frac{N_1+1}{N_0+1}\right) = \beta$ and $N_1 = exp(\beta)(N_0+1) - 1$. We can then calculate N_1 given the values of N_0 and β. We measure the effect size as the percentage change in the daily number of ratings, $\frac{N_1-N_0}{N_0}$, which depends on the daily number of ratings prior to the free version's launch, N_0. We focus on the estimate obtained in Table 5, Column (2), which is 0.0291. It suggests that if N_0 is at the mean (0.505), then all else equal, after the launch of the free version, N_1 increases to 0.550, that is an increase by 8.9%.

In order to check that our results are robust to the specification of the observation window around the time when the free version was launched, Column (3) reports results based on a time window of one week before the launch and one week after the launch. We find the effect holds, and the effect size is not statistically different from the effect size under the two-week window (p=0.62).

To deal with the concern that many apps receive no rating on a given day, we replicate Column (2) with a Type I Tobit model. The results reported in Column (4) are qualitatively similar to those in Column (2).

3.3.2 Difference-in-Difference Results

Table 6 shows the difference-in-difference estimation results for the two-week window (Column 1) and the one-week window (Column 2). We use the same set of apps for the one-week analysis as for the two-week analysis to take advantage of the more granular matching based on two weeks of pre-launch data. Effect sizes are similar to those in Table 5. The estimate in Column (1) is very similar to that in Column (2) of Table 5. It suggests that if the daily number of ratings before the free version's launch is at the mean (0.505), then all else equal, after the launch of the free version, the daily number of ratings increases to 0.550 that is an increase by 8.9%. In a further specification, we estimate separately the effect for each day within the two weeks before and after the free version's launch. Figure 4 plots the coefficients when allowing the effect to vary by day, and consistent with Figure 3, during the two weeks after the free version's launch, the daily number of ratings for its paid counterpart increases, and the effect seems stable.

As discussed, the interpretation that this effect is causal holds under the assumptions that the firm cannot time precisely the date for the free version to be approved and launched by App Store and that, within this brief time period, there are no other factors correlated with the launch which only affect demand of the freemium app but not the demand of the similar control apps. The first assumption is supported by Figure 2. We will discuss the validity of the second assumption in §3.4.

Table 6: Difference-in-difference Results

Variables	ln (Number of Ratings of Paid Version+1)	
	(1) 2-week window	(2) 1-week window
T (After new version launch)$\times$ Freemium app	0.0289***	0.0333***
	(0.00429)	(0.00497)
Lag (Ranked on top 10)	1.311***	1.197***
	(0.184)	(0.211)
Lag (Ranked on top 11-20)	1.064***	0.754***
	(0.139)	(0.179)
Lag (Ranked on top 21-50)	0.716***	0.654***
	(0.0901)	(0.101)
Lag (Ranked on top 51-100)	0.583***	0.598***
	(0.0671)	(0.0779)
Lag (Ranked on top 101-)	0.0977***	0.0703***
	(0.00943)	(0.0115)
Lag (Price ($))	–0.00225***	–0.00202**
	(0.000840)	(0.000841)
Lag (ln (Total # ratings+1))	–0.0454	–0.221***
	(0.0440)	(0.0756)
Lag (If rated)	–0.107	0.164
	(0.123)	(0.208)
Lag (Average star rating)	0.0235	0.0450
	(0.0189)	(0.0349)
App-App group FE	Yes	Yes
App group-Day FE	Yes	Yes
Number of Observations	402,733	222,640
Number of App Groups	1,594	1,594
Adj. R-squared	0.725	0.741

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

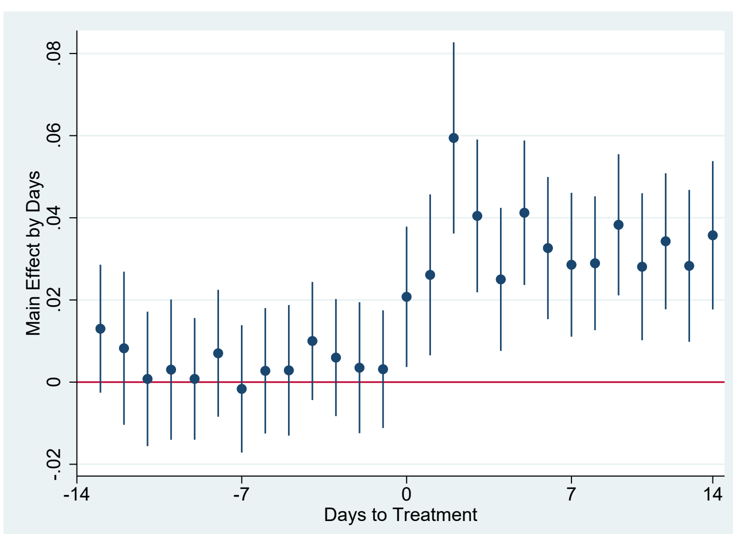

Figure 4: Coefficients by Day

3.3.3 Difference-in-Difference Results with Alternative Dependent Variables

We replicate the difference-in-difference analysis with an app's ranking, instead of the number of ratings, as an alternative outcome variable. We focus on the 26 freemium apps for which we observe a ranking on each of the fifteen days within the window of one week before and one week after the launch.[16] Column (1) of Table 7 illustrates that ranking improves after the launch of the free version, consistent with our prior finding.[17]

Next, we replicate the difference-in-difference analysis with data from Google Play. We first implement the same matching procedure as described in §3.2 to identify control apps for each freemium app. We then estimate Equation (1) with the 174 freemium apps in Google Play along with their respective control apps. Column (2) of Table 7 demonstrates that the results are consistent, and the effect size is not statistically different from the effect size for App Store.

Table 7: Robustness Checks on Alternative Dependent Variables

Variables	(1) ln (ranking)	(2) ln(Number of Ratings of Paid Version+1) in Google Play
T (After new version launch)× Freemium app	–0.200***	0.0269***
	(0.0640)	(0.00972)
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes (except ranking)	Yes
App-App group FE	Yes	Yes
App group-Day FE	Yes	Yes
Number of Observations	2,805	41,791
Number of App Groups	26	174
Adj. R-squared	0.913	0.765

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

3.4 Robustness Checks

In this subsection, we explore whether the effect could arise from developers systematically coordinating other marketing activities (e.g., changes to existing product's design or price, advertising and App Store

[16]As described in §2.6, there is a great variation in the ranking of apps. This means that many apps are ranked among the top 1000 apps on one day and drop out of the list on the subsequent day. But for the purpose of the estimation, we need a complete ranking history within the window. If we were to use the window of two weeks prior to and after the launch, the data set would reduce to only 18 apps.

[17]Note that to conserve space, in Table 7 and subsequent tables, we do not report the estimates associated with other control variables, which are qualitatively similar to those in Column (1) of Table 6.

promotion) with the launch of the free version.

As discussed in §3.1, while the decision to introduce a free version may be endogenous, the precise timing of the launch of the app cannot be determined directly by the developer. Therefore, it would be difficult for the developer to coordinate other actions to precisely fit into the short window following the launch. Nevertheless, we provide additional evidence supporting that the effect we observe is due to the launch of the free version and unlikely to be primarily driven by other marketing actions correlated with the launch.

Change of existing version's design: One alternative explanation for the positive spillover effect is that app developers simultaneously improve the existing version when launching the new version, thus enhancing the existing version's demand. If this were the case, our estimates might reflect an improvement in the paid version rather than an effect of the addition of the free version. To alleviate this concern, we test whether the result also holds for apps without updates in the four-week window around the free version's launch by replicating the analysis in Column (1) of Table 6 with this subset of apps. Results are presented in Column (1) of Table 8 and the effect holds.[18]

Change of existing version's price: We also check whether the results hold for apps where there is no price change for the paid app. If the developer lowered the price of the paid version at the same time that they launched a free version, we might observe increased sales of the paid version coinciding with the launch of the free version simply due to the price change of the paid version. During the period of two weeks before and two weeks after the launch of the free version, 27.2% of apps in Table 6 had at least one price change. To ensure that price changes did not influence our results, we replicate the analysis in Column (1) of Table 6 excluding any apps that had a price change within the four-week window around the launch. Table 8, Column (2) shows that the results hold and the effect size is not significantly different from that in Column (1) of Table 6 (p=0.52). Table 8, Column (3) shows the results hold when focusing on apps that had neither

[18]We note that the effect size is lower than that in Column (1) of Table 6, possibly because apps that update more frequently also differ from other apps along other characteristics (such as star rating and number of ratings). To explore this possibility, we calculate app updating frequency at the app-week level, and regress the average star rating on update frequency, controlling for app and week fixed effects . The coefficient of update frequency is positive and significant (est=0.250, s.e.=0.041). Similarly, we regress the logarithm of the number of ratings on update frequency, controlling for app and week fixed effects. Again, the coefficient of update frequency is positive and significant (est=2.210, s.e.=0.075). This pattern suggests that apps that did not update during the 4-week window around the free version's introduction might be less attractive to consumers, resulting in a smaller effect size compared with the effect size for the entire sample. §4.1further discusses how the average star rating and the number of ratings moderate the effect size.

updates nor price changes within the four-week window around the launch.[19]

Table 8: Robustness Checks on Changes on Existing Version

Variables	ln (Number of Ratings of Paid Version+1)		
	(1) Exclude apps with updates on existing version	(2) Exclude apps with price changes	(3) Exclude apps with updates or price changes
T (After new version launch) × Freemium app	0.0118**	0.0253***	0.00861**
	(0.00488)	(0.00417)	(0.00342)
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes	Yes	Yes
App-App group FE	Yes	Yes	Yes
App group-Day FE	Yes	Yes	Yes
Number of Observations	170,511	267,433	127,645
Number of App Groups	712	1,161	571
Adj. R-squared	0.764	0.737	0.801

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Advertising: If developers typically advertise the free version at its release, this could likewise lead to greater awareness of the paid version. We provide five pieces of evidence from the industry and our data that suggest our results are unlikely an outcome of the change in advertising levels.

First, during the period of our data, advertising was not yet a popular means for app developers to promote apps. In October 2012, Facebook was the first firm to enter the app install ad market.[20] Google and Twitter both entered this market only in 2014 and Pinterest in March 2017.[21] eMarketer did not start monitoring app install ad spending until 2013[22] while Sensor Tower, a leading company in app intelligence, did not start monitoring app advertising until March 2015.[23] Together these facts suggest that the app-install ad market was insignificant in the period we study and it appears extremely unlikely that it played a significant role in the effect we observe.

Second, we analyze the period before Facebook, the first company to enter the app-install ad market, entered this market in October 2012 separately from the subsequent period. If advertising contributed significantly to the effect, we should observe a significantly higher effect size between October 2012 and October

[19]The effect size is lower than in Column (1) of Table 6, presumably due to similar reasons that apply for Column (1) of Table 8.

[20]An app install ad is a mobile ad that offers a direct click to an app's page in the app store. When people click on the ad on a mobile device, they will be directed to the app store to install the app.

[21]`https://techcrunch.com/2014/11/30/like-advertising-a-needle-in-a-haystack/`; `http://www.adweek.com/digital/pinterest-is-now-offering-app-install-ads-to-all-marketers-will-it-see-sales-growth-like-facebook-and-google/`

[22]`http://www.emarketer.com/Chart/US-Mobile-App-Install-Ad-Spending-2013-2015/167985`

[23]Based on email correspondence with Sensor Tower's Head of Mobile Insights from June 13 to June 21, 2017.

2013 than from September 2011 to September 2012. In Table 9, Column (1) we interact our coefficient of interest with an indicator for whether the observation falls into the first or second year of the data and find that there is no significant difference. Similarly, if we interact the coefficient of interest with a linear weekly time trend, we find this to be insignificant (p=0.54). The fact that there was little change in the effect size over time further supports that advertising is unlikely to have played a significant role.

Table 9: Robustness Checks on Advertising

Variables	ln (Number of Ratings of Paid Version+1)	
	(1) Add time indicator	(2) Smaller developers
T (After new version launch) $\times$ Freemium app	0.0303***	0.0202***
	(0.00599)	(0.00381)
$T\times$ Freemium app $\times$First year of data	–0.00341	
	(0.00847)	
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes	Yes
App-App group FE	Yes	Yes
App group-Day FE	Yes	Yes
Number of Observations	402,733	241,591
Number of App Groups	1,594	977
R-squared	0.725	0.695

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Third, we use the website `https://www.spyfu.com/` to check whether the top 50 developers in our data (measured by their number of apps) that had launched a free version of at least one paid app did search advertising during our data period. We found that none of these developers did search advertising during that time period.

Fourth, we replicate the analysis with a subset of developers that we expect to have particularly limited resources for advertising. Specifically, we focus on apps by developers who at the time of the launch of the free version either had only one app, or had other apps but none of these apps had any ratings. These app developers are presumably small in size, had limited budget, and thus had little capacity for cross promotion. Column (2) of Table 9 illustrates a similar positive spillover effect to that of Column (1) in Table 6. We additionally limit the analysis to apps from developers that have a single app only and again find very similar results (est=0.0276, s.e.=0.0062). These effect sizes are not significantly different from that in Column (1) of Table 6.

Fifth, as a final piece of evidence, we contacted 14 developers of apps in our sample. They all indicated they did little or no advertising during the period of our data.

Together, these results lead us to conclude that it is highly unlikely that developers systematically coordinated advertising strategies to coincide with the launch of the free version, leading to an effect of the free version's launch on the paid version's demand.

Featuring in App Store: If the launch of the free version would lead the app to be featured in App Store, then it would enhance the app's visibility and in turn increase demand (Liang et al., 2019). In our data period, App Store had a section entitled "New and Noteworthy". This section was separate from rankings and Apple chose the apps to be featured there. As a result of being featured, apps typically receive a lot of attention and a significant increase in demand. If the launch of a free version would motivate Apple to feature the app, then we might observe an increase in the number of ratings that is correlated with the launch of the free version but not a result of a spillover effect from users of the free version. We do not observe whether apps were featured. However, to be selected by Apple for this section, an app needs to be original and outstanding and receive media attention.[24] Because the free version is inferior to the existing paid version, it is extremely unlikely that Apple would ever feature the free version and not the paid version or that Apple would feature the paid version simply because of the release of the free version. It is therefore not plausible that the effect we observe would be a result of the free version being featured in App Store. Note also that anecdotal evidence suggests that being featured as "New and Noteworthy" typically leads to immediate big jumps in demand, larger than those we observe following the launch of a free version.[25]

Solicited Ratings: Another possible driver to our result would be that concurrently with the launch of the free version, developers systematically solicit consumers to rate apps. Such a strategy could lead to an increase in the number of ratings, holding constant the number of downloads. We check this possibility. First, if developers were more likely to solicit ratings shortly after the launch of the free version, then we might expect different types of consumers to rate the app than previously. If users with a different usage experience of the app evaluate it, the valence of ratings should change. We estimate Equation (1) with the star

[24] See, for example: `http://www.appfreak.net/how-to-get-your-app-into-the-new-and-noteworthy/`, `https://growthtower.com/7-ways-to-get-featured-on-the-app-store/` or `https://www.quora.com/How-do-you-get-onto-new-noteworthy-on-the-App-Store`.

[25] One developer stated that "My app, called Citrus, was featured once on the German iOS AppStore... Our App Store page got around 50,000 views and 4,000 downloads for the week that it was up. Before that we were getting about 2–4 downloads a day." (see `https://www.quora.com/What-is-it-like-to-have-your-app-featured-on-the-App-Store`). For similar evidence, see `http://blog.anylist.com/2012/08/app-store/`. We acknowledge that while such anecdotal evidence supports the notion that our results are unlikely driven by app being featured in App Store shortly after the free version's introduction, it does not completely rule out this possibility.

rating as the dependent variable, and the estimated coefficient of free version launch is 0.0019 (s.e.=0.0045), indicating there is little difference in the valence of ratings before and after the launch of the free version. In addition, we tested whether the distribution of high (4 or 5)/low (1 to 3) ratings is the same before and after the launch of the free version. The average share of high ratings is 81.2% before the launch and 82.4% after the launch. A t-test indicates that the null hypothesis that the two shares are equal cannot be rejected (p=0.31). Overall we find that there is little difference in the valence of ratings before and after the launch of the free version.

Further, we were unable to find any direct evidence suggesting that a developer would benefit from requesting consumers to rate an app specifically around the point in time when a new version is launched. Indeed, recommendations in online forums that suggest strategies when firms should request reviews largely relate to usage patterns of a specific user and not to the broader product strategy of the developer. None of the recommendations relates to the launch of a free version.[26]

Finally, during our observation period, consumers were only able to review apps that they had downloaded,[27] thus all the ratings on the paid version are provided by consumers who have purchased the paid version, instead of being provided by consumers who have only downloaded the free version and accidentally rated the paid version.

In sum, it seems unlikely that the effect we observe across a large number of developers and apps can be fully attributed to increased promotion or advertising when the free version was released.

4 Exploring the Mechanism

The result that the introduction of a free version increases the demand of the paid version is consistent with prior literature on sampling: the free version allows consumers who are uncertain about the quality of the product to sample before purchasing the full version. At the same time, the availability of a free version may enhance visibility of the paid version in App Store, and thus facilitate product discovery. We explore sampling and enhanced product discovery by consumers as potential mechanisms that can explain why the

[26] For example, it is suggested to consider how long and how extensively a user has been using an app, such as making a request at the moment when a user has finished a positive interaction, or when updating the app but not immediately when an app is opened. See, for example: `https://www.quora.com/What-analytics-are-iOS-app-developers-using-to-identify-the-best-time-to-prompt-for-a-user-rating-in-order-to-generate-positive-results`; `https://www.quora.com/Whats-the-best-way-to-ask-an-app-user-to-rate-this-app-without-berating-them`.

[27] See, for example: `https://www.macrumors.com/2011/05/03/apple-no-longer-accepting-app-store-reviews-for-redeemed-promo-codes/`.

launch of the free version increases demand for the paid app.

4.1 Sampling

Before purchasing a product, consumers are often uncertain about product attributes and the product's quality. Such uncertainty is a major impediment to sales in online markets (Appel et al., 2019; Hong and Pavlou, 2014). One way to resolve this fit uncertainty is for consumers to try the product and learn about the product quality prior to purchasing, thus reducing the uncertainty around the product value.

Several patterns in our data are consistent with sampling as a mechanism behind the spillover effect on the paid version. First, if consumers upgrade to the paid version of an app after sampling and enjoying the free version, then we should find a positive correlation between the free version's incremental star ratings and the paid version's incremental number of ratings as the latter proxies for demand. We indeed find a correlation of 0.17 ($p < 0.0001$). In addition, we also find a correlation of 0.24 ($p < 0.0001$) between the paid version's daily incremental number of ratings and the free version's daily incremental number of ratings, a pattern which is consistent with users purchasing the full version after sampling the free version. Note that if the free version would solely aid product discovery – that is to help consumers find the paid version in the store but without consumers sampling the free version – then we would not expect these relationships.

Next, we replicate the analysis in Column (1) of Table 6 with a subset of apps that have identical descriptions for the paid version and the free version. The point estimate is 0.0327 with a standard error of 0.0058, which is very close to that in Column (1) of Table 6. For these apps, consumers cannot observe the difference between the two versions before downloading, and we would expect them to first download the free version. While playing the game, they would be prompted to download the enhanced paid version. The fact that for this sample we still find a similar effect further supports that sampling is a mechanism at work.

We next conceptualize when a consumer samples and the effect of sampling on purchase of the paid version. We then take this framework to our data.

4.1.1 Conceptual Framework

While sampling may help consumers learn the value of a product, sampling is not without cost. Consumers may incur cost of time in acquiring the free sample (e.g. downloading the app) and in getting familiar with its functionalities (Cheng and Liu, 2011). Further, downloading even a free app uses data allowance and storing it uses space on a consumer's device, introducing further opportunity cost for a consumer. As a result, a consumer needs to weigh the opportunity cost associated with acquiring, setting up and learning about the product with the expected utility from sampling (Lee and Tan, 2013). This expected utility of sampling the free version accounts for utility that may arise from upgrading to the paid version.

As sampling is not costless, a consumer's decision on whether or not to sample depends on the expected benefits. Assume the cost for the consumer to sample is c which captures any opportunity cost, such as cost related to time or storage requirements. Also, assume the cost of purchasing the paid version to be p, which includes the price of the product as well as any opportunity cost, such as that associated with time or storage.[28] Assume that a consumer knows the value of the free version, v_f, which varies across consumers. Specifically, we assume $v_f = \theta \bar{v}_f$, where $\theta \in (0,1)$ represents consumer "type". However, prior to sampling, the consumer has uncertainty about the paid version's value, v_p.[29] Instead, the consumer only knows that v_p is uniformly distributed on $\left[\theta \underline{v}_p, \theta \bar{v}_p\right]$, where $\underline{v}_p < p < \bar{v}_p$, and $\bar{v}_p > \bar{v}_f$. After the consumer has used the free version, the value of v_p is revealed. We assume utility is a linear function of the valuation, and for simplicity assume the consumer obtains utility $U_f = v_f - c$ from the free version and utility $U_p = v_p - p$ from the paid version.

In the absence of a free version, a consumer will purchase the paid version if the expected utility of the paid version is larger than zero, that is, $E(U_p) = E(v_p) - p > 0$. $E(U_p)$ is also the expected utility of purchasing the paid version directly in the presence of a free version. Under the assumption that v_p is uniformly distributed on $\left[\theta \underline{v}_p, \theta \bar{v}_p\right]$, a consumer of type θ would purchase if $\theta > \frac{2p}{\underline{v}_p + \bar{v}_p}$.

When the firm offers a free version, a consumer can sample the free version to learn about the value of v_p. After sampling, the consumer upgrades to the paid version if the difference between the value of the paid version and the value of the free version exceeds the price of the paid version, $v_p - v_f > p$. Thus, the probability of upgrading is $Prob(v_p - v_f > p) = \frac{\theta \bar{v}_p - \left(p + \theta \bar{v}_f\right)}{\theta \bar{v}_p - \theta \underline{v}_p}$. Upon upgrading, the consumer obtains utility $v_p - p$. If the difference between the value of the paid version and the free version is smaller than the price of the paid version, $v_p - v_f < p$, then the consumer does not upgrade and only receives utility v_f from the

[28]While we take price as exogenous without modeling the app developer's pricing decision, our results hold as long as the market is not saturated, i.e., when the price of the paid version is higher than the expected valuation for at least a fraction of consumers.

[29]The consumer knowing the value of the free version, v_f, but not the value of the paid version, v_p, is a simplification for ease of analysis. The underlying assumption is that the consumer has a more accurate belief of v_f than of v_p, that is a tighter distribution around v_f than around v_p. By definition, the free version is a simplified version of the paid version and would have a lower maximum value it could possibly provide, thus leading to a tighter distribution.

free version.

Thus, the expected utility from sampling is:

$$E(U_s) = v_f Prob(v_p - v_f \leq p) + E(v_p - p \mid v_p - v_f > p) Prob(v_p - v_f > p) - c \tag{2}$$

For consumers to sample the free version, two conditions need to be satisfied. First, the expected utility of sampling needs to be greater than zero ($E(U_s) > 0$). Second, the expected utility of sampling should exceed the expected utility from purchasing the paid version directly ($E(U_s) > E(U_p)$). For consumers to upgrade to the paid version after sampling, v_p needs to satisfy $v_p - v_f > p$. The condition $v_p - v_f > p$ will be met for consumers with $\theta < \frac{p - v_p}{\bar{v}_f}$. Thus, for these consumers, the introduction of a free version can increase the paid version's demand through sampling if $E(U_s) > 0$ and $E(U_s) > E(U_p)$.[30]

It can be shown that $E(U_s)$ increases with $\underline{v}_p$, and $E(U_s) - E(U_p)$ decreases with $\underline{v}_p$ (for details, see Online Appendix C). Therefore, the conditions for sampling, $E(U_s) > 0$ and $E(U_s) > E(U_p)$, are likely to be met simultaneously when $\underline{v}_p$ is neither too large nor too small.

If $E(U_s) > 0$ but $E(U_s) < E(U_p)$, then consumers would directly purchase the paid version regardless of whether there is a free version. The introduction of a free version would not affect the demand for the paid version.

4.1.2 App Quality

One factor that plausibly affects whether a consumer expects sampling to be beneficial is the quality information they receive about a product. Such quality information may shift the distribution of the consumer's belief about the value of the paid version, v_p. We discuss how publicly available information on app quality in the marketplace may shift the lower bound of the consumer's belief ($\underline{v}_p$) which then in turn may affect the consumer's decision to sample.

Our result in §4.1.1 indicates that sampling is more effective when $\underline{v}_p$ is neither too large nor too small. The intuition is as follows: When a consumer receives information on app quality indicating a very low value of $\underline{v}_p$, they expect to receive a low utility, $E(U_p)$, from the product, and the cost of sampling may

[30]For consumers with $\theta > \frac{v_p - p}{\bar{v}_f}$, the introduction of a free version will not increase the paid version's demand because the free version provides enough utility so that an upgrade to the paid version is not justified. Note that it is also possible that a consumer downloads the free version without the intention to upgrade and then is enticed to upgrade. This would happen if $v_f - c > 0$. Because $E(U_s) > v_f - c$, when $v_f - c > 0$, the condition $E(U_s) > 0$ is also met. Therefore, this can be considered as a special case which is in line with our discussion.

exceed the marginal benefit of sampling. It may then not be optimal for the consumer to sample. Put simply, the consumer has such low expectations of product quality that it is not worthwhile to incur the cost of sampling, and as a consequence, offering a free sample will have little effect on the consumer's purchase of the paid version.

A similar pattern occurs when the consumer receives information that indicates a very high value of the lower bound $\underline{v}_p$. If $\underline{v}_p$ is very high, then the consumer's expected utility, $E(U_p)$, is very high and the marginal benefit of sampling may be below the marginal cost of sampling. Put differently, if the consumer expects a product to be of high quality, they may prefer to purchase the paid version directly and avoid the opportunity cost associated with sampling. Again, offering a free sample will have little effect on the consumer's purchase of the paid version.

One piece of public information on app quality that is readily available to consumers in the mobile app market is an app's star rating. If an app is poorly rated, consumers may conclude that its quality is too low to be a good fit and that sampling is not worthwhile. On the other hand, a very high star rating indicates that prior users found the app useful, making it more efficient for consumers to directly purchase the paid version and avoid the cost of sampling. In consequence, sampling is more beneficial for apps that have a moderate star rating.

We thus expect an inverted-U shaped relationship between the effect of sampling and the average star rating of an app. We examine empirically how the treatment effect is moderated by the paid version's average star rating on the day prior to its free counterpart's launch. That is, we estimate Equation (1) adding interaction terms between the free version's introduction and the average star rating and its squared term. The estimation results in Column (1) Table 10 demonstrate indeed an inverted U-shaped relationship between the effect size and the average star rating.

The number of ratings a product has obtained has been suggested as an alternative measure of product quality (Schoenmueller et al., 2020). We therefore present in Column (2) a robustness check to the results in Column (1) of Table 10, using the number of ratings of an app as a quality measure. Again, we find similar results suggesting an inverted U-shaped relationship between the effect size and the number of ratings.

Table 10: Moderating Effect of Star Rating and Number of Ratings

Variables	ln (Number of Ratings of Paid Version+1)	
	(1)	(2)
T(After new version launch) $\times$ Freemium app $\times$Star rating prior to free version release	0.176**	
	(0.0870)	
$T\times$ Freemium app $\times$Star rating prior to free version release^2	–0.0225*	
	(0.0121)	
T $\times$ Freemium app $\times$ ln(Number of ratings prior to free version release+1)		0.0805**
		(0.0374)
$T\times$ Freemium app $\times$ ln(Number of ratings prior to free version release+1)^2		–0.00822**
		(0.00402)
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes	Yes
App-App group FE	Yes	Yes
App group-Day FE	Yes	Yes
Observations	402,733	402,733
Number of App Groups	1,5954	1,594
Adj. R-Squared	0.725	0.725

Robust standard errors clustered at app level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

4.1.3 Difference between Free and Paid Versions

Prior research has demonstrated that the effect of sampling depends on the quality difference between the free and the paid products. For instance, Li et al. (2019) find that when a publisher provides free digital samples to print copies of books, it is beneficial to increase the functionality difference between the high-quality digital sample and the print format. Cheng and Liu (2011) find that offering a feature limited free sample is more suitable for software products that exhibit strong network effect.

In our conceptual framework, the consumer will upgrade to the paid version after sampling if the difference between the paid version and the free version ($v_p - v_f$) is larger than the price of the paid version. Thus, the paid version may provide sufficiently greater utility for consumers to upgrade once consumers are satiated with the free version (Appel et al., 2019). To see if this holds empirically, we examine the moderating effects of version design on the treatment effect. Specifically, we estimate Equation (1) adding interaction terms between treatment and the app description-based design differences described in §2.3: more levels, more modes/themes, more functions/features, social interactions, ad free, and better support. We group apps with the same combination of design differences, and compute robust standard errors clustered at the group level. Table 11 shows the results. Consistent with the sampling explanation, compared with apps with identical descriptions for the paid version and the free version, this effect is more pronounced for apps where the paid version appears to offer greater additional benefits. It is greatest when the paid version offers "more levels", "more functions/features", or "social interactions". It seems plausible that the free version provides consumers an opportunity to try the game on initial levels while additional levels provide sufficient utility for consumers to upgrade. Similarly, "more functions/features" (such as more powerful weapons) attract consumers to upgrade. Finally, the ability to compete, compare scores and communicate with others may give consumers who found they like the initial version significant extra utility, thus making a purchase of the full version attractive. The moderating effect of "more modes/themes" (e.g. a different background picture) is not statistically significant, indicating that consumers value having more modes/themes such as background colors (horizontal differentiation) relatively less than additional levels in the game (vertical differentiation), a result which appears plausible as additional modes/themes do relatively little to change the gaming experience.

Table 11: Moderating Effects of Design Differences based on App Description

Variables	ln (Number of Ratings of Paid Version+1)
T (After new version launch)$\times$ Freemium app	0.0223***
	(0.00285)
$T\times$ Freemium app $\times$ More levels	0.0114*
	(0.00637)
$T\times$ Freemium app $\times$ More modes/themes	–0.00799
	(0.00792)
$T\times$ Freemium app $\times$ More functions/features	0.0189**
	(0.00833)
$T\times$ Freemium app $\times$ Social	0.0285**
	(0.0118)
$T\times$ Freemium app $\times$ Ad free	–0.0162**
	(0.00625)
$T\times$ Freemium app $\times$ Support	–0.0493***
	(0.0156)
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes
App-App group FE	Yes
App group-Day FE	Yes
Number of Observations	402,733
Number of App Groups	1,594
Adj. R-squared	0.725

Robust standard errors clustered at app group level in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Interestingly, the negative moderating effect of "ad free" suggests that for the average consumer, the negative utility associated with advertising significantly reduces the potential benefit of the free version's launch on demand for the paid version. The result seems plausible in view of the broader evidence that many consumers prefer ads rather than paying for content (Papies et al., 2011; Chiou and Tucker, 2013).[31] Finally, the moderating effect of "support" is negative and its effect size exceeds that of the main effect. The estimate suggests that the average consumer does not place much value on support offered by the developer, and the introduction of a free version that is mainly differentiated by providing less support may cannibalize demand of the paid version.

In sum, this set of results supports the explanation that consumers' sampling of the free version leads to the purchase of the paid counterpart. The strategy of releasing a free version to let consumers sample is effective when apps have a moderate level of the average star rating or a moderate number of ratings. In addition, consumers are more likely to upgrade from the free to the paid version when the two versions differ substantially in the value they provide to consumers.

4.2 App Discovery

With the rapidly increasing number of apps on app stores, "the gap between search and find is getting bigger and bigger",[32] making app discovery a problem as it is difficult for any individual app to be discovered by a consumer (Li et al., 2016). At the time of our data, App Store offered close to 200,000 game apps, most of which were by small independent developers. In such a setting, having two versions instead of one increases visibility of the app and makes it more likely to be discovered by a consumer. Thus, the availability of a free version may increase the likelihood of discovery of the paid version. This can impact demand for the paid version in two ways: First, the enhanced visibility may lead consumers to discover the free version, sample the free version and later purchase the paid version; that is discovery may lead to sampling. Second, the availability of the free version may increase a consumer's probability of discovering the paid version and then purchasing the paid version directly, without first sampling the free version.[33]

Ideally, one would explore app discovery based on consumer browsing history in App Store. However,

[31] A study by the Digital Advertising Alliance (DAA) finds more than 85% of respondents prefer an ad-supported Internet instead of having to pay for the content they now get for nothing (see `https://www.consumeraffairs.com/news/survey-finds-85-of-consumers-prefer-an-ad-supported-internet-051216`).

[32] `https://www.forbes.com/sites/allbusiness/2013/08/15/app-discovery-why-cant-anyone-figure-this-out-yet/`

[33] Note that on a high level, the question of discovery is related to research which documents that music streaming services serve as an alternative channel for music discovery, which increases downloads at alternative consumption channels (Aguiar, 2017) or recorded music sales (Kretschmer and Peukert, 2020), or increases volume and diversity of music consumption (Datta et al., 2017). Both their research and ours argue that a wider availability may facilitate discovery and thus increase sales. However, in our context we argue that more variants of a product promote discovery, while prior research focused on the number of channels that made a single product available for purchase or consumption.

such data are not publicly accessible. Instead, we use two features of our data to shed light on whether app discovery is a likely mechanism. First, we turn to a setting where we would expect sampling to matter less. We focus on apps that initially offered a free version and later added a paid version. If app discovery matters, then the introduction of a paid version should increase the probability for the app to be discovered, and thus demand for the original free version. However, sampling matters less in this context because consumers who are interested in this app could have downloaded the free version before the paid version becomes available. For this analysis, we focus on apps where the free version was launched first and the paid version was launched later during our sample period. For each of these free apps, we select a set of control apps from free apps that did not have a paid version introduced in our data period using the procedure described in §3.2. We then estimate the effect of introducing the paid version on the free version's number of ratings in a difference-in-difference framework. The results reported in Table 12, Column (1) indeed indicates a positive effect, suggesting that the availability of an additional version may enhance discovery.

Second, the app discovery mechanism suggests that when the number of apps in a category increases, providing an additional version of an app has increasingly less effect on the probability for an app to be discovered, simply because the "haystack" gets larger. Thus, the marginal benefit of providing an additional version of an app becomes smaller as the genre size increases. In our data, the number of apps per genre varies significantly between 1,405 (Dice) and 40,704 (Puzzle). We estimate the moderating effect of the number of apps in the same genre. Column (2) of Table 12 shows the results for apps where the free version was launched first and Column (3) shows the corresponding results for instances where the paid version was launched first. In both instances, the estimates indicate that adding an additional version of an app has a smaller effect on the demand for the original version when a genre comprises more apps. This is in line with the mechanism related to app discovery, that is apps finding it more challenging to get any consumer's attention if they are among a large number of available options.[34]

In sum, our results support that offering a free version in addition to a paid version can enhance visibility of the app and thus increase demand for the paid version.

[34] We acknowledge that if a larger choice of apps were to go along with a greater value of the outside option available within App Store, consumers facing a larger choice in a genre might be less likely to sample the focal paid app. However, we do not have any empirical evidence to support that the value of the outside option (if that were confined to App Store) increases with the number of apps in a genre. Further, it is less clear how precisely the outside option should be defined here as consumers may allocate their time and money not only across apps but also across other devices and means of entertainment.

Table 12: App Discovery

Variables	ln (Number of Ratings of Free Version+1) for free-first apps		ln (Number of Ratings of Paid Version+1) for paid-first apps	
	(1)	(2)	(3)	(4)
T (After new version launch)× Freemium app	0.0285** (0.0114)	0.0747*** (0.0235)	0.0451*** (0.0102)	
T × Freemium app × Number of Apps in the Genre (100,000)		–0.228*** (0.0787)	–0.0664* (0.0342)	–0.0611* (0.0342)
T× Freemium app ×Star rating prior to free version release				0.175** (0.0870)
T× Freemium app ×Star rating prior to free version release^2				–0.0222* (0.0121)
Other control variables (ranking, price, the total number of ratings, the average star rating)	Yes	Yes	Yes	Yes
App-App group FE	Yes	Yes	Yes	Yes
App group-Day FE	Yes	Yes	Yes	Yes
Observations	80,944	80,944	402,733	402,733
Number of App Groups	281	281	1,594	1,594
Adj. R-Squared	0.776	0.776	0.725	0.725

4.3 Relative Importance of the Two Mechanisms

We aim to provide some insight on the relative importance of the sampling and the discovery mechanisms. At the same time, we acknowledge that any such discussion is a rough approximation. Recall that in the empirical exploration of sampling as a mechanism, we are unable to measure the precise value of $\underline{v}_p$. Instead we use proxies to show that empirical patterns are aligned with our predictions. For the purpose of contrasting here the relative importance of sampling and discovery, we focus on using an app's average star rating prior to the free version's introduction, which we believe to be the most direct proxy on sampling. We proxy discovery through the number of apps in a genre.

To assess the relative importance of sampling and discovery, we present in Column (4) of Table 12 an estimation result that includes as independent variables both star rating and its squared term (which proxy the effect of sampling) and the number of apps in the genre (which proxies the effect of app discovery). For an app whose star rating is at the average level across all apps, we calculate how a change in star rating and a change in genre size would affect our key outcome variable, the logarithm of the number of ratings.

We illustrate the calculation by estimating the effect of a change in one standard deviation of the variables that proxy sampling or discovery. With respect to sampling, the mean and standard deviation of star ratings across the 1,594 freemium apps in our analysis, are, respectively, 4.071 and 0.758. Therefore, for an app with an average star rating, an increase in star ratings by one standard deviation would change the logarithm of the number of ratings by $0.758 \times 0.175 - \left((4.071+0.758)^2 - 4.071^2\right) \times 0.0222 = -0.0171$. Turning to app discovery, the standard deviation of genre size (in 100,000) is 0.119. Therefore, a one standard deviation increase in the genre size would change the logarithm of the number of ratings by $-0.119 \times 0.0611 = -0.00727$. At these points, the relative share of the effect size change induced by sampling is 0.0171/(0.0171+0.00727)=70%.

To evaluate the effect of changes smaller than one standard deviation, we next allow the scale of change to vary from 1% to 100% of the standard deviation, and plot in Figure 5 the changes in effect size that are induced by changes in the two variables. We then plot in Figure 6 the relative share of the effect size change attributed to the star rating, the variable that proxies for the effect of sampling. These figures indicate that, based on our proxies, at small levels of changes, the discovery effect has a somewhat larger contribution than the sampling effect. However, when the changes from the mean values get larger, the sampling effect contributes significantly more to the overall effect. On average, at least for our data and the proxies for sampling and discovery that we use in this study, the sampling effect appears to be larger than the discovery effect because more than half of Figure 6 has the sampling effect contributing to over 50% of the effect change.

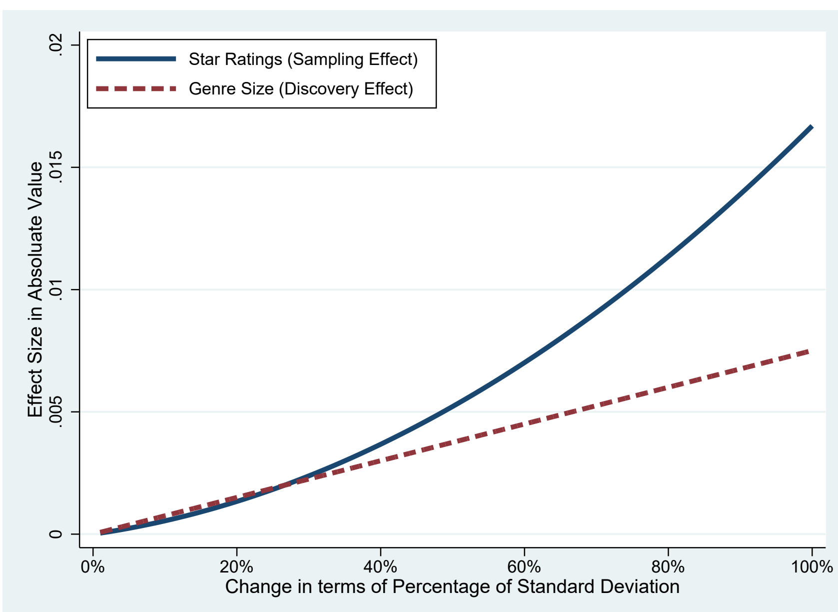

Figure 5: Effect of Changes in Star Rating and Genre Size

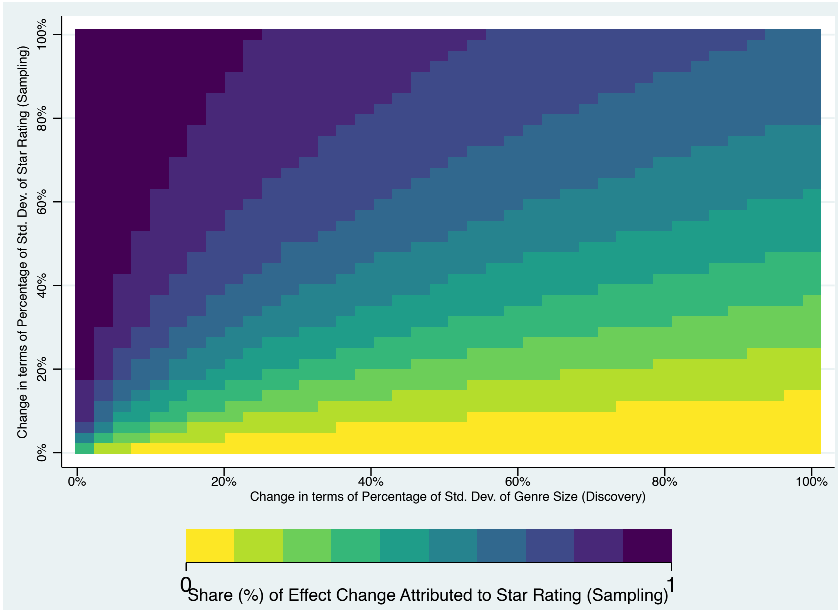

Figure 6: Share of Effect Change Contributed by Star Rating

We emphasize that while this analysis sheds light on the relative importance of the two mechanisms, it relies on the empirical proxies for the mechanisms and functional form assumptions.

5 Conclusion

Freemium pricing has become an increasingly popular business model in the digital economy. While the free version may tempt consumers to purchase the paid version, it may likewise cannibalize the sales of the paid version. In this paper, we aim to measure the overall impact of a freemium strategy using granular data on game apps available in App Store when a developer offers both a free and a paid versions of the same game. As app download data are not publicly available, we use the number of ratings available for an app to proxy app demand. We focus on apps that introduced a free version after initially offering only a paid version. We estimate in a difference-in-difference framework, how this free version's introduction changed the number of ratings of the paid version. We find that, on average, demand for the paid version increased following the launch of the free version. This result holds in a battery of robustness checks and is unlikely to be an outcome of other marketing actions correlated with the launch.

We then explore the mechanisms at work. First, we develop a conceptual framework for sampling of the free version and the effect of sampling on purchase of the paid version, and put forward predictions on when sampling is more effective. We then show evidence from the data that is in line with these predictions, supporting that consumers use free versions to sample and later upgrade to the paid versions. Second, we demonstrate that a new version of an app may increase the initial version's demand by enhancing the app's visibility (and thus its chance to be discovered). Finally, we evaluate the relative importance of the two mechanisms.

Our results are relevant for app developers and, more broadly, digital firms who offer freemium pricing. First, they confirm that a freemium strategy can indeed increase demand for the paid version of a product. At least on average, this spillover effect outweighs any possible cannibalization in the market we study. Second, our results indicate that a freemium strategy would be more effective for products that prior users evaluated as moderately good. Third, our findings suggest that to truly benefit from a freemium strategy, firms need to ensure that the difference in utility for consumer between the free product and the paid product is sufficiently high to induce upgrades. At least in the market for game apps, this appears to be the case when the user can have a substantially enhanced usage experience by playing more levels, enjoying more features, or through social interaction. Fourth, we document that simply the fact that an additional version is available can increase demand of the paid version by making it easier for consumers to discover the app.

While the focus of our empirical analysis is on a setting where the firm offers two separate versions of an app, we expect that many of our results on consumers' decisions to first access a free version and then a paid version translate to other forms of freemium pricing, such as when in-app purchases can be made within a free app. While the technology to implement freemium pricing differs across domains and even within domains changes over time, we believe that the underlying economics and implications for consumers are similar. In all instances, consumers can access a certain set of features or levels for free, or use the product for free for a certain amount of time. Full access to all features or further usage requires payment. Our results suggest that even firms that offer in-app purchases should take care to design the free component of an app in a way that it is still sufficiently attractive for consumers to download, even in the absence of in-app purchases, while at the same time in-app purchases need to provide significant additional value to incentivize consumers to spend money.

That said, it is likely that the respective importance of the sampling and discovery mechanisms may be different in a setting where "freemium" is implemented in a way that in-app purchases can be made within a free app. We expect the sampling effect to generally translate to such a setting. Sampling the "free version" in fact should become even easier for consumers as they no longer need to download a second version of the app. By contrast, the discovery effect may be much less important, or even absent, when only a single version of the app is offered – simply because there is not a second version that captures consumer attention. This suggests that app developers may need to resort to other strategies to direct consumer attention to their apps.

There are several limitations to the present study that represent opportunities for future research. First, we focus on a specific empirical application – the market for game apps during a time period when developers marketing freemium apps offered both a free and a paid version, and in-app purchases within the same app were less common. Our estimation results most directly apply to this specific empirical setting and data period. Exploring in-app purchases may widen our understanding of how consumers respond to freemium offerings and of how firms should best design freemium products. Second, as we lack data on app downloads, we use an app's number of ratings as a proxy for its demand. Given the downward trend in rating accumulation, this approach appears conservative as it likely underestimates incremental downloads as the number of cumulative downloads increases. Nonetheless, using a more precise measure of demand could yield more precise results. Third, we focus on the short-term effect of the free version introduction, as identification of a long-term effect is more difficult and would be significantly less precise due to changes in the app itself, in the market for apps and in patterns of consumer demand over time. It would be interesting to explore in future research the long-term effect of a freemium strategy while accounting for such potential shifts over time. Fourth, while we provide ample evidence that correlated marketing activities are unlikely the driver of the spillover effect as a rule, we cannot fully rule out that in some instances firms may still have tried to support the launch of a free version with marketing activities, which might have had some effect on our estimates. Finally, our findings related to the mechanisms are descriptive instead of causal.

References

Aguiar, Luis. 2017. Let the music play? free streaming and its effects on digital music consumption. *Information Economics and Policy* **41** 1–14.

AppAnnie. 2015. The state of play: A look at the growth of Google Play. Available at https://s3.amazonaws.com/files.appannie.com/reports/The-State-of-Play-Google-IO-2015-EN.pdf.

Appel, Gil, Barak Libai, Eitan Muller, Ron Shachar. 2019. On the monetization of mobile apps. *International Journal of Research in Marketing* .

Arora, Sandeep, Frenkel ter Hofstede, Vijay Mahajan. 2017. To free, or not to free: the implications of offering free versions for the performance of paid mobile apps. *Journal of Marketing* **81**(6) 62–78.

Bawa, Kapil, Robert Shoemaker. 2004. The effects of free sample promotions on incremental brand sales. *Marketing Science* **23**(3) 345–363.

Boudreau, Kevin J. 2018. Amateurs crowds & professional entrepreneurs as platform complementors. Working Paper 24512, National Bureau of Economic Research.

Brecko, Kristina. 2020. New features free of charge? using price to sort consumers among legacy software versions. *Working Paper* .

Chellappa, Ramnath K., Amit Mehra. 2017. Cost Drivers of Versioning: Pricing and Product Line Strategies for Information Goods. *Management Science* .

Chen, Hong, Wenjing Duan, Wenqi Zhou. 2017. The interplay between free sampling and word of mouth in the online software market. *Decision Support Systems* **95** 82–90.

Cheng, Hsing Kenneth, Yipeng Liu. 2011. Optimal Software Free Trial Strategy: The Impact of Network Externalities and Consumer Uncertainty. *Information Systems Research* **23**(2) 488–504.

Chevalier, Judith, Austan Goolsbee. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing & Economics* **1**(2) 203–222.

Chiou, Lesley, Catherine Tucker. 2013. Paywalls and the demand for news. *Information Economics and Policy* **25**(2) 61–69.

Datta, Hannes, George Knox, Bart J. Bronnenberg. 2017. Changing their tune: How consumersâ adoption of online streaming affects music consumption and discovery. *Marketing Science* **37**(1) 5–21.

Dey, Debabrata, Atanu Lahiri. 2016. Versioning: Go vertical in a horizontal market? *Journal of Management Information Systems* **33**(2) 546–572.

Dey, Debabrata, Atanu Lahiri, Dengpan Liu. 2013. Consumer learning and time-locked trials of software products. *Journal of Management Information Systems* **30**(2) 239–268.

Foubert, Bram, Els Gijsbrechts. 2016. Try it, you'll like it - or will you? the perils of early free-trial promotions for high-tech service adoption. *Marketing Science* **35**(5) 810–826.

Garg, Rajiv, Rahul Telang. 2013. Inferring app demand from publicly available data. *MIS Quarterly* **37**(4) 1253–1264.

Ghose, Anindya, Sang Pil Han. 2014. Estimating demand for mobile applications in the new economy. *Management Science* **60**(6) 1470–1488.

Godes, David, Jose C. Silva. 2012. Sequential and temporal dynamics of online opinion. *Marketing Science* **31**(3) 448–473.

Gu, Xian, PK Kannan, Liye Ma. 2018. Selling the premium in freemium. *Journal of Marketing* **82**(6) 10–27.

Halbheer, Daniel, Florian Stahl, Oded Koenigsberg, Donald R. Lehmann. 2014. Choosing a digital content strategy: How much should be free? *International Journal of Research in Marketing* **31**(2) 192–206.

Heiman, Amir, Bruce McWilliams, Zhihua Shen, David Zilberman. 2001. Learning and forgetting: Modeling optimal product sampling over time. *Management Science* **47**(4) 532–546.

Hoang, Ai-Phuong, Robert J. Kauffman. 2018. Content sampling, household informedness, and the consumption of digital information goods. *Journal of Management Information Systems* **35**(2) 575–609.

Hong, Yili (Kevin), Paul A. Pavlou. 2014. Product fit uncertainty in online markets: Nature, effects, and antecedents. *Information Systems Research* **25**(2) 328–344.

Jain, Dipak, Vijay Mahajan, Eitan Muller. 1995. An approach for determining optimal product sampling for the diffusion of a new product. *Journal of Product Innovation Management* **12**(2) 124–135.

Kannan, P. K., Barbara Kline Pope, Sanjay Jain. 2009. Practice prize winner - pricing digital content product lines: A model and application for the National Academies Press. *Marketing Science* **28**(4) 620–636.

Kosner, Anthony Wing. 2015. Apple App Store revenue surge and the rise of freemium app pricing. Available at `https://www.forbes.com/sites/anthonykosner/2015/01/11/apple-app-store-revenue-surge-and-the-rise-of-the-freemium`.

Kretschmer, Tobias, Christian Peukert. 2020. Video killed the radio star? online music videos and recorded music sales. *Information Systems Research* .

Kummer, Michael, Patrick Schulte. 2019. When private information settles the bill: Money and privacy in Google's market for smartphone applications. *Management Science* **65**(8) 3470–3494.

Lahiri, Atanu, Debabrata Dey. 2018. Versioning and information dissemination: A new perspective. *Information Systems Research* **29**(4) 965–983.

Lambrecht, Anja, Kanishka Misra. 2016. Fee or free: When should firms charge for online content? *Management Science* **63**(4) 1150–1165.

Lee, Clarence, Vineet Kumar, Sunil Gupta. 2017. Designing freemium: Strategic balancing growth and monetization. SSRN Scholarly Paper ID 2767135, Social Science Research Network, Rochester, NY.

Lee, Seoungwoo, Jie Zhang, Michel Wedel. 2021. Express: Managing the versioning decision over an app's lifetime. *Journal of Marketing* .

Lee, Young-Jin, Yong Tan. 2013. Effects of different types of free trials and ratings in sampling of consumer software: An empirical study. *Journal of Management Information Systems* **30**(3) 213–246.

Li, Hongshuang (Alice), Sanjay Jain, P. K. Kannan. 2019. Optimal design of free samples for digital products and services. *Journal of Marketing Research* **56**(3) 419–438.

Li, Xing, Timothy Bresnahan, Pai-Ling Yin. 2016. Paying incumbents and customers to enter an industry: Buying downloads. *Available at SSRN 2834564* .

Liang, Chen, Zhan (Michael) Shi, T. S. Raghu. 2019. The spillover of spotlight: Platform recommendation in the mobile app market. *Information Systems Research* **30**(4) 1296–1318.

Liu, Charles Zhechao, Yoris A. Au, Hoon Seok Choi. 2014. Effects of freemium strategy in the mobile app market: An empirical study of Google Play. *Journal of Management Information Systems* **31**(3) 326–354.

Niculescu, Marius F., D. J. Wu. 2014. Economics of Free Under Perpetual Licensing: Implications for the Software Industry. *Information Systems Research* **25**(1) 173–199.

Papies, Dominik, Felix Eggers, Nils Wlömert. 2011. Music for free? How free ad-funded downloads affect consumer choice. *Journal of the Academy of Marketing Science* **39**(5) 777–794.

Pauwels, Koen, Allen Weiss. 2008. Moving from free to fee: How online firms market to change their business model successfully. *Journal of Marketing* **72**(3) 14–31.

Runge, Julian, Stefan Wagner, Jorg Claussen, Daniel Klapper. 2016. Freemium pricing: Evidence from a large-scale field experiment. SSRN Scholarly Paper ID 2888471, Social Science Research Network, Rochester, NY.

Schoenmueller, Verena, Oded Netzer, Florian Stahl. 2020. The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research* 0022243720941832.

Shoemaker, Robert W., F. Robert Shoaf. 1975. Behavioral changes in the trial of new products. *Journal of Consumer Research* **2**(2) 104–109.

Statista. 2019. Worldwide mobile app revenues in 2014 to 2023. Available at `https://www.statista.com/statistics/269025/worldwide-mobile-app-revenue-forecast/`.

Statista. 2020a. Number of available applications in the Google Play store from December 2009 to September 2020. Available at `https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/`.

Statista. 2020b. Number of available apps in the Apple App Store from 2008 to 2020. Available at https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/.

Sunada, Takeaki. 2020. Consumer learning in a durable-goods environment and profitable free trials. *Working Paper* .

Wang, Chong (Alex), Xiaoquan (Michael) Zhang. 2009. Sampling of information goods. *Decision Support Systems* **48**(1) 14–22.

Wu, Fang, Bernardo A. Huberman. 2008. How public opinion forms. Christos Papadimitriou, Shuzhong Zhang, eds., *Internet and Network Economics*. No. 5385 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 334–341.

Online Appendix

A Identify Freemium Pairs

We use Python to identify freemium pairs in three general steps described below.

Step 1: Match

For each publisher, we generate a group comprising of all apps developed by this publisher.

For each app, we compute its similarity to all other apps in the same groups, using app name as the basis of comparison. For each app pair, we first compute a Jaro-Winkler distance between the full strings of the names. After this, we tokenize app names, leading to two lists of tokens (words) for each app pair. For instance, “Angry Birds” and “Angry Birds Lite” are tokenized into [“angry”, “birds”] and [“angry”, “birds”, “lite”]. We then remove paid/free keywords (e.g., “lite”, “pro”, free”, “premium”) . In the the previous example, the lists become [“angry”, “birds”] and [“angry”, “birds”]. For each pair of words in order (meaning “angry” will only compare to “angry”), we compute a new Jaro-Winkler distance which is added to the previous measure of the full string distance. The aggregated Jaro-Winkler distance is then normalized to a scale of 0-1.

Using this established Jaro-Winkler distance, each app pair gets sent into a series of heuristics using cut-off values to determine the quality of match. Very confident matches (>=0.9) are added as a “NoFlag” match indicating that they are very close matches. Matches that score a distance between 0.9-0.8 are added but with an added “Flag” variable indicating that they are closely related but may contain some false positives. After these two cut-offs, two more samples are added at 0.8-0.75 and 0.75-0.8 respectively. These last two cut-offs add an “OtherSample” variable to the match.

Step 2: Remove False Matches

We remove the following types of false matches:

- We remove matches between an iPad app and an iPhone app using a set of iPad specific keywords (e.g. “Angry Birds” and “Angry Birds iPad” or “Angry Birds HD”).
- Sequels such as “angry birds” and “angry birds 2” would have a very high similarity score but are not a true freemium pair. We remove sequels by comparing suffix numbers of app names. If app1 has a

"2" at the end whilst app2 has nothing, app1 is considered a sequel to app2 or vice versa. Sequels up to the number 4 are considered.

- If a free (paid) app matched with another free (paid) app respectively (i.e. paid-paid or free-free matches only), this pair is marked for removal (e.g. "angry birds lite" and "angry birds free").

Step 3: Manual Check

We order all the identified freemium pairs so that app1 is the paid version and app2 is the free version. We then manually check to ensure that they are indeed a freemium pair.

B Summary Statistics

Table A.1 reports summary statistics for the paid version of freemium apps in the panel setting. For each variable x, we compute the overall, within and between variance as follows.

Denote the value of variable x for app i on day t as x_{it}, $i = 1, ..., N$; $t = 1, ..., T$. Denote the mean value of x across apps and days as $\bar{x}$, and the mean value of x for app i as $\bar{x}_i$. The overall variance of x is computed as $\frac{\sum_i \sum_t (x_{it} - \bar{x})^2}{NT-1}$. The between variance of x is computed as $\frac{\sum_i (\bar{x}_i - \bar{x})^2}{N-1}$. The within variance of x is computed as $\frac{\sum_i \sum_t (x_{it} - \bar{x}_i + \bar{x})^2}{NT-1}$.

Table A.1: Summary Statistics on Freemium Apps (with Both Within- and Between- Variations)

Variable	Mean	Std. Dev.			Min	Max
		Overall	Between apps	Within app		
Incremental daily No. of ratings	0.54	9.44	4.68	8.20	0	1316
ln (incremental daily No. of ratings+1)	0.07	0.40	0.28	0.28	0	7.2
Price	1.33	1.51	1.13	1.01	0	100
If rated	0.30	0.46	0.44	0.15	0	1
No. ratings	105.18	1209.93	1243.15	39.35	0	29717
ln (No. ratings+1)	1.02	1.78	1.73	0.41	0	10.3
Average star rating	4.26	0.69	0.68	0.06	1.2	5
Ranked on top 10	0.00	0.03	0.01	0.03	0	1
Ranked on top 11-20	0.00	0.02	0.01	0.02	0	1
Ranked on top 21-50	0.00	0.04	0.03	0.03	0	1
Ranked on top 51-100	0.00	0.04	0.02	0.03	0	1
Ranked on top 101-	0.05	0.21	0.17	0.14	0	1
Age (Days)	147.96	214.14	213.92	8.32	1	1694

Table A.2 shows the descriptive statistics on the propensity score matching results, which indicate that the freemium and control apps are not significantly different in key variables.

Table A.2: Descriptive Statistics on the Matching Results

	Means Treated (s.e.)	Means Control (s.e.)	p-value of t-test
Propensity Score	0.0007	0.0007	0.9995
Ave. Price of Week	1.34(1.38)	1.32(1.54)	0.7714
Ln(No of 1-Star Ratings)	0.33(0.95)	0.34(0.98)	0.5297
Ln(No of 2-Star Ratings)	0.25(0.82)	0.26(0.82)	0.6010
Ln(No of 3-Star Ratings)	0.31(0.94)	0.31(0.94)	0.9827
Ln(No of 4-Star Ratings)	0.41(1.09)	0.39(1.1)	0.4985
Ln(No of 5-Star Ratings)	0.77(1.54)	0.7(1.55)	0.1249
Ranked on top 10	0.05%(1.35%)	0.06%(1.83%)	0.8225
Ranked on top 11-20	0.02%(0.39%)	0.05%(1.46%)	0.4372
Ranked on top 21-50	0.15%(2.82%)	0.12%(2.47%)	0.6631
Ranked on top 51-100	0.12%(1.94%)	0.22%(3.35%)	0.2499
Ranked on top 101-	5.00%(18.15%)	5.27%(19.59%)	0.5903
Genre ID	Exact match		
Whether Updated in the 2-week window before free version introduction	Exact match		
Age	Difference within 6 months		

C Analyses on the Effect of Sampling

Effect of $\underline{v}_p$ on $E(U_s)$:

$$\begin{aligned}
\frac{dE(U_s)}{d\underline{v}_p} &= \frac{-2\theta^2\bar{v}_f 2\theta\left(\bar{v}_p-\underline{v}_p\right)+2\theta\left[(p+\theta\bar{v}_f)^2-2\theta^2\bar{v}_f\underline{v}_p-2p\theta\bar{v}_p+(\theta\bar{v}_p)^2\right]}{4\theta^2\left(\bar{v}_p-\underline{v}_p\right)^2}\\
&= \frac{-2\theta^2\bar{v}_f\left(\bar{v}_p-\underline{v}_p\right)+(p+\theta\bar{v}_f)^2-2\theta^2\bar{v}_f\underline{v}_p-2p\theta\bar{v}_p+(\theta\bar{v}_p)^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)^2}\\
&= \frac{(p+\theta\bar{v}_f)^2-2\theta\bar{v}_p(p+\theta\bar{v}_f)+(\theta\bar{v}_p)^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)^2}\\
&= \frac{(p+\theta\bar{v}_f-\theta\bar{v}_p)^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)^2}>0
\end{aligned}$$

Therefore, $E(U_p)$ increases with $\underline{v}_p$.

Effect of $\underline{v}_p$ on $E(U_s)-E(U_p)$:

$$E(U_s)-E(U_p) = v_f\int_{\theta\underline{v}_p}^{p+\theta\bar{v}_f} f(v_p)\,d(v_p)+\int_{p+\theta\bar{v}_f}^{\theta\bar{v}_p} f(v_p)(v_p-p)\,d(v_p)-c-\left[\int_{\theta\underline{v}_p}^{\theta\bar{v}_p} v_p f(v_p)\,d(v_p)-p\right]$$

$$
\begin{aligned}
&= \frac{(p+\theta\bar{v}_f)^2-2\theta^2\bar{v}_f\underline{v}_p-2p\theta\bar{v}_p+(\theta\bar{v}_p)^2}{2\left(\bar{v}_p-\underline{v}_p\right)}-\frac{\theta\left(\underline{v}_p+\bar{v}_p\right)}{2}+p-c \\
&= \frac{(p+\theta\bar{v}_f)^2-2\theta^2\bar{v}_f\underline{v}_p-2p\theta\bar{v}_p-\theta^2\underline{v}_p^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)}+p-c
\end{aligned}
$$

Thus, we have:

$$
\begin{aligned}
\frac{dE\left(U_s-U_p\right)}{d\underline{v}_p} &= \frac{dE\left(U_s\right)}{d\underline{v}_p}-\frac{dE\left(U_p\right)}{d\underline{v}_p} \\
&= \frac{(p+\theta\bar{v}_f-\theta\bar{v}_p)^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)^2}-\frac{\theta}{2} \\
&= \frac{(p+\theta\bar{v}_f-\theta\bar{v}_p)^2-\theta^2\left(\bar{v}_p-\underline{v}_p\right)^2}{2\theta\left(\bar{v}_p-\underline{v}_p\right)^2}
\end{aligned}
$$

Note that:

$$
\begin{aligned}
(\theta\bar{v}_p-p-\theta\bar{v}_f)^2-\theta^2\left(\bar{v}_p-\underline{v}_p\right)^2 &= \left(p+\theta\bar{v}_f-\theta\bar{v}_p+\theta\left(\bar{v}_p-\underline{v}_p\right)\right)\left(p+\theta\bar{v}_f-\theta\bar{v}_p-\theta\left(\bar{v}_p-\underline{v}_p\right)\right) \\
&= \left(p+\theta\bar{v}_f-\theta\underline{v}_p\right)\left(p+\theta\bar{v}_f-2\theta\bar{v}_p+\theta\underline{v}_p\right)
\end{aligned}
$$

Because $\underline{v}_p < p$, $p+\theta\bar{v}_f-\theta\underline{v}_p > 0$

Because $\theta\underline{v}_p \le v_p \le \theta\bar{v}_p$, $p+\theta\bar{v}_f-2\theta\bar{v}_p+\theta\underline{v}_p \le p+\theta\bar{v}_f-2v_p+v_p = p+\theta\bar{v}_f-v_p = p-(v_p-\theta\bar{v}_f)$. If the consumer updates to paid version, then $v_p-\theta\bar{v}_f > p$. Thus $p+\theta\bar{v}_f-2\theta\bar{v}_p+\theta\underline{v}_p < 0$.

Therefore, $\frac{dE(U_s-U_p)}{d\underline{v}_p} < 0$, indicating that $E\left(U_s\right)-E\left(U_p\right)$ decreases with $\underline{v}_p$.

Effect of $\underline{v}_p$ on the sampling effect: Sampling is effective when $E\left(U_s\right) > 0$ and $E\left(U_s\right) > E\left(U_p\right)$. Assume $E\left(U_s\right) = 0$ when $\underline{v}_p = \underline{v}_p^{\star}$; and $E\left(U_s\right)-E\left(U_p\right) = 0$ when $\underline{v}_p = \underline{v}_p^{\star\star}$.

Because $E\left(U_s\right)$ increases with $\underline{v}_p$ but $E\left(U_s\right)-E\left(U_p\right)$ decrease with $\underline{v}_p$, the consumer both samples and also upgrades to the paid version when $\underline{v}_p^{\star} \le \underline{v}_p \le \underline{v}_p^{\star\star}$.

For a piece of information on app quality that shifts $\underline{v}_p$, condition $\underline{v}_p^{\star} \le \underline{v}_p \le \underline{v}_p^{\star\star}$ is more likely to meet when the signal indicates that app quality is neither too low nor too high, thus the consumer is most likely to sample and to upgrade when this quality signal is moderate.