The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market

Daniel Ershov*
Toulouse School of Economics
daniel.ershov@tse-fr.eu

April 18, 2018

Click here for current version

Abstract

This paper examines the effects of consumer search costs on entry, product design, and quality in online markets. Using 2012-2014 data from the Google Play mobile app store, I take advantage of a natural experiment that reduced search costs for one product type (game apps) in early 2014. Difference-in-differences estimates show that entry increased by 33% relative to the control group (non-games), and that most additional entry was by "niche" products. These estimates also show that lower search costs reduced the quality of new entrants. To separate out the different welfare effects of this change, I develop and estimate a structural model of demand and supply. I show that there are large welfare gains from reduced marginal search costs, smaller gains from increased product variety, and very small losses from lower product quality.

*I would like to thank Victor Aguirregabiria, Avi Goldfarb, Heski Bar-Isaac and Michelle Sovinsky for their time spent reading and discussing this paper and for their helpful comments. This paper also substantially benefitted from discussions with Joel Waldfogel, Ben Shiller, Philipp Schmidt-Dengler, Jason Walter, Eduardo Souza Rodrigues, and seminar participants at the University of Toronto, TSE, Tilburg, Sciences Po, KU Leuven, Mannheim, the 2016 EARIE and JEI conferences, the 2017 IIOC, the 8th Annual Workshop on Consumer Search and Switching Costs, and the 8th Annual Conference on Internet Commerce and Innovation. All errors are my own.

1 Introduction

Contrary to expectations, the internet did not eliminate consumer search costs.[1] With the proliferation of new product varieties, consumers can search through thousands, or hundreds of thousands of products online (Sorensen 2007, Hendricks and Sorensen 2009). Indeed, the difficulty of product discovery or "discoverability" online is a major concern for consumers, firms, and regulators.[2] Firms are concerned about investing sunk costs and entering into markets where they cannot be discovered by consumers. Regulators are concerned that platforms, by changing search algorithms and search costs, can influence firm entry, investment incentives, and the degree of online competition.

While there is a large existing literature on search costs and competition (see more below), it mostly focuses on the effects of search costs on prices and largely ignores non-price effects such as entry or product quality. These non-price effects are important, particularly in the numerous online markets where prices are uniform (e.g., iTunes) or are zero (e.g., SoundCloud).

This paper examines how consumer search costs in online markets affect market structure, product variety, quality, and consumer welfare. I empirically study these effects using new 2012-2014 data from the Google Play mobile app store, a large online market where most products are free to download. App stores illustrate the main characteristics that distinguish online markets (Ellison and Ellison 2005, Levin 2011). In particular, app stores have a large number of products. Thousands of new apps appear every week, and it is costly for consumers to search for new products.

According to industry surveys (Nielsen 2011, Ipsos 2013), consumers' primary search method in app stores is to browse through category areas (e.g.,"Productivity Apps"). These category areas include "bestseller" lists, as well as other "featured" apps (see Section 2.1). I take advantage of a natural experiment: a re-categorization of part of the Google Play store. App stores are generally split into a "game" and "non-game" areas. In March 2014, Google Play re-categorized its 6 game categories into 18 new categories.[3] Industry observers believe that the re-categorization reduced consumer search costs.[4] Before the change, consumers browsing through the categories would see different app types together (e.g., Family Games and Action Games). Consequently, consumers looking for a particular app type would not necessarily find it easily.[5]

[1]e.g., Levin (2011), Ghose, Goldfarb, and Han (2012), Athey and Imbens (2015).

[2]See, respectively, iMore.com , iMediaConnection.com, TheVerge.com.

[3]See Appendix A for a full list of categories before and after the recategorization.

[4]AndroidCommunity.com

[5]Increasing the number of categories should not *always* reduce search costs. At some point,

The re-categorization of the store was explicitly designed by Google to reduce consumer search costs for games and took developers by surprise.[6] Google's split game categories matched iTunes' existing game categories, suggesting that category selection was not driven by pre-existing trends. The alternative view, that something particular to competition or entry in games caused Google to change them, does not hold in the data.

My data consists of weekly and monthly snapshots of the US Google Play store from January 2012 to December 2014. This data includes observable characteristics for all apps available in the market. Since non-game categories were not split, I treat non-games as a control group and use difference in differences (DiD) to capture three key effects. First, entry increases in games relative to non-games by 33%. Second, most of the entry effects are driven by "niche" app types that were more difficult to find before the category split. Lastly, the quality of new games - as measured both by consumer ratings and app size in MB[7] - fell after the split.

The overall impact of the re-categorization on consumer welfare is not easily measured since most apps are free. In addition, the main effects that I capture can point in different welfare directions. If consumers like variety, entry should be welfare enhancing. As well, the fall in *marginal* search costs should benefit consumers. On the other hand, with a larger number of products available consumers may potentially end up searching through more objects and increase their *total* search costs. They would then not fully benefit from the additional variety in the market. Moreover, consumers should also like quality. Conditional on the number of products, a greater share of low quality products would reduce consumer welfare. In the presence of search costs, a larger share of low quality products could make it harder to find high quality products. This would also offset welfare gains.

To measure and decompose the welfare implications of the re-categorization, I set up a structural model of demand and supply. On the demand side, following Moraga-Gonzalez, Sandor and Wildenbeest (2015), I estimate a model that merges a logit model of differentiated product choice with a search model where consumers search through products in any category. This demand model estimates consumer utility parameters and consumer search costs, using the variation from the natural experiment to identify the change in marginal search costs. I estimate two specifica-

having too many categories can result in a more difficult search process. There is likely an optimal number (or range) of categories to display in this market, but this paper cannot say what this number (or range) may be.

[6]Google's announcements about such re-categorizations do not suggest considerations about developer entry decisions: see GoogleBlog.com

[7]An indicator of the number of features in the app, and correlated with the other quality measure.

tions - a static specification, and a dynamic specification where apps' past downloads affect their current demand.

To capture the effect of the re-categorization on marginal search costs, I allow the search cost parameters (which depend on search technology) to vary before and after the split. I find large differences in these parameters, showing that marginal search costs fell by as much as 50%. The demand estimates also suggest that a 1% increase in search costs reduces consumer utility by approximately 30 cents - or 20% of a paid app's average price.

The supply side of the model is an incomplete information oligopoly game of market entry and choice of product characteristics (Seim 2006, Augereau, Greenstein and Rysman 2006). Unlike most previous papers in this literature, firms choose both horizontal characteristics (which categories to enter, if any) and vertical characteristics (quality level). The estimation of this game identifies entry cost parameters and variable profit parameters. In specifying firm profits, I take into account that most of the apps are free to download and firms make money from a combination of in-app purchases and in-app advertising. I model variable profits as a function of the number of downloads.[8] The number of downloads, in turn, depends on the entry decisions of other apps and consumer search costs. The specification of entry costs take into account that these costs non-linearly depend on product quality.

After estimating the demand and supply models, I measure the change in consumer welfare due to the category split. I also decompose the total welfare effect into the changes due to product variety, product quality, and falling marginal search costs. These decompositions show that welfare increases by 56% after the split in the categories, relative to the pre-split baseline. This is equivalent to an increase of 200 million dollars per month over all Android consumers. Most of the increase in consumer surplus comes from the reduction in marginal search costs. There is also an increase in consumer surplus due to greater product variety. This effect is larger than the fall in consumer surplus due to lower quality.

This paper is part of a large literature on the effects of consumer search costs on competition and market outcomes - starting with Stigler (1961), Nelson (1970), and Diamond (1971).[9] Most of this literature focuses on price effects, but there are several recent theory papers that examine the interaction between search costs,

[8]A more downloaded app would likely have more consumers to make in-app purchases, and more consumers to view ads.

[9]See Stahl (1989), Anderson and Renault (1999), Armstrong, Vickers, and Zhou (2009), Chen and He (2011), and Zhou (2014), for more recent theory papers. See Syverson and Hortacsu (2004), Wildenbeest (2011), Chandra and Tappata (2011), De los Santos et al (2012), Koulayev (2014), Bronnenberg et al (2016), and Koulayev and De los Santos (2017) for recent empirics.

entry incentives, investment in quality and product design.[10] The existing empirical literature on these topics is more sparse.[11] Data suggests that the introduction of the internet in the 1990s reduced consumer search costs for books, movies, and music. At the same time, the number of these products available to consumers increased. The implication is that lower search costs increased entry. However, this evidence is primarily descriptive. It also does not clearly distinguish changing search costs from other effects that could increase entry.[12]

This paper has three key contributions to the existing literature. First, no other paper, to my knowledge, uses a plausibly exogenous change to search technology to identify the effects of consumer search costs on firm entry. As stated above, past literature relies on descriptive evidence. Nonetheless, my results are consistent with past theoretical and empirical findings (e.g., Cachon, Terwiesch, and Xu 2008).

Second, this paper presents evidence on questions that have not yet been examined empirically. I show that lower search costs have an effect on product design (consistent with Bar-Isaac, Caruana and Cunat 2012, Yang 2013), as well as on product quality. The effects on quality are ambiguous in theory (Fishman and Levy 2015): on the one hand, when search costs fall in a market with vertical and horizontal differentiation, consumers can find the highest quality products, which should provide incentives to improve quality. On the other hand, if a firm invests in high quality, consumers who find their product can also find competing high quality products that are better horizontal matches. Thus, lower search costs can drive quality down. My results show that average quality falls, suggesting that the second effect dominates in this application.

Lastly, the structural model allows me to measure consumer welfare arising from free apps. As such, this paper follows up on a small literature that captures consumer surplus arising out of the digital economy (Brynjolfsson, Hu and Smith 2003, Brynjolfsson et al 2018). Moreover, the model allows me to measure the change in welfare coming from the re-categorization and consumer search costs. From a policy perspective, the results suggest that policies which increase consumer search costs in online markets can have two separate negative effects on consumer welfare. First, there is a direct effect: consumers have higher marginal search costs that reduce their welfare. Second, there is an indirect foreclosure effect: higher consumer search costs discourage firms from entering the market, reducing product variety and welfare.

[10] e.g., Bar-Isaac, Caruana, and Cunat (2012), Yang (2013), Cachon, Terwiesch, and Xu (2008), Larson (2013).

[11] Goldmanis, Hortacsu, Syverson and Emre (2010), Waldfogel (2011), Brynjolfsson, Hu, and Simester (2011), Zentner, Smith, and Kaya (2013), and Aguiar and Waldfogel (2016).

[12] The costs of producing movies, books, and music fell in the same time period.

The paper proceeds as follows: Section 2 provides an overview of the mobile app market. Section 3 describes the data and presents some summary statistics. The fourth section presents the reduced form results. The fifth section presents the specification and estimation of the structural model, and the counterfactuals. The final section concludes.

2 App Market Background

2.1 Users

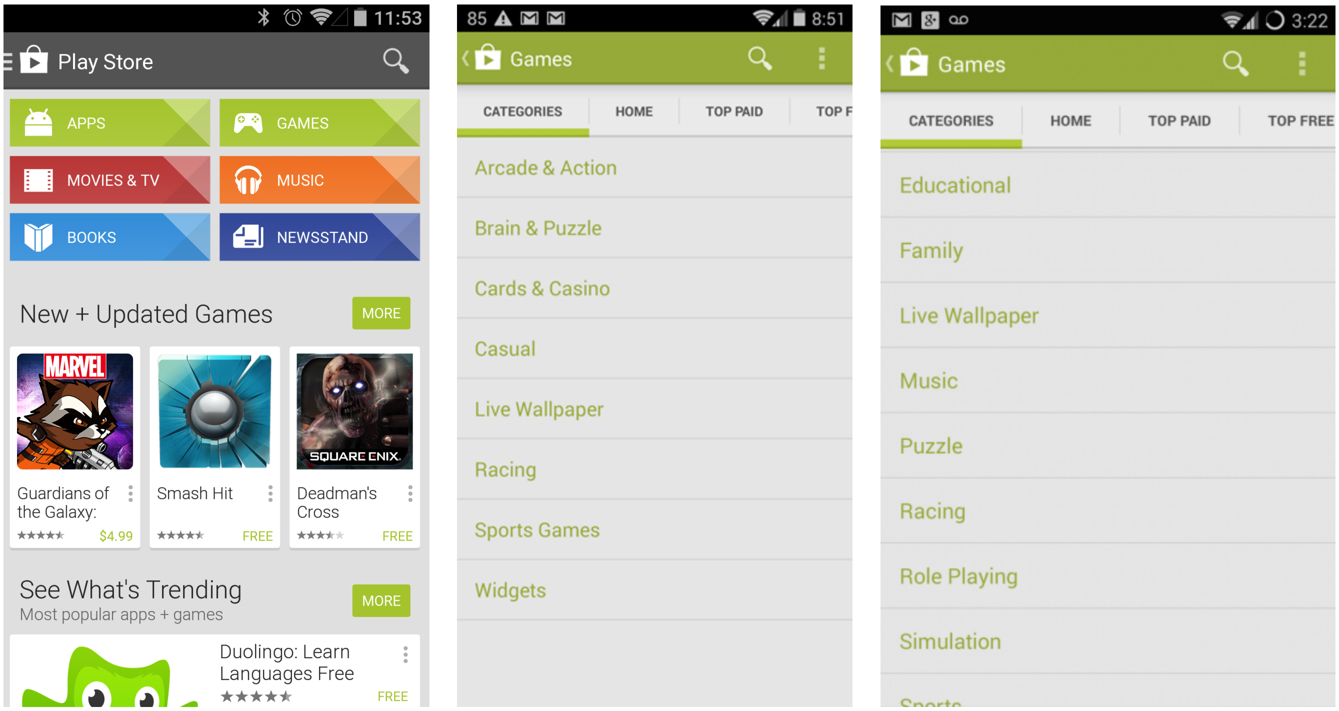
The Google Play store comes pre-installed on any phone that runs Android OS. The first screen of the Google Play store (c. 2014) appears below in the left panel of Figure 1. When a user open the Google Play store on their phone, they see a number of tabs. They can choose to look at Games, or Apps (non-game apps), or they can choose to look at individual "featured" games/apps/music/movies/books that appear directly on the first page (this would be the "New + Updated Games" in the Figure). Once they choose a product type, Games for example, users can choose to look for more specific product types by choosing a "category." The middle panel of Figure 1 shows the choice of game categories in late 2013, and the right panel of Figure 1 shows the choice of game categories in mid 2014. I am exploiting the change in this feature as the natural policy experiment. Rather than 6 categories (in addition to the "widget" and "live-wallpaper" category), Google Play split their game categories into 18 different types in March 2014.[13]

Once users choose a category, they can either look at a panel of "featured" products from that category, or they can look at top-lists, which display the apps with the largest number of downloads in approximately the past week. The exact algorithm determining the position of an app in the top lists is unknown, but it is correlated to downloads.[14] The top lists are Top Paid, Top Free, Top Grossing, and Top New Paid and Top New Free, arranged in that order horizontally. The left panel of Figure 2 below shows the top list for all free Apps. At that point, users only observe apps' names, their icons, their position in the list, their average user ratings (the average "star" rating of the app), and their price. They observe the same information about featured apps. Once they click on a particular app listing they obtain more information about the app (see the right panel of Figure 2). In particular, users get to observe a number of screenshots from the app, the average rating of the app, how

[13] For a full list of categories before and after the split, see Appendix A.

[14] Adweek.com

Figure 1



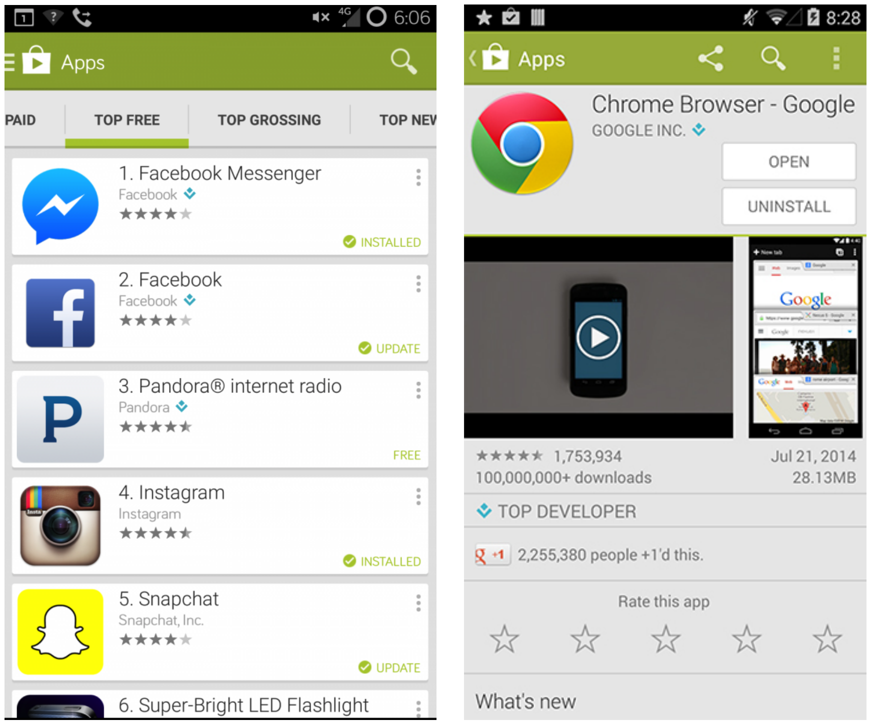
many people have downloaded this app, the size of the app in MB, and a textual description of the app. It is at that point that they choose whether they want to download (or purchase) the app.

2.2 Developers

The costs of publishing an app on the Google Play market are zero - each developer has to pay a one time fee of $25 to register with Google Play and then can publish apps for free.[15] On the other hand, developing an app can be costly. At the low end, the cost of programming an app (in computer equipment, software, and wages for programmers, designers, and debuggers) can be as low as a few thousand dollars. Companies that develop those low cost apps generally consist of a few individuals who work together for a few days or weeks. On the higher end, apps that link up to databases (e.g., calendar applications, notes applications, texting applications), can cost up to and above 50,000 dollars.[16] At the very high end, apps that require

[15] TechRepublic.com
[16] ApplicoInc.com

Figure 2



more inputs can cost as much as hundreds of thousands or millions of dollars (e.g., mobile games with advanced 3D graphics, or a social media application with video chatting).[17] Instagram, for example, spent over $500,000 in venture capital money on various technologies designed to accommodate a scalable photography-based social network.[18]

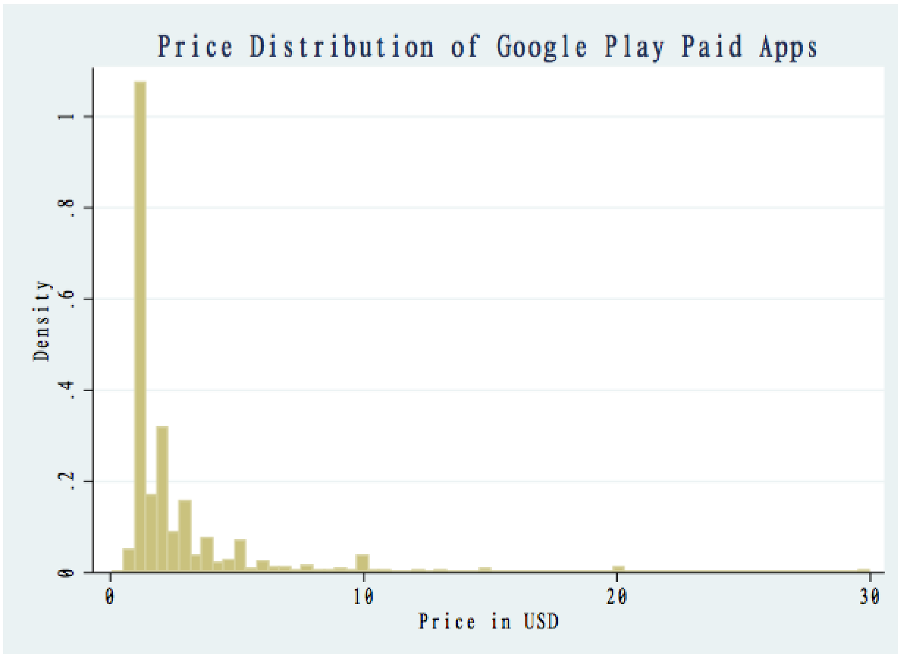
When developers introduce a new app into the Android market, their key choice is the category into which they introduce their app into, since it matters for consumers' search.[19] On Android (unlike the Apple App store), the category choice is mutually exclusive and you can only choose one category at a time.

Developers make money from their apps in one of three ways. First, they can charge consumers upfront for downloading their app. These are the "paid apps" and in Google Play they constitute a minority, about 20%, of the total number of apps. Median paid app price is $1.5, the price of an app in the 90th percentile is $5, and the

[17] SavvyApps.com
[18] Instagram Tumblr
[19] See previous subsection for more detail on the way consumers search the app market.

price of an app in the 99th percentile of the price distribution is $29.99 (see Figure 3 below for a more detailed price distribution).

Figure 3

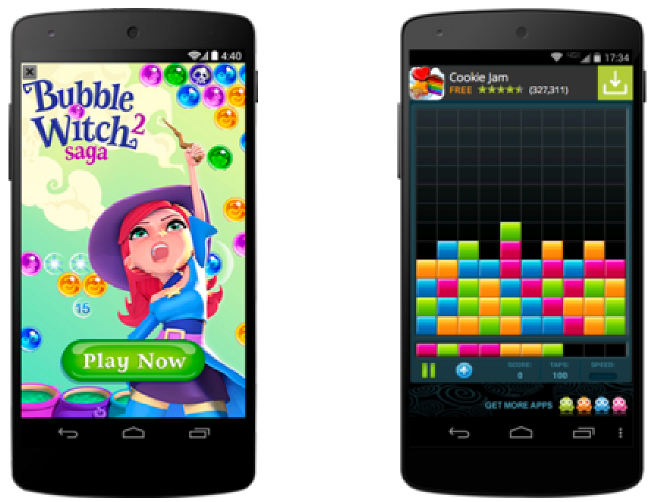


Second, the developers can allow users to download their apps for free, but place advertisements in the app. These are full-page pop-up ads (see left panel of Figure 4), banner ads (see right panel of Figure 4), or video ads. The revenues from in-app advertising are similar to the revenues from search advertising - both depend on the number of users, the type of the ads, the frequency with which ads are shown, and other factors. There is a competitive market for advertisers, who bid in second price style auctions to display their advertisements in different types of apps.[20] This auction system is similar to the Google search advertising system, which suggests that the revenues for apps highly depend on their downloads and their type. Varian (2007) shows that bids in Google search auctions are increasing in rank and are convex (sometimes exponential), meaning that the bids for top slots are much higher than bids for lower slots. Further anecdotal evidence suggests that the average price paid for 1,000 shown ads ("impressions") in the US is approximately $2-3.[21] It is likely

[20]This side of the market is beyond the scope of this paper and is not modelled.
[21]Quora

that there is substantial variation across apps - with top apps receiving substantially higher rates, as they are more attractive to advertisers, whereas apps which are lower ranked faring much worse.

Figure 4



In-App Interstitial: Full-Page Ad

In-App Banner: App Install Ad

Source: https://www.google.com/admob/

Lastly, the developers can allow users to download their apps for free, but sell them features or products within the app (so-called "in-app purchases"). Those could include additional levels for games, subscriptions for magazines (or particular magazine columns) and so on. The popularity of these business models reflects their profitability - indeed, many popular games such as Clash of Clans are free but can make hundreds of millions of dollars from a combination of in-app purchases and advertising.

As the platform, Google also makes money from the app market. First, it obtains \$25 for every new developer who enters the market. Second, Google takes a 30% cut of paid app downloads. It also takes 30% of the in-app-purchases made on free apps.[22] Of the remaining free apps who use app advertising to make money, about 50% are using Google's advertising platform, Admob, meaning that Google gets a cut from the advertising revenues.[23] Lastly, even if Google does not earn direct money

[22] Google

[23] AppBrain.com

from other advertising platforms, they do get valuable data from all apps that helps them optimize their own search results and advertising.

3 Data

3.1 Data Description

My data comes from AppMonsta.com and consists of daily snapshots of all apps on the Google Play store, aggregated at the weekly or monthly level, starting from January, 2012, up to December, 2014.[24] This is the first paper to use this particular version of the dataset.[25] This dataset contains all information that consumers observe on the Play store - app price (in USD), a histogram of the ratings the app received (ranging from 1 to 5), app size (in MB), the number of screenshots the app shows, the number of video previews the app shows, and a download range for the app (referring to the total number of downloads the app received over its lifetime). Additionally, I observe the app's category, the name of the app's developer, and a textual description of the app, explaining what it "does." Lastly, I observe the "top lists" for every category, which are the top 500 best-selling free and paid apps in each category for each week.

3.2 Data Management

3.2.1 Predicted Reduced Form Categories

For the reduced form analysis, I want to estimate effects at more granular levels than "game" and "non-game." Since the categories are not stable over time, I run a machine learning algorithm on app descriptions in order to re-classify game apps from the period before the split into post re-categorization categories. Section 8.1 in Appendix B provides more details about this re-classification procedure. I do not use these predicted categories in the structural model.

The following table shows some summary statistics at the predicted category level. There are 42 categories, 18 of which are game categories. The game categories, on average, are much smaller than the non-game categories in terms of the number of

[24] Weekly aggregation is used to predict the downloads of each app (see following section), and the monthly aggregation for the rest of the analysis.

[25] Liu, Nekipelov, and Park (2014) use a similar dataset (for Google Play and iTunes) from the same source for 2011-2012.

apps - the average game category is less than a quarter of the size of the average non-game category.

Table 1: **Summary Statistics at the Category-Month Level**

Variable	Mean	Std. Dev.	N
Game Categories			
Number of Apps	7,256	12,990	630
Number of New Apps	452	954	630
Non-Game Categories			
Number of Apps	32,656	30,000	840
Number of New Apps	1,522	1,902	840

3.2.2 Predicting Downloads

Most of the reduced form results in Section 3 (e.g., the effects of the policy change on entry, quality, and prices) do not rely on measuring weekly or monthly app downloads. However, per-month app downloads are necessary for the estimation of the structural demand model in Section 4. To recover them, I do the following: I use information about the rankings of the apps in each category in each week and the the download bandwidth of new apps arriving in the market to recover a relationship between rankings and downloads.[26] Then, I predict the downloads of all other apps in the market. Section 8.2 in Appendix B provides more detailed explanation about this procedure. I predict zero downloads for about 20% of apps in a given period. In order to avoid the "zeroes" problem in demand estimation, under some Bayesian assumptions about the underlying distribution of consumer downloads I can apply the techniques of Gandhi et al (2014).

3.3 Descriptive Statistics

Table 2 shows some summary statistics at the app level, where there are approximately 33.7 million app-month observations in the dataset, consisting of 2.6 million unique apps. Of these, approximately 17% belong to game categories, whereas the rest are non-game apps.

[26] Intuitively, I observe a new app ranked 1st in a given category with a lower bound of downloads of 50,000, a new app ranked 10th with a lower bound of downloads of 10,000, and a new app ranked 100th with a lower bound of downloads of 500. With certain distributional assumptions, I can then recover the relationship between the lower bound of downloads and ranking.

Table 2: **Summary statistics at the App Level**

Variable	Mean	Std. Dev.	N
App Level			
Game App	0.168	0.374	2.6 million
Paid App	0.2	0.4	2.6 million
App-Month Level			
Lifetime Downloads (Min.)	38,261	1.9 million	33.7 million
App Size (in MB)	21.99	29.75	33.7 million
Monthly Predicted Downloads	559	25,169	33.7 million
Number of Screenshots	4.71	3.54	33.7 million
Number of Videos	0.09	0.28	33.7 million
Mean Rating	4.0	0.66	27 million
Price (for Paid Apps)	3.27	8.93	6.8 million

Note: Mean app rating calculated for apps with 5 or more ratings.

4 Reduced Form Evidence

4.1 Entry

Figure 5 below is a simple plot showing the number of new apps appearing each month in the game categories and in the non-game categories. The left hand panel shows absolute monthly (log) entry levels for games and non-games. The right hand panel shows the ratio of the number of games entering the market relative to the number of non-games.

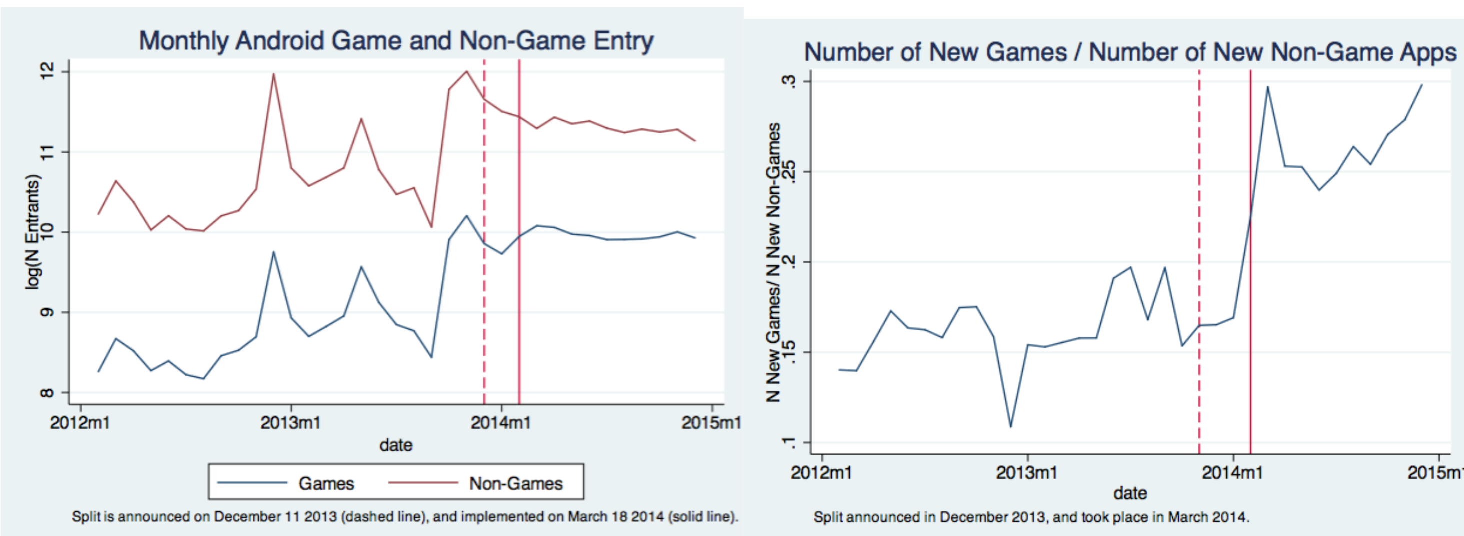
The left hand panel shows that there are large changes in absolute entry for games and non-games over the sample period. There are several spikes and troughs over the sample period, and a sharp absolute increase in the number of entering games and apps in mid-2013. There can be many drivers of this variation. For example, Christmas or "back-to-school" shopping appears to be responsible for the peaks in 2012. The release of new phones or OS can also drive absolute app entry. The large absolute increase in entry in mid-2013 appears to be driven by the release of the Samsung Galaxy S4 - a flagship Android phone which became the best-selling smartphone in the world.[27] This dramatically increased the market size for Android apps and generated much higher returns for entry.

The right hand panel shows that despite the large absolute changes in game and

[27] GSMArena.com

non-game entry over time, the relative trends of game and non-game entry in 2012 and 2013 were nearly constant. The ratio of the number of new games to the number of new apps was around 0.15 (approximately 7 new apps entered for every new game) for this entire period. This changed following the announcement of the split in the game categories in late December 2013. The left hand panel shows that starting February 2014 there is an increase in entry by game apps but not by non-game apps. This increase in entry persists and peaks right after the actual split in the game categories in March 2014, but it remains afterwards as well.[28] On the right hand panel, this is a clear shock to the ratio. Starting in February 2014, the ratio of new games to new non-games experienced a shock, moving to about 0.25 for the rest of the sample period - meaning that in 2014, only 4 new apps entered for every new game that entered.

Figure 5



It is reasonable to interpret the change in entry around the split in the categories as a response by developers who were incentivized by the announcement (and then creation) of new categories to produce more games rather than non-game apps.[29] This new entry could have come from multiple sources - either the developers started creating completely new apps from scratch to compete in the new categories, or the

[28] This is also true at the weekly level: see Figure 11 is in Appendix C.

[29] The announcement did not set a strict date for the implementation of new categories, but mentioned that the change will happen in the first quarter of 2014: 9to5Google.com

developers quickly released already developed products into the market early in order to take advantage of the impending changes.

The relatively constant difference between game and non-game entry suggests that a difference-in-differences estimation strategy is appropriate to look more carefully at developers' entry decisions - comparing the decisions of the developers to produce game vs non-game apps before and after the split in the game categories occurs. The non-game categories thus serve as a control group, and the game categories serve as the treatment. As well, the panel nature of the data allows me to do a formal timing test of the treatment (see Section 9.2 in Appendix C).

More formally, the logarithm of the number of apps that enter into category group $c \in \{GAME, NONGAME\}$ at time t - Y_{ct} - can be represented as:

$$Y_{ct} = \tau(Game_c \times Post_t) + Game_c + \delta_t + \epsilon_{ct} \tag{1}$$

where δ_t are time fixed effects, $Post_t$ is a dummy that is equal to one after the game categories have been split, and $Game_c$ is a dummy variable equal to one for the game category group and zero for non-game category group (including category and time fixed effects causes $Post_t$ to drop out of the estimating equation). The coefficient of interest in this regression is τ, which captures the treatment effect on entry for the game categories, relative to non-game (app) categories. If this coefficient is positive, then it means that developers produced more game apps as compared to non-games apps after the split.

Estimating Equation (1) pools together entry in all game and non-game categories into "category groups" (it is Figure 5 with standard errors). If I use app descriptions to allocate apps into post-split game categories for the pre-split period (see Section 3.2), I can also run this regression at the category level. That is, c can refer to different categories (e.g., "Productivity Apps", "Sports Games", etc) as follows:

$$Y_{ct} = \tau(Game_c \times Post_t) + \delta_c + \delta_t + \epsilon_{ct} \tag{2}$$

where $Game_c$ is a dummy representing the categories which are game categories, and δ_c is a category specific fixed effect for the game categories (the average non-game category is the reference group).

Table 3 shows results from the first three regressions. Column (1) shows the regression with the data pooled across category groups, Column (2) shows the regression for separated categories without group fixed effects, and Column (3) shows the same regression as (2) but with category fixed effects, where the reference group is the average non-treated category.

The key treatment coefficient in the first column is approximately 0.3 which suggests that the following the split, developers entered 30% more apps in the game

Table 3: **Regression Estimates for log(N Entrants)**

VARIABLES	(1) Category Groups	(2) Separate Cats	(3) w/ Cat FE
Games × Post Split Period	0.334***	1.466***	1.472***
	(0.072)	(0.130)	(0.101)
Games (Treated Group)	-1.706***	-2.561***	
	(0.093)	(0.092)	
Time FE	YES	YES	YES
Category FE		NO	YES
Observations	70	1,469	1,469
R-squared	0.965	0.541	0.794

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

categories than in the non-game categories relative to the time before the split. This coefficient is also highly statistically significant at the 99 percent confidence level. The treatment coefficient in Column (2) is approximately three times larger, since it looks at individual categories, which are smaller in size and thus produce higher relative changes (recall, the dependent variable is measured in logarithms). That is, 150 percent more apps entered into the average game category as compared to the average non-game category.

Appendix A, which shows the complete list of pre-split and post-split categories suggests another potential DiD regression candidate. In particular, three of the game categories that existed in the pre-split period - "Sports Games", "Casual Games", and "Racing Games" - remain the same in the post-split period. As a result, they appear to be natural control group candidates for the game categories that were split. This is not the case. While the split in the other game categories may not have affected them as directly, it still changed their definition.[30] In other words, the treatment does affect the control group (in this case). The category specific treatment effects in the next two subsections suggest that the treatment did affect

[30] For example, some apps that were previously considered to be "Sports Games" or "Casual Games" may now consider themselves to be something else and change to one of the new categories. Then, new apps that enter the market later observe the changed roster of apps in the previously existing categories and may make different decisions as a result.

the three previously existing categories less on average. However, this appears to be related to their size (in terms of the number of apps), as new game categories of similar size had similar treatment effects.

4.1.1 Possible Alternative Explanations

A potential concern about these treatment effects is that the change in entry is not driven by the split in the categories but by another factor that is then confounded with the re-categorization of games. I do a formal placebo timing test to check whether the τ coefficient captures something other than the split in the categories by interacting the treatment effect with the monthly time dummies. Results from the timing test, presented in Section 9.2 of Appendix C, show that the difference in entry between games and non-games becomes statistically significantly positive around the month of the split. This means that the treatment effect would have to confound an event or a change that occurs in the month of the treatment, which is unlikely.

Possible confounding factors could also include the release of new phones or changes in the demographics of consumers or in their app usage habits. There was no significant Android phone release in early 2014. As mentioned above, the previous significant scale phone release was the Samsung Galaxy S4 in mid 2013. As well, there is no evidence that player attitudes were changing differentially for games and non-games. Aggregate data from the Nielsen Insights and Mobile NetView Surveys from 2012 to 2014 show that while per-consumer monthly game usage is increasing from 2013 to 2014, it was also growing as rapidly from 2012 to 2013. Non-game per-consumer usage was also growing at the same rate.[31] The same survey also suggests that there are no obvious changes in the demographics of Android users between 2013 and 2014.

A last concern is that the re-categorization and the subsequent change in entry is not the result of changes in search costs, but of consumer preferences. That is, when the re-categorization happened consumers realized that they might like different types of products which they were not aware of before, and firms entered those products into the market. This is an unlikely explanation because change in entry is immediate. Indeed, it starts happening in anticipation of the re-categorization. If the effect of the change in categories was on consumer preferences, then it is reasonable to expect that it takes some time for firms to observe the new consumer preferences and respond. This is not what happens, which is more consistent with firms anticipating changes in consumer search costs.

[31] Source: Nielsen.com

4.2 Product Design

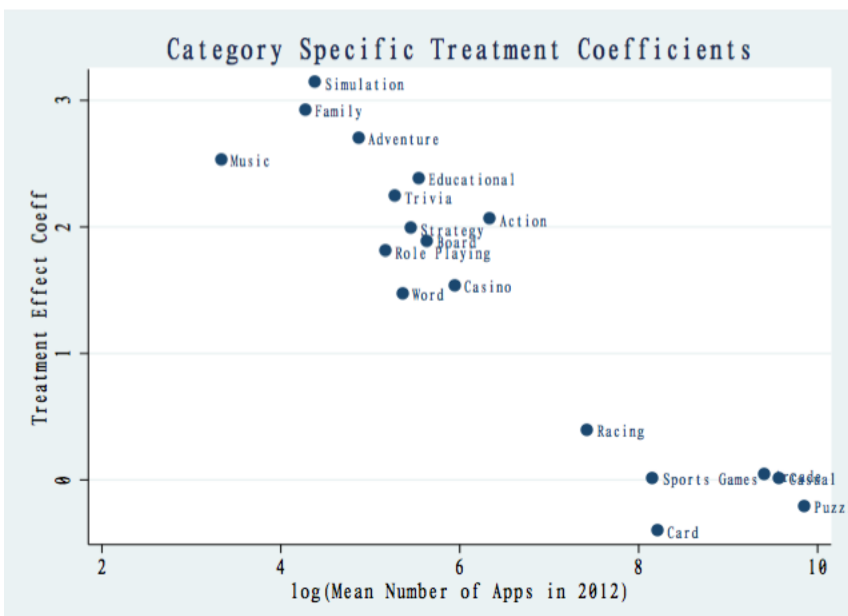
I can examine changes in product design decisions by firms by looking at the heterogenous effects of the game category split. It is reasonable to expect that more niche game genres should benefit more from the re-categorization than the larger (more mainstream) game genres. The reason is that it was harder for consumers to find more niche products (e.g., Music Games) as compared to more popular products (e.g., Arcade Games).

The category specific treatment effects regression is as follows:

$$Y_{ct} = \sum_c \tau_c \left(Post_t \times \delta_c\right) + \delta_c + \delta_t + \epsilon_{ct} \tag{3}$$

where the τ_cs are the game category specific treatment coefficients, and the rest of the notation is similar to that in the previous section. The (game) category specific treatment coefficient estimates from that regression are presented in Appendix C and in Figure 6 below.

Figure 6



With the exception of a few categories, most of the treatment effects are statistically significant at the 99 percent confidence level and positive. The results make it clear that the change in entry was primarily driven by the change in search, since it is the smaller niche categories that have the largest coefficients. Coefficient estimates are monotonically falling with category size. Based on category size in 2012, the categories in the top quartile (e.g., Card, Casual, Puzzle) have coefficient estimates that are zero (or negative), meaning that the increase in the number of categories did not incentivize more developers to enter products there. This makes sense, since a developer who wanted to enter an "arcade" app (that is an app with characteristics which fit best into the "Arcade" category) she created before the category split would be likely to enter it into the very large category "Arcade & Action", of which "arcade" games constituted a large part. There is likely little difference in incentives between this category and the "Arcade" category in the probability of the relevant consumers finding her app. For more niche categories, the change in consumer search costs would have been large, as would the incentives of producers.

The treatment coefficients are relative, since the dependent variable is in logarithms. As such, a large treatment coefficient for the smaller categories may mean less in absolute terms than a small treatment coefficient for large categories. The treatment coefficients are monthly - suggesting that the Family app group, for example, grows at 300% per month as compared to the average non-game category - but it starts from a low baseline of 100 or so apps. Nonetheless, compared to a growth rate of the Casino category, which grows at 150% per month but started at a higher baseline of about 1,000 apps, it is conceivable that in a short period of time the number of Family apps would overtake the number of Casino apps.

These results are entirely consistent with past theoretical predictions (Bar-Isaac, Caruana and Cunat 2012, Yang 2013) which suggest that when search costs fall, producers will enter more niche products into the market as opposed to more mainstream products.

4.3 Quality

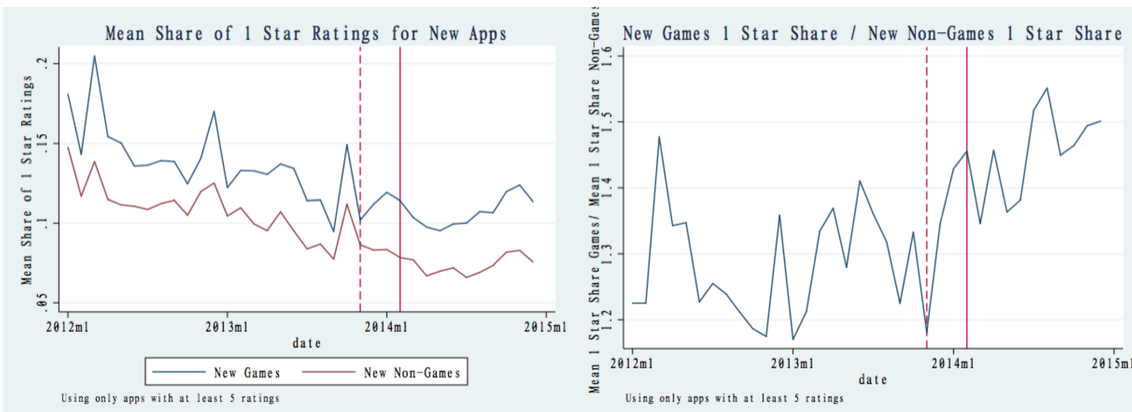
There are multiple proxies for app quality in the data, including the ratings of the apps, the size of the apps, and the information that the apps reveal about themselves (e.g., number of photo or video previews). These proxies are all correlated with one another, so I will use a direct measure of quality - the ratings that consumers give to apps. I use the app's percentage of one star ratings, following the methodology of Chevalier and Mayzlin (2006).[32] I limit the sample to all apps that have at least

[32] I also use an indirect proxy of quality - the size of the app in MB. To give an app more features, developers have to write more code, which requires additional investment. It is possible to argue that better apps are more efficiently coded and would therefore be smaller. On the other hand, anecdotally, apps that are described as high quality integrate many features - such as Facebook

5 ratings.[33]

A time series of the mean share of one star reviews of new game and non-game apps (left panel of Figure 7 below) suggests that while the two types of apps do not have the same share of 1-star ratings - game apps have more - these shares move very similarly over time in 2012 and 2013. Both shares fall over time. Then, after the announcement of the split in the categories the mean share of 1-star game reviews starts increasing, while it continues falling (or stays constant) for new non-game apps. The ratio of the mean shares (games over non-games) is constant over time, until the announcement at which point it begins increasing (right panel of Figure 7).

Figure 7



As for entry in the previous section, I run three sets of regressions, the results of which are in Table 4. Column (1) shows the regression with the data pooled into category groups, Column (2) shows the regression for separated categories without category fixed effects, and Column (3) shows the same regression as (2) but with category fixed effects, where the reference group is again the average non-treated category:

The results point in the same direction as the raw data above. The average share of 1 star ratings increased after the split for games relative to non-games. Although the results at the game/non-game level are not statistically significant, the results at the category level are, and show that the share of 1 star ratings increased for game

or Twitter linking, or databases - that require additional lines of code and that would therefore increase the size of the app. The direct and indirect measures are positively correlated - with a correlation coefficient of approximately 0.3. The results are in Appendix C.

[33]The estimates are qualitatively and quantitatively similar when I either include all apps, or limit the sample further to apps with 10, 20, 100 or more reviews.

Table 4: **Regression Estimates for Mean App Share of 1 Star Ratings**

VARIABLES	(1) Category Groups	(2) Separate Cats	(3) w/ Cat FE
Games × Post Split Period	0.003	0.033***	0.033***
	(0.004)	(0.003)	(0.003)
Games (Treated Group)	0.031***	0.007***	
	(0.003)	(0.002)	
Time FE	YES	YES	YES
Category FE		NO	YES
Observations	70	1,468	1,468
R-squared	0.932	0.232	0.401

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

categories relative to the average non-game category by approximately 3 percentage points - suggesting a decline in quality. Considering that the average share of 1 star ratings is approximately 10%, a 3 percentage point increase is economically significant.

There are two possible explanations for this: First, the reduction in entry barriers could have resulted in the entry of infra-marginal apps of lower quality.[34] Second, Fishman and Levy (2015) suggests that the effect could be due to the search costs themselves. In their model, as search costs effectively fall for game apps, the incentives to produce additional high quality products decline somewhat because consumers are more likely to look through a number of high quality products and find the best match, so the downloads for each high quality product potentially decline to the extent that it does not make sense for firms to make the investment.

I run an additional regression with category specific treatment effects, which provide further potential evidence to both of these effects in the data. The category specific treatment effects are pictured in Figure 8, where they are plotted against the average number of apps for each category in 2012 (left panel), and the average share of 1 star ratings in 2012 (right panel).

[34]FORTHCOMING RESULTS suggest that this is unlikely. The variance of quality of new products did not increase, suggesting that the reduction in average quality indicates a shift down in the distribution of entrants, rather than an increase of additional low quality entrants.

Figure 8

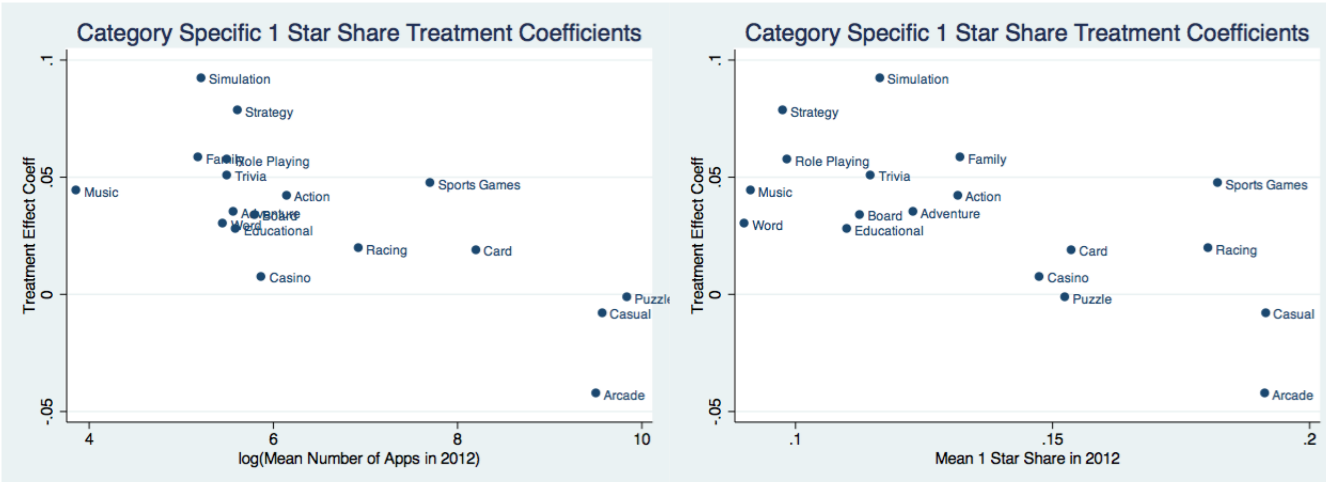


Figure 8 suggests that most of the effects came from the smaller game categories, with relatively small (or negative) average effects for the larger game categories. The largest quartile of categories includes Puzzle, Casual, and Card games. The treatment coefficients for those categories are on average close to zero or negative, as compared to the treatment coefficients for the smaller categories, which are mostly positive and statistically signifiant. Indeed, in the actual regression results the treatment coefficients for the largest four categories are also not statistically significant, suggesting that treatment has no effect on average entrant quality.[35] Similarly, the categories with the biggest treatment effects are the ones with the lowest average number of 1 star ratings in the pre-split period (2012) - i.e. the categories with the highest initial quality are the ones that are seeing the biggest subsequent decline in quality as a result of the split.[36]

4.4 Additional Reduced Form Results

In addition to entry, product design, quality and engagement, I test for changes in two other measures: prices and downloads. Although paid apps are a small part of the market, it is reasonable to test whether the average prices of apps change as well.[37] It is not ex-ante obvious what the price effects should be. Some theory predicts prices

[35] Full coefficient estimates are in Appendix C.

[36] This is not due to mean reversion for the small categories. See Appendix C for a placebo test.

[37] The entry and quality patterns are the same for paid and non-paid products.

to fall when search costs fall, as competition between products intensifies. Other theory, such as Bar-Isaac et al (2012) suggests that prices will increase as search costs fall, since consumers will be able to find products which they prefer more.

Appendix C shows the price patterns, but broadly I find that the latter explanation is consistent - prices weakly rise in games relative to non-games after the re-categorization. However, the lack of parallel trends between game and non-game prices in the pre re-categorization period means that it is difficult to make causal claims. In addition, I test whether the share of paid games and paid non-games changes after the re-categorization. I find that while the share of new paid products falls over my sample period, there is no evidence that it changes for games relative to non-games after re-categorization.

Lastly, I look at the total number of downloads for game and non-game apps in the market over time. In Appendix C, I show that the number of downloads for games increases relative to the total number of downloads for non-games (and in the absolute) very soon after the re-categorization. This has two implications. First, it suggests that the consumer search mechanism is likely driving the results - as search becomes easier, consumers download more products. Second, it suggests that there is a market expansion effect and potential overall welfare increases from the additional entry.

5 Structural Model

The main reduced form effects of the re-categorization are an increase in entry and a reduction in the quality of the entrants. It is not clear how important these changes are to consumer welfare. As mentioned before, over 80% of the apps in the market are free, meaning that the benefits are not easily measured. As well, there were already a large number of products in the market, and additional products may not add much to consumer welfare. Due to search costs, consumers may also not enjoy the full effects of the increase in product variety. The reduction in search costs may increase welfare in and of itself (as evidenced by the increase in downloads for games), but it is also not clear by how much. The decrease in the quality of new products could also reduce consumer welfare relative to a counterfactual where the quality of new products is higher.

For these reasons, a structural model is necessary. A demand model allows me to measure changes in consumer welfare as a result of changes in the market. A supply model allows me to compute counterfactual entry outcomes in order to decompose the different welfare effects.

5.1 Demand

Search plays an important role in the mobile app market as it is highly unlikely that consumers are perfectly aware of all products in this market and their characteristics. Most papers integrating search into demand models make one of two assumptions about consumer search in order to identify search parameters as well as utility parameters using aggregate data: they assume that the order in which consumers examine products is known (ordered sequential search, e.g., Hortacsu and Syverson 2004, Kim et al 2010, Bronnenberg et al 2016), or they assume that consumers arrive at products randomly (random search).

In this case, these assumptions are not reasonable. It is difficult to justify assumptions about the order in which consumers look at products.[38] It is possible that consumers look at several apps in one category, and then switch to look at other apps in another category. It is similarly unreasonable to assume that consumers look at apps completely at random. App stores are specifically organized to direct consumers towards certain apps - for example, by having bestseller lists.

A small recent literature allows for demand estimation using aggregate data without making assumptions about consumer search order by focusing on the consideration sets of consumers (Sovinsky/Goeree 2008, Ching, Erdem, and Keane 2009, Moraga Gonzalez, Sandor, and Wildenbeest 2015, Honka, Hortacsu, and Vitorino 2015). The following derivation of a demand model follows the approach of Moraga Gonzalez, Sandor and Wildenbeest (2015).

Suppose consumers search through N apps, where the utility that consumer i receives from downloading app $j \in \{1,..,N\}$ is:

$$\begin{aligned} u_{ij} &= \delta_j + \epsilon_{ij} \\ &= X_j\beta + \xi_j + \epsilon_{ij} \end{aligned} \tag{4}$$

where X_j are observable product characteristics (including price), ξ_j are unobservable product characteristics, and ϵ_{ij} is a random demand shock that is consumer and product specific and that is distributed iid with an extreme value type 1 distribution (mean zero, standard deviation normalized to 1). The consumer also has the choice to not buy anything (or buy the outside option), in which case they obtain:

$$u_{i0} = \epsilon_{i0} \tag{5}$$

[38] Additionally, De los Santos et al (2012) suggests that online search is unlikely to be sequential and ordered.

Initially, consumers are not fully informed about all products: they do not know their ϵs. Search resolves this uncertainty. Consumers search the market by first choosing a consideration set A of products while incurring some search cost, finding out about the ϵs for those products, and then picking product j out of the subset A. In my application, the products in subset A can be located within one category or across categories. The subsets are unobserved to the econometrician. The expected benefit from choosing subset A is the maximum expected utility the consumer would obtain from those products, which due to the logit error term can be expressed as following inclusive value:

$$E[\max_{r\in A} u_{ir}] = log[1 + \sum_{r\in A} exp(\delta_r)] \tag{6}$$

where δ_r is simply $X_r\beta + \xi_r$. The consumer always has the outside option, regardless of the set they consider. To find set A, the consumer had to incur the following search costs:

$$c_{iA} = \sum_{r\in A} \gamma\psi_r + \lambda_{iA} \tag{7}$$

where ψ_r reflects the "distance" between the consumers and each product r in set A, and λ_{iA} is a consumer/choice set specific search cost shock, which is logit distributed mean zero with a standard error of 1.[39] This logit shock can be interpreted as an information shock (e.g., links) that gives consumers better access to a certain subset of products. Renaming $\sum_{r\in A}\psi_r$ as ψ_A, the utility of a consumer i of picking subset A is:

$$u_{iA} = log[1 + \sum_{r\in A} exp(\delta_r)] - \gamma\psi_A - \lambda_{iA} \tag{8}$$

Due to the logit error term, the probability consumer i picks subset A from the set of all possible subsets $\mathbb{A}$ is:

$$P_A = \frac{exp(\overline{U_A})}{\sum_{A'\in\mathbb{A}} exp(\overline{U_{A'}})} \tag{9}$$

where $\overline{U_A} = log[1 + \sum_{r\in A} exp(\delta_r)] - \gamma\psi_A$.

[39] The γ coefficient on ψ can vary across products or product groups. However, the assumption that the "distance" of products in a consideration set is additive in the set's search costs is key for obtaining a closed form expression for the choice probabilities.

Similarly, the probability that a consumer picks product j from subset A can be expressed as follows:

$$P_{j|A} = \frac{exp(\delta_j)}{1 + \sum_{r \in A} exp(\delta_r)} \tag{10}$$

Product j belongs to a number of subsets - denoted as A_j. As a result, I can express the unconditional probability of picking product j as follows:

$$s_j = \sum_{A \in A_j} P_A P_{j|A} = \sum_{A \in A_j} \frac{exp(\delta_j)}{1 + \sum_{r \in A} exp(\delta_r)} \frac{exp(\overline{U_A})}{\sum_{A' \in \mathbb{A}} exp(\overline{U_{A'}})} \tag{11}$$

$exp(\overline{U_A})$ can be expressed as $(1 + \sum_{r \in A} exp(\delta_r))exp(-\gamma\psi_A)$. It is possible to show that this expression is equivalent to the following closed form expression:[40]

$$s_j = \frac{\frac{exp(\delta_j)}{1+exp(\gamma\psi_j)}}{1 + \sum_{k \in N} \frac{exp(\delta_k)}{1+exp(\gamma\psi_k)}} \tag{12}$$

In this market share equation s_j does not depend on the subsets anymore. This is important since both the subsets and which subsets product j belongs to are unobservable to the econometrician.

5.2 Demand Estimation - Logit

The market share of the outside option is:

$$s_0 = \frac{1}{1 + \sum_{k \in N} \frac{exp(\delta_k)}{1+exp(\gamma\psi_k)}} \tag{13}$$

Following Berry (1994):

$$ln(\frac{s_j}{s_0}) = \delta_j - ln(1 + exp(\gamma\psi_j)) \tag{14}$$

or:

$$ln(\frac{s_j}{s_0}) = X_j\beta - ln(1 + exp(\gamma\psi_j)) + \xi_j \tag{15}$$

It is straight-forward to estimate this model using GMM, with BLP instruments accounting for the endogeneity of prices in the model. To properly specify this

[40] A full derivation can be found in Moraga Gonzalez, Sandor and Wildenbeest (2015).

estimating equation, I need to decide which variables affect consumer purchasing utility but not search costs, and which variables affect search costs but not purchasing utility.

5.3 Specification A

The number of products in the category of app j is an example of a search cost shifter. Consider products in categories with more apps (larger categories). The number of apps in these categories should not affect consumer utility conditional on consumer discovery - it should only affect the probability that consumers find these products. For example, if there are 10 products in a category bestseller list, it is easier for consumers to look through all 10. If there are 50 products, it takes longer for consumers to scroll through to discover all the products. Additionally, it is also easier for products to be featured in their category if there is a relatively smaller number of products in that category.

There are a number of utility shifting product characteristics that can be included in the model. The size of the app (in MB) could be such a shifter. As well, the average star rating of the app (e.g., 2 star, 3.5 star) would reveal to potential buyers the quality of the app. These can be entered as dummy variables. For example, relative to a baseline of very low rated apps (those with an average rating of 2 stars or less), a 4 star dummy would show how much more utility consumers obtain from a higher rated app. Similarly, past literature on online markets considered information disclosure about a product as a signal of quality (Dranove and Jin 2010, Lewis 2011). With that in mind, the number of screenshots and video previews a developer includes would reveal information about the app to potential consumers. These would be two additional utility shifters.

A particularly simple demand model specification (Specification A) includes the product characteristics discussed above and the number of apps in the category as the search cost parameter. The estimating equation for app j in category c at time t would be as follows:

$$\begin{aligned} ln(\frac{s_{jct}}{s_{0t}}) &= \delta_{jt} - ln(1 + exp(\gamma ln(N_{ct}))) \\ &= X_{jt}\beta - ln(1 + exp(\gamma ln(N_{ct}))) + \xi_{jt} \end{aligned} \tag{16}$$

where X_{jt} are product characteristics, and N_{ct} is the number of apps in category c. In addition to the utility shifters described above, product characteristics also include the prices for paid products, as well as different utility intercepts for free products.[41]

The intuitive argument for the identification of the β parameters is similar to the standard discrete choice demand model - the X characteristics shift consumer utility and affect the purchase decisions of consumers which then affect the market shares. The intuitive argument for the cross-sectional identification of the γ parameter is as follows: two apps with the same observable utility shifting characteristics are located in categories that have a different number of apps (and different search costs). The demand for the app with more other apps in its category should be lower, and their market share should be lower as well. Alternatively, it is also possible to think about identifying the search costs over time: a given app has identical quality over time, but in different time periods there is a different number of apps in its category. Holding everything else constant, it is harder to find the app in more populous category, which reduces its demand and shifts down its market share.

Since I use observational data on app demand and characteristics, prices are likely endogenous with respect to unobservable product quality (ξ). Products with higher ξ will likely have more demand and will have higher prices, meaning that the coefficient on the price variable will not reflect the effect of prices on demand and utility. In order to control for this, I use characteristics based instruments: The average size of all other apps in the category of app j by the same developer of app j, and the average rating of all other apps in the category of app j by the same developer of app j. The characteristics of other products should be correlated with the price of product j through the supply side, but uncorrelated with the demand shock for product j.

The specification of search costs as the number of products is similar to the demand specification of Ackerberg and Rysman (2005). Their model uses the number of products in the category to account for the mechanical increase in consumer utility with a larger array of products. The intuition is that with a larger variety, some products are "squeezed out" - for example, due to limited shelf space. This "squeezing out" effect can be related to search, but it does not necessarily reflects search costs. It can also represent some other consumer costs related to the number of products that shift the market shares. Therefore, it is not clear if the γ parameter represents marginal search costs or the Ackerberg and Rysman (2005) logit model correction.

The utility parameters (β and α) are the underlying "deep" parameters that depend on consumer preferences and should be immutable. By contrast, the γ parameter is not a preference parameter, but rather a parameter that depends on the search technology available to consumers. A change in the search technology should change γ.

[41] I also run the model for free products only, since they represent 80% of the market.

The Android app market setting has the benefit of a natural experiment to disentangle the search costs and other components of γ. Marginal consumer search costs should decrease for games when the split in the categories takes place, but they should stay constant for non-games. The non-search component of γ should remain stable over time. Then, the change in γ between the pre- and post- split period for games identifies the change in consumer search costs as a result of the split.[42] In the non-game categories, γ should remain stable over time.

Table 5 shows estimates of Specification A for Games and Non-Games before and after the split. Columns (1) and (2) show results for Games - Column (1) for free apps only, Column (2) for free and paid apps. Columns (3) and (4) show the results for non-games.

The results from Specification A largely confirm the main hypothesis. The coefficient on the number of apps in the category falls from 0.6 to approximately 0.35 after the split for games in Table 5 in all specifications. This suggests that search costs fell in the games market. The coefficient on the number of apps in the category for Non-Games remains roughly constant between the two periods.

The price coefficients in Columns (2) and (4) suggest that demand for mobile games in this market is relatively inelastic. The median paid game has negligible market share, and a price of \$1.5, meaning that demand elasticity around 1.8. Elasticity for non-games is around 1.[43]

The results from Specification A suggest that the value of increasing the size of a game by 1% (equivalent to 220 KB for the median game) is between 0.5 cents and 0.8 cents for consumers; equivalently, increasing the size of an median app by 5 MB would increase the utility of consumers by 75 cents, or half the price of the median app.

Similarly, the cost to consumers of increasing the number of apps in the category by 1% was approximately 30 cents (20% of a median app's worth). This can reflect the marginal cost of "scrolling" through the app store.[44]

The coefficients on average rating dummies are positive for both games and non-game apps (relative to a baseline of apps with an average star rating of 2 or less).

[42]In fact, the change in γ would help recover a lower bound on the "search cost" component of the parameter.

[43]There may also be identification issues with having a large number of very small firms and few markets and using characteristics based instruments. As Armstrong (2016) shows, BLP instruments can be inconsistent in these environments, since they end up having very little correlation with market shares. The weak instruments could also account for the relative instability of the utility parameters in the samples which include paid apps and games.

[44]Search costs do not enter the utility function linearly, so the marginal rate of substitution between search cost variable ψ and price is: $\frac{MU_\psi}{MU_{price}} = \frac{exp(\gamma\psi)}{1+exp(\gamma\psi)}\frac{\gamma}{\alpha}$

Table 5: **Specification A - Demand Coefficients Estimates**

	Games		Non-Games	
	Free	Free & Paid	Free	Free & Paid
ln(N Apps in Category) - γ	0.693***	0.604***	0.344***	0.521***
	(0.002)	(0.002)	(0.001)	(0.002)
γ × Post Split	-0.361***	-0.294***	-0.011***	0.051***
	(0.002)	(0.002)	(0.001)	(0.003)
2.5 Stars Avg. Rating	2.361***	2.162***	2.69***	2.183***
	(0.005)	(0.005)	(0.002)	(0.004)
3 Stars Avg. Rating	2.352***	2.133***	2.529***	1.200***
	(0.004)	(0.004)	(0.002)	(0.003)
3.5 Stars Avg. Rating	2.795***	2.598***	2.947***	2.472***
	(0.004)	(0.003)	(0.001)	(0.003)
4 Stars Avg. Rating	2.830***	2.624***	2.992***	2.488***
	(0.003)	(0.003)	(0.001)	(0.003)
4.5 Stars Avg. Rating	2.572***	2.402***	2.896***	2.392***
	(0.004)	(0.003)	(0.001)	(0.003)
5 Stars Avg. Rating	1.014***	0.886***	1.310***	0.855***
	(0.004)	(0.002)	(0.001)	(0.003)
Price		-1.221***		-0.781***
		(0.03)		(0.013)
ln(Size)	0.050***	0.048***	0.002***	0.075***
	(0.001)	(0.002)	(0.0002)	(0.001)
N Screenshots	0.025***	0.023***	0.020***	0.027***
	(0.0002)	(0.0001)	(0.0001)	(0.0003)
Video Preview Dummy	0.287***	0.246***	0.285***	0.274***
	(0.002)	(0.002)	(0.001)	(0.003)
Paid App		0.092		0.096**
		(0.327)		(0.039)
Date FE	YES	YES	YES	YES
Observations	3,898,993	4,790,522	22,731,211	28,093,647

Columns 1-2 show results for Games. Columns 3-4 show results for Non-Games.
Columns 1 and 3 include only free apps.
Columns 2 and 4 include both free and paid apps.
Apps with 2 stars or less are the "baseline" category for the star rating dummies.
Price instruments for Columns 2 and 4
include the average characteristics of all other apps in the same category.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Apps with higher average ratings give consumers more utility. The exception to this are apps with an average rating of 5 (out of 5). The coefficient for the 5 star dummy is positive, meaning that consumers receive higher utility from a "5 star app" than a "2 star app." It is, however, lower than all other coefficients, including of a "2.5 star app." A possible reason for this is a popular strategy among firms in online markets of faking 5 star ratings. Luca and Zervas (2016) show that, in the context of Yelp, restaurants with low quality are more likely to buy fake reviews.[45] This is likely what is happening in this market as well: low quality apps buy a large number of fake 5 star reviews which pushes their average rating up to being close to 5.

5.4 Period Specific Search Cost Coefficient Estimates

In addition to the regressions that pool together observations from different months, I also estimate *period-specific* search cost coefficients using a series of cross sectional regressions separately for every month.[46] Figure 9 plots the search cost coefficients for games and non-games over time. For this estimation I use Specification A and the sample of free games and non-games - equivalent to Columns (1) and (3) in Table 5.

The figure shows that the change in the game coefficients happens sharply at the moment of the split in the categories, rather than gradually over time.[47] The search cost coefficient for games hovers around 0.5-0.6 for the year prior to the split in the categories, and drops to less than 0.3 right as the categories are split.[48] By comparison, the non-games search cost coefficient does not show any drastic movement around the split. There is some general instability in the coefficients over time. The figure suggests that search costs for both games and non-games were increasing in mid to late 2014. Unlike the change right after the period of the split, this change in the coefficients is common.

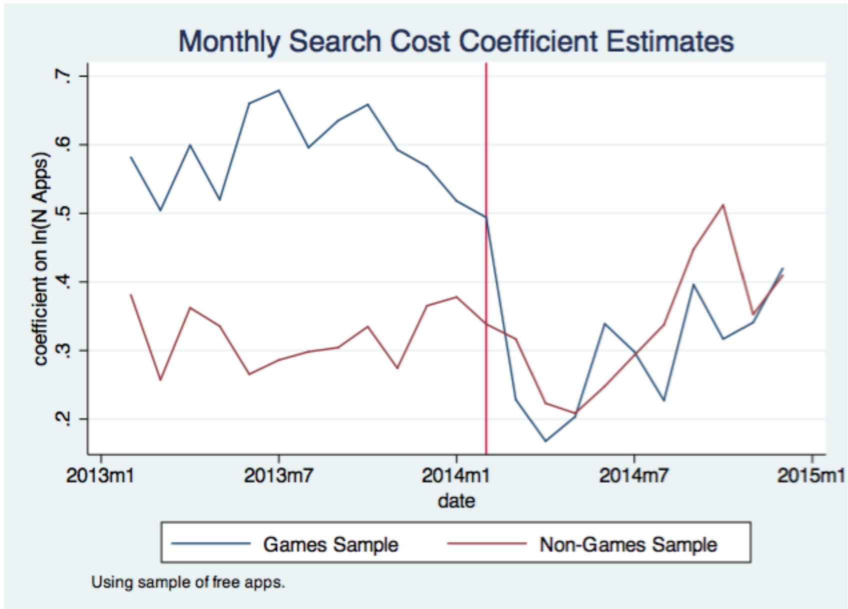
[45] Li, Bresnahan and Yin (2016) show evidence that mobile apps buy fake downloads, though they do not discuss reviews.

[46] I can also do the same exercise by pooling observations from all periods together and estimating interactions between the variables and time dummies. However, estimating a non-linear model with a large number of observations and fixed effects tends to present a computational problem, as the GMM criterion function becomes relatively flat and the discrete jumps from the fixed effects make convergence to the global minimum difficult (see Greene 2004 for a similar discussion about MLE estimation).

[47] Two representative monthly coefficients are plotted in Appendix D, and although they show some instability over time, the average coefficient level before the split is comparable to the average coefficient level after the split.

[48] There is a small estimated drop in search costs the period before the split. This may be driven by my inability to completely capture the the unobservable heterogeneity of the products. For example, if many high quality products enter the market in this period and obtain high downloads, this could appear as an overall decrease in search costs, as more products in the category will have higher sales.

Figure 9



5.5 Specification B

Although Specification A captures many key features of the market, it does not capture the effect of product popularity in period t on the probability of finding the product in period $t + 1$. If a product sold more in the past, it is more likely to be on category best-seller lists and to be featured on the store, making it easier for consumers to find. Keeping the utility variables the same as Specification A, demand Specification B includes the lag of the downloads of product j as part of the search costs in addition to the previously introduced search cost shifter:

$$ln(\frac{s_{jct}}{s_{0t}}) = X_{jt}\beta - ln(1 + exp(\gamma_1 ln(N_{ct}) + \gamma_2 ln(q_{jt-1})) + \xi_{jt} \quad (17)$$

where q_{jt-1} are the downloads of product j at period $t-1$. There are several reasons why I include the lag of the downloads of product j rather than past best-seller rank. Ranks change on a weekly basis and my estimation is done on a monthly basis to limit the number of observations for computational reasons. Also, the implications of ranking as number 1 may be different for different categories. The downloads measure accounts for that - the top ranked app in a less popular category can have fewer downloads than a 10th ranked app in a more popular category. Lastly, the categories change for games - in order to properly account for the effects of changes in categories, I would need to have category specific past ranking parameters. Using past downloads allows me to limit the number of parameters I have to estimate.

The model is now dynamic, and it has standard dynamic panel data estimation problems as described in Nickell (1981) and Arellano and Bond (1991). Namely, if there are persistent product level effects (i.e. unobserved app quality), then OLS or GMM estimates are inconsistent. Even if fixed effects are included (or the model is differenced), there is still correlation between the lagged q and the differenced error term.

I estimate this model as follows: first, I difference Equation (17) above, obtaining:

$$\Delta ln(\frac{s_{jct}}{s_{0t}}) = \Delta X_{jt}\beta - ln\left(\frac{1 + exp(\gamma_1 ln(N_{ct}) + \gamma_2 ln(q_{jt-1})}{1 + exp(\gamma_1 ln(N_{ct-1}) + \gamma_2 ln(q_{jt-2})}\right) + \Delta\xi_{jt} \tag{18}$$

where the Δ represents the difference between the variable in period t and the variable in period $t-1$.

Standard instruments in the literature include the second lag of the downloads of product j: $ln(q_{jt-2})$.[49] This instrument is correlated with the non-linear function of $N_{ct}, N_{ct-1}, ln(q_{jt-1}), ln(q_{jt-2})$ by construction. However, product downloads are imputed in this setting, likelly with some measurement error. If this measurement error is correlated over time, then lagged downloads are going to be correlated with the differenced error term (which will also include the differenced measurement error between periods $t-1$ and $t-2$).

Instead of lags of the downloads, I use lags of BLP instruments as IVs. The intuition is straight-forward: the characteristics of product j's competitors in period $t-2$ should be correlated with the demand for product j in that period, but they should not be correlated with the demand shock for product j (or the difference in the demand shocks). In addition to these instruments, I use ΔX_{jt}.

A standard problem with estimating dynamic panel data models in differences is that of weak instruments. This may be an even greater problem in this application, since the instruments need to be correlated with a non-linear function of variables. To improve the estimates, I follow Blundell and Bond (1998) and add to the GMM system the levels estimating equation (Equation 17). The *differenced* lag of BLP instruments as IVs for the new moment conditions. Again, this instrument is by construction correlated with the lag of the downloads of product j, but it should be uncorrelated with the error term.

Table 6 shows estimates of Specification B for Games and Non-Games before and after the split. As in Table 5, Columns (1) and (2) show results for a subsample of

[49] Sweeting (2013) and Aguirregabiria and Ho (2012).

Games apps, and Columns (3) and (4) show results for a subsample of Non-Game apps. The sample includes only free apps in Columns (1) and (3), and free as well as paid apps in columns (2) and (4).

Table 6: **Specification B - Demand Coefficients Estimates**

	Games		Non-Games	
	Free	Free & Paid	Free	Free & Paid
ln(N Apps in Category) - γ	0.532***	0.658***	0.360***	0.388***
	(0.003)	(0.03)	(0.001)	(0.002)
$\gamma \times$ Post Split	-0.114***	-0.232***	-0.043***	0.124***
	(0.003)	(0.008)	(0.001)	(0.001)
lag Downloads	-0.316***	-0.324***	-0.162***	-0.135***
	(0.005)	(0.012)	(0.001)	(0.001)
lag Downloads $\times$ Post Split	0.082***	0.104***	-0.199***	-0.198***
	(0.009)	(0.019)	(0.002)	(0.002)
2.5 Stars Avg. Rating	1.505***	1.635***	1.958***	1.588***
	(0.01)	(0.087)	(0.003)	(0.003)
3 Stars Avg. Rating	1.527***	1.574***	1.862***	1.474***
	(0.009)	(0.066)	(0.003)	(0.002)
3.5 Stars Avg. Rating	1.812***	1.858***	2.153***	1.798***
	(0.010)	(0.057)	(0.003)	(0.002)
4 Stars Avg. Rating	1.841***	1.836***	2.199***	1.827***
	(0.011)	(0.050)	(0.003)	(0.002)
4.5 Stars Avg. Rating	1.688***	1.756***	2.161***	1.813***
	(0.01)	(0.068)	(0.003)	(0.002)
5 Stars Avg. Rating	0.662***	0.411***	0.971***	0.663***
	(0.006)	(0.012)	(0.002)	(0.002)
Price		-3.212***		-0.290***
		(0.901)		(0.005)
ln(Size)	0.039***	0.085***	0.015***	0.040***
	(0.001)	(0.01)	(0.0002)	(0.0005)
N Screenshots	0.023***	0.024***	0.026***	0.029***
	(0.0002)	(0.001)	(0.0001)	(0.0002)
Video Preview Dummy	0.225***	0.229***	0.214***	0.202***
	(0.002)	(0.012)	(0.002)	(0.002)
Paid App		3.867**		-1.313***
		(1.696)		(0.015)
Date FE	YES	YES	YES	YES
Observations	2,450,174	3,098,476	14,714,777	18,964,181

Columns 1-2 show results for Games. Columns 3-4 show results for Non-Games.
Columns 1 and 3 include only free apps.
Columns 2 and 4 include both free and paid apps.
Apps with 2 stars or less are the "baseline" category for the star rating dummies.
Price instruments for Columns 2 and 4
include the average characteristics of all other apps in the same category.
Instruments for lagged app downloads include the 2nd and 3rd lags
of other app characteristics.
Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The key result from Table 6 is the persistence of the fall in category specific consumer search costs even when additional search cost controls are included. The coefficient on the number of apps in the category still falls for the Game apps regression - in both Columns (1) and (2)- but stays relatively constant for Non-Game apps.

The coefficient on the natural log of lagged downloads is negative for both games and non-games. It suggests that apps with more past downloads are easier to find by consumers. This is consistent with structure of the store, where apps are featured in best-seller lists. After the re-categorization, the coefficient on past downloads for games becomes smaller. This suggests that the benefits that past bestsellers experiences lessens. This is consistent with what happens in the market - with a larger number of categories there are more opportunities for apps to showcase themselves, meaning that less successful apps can showcase themselves more successfully. The coefficient on past downloads also changes for non-games, although in the opposite direction. This suggests that the benefits of being a bestseller increase over time for non-games. Again, this makes sense. There are many thousands of apps in each category and the total number of consumers is increasing over time. It is intuitive that the benefits of being more visible would increase (a la Rosen 1981).

5.6 Robustness Tests

I test the robustness of this model in several ways. A concern may be that that search costs are not specified flexibly enough. For example, paid apps may be harder to find than non-paid apps, or older apps on the market may be harder to find than new apps. In order to address such concerns, I re-estimate the model with more flexible search cost specifications. The results of these additional regressions are in Appendix D. Column (1) of Table 18 shows that the γ coefficient is lower for new apps - that is, they are easier to find. This is consistent with the structure of the app store, which often promotes new products. However, they experience a smaller change in marginal search costs as a result of the re-categorization. Column (2) finds similar effects (though much smaller) effects for paid apps. On average, there are fewer paid apps per category, so they are easier to find than free apps. They also experience lower benefits of due to the re-categorization (although those are still substantial).

An additional possible concern is that the results of search costs are coming through due to the nonlinear specification of demand - due to the functional form of the model. To account for that, I estimate a linear model of demand with the search cost shifters. I use the same set of instruments and estimate both Specification A for game apps with both free and paid products. The results are in Column (3) of Table 18 in Appendix D. They are qualitatively and quantitatively similar to the main results. In particular, the coefficient on search costs declines in the period after the re-categorization. This suggests that the non-linear specification of demand is not driving the results. I continue to primarily use the non-linear specification since it is derived from the theoretical model and the coefficients in it have clear structural interpretations.

5.7 Demand Estimation - Random Coefficients

I introduce three sets of random coefficients into the model. I allow consumers have a random preference for all vertical shifters (δ) as a whole - a random preference for app quality - in the spirit of Fan and Yang (2017). Although it's plausible that consumers have different random preferences for each characteristic, the single random preference for quality facilitates estimation and should broadly capture the main heterogeneity in consumer tastes. This also fits better with the entry model, where developers choose to invest into different levels of δ rather than individual vertical shifters (see section below). The second random coefficient is for price. The last random coefficient is for search costs. Here, again, I group the search costs together, capturing a general individual unobservable heterogeneity in search costs.

I assume that the unobservable heterogeneity is normally distributed, with some mean value (which is normalized to 1), and a variance covariance matrix Σ (to be estimated). I allow for the unobservable preference for quality to be correlated with the unobservable heterogeneity in search costs. This is done to capture the possibility that some consumers like high quality apps and also have high search costs. Ignoring this possible correlation, if it exists, would greatly underestimate the benefits of any changes in search costs. Consumers who have high search costs and who value quality would benefit the most from any reduction in search costs since they would be able to pick better products. I estimate the model using the standard BLP algorithm - simulate the unobservable heterogeneity of consumers via random draws (100 Halton draws), inverting the market shares to identify the δs with the Berry (1994) contraction mapping, and estimating the non-linear parameters - Σ and the search cost parameters - in the outer loop using GMM.

With a relatively small number of markets (35 time periods) and up to a million products in a given market, estimation using the standard BLP algorithm and a full set of products would be much too computationally intensive. To estimate the random coefficients model, I restrict my sample to apps which have made it to the top 500 app list in a given category at least once.[50] Note that since I observe the full universe of apps in every time period, the apps in the restricted sample are observable even when they are not in the top list. This reduces the number of games from nearly 400,000, to 22,000. In order to make sure that the sample is representative and generates similar results to the full sample, I re-estimate the main model with no unobservable heterogeneity. The estimates are in Table 19 of Appendix D, and they are qualitatively and quantitatively similar, suggesting that the sample is representative of the population.

The results of the full model with unobservable heterogeneity are in Tables YYY. Column (1) shows Specification (1) for the sample of free games, Column (2) shows Specification (1) for the sample of both free and paid games, and Columns (3) and (4) show the results using Specification (2). The variance covariance of the

TABLE YYY HERE

The results for the mean coefficients are qualitatively similar as those in the estimates without unobservable heterogeneity. In addition to that, the results suggest that consumers do indeed have positively correlated unobservable heterogeneity in search costs and in quality.

5.8 Entry Model

As described in Section 2.2, key strategic choices by developers include app characteristics and quality (how much to invest in an app), whether to enter the app into the market, whether to make the app free or paid, and what (mutually exclusive) category to enter the app into. For example, a developer who develops a non-game app can technically enter it in 24 potential categories. These categories appeal to potential entrants based on the number of other firms that the app expects to compete with in the category. To model firm entry decisions, I use a static entry model under incomplete information.[51] I will not model the choices of firms to produce free or paid products. I will focus only on free games, using the demand model from Column (1) in Table 5.

In each period, there are NP_t potential free new game apps choosing whether to enter the market. Consider the developer of game j, who chooses whether to invest in quality $w \in \{1, 2, ..., W\}$, as well as entry into categories $c \in \{1, 2, ..., C\}$ in period t. I define quality via the δ measure recovered from the demand model, which is an

[50] This is a standard sample restriction in the empirical literature examining mobile apps - for example, see Leyden (2018).

[51] I assume throughout that firms enter each app independently of the other apps that they own.

estimated weighted index of various app characteristics.[52] Apps are subdivided into groups based on the distribution of δ - low quality apps with the bottom 33% of δ values, medium quality apps with the middle 33% of values, high quality apps with the top 33% of values, and very high quality apps with the top 1% of δ.

I will call the entry variable a_{jcwt}, which is equal to 0 or 1 if the firm chooses to enter an app of type w into category c at time t. The profits of an app j of type w from entering into category c at time t can be expressed as follows:

$$\pi_{jcwt} = \theta_1 \hat{q}(\mathcal{N}_t, w, \boldsymbol{w_{-jt}}, \boldsymbol{a_{-jt}}) - FC_w - \epsilon_{jcwt} \tag{19}$$

where $\hat{q}$ are the predicted downloads from the demand estimates, and the θ parameter reflects how the downloads are converted into profits.[53]

The downloads are a function of the decisions of all other entrants ($\boldsymbol{a}_{-jt}$), as well as the quality of the app (w) and the quality of all other entrants into the category ($\boldsymbol{w_{-jt}}$). $\mathcal{N}_t$ captures the number (and quality) of existing firms which entered in previous periods.[54] As mentioned above, the mapping of the entry and quality decisions into downloads is done via the demand model estimated in the previous section.

The predicted downloads from the demand model are monthly. θ then translates those monthly downloads into lifetime profits. The assumption is that when deciding on entry, the average app developer places most of the weight on their expected downloads in the first month. This is consistent with recent studies of the app market (e.g., Li, Bresnahan and Yin 2016) which suggest that downloads in the first month are extremely important for developers.[55]

FC_w is a parameter representing the fixed costs of entry for an app of quality w, and ϵ_{jcwt} is an idiosyncratic cost shock, assumed to be drawn iid from an Extreme

[52]This is a simplifying assumption. An alternative specification would define groups around different interactions of product characteristics: e.g., apps which are less than 1 MB but have an average rating of 4 stars, and apps which are 1 to 10 MB and have an average rating of 3.5 stars. However, this form of grouping quickly increases the state space and generates a dimensionality problem.

[53]Downloads only partially determine the value of apps to advertisers. The other big component is the engagement of downloading consumers with the apps they downloaded. They are likely correlated - apps with more downloads are those that consumers engage with more. With observational data at the app level, I can only focus on the downloads.

[54]Ignoring the behaviour of apps which have entered the market is a simplifying assumption. It is reasonable in this setting since conditional on entry, apps rarely exit the market and almost never change categories.

[55]An alternative would be to estimate a full dynamic model where developers consider the expected stream of lifetime variable profits. With a large number of developers, this is computationally complex and not particularly feasible.

Value Type 1 distribution. I assume that the shock of each firm is known to that firm, but unknown to other firms or to the econometrician (private information).

I normalize the profits of an app from staying out of the market to zero. Since I only focus on free firms in this analysis, this outside option for the potential free firms also includes entering into the paid market.

The number of firms in the market is correlated with the structural error term. A high idiosyncratic draw for a particular market for a particular firm (high ϵ_{jcwt}) could be due to some unobservable (to the econometrician) characteristics in market c that would encourage other firms to enter as well. Then, markets with high draws of ϵ would also have more competitors - and the number of competitors would be correlated with the error term. In a game of *incomplete* information with iid draws, however, the firm's information set is by assumption closer to the econometrician's information set.

Due to the presence of the idiosyncratic information shock ϵ, firm j's decision is unknown to other firms $(-j)$. The other firms do know the distribution of the cost shock G_ϵ. A firm k can then form expectations about the behaviour of firm j by integrating its' decision over the distribution of private information. These expectations are in the form of a probability P_{jcwt} that firm j will enter category c at time t as type w.

I can then re-write firm j's profit function as follows:

$$\pi^e_{jcwt}(\boldsymbol{P_{-jt}}) - \epsilon_{jcwt} \tag{20}$$

where π^e_{jcwt} are the expected profits of firm j, which are a function of other firms' choice probabilities $(\boldsymbol{P_{-jt}})$.[56]

[56] Expected profits can be expressed as:

$$\pi^e_{jcwt} = \theta_1 E[\hat{q}(\mathcal{N}, w, \boldsymbol{w_{-jt}}, \boldsymbol{a_{-jt}})] - FC_w$$

where $E[\hat{q}]$ are the *expected* sales in category c at time t. The expectation is conditional on the exogenous variables including market characteristics. The expected number of downloads in a market is a function of an expectation about the number of products. It is possible to show that the expected downloads of a firm (as a function of the number of competitors) is the following binomial expression:

$$E[\hat{q}(w, N_{ct})] = \sum_{n=1}^{NP} \binom{NP}{n} (\hat{q}(w,n)) (P_{cwt})^n (1 - P_{cwt})^{NP-n}$$

where P_{cwt} is the probability that a given firm enters category c as type w at time t. I can abstract from the individual firm subscript, since all firms are ex-ante identical aside from their idiosyncratic fixed costs. NP is the total number of potential entrants in the market. This can be a very large number (e.g., 2 times the maximum number of observed firms). Since the number of potential

From the perspective of firm i, firm j chooses to enter a category conditional on their beliefs about the entry probabilities of other firms in the market. Their best response is to enter category cwt ($a_{jcwt} = 1$) if:

$$\pi^e_{jcwt}(\boldsymbol{P_{-jt}}) - \epsilon_{jcwt} > max_{n\in\{1,...,C\},w'\in\{1,...,W\}}\pi^e_{jnw't}(\boldsymbol{P_{-jt}}) - \epsilon_{jnw't} \tag{21}$$

Under the EV Type 1 distribution of the ϵ, the best response probability function of firm j is:

$$Pr(a_{jcwt} = 1) = \frac{exp(\pi^e_{jcwt}(\boldsymbol{P_{-jt}}))}{1 + \sum_{w'\ \in\{1,...,W\}}\sum_{n\in\{1,...,C\}} exp(\pi^e_{jnw't}(\boldsymbol{P_{-jt}}))} \tag{22}$$

This is an incomplete information Bayes-Nash game where each firm has a belief about the probability that other firms enter the market and make their own decisions based on the probability. A Bayesian Nash Equilibrium (BNE) can be defined as a vector of choice probabilities (P) that solves the following fixed point problem:

$$P_{jcwt} = \frac{exp(\pi^e_{jcwt}(\boldsymbol{P_{-jt}}))}{1 + \sum_{w'\ \in\{1,...,W\}}\sum_{n\in\{1,...,C\}} exp(\pi^e_{jnw't}(\boldsymbol{P_{-jt}}))} \tag{23}$$

In the equilibrium, the beliefs of firms about their opponents' entry probabilities are the opponents' best responses to their own beliefs. The beliefs are therefore consistent with equilibrium actions.

5.9 Entry Model Estimation

The idiosyncratic error component is the only difference across firms trying to enter the same market. Therefore, the probabilities of entry for two firms in the same market are the same: $P_{jcwt} = P_{icwt}$. I can then generalize the probability for a given market as P_{cwt}, and write a log likelihood function which is as follows:

$$ln(\mathcal{L}) = \sum_c \sum_w \sum_t [Napps_{cwt}\ ln(P_{cwt}))] \tag{24}$$

where $Napps_{cwt}$ is the actual number of apps of type w in category c that enter in time t. The probability of entry is a function of the parameter vector as well as of all of the other optimal predicted entry probabilities ($\hat{P}_{cwt}$). By taking first order

entrants is very large, and for computational simplicity, I approximate expected sales as follows: $E[\hat{q}(w, N_c t)] = \hat{q}(\mathbf{w}, E[N_c t]) = \hat{q}(\mathbf{w}, P_{cwt}NP)$. That is, I approximate the expectation of a function as a function of the expectation. Although Jensen's inequality applies, preliminary simulations suggest that as the number of products increases, the approximation is close.

conditions with respect to the parameter vector, it is possible to obtain parameter estimates that maximize the probability of obtaining the observed data.

The model can be estimated using an NFXP algorithm. However, it is probable that there are multiple equilibria in the data, meaning that the likelihood function is in fact a likelihood correspondence, and the estimated set of parameters is only one of a number of possible parameter values in the model. As is standard in the literature (Sweeting 2009, Bajari, Hong, Krainer and Nekipelov 2010), I assume that only one equilibrium is played in the data.[57]

I estimate the model using a two step algorithm following Hotz and Miller (1993). The first step generates an estimate of the choice probabilities. Since I do not have characteristics at this stage, I can generate the choice probabilities using simple count data (the average share of apps of a particular type that enter into a category). The second step uses the estimates of the CCPs to back out the parameters from the log likelihood function. I include date and category fixed effects in the second step in order to control for time varying or category varying factors that can influence firm entry decisions.

As mentioned above, I do not directly model the paid part of the market, although it could interact with the free part of the market. I include the number of paid apps in category c at period t as a reduced form control for the effect of the paid market on the free app entry decision. Lastly, the reduced form results suggest that after the announcement of the re-categorization in December 2013, there is entry by developers in anticipation of the actual change in the market. It is difficult to model entry decisions with this component of firm expectations about the future, so I drop the three months between the announcement and the actual change in the market from my estimation.

Table 7 presents supply side parameter estimates. The fixed cost coefficients are monotonically increasing in the quality of the app, as expected. While the fixed costs of producing a low and medium quality app are pretty close, the fixed costs of producing a very high quality app are roughly double the fixed cost of a low quality app.[58]

[57] de Paula and Tang (2012) and Aguirregabiria and Mira (2015) are among recent papers that drop the "single equilibrium" assumption. However, dropping this assumption forces the econometrician to make distributional assumptions about the potential equilibria in the model. This method would be highly computationally intensive, considering the large number of firms in the sample and the large number of potential equilibria.

[58] A timing test suggests that these parameters do not significantly change over time.

Table 7: **Supply Estimates**

	(1)
Downloads (θ_1)	0.00064***
	(0.00001)
Low Quality App FC	3.7***
	(0.24)
Medium Quality App FC	5.7***
	(0.25)
High Quality App FC	6.4***
	(0.25)
Very High Quality App FC	8.5***
	(0.27)
N Paid Apps in Category	-0.0001
	(0.003)
Time FE	YES
Category FE	YES
Observations (months × categories × sizes)	1,302

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.10 Consumer Welfare Decomposition and Counterfactuals

Consumer welfare from a set of products in a standard logit demand model can be calculated (up to a constant) as the expected value of the products:

$$CS = ln(\sum_{j \in N} exp(\delta_j - ln(1 + exp(\gamma\psi_j)))) \tag{25}$$

this measure of consumer surplus is not just the purchase utility, but also includes the search costs associated with the different products.[59] This allows me to simply examine the change in consumer surplus as the difference between the expected

[59] This does not necessarily fully capture the search costs that consumers experience since those depend on the consideration set. It effectively assumes that consumers have consideration sets of single products, or alternatively a large consideration set of all the products.

utilities:

$$\Delta CS = ln(\sum_{j \in N} exp(\delta_j^{powt} - ln(1 + exp(\gamma^{powt}\psi_j)))) \\ - ln(\sum_{j \in N} exp(\delta_j^{pre} - ln(1 + exp(\gamma^{pre}\psi_j)))) \tag{26}$$

In order to convert the change in consumer surplus from "utils" to dollars, I need a measure of the value of money. Recall that the estimates used to calculate the surplus measures come from Column (1) of Table 5, which only looks at free games. To get the value of money I use the price coefficient from Column (2) of Table 5, which includes both free and paid apps.[60] I calculate the dollar value of consumer surplus by dividing Equation (26) by the absolute value of the price coefficient.

The supply model allows for a decomposition of changes in consumer welfare. First, it is possible to fully decompose changes in search costs from the change in product variety. Second, I can additionally decompose the welfare effects of the change in product variety into the welfare effects from the increase in product variety *while holding the quality distribution constant*, and the welfare effects from the change in the quality distribution.

I compute the counterfactuals by simulating the demand and the supply model under different parameter values for 10 periods off the same starting number of firms - the market as it was in November 2013 (the last period before the announcement of the re-categorization).[61] These simulations then produce an equilibrium number of entrants and an equilibrium quality distribution. For example, the "baseline" equilibrium is a market where the search costs of consumers are high. These search costs result in an equilibrium number of competitors across the available categories and an equilibrium quality distribution. It is then possible to compute the welfare of consumers in this market. I can also generate a "post- re-categorization" equilibrium where consumers have low search costs. These demand conditions result in a different number of equilibrium entrants in the available categories and a different distribution of the type of apps that enter into the categories.

For the counterfactuals looking at the welfare effects of the changes in the quality distribution, I define the share of each quality group of the total number of apps in the "baseline" equilibrium as the "baseline" distribution of quality. The share

[60] I do not use Column (2) parameter estimates in the supply side since I cannot account for the decision new entrants to enter as a free or paid product. The parameter estimates are qualitatively the same as in Column (1), and quantitatively very close as well.

[61] Longer simulation periods produce results which are similar qualitatively.

of each quality group of the total number of apps in the "post-split" equilibrium is the new distribution of quality. I can then calculate consumer welfare in a market where the total number of apps is the same as in the "post-split" equilibrium, but the quality distribution is the same as in the "baseline" equilibrium.

On the supply side, I start each simulation of the entry probabilities from the estimated entry probabilities in the data (in the last period before the announcement of the re-categorization) and allow the model to converge to the new equilibrium probabilities. This means that I am looking for a stable equilibrium that is around the equilibrium in the data. There may be multiple counterfactual equilibria for any given set of demand conditions. I do not formally test for the presence of multiple equilibria. I have iterated on a grid of starting probabilities that is centred around the probabilities in the data and have only managed to find single equilibria for every set of demand conditions.

As mentioned above, each firm only takes into account other firms' entry decisions in the free market and not the paid market. The relative separation of the two markets (with different best-seller lists) makes this a reasonable assumption. There may still be strategic linkages between the two parts of the app store - firms considering entering into the free markets may still care about the number of paid apps in the same category (and vice versa). In that sense, the counterfactual equilibria that I find are holding the actions of paid apps constant. Similarly, the consumer welfare changes I examine only reflect the changes in welfare that consumers obtain from the free market.

Table 8 shows the consumer surplus change decompositions.

The results in Columns (1) and (2) show that the increase in product variety had a positive effect on consumer surplus - increasing it by 10% from the baseline equilibrium. In dollars this adds up to 47 million additional dollars of utility per month, or approximately 564 million dollars in additional welfare per year (since there are 100 million US Android users in 2014). There is an even larger effect from the reduction in in marginal search costs for consumers. On average, consumers gain an additional 46.5% in welfare relative to their baseline equilibrium welfare - over 200 additional million dollars per month (approximately 2 dollars per month per consumer). Overall, the net impact of the change in the number of game categories on consumer surplus is strongly positive.

The reduced form results raise a question about the relative welfare effects of changes in product variety and product quality. Counterfactual simulations using the structural model show that while the increase in the share of low quality products reduces consumer surplus by 0.8% (or 3.8 million dollars) relative to its pre-equilibrium baseline, it is smaller than the effect of the increase in product variety.

Table 8: **Estimated Changes to Total Monthly US Consumer Surplus**

	(1) % Change	(2) $ Change
Pre-Split CS	100%	
Change in Variety	+10.7%	+47 million
Change in Quality Distribution	-0.8%	-3.8 million
Reduction in Search Costs	+46.5%	+223.5 million
Overall Increase in CS	+56.4%	266.7 million

There are approximately 100 million Android consumers in 2014.

6 Conclusion

This paper presents evidence regarding the effect of consumer search costs on competition, entry, quality, and product design in online markets, using new detailed data from the Google Play mobile app store. I take advantage of an exogenous shock whereby consumer search improved for some product types (games) in the Google Play store but not for others (non-games).

The data shows that lower search costs increase product entry and that niche products drive most of the entry effects. This is consistent with findings from IO theory (e.g., Bar-Isaac et al 2012), and it is the first clear evidence on this topic in the empirical literature. I also find that the average quality of the affected products falls relative to the unaffected products following the split. This is the first direct empirical evidence of the effect of changes in consumer search costs on quality.

I set up a structural demand model which estimates consumer choice coefficients related to utility and to search in order to measure the overall welfare effects, and the contribution of different factors to welfare. I find that, consistently with the reduced form results, the search costs associated with game apps declined after the split in the game categories. I also set up an entry model with an incomplete information framework where apps choose to enter into different categories and invest in quality based on expected profits (and downloads).

I show that the re-categorization of games increased consumer welfare in the mobile app market; a decomposition of the welfare changes shows that most of the increase came from a reduction in marginal search costs. There is also an increase in welfare due to the increase in greater product variety, and a smaller fall in welfare due to lower product quality.

The main policy concern related to search is not reducing search costs for consumers, but rather increasing search costs.[62] The empirical evidence in this paper suggests that there are two consumer welfare implications to policies that *increase* search costs online: First, consumer welfare falls because marginal search costs increase. This is the direct negative effect of increasing search cost. Second, consumer welfare falls because of foreclosure; firms that would have entered the market in the presence of low search costs now choose not to enter the market, which reduces the product variety available to consumers. Overall, the results of this paper suggest that policies which increase search costs in online markets may have strongly detrimental effects to consumer welfare.

For mobile app stores, the incentives of the platform and welfare incentives are aligned. App stores should want to reduce consumer search costs in order to stimulate search, entry and downloads. However, other online platforms may not have the same incentives. The EUR2.4 billion fine to Google in the EU provides an interesting case.[63] Google's policy in the search engine market (e.g., hotels, flights) was judged to have increased consumer search costs. This clearly benefitted Google, but the harm was not limited to existing competing platforms - e.g., Expedia, or Kayak. There was also potential consumer harm due to the foreclosure of future competitors from the market.[64]

[62] Wall Street Journal

[63] TheVerge.com

[64] There are also potential benefits to Google's search engine policies that my model does not capture. For example, the quality of Google's products is arguably higher than their competitors' qualities. Thus, by presenting the highest quality product first, Google minimizes the total search costs consumers pay.

References

Ackerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects. *RAND Journal of Economics 36*(4), 771–789.

Aguiar, L. and J. Waldfogel (2016). Even the losers get lucky sometimes: New products and the evolution of music quality since napster. *Information Economics and Policy 34*, 1–15.

Aguirregabiria, V. and C.-Y. Ho (2012). A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics 168*(1), 156–173.

Aguirregabiria, V. and P. Mira (2015). Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity. *University of Toronto Department of Economics Working Paper 474*.

Anderson, S. P. and R. Renault (1999). Pricing, product diversity, and search costs: A bertrand-chamberlin-diamond model. *The RAND Journal of Economics*, 719–735.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies 58*(2), 277–297.

Armstrong, M., J. Vickers, and J. Zhou (2009). Prominence and consumer search. *The RAND Journal of Economics 40*(2), 209–233.

Armstrong, T. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica (Forthcoming)*.

Athey, S. and G. Imbens (2015). Machine learning methods for estimating heterogeneous causal effects.

Augereau, A., S. Greenstein, and M. Rysman (2006). Coordination versus differentiation in a standards war: 56k modems. *The RAND Journal of Economics 37*(4), 887–909.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2009). Estimating static games of incomplete information. *Journal of Business and Economic Studies*.

Bar-Isaac, H., G. Caruana, and V. Cuñat Martinez (2012). Search, design and market structure. *American Economic Review 102*(2), 1140–1160.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T. (1999). Free entry and social inefficiency in radio broadcasting. *The RAND Journal of Economics 30*(3), 397–420.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics 87*(1), 115–143.

Bronnenberg, B., J. B. Kim, and P. Albuquerque (2016). The probit choice model under sequential search with an application to online retailing. *Management Science*.

Brynjolfsson, E., F. Eggers, and A. Gannamaneni (2017). Using massive online choice experiments to measure changes in well-being. In *5th IMF Statistical Forum.*

Brynjolfsson, E., Y. Hu, and D. Simester (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science 57*(8), 1373–1386.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science 49*(11), 1580–1596.

Brynjolfsson, E. and M. D. Smith (2000). Frictionless commerce? a comparison of internet and conventional retailers. *Management science 46*(4), 563–585.

Cachon, G. P., C. Terwiesch, and Y. Xu (2008). On the effects of consumer search and firm entry in a multiproduct competitive market. *Marketing Science 27*(3), 461–473.

Chandra, A. and M. Tappata (2011). Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics 42*(4), 681–704.

Chen, Y. and C. He (2011). Paid placement: Advertising and search on the internet. *The Economic Journal 121*(556), F309–F328.

Chevalier, J. and A. Goolsbee (2003). Measuring prices and price competition online: Amazon. com and barnesandnoble. com. *Quantitative Marketing and Economics 1*(2), 203–222.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research 43*(3), 345–354.

Ching, A., T. Erdem, and M. Keane (2009). The price consideration model of brand choice. *Journal of Applied Econometrics 24*(3), 393–420.

Chiou, L. (2015). Vertical integration and antitrust in search markets. *Available at SSRN 2663557*.

Crawford, G. S., R. S. Lee, M. D. Whinston, and A. Yurukoglu (2015). The welfare effects of vertical integration in multichannel television markets. Technical report, National Bureau of Economic Research.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review 102*(6), 2955–2980.

De los Santos, B. and S. Koulayev (2017). Optimizing click-through in online rankings with endogenous search refinement. *Management Science (forthcoming)*.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Diamond, P. A. (1971). A model of price adjustment. *Journal of economic theory 3*(2), 156–168.

Dranove, D. and G. Z. Jin (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature 48*(4), 935–963.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 1429–1451.

Ellison, G. and S. F. Ellison (2005). Lessons about markets from the internet. *Journal of Economic Perspectives*, 139–158.

Ferreira, F. and J. Waldfogel (2013). Pop internationalism: has half a century of world music trade displaced local culture? *The Economic Journal 123*(569), 634–664.

Fishman, A. and N. Levy (2015). Search costs and investment in quality. *The Journal of Industrial Economics 63*(4), 625–641.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives 30*(1), 185–206.

Gandhi, A., Z. Lu, and X. Shi (2014). Demand estimation with scanner data: Revisiting the loss-leader hypothesis. In *URL http://economics. uchicago. edu/workshops/pdfs/Gandhi_Scanner_Demand_v2. pdf.*

Garg, R. and R. Telang (2012). Inferring app demand from publicly available data. *MIS Quarterly, Forthcoming*.

Ghose, A., A. Goldfarb, and S. P. Han (2012). How is the mobile internet different? search costs and local activities. *Information Systems Research 24*(3), 613–631.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Goldmanis, M., A. Hortaçsu, C. Syverson, and Ö. Emre (2010). E-commerce and the market structure of retail industries*. *The Economic Journal 120*(545), 651–682.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal 7*(1), 98–119.

Hendricks, K. and A. Sorensen (2009). Information and the skewness of music sales. *Journal of Political Economy 117*(2), 324–369.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2015). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *Consumer Awareness, and Choice: Evidence from the US Banking Industry (September 16, 2015)*.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *The Quarterly Journal of Economics 119*(2), 403–456.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies 60*(3), 497–529.

Ipsos (2013). Mobile app marketing insights: How consumers really find and use your apps. Technical report, Ipsos and Google.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing science 29*(6), 1001–1023.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics 45*(3), 553–575.

Larson, N. (2013). Niche products, generic products, and consumer search. *Economic Theory 52*(2), 793–832.

Levin, J. D. (2011). The economics of internet markets. Technical report, National Bureau of Economic Research.

Lewis, G. (2011). Asymmetric information, adverse selection and online disclosure: The case of ebay motors. *The American Economic Review 101*(4), 1535–1546.

Li, X., T. Bresnahan, and P.-L. Yin (2016). Paying incumbents and customers to enter an industry: Buying downloads. *Available at SSRN*.

Liu, Y., D. Nekipelov, and M. Park (2014). Timely versus quality innovation: The case of mobile applications on itunes and google play. *NBER Working Paper*.

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. *The RAND Journal of Economics*, 48–58.

Moraga Gonzalez, J. L., Z. Sándor, and M. R. Wildenbeest (2015). Consumer search and prices in the automobile market.

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy 78*(2), 311–329.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

Nielsen (2011). State of the media: Mobile usage trends. Technical report.

Rosen, S. (1981). The economics of superstars. *The American economic review 71*(5), 845–858.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, 619–640.

Sorensen, A. T. (2007). Bestseller lists and product variety*. *The journal of industrial economics 55*(4), 715–738.

Stahl, D. O. (1989). Oligopolistic pricing with sequential consumer search. *The American Economic Review*, 700–712.

Stigler, G. J. (1961). The economics of information. *The journal of political economy*, 213–225.

Sweeting, A. (2009). The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *The RAND Journal of Economics 40*(4), 710–742.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica 81*(5), 1763–1803.

Varian, H. R. (2007). Position auctions. *international Journal of industrial Organization 25*(6), 1163–1178.

Waldfogel, J. (2011). Bye, bye, miss american pie? the supply of new recorded music since napster. Technical report, National Bureau of Economic Research.

Wildenbeest, M. R. (2011). An empirical model of search with vertically differentiated products. *The RAND Journal of Economics 42*(4), 729–757.

Yang, H. (2013). Targeted search and the long tail effect. *The RAND Journal of Economics 44*(4), 733–756.

Zentner, A., M. Smith, and C. Kaya (2013). How video rental patterns change as consumers move online. *Management Science 59*(11), 2622–2634.

Zhou, J. (2014). Multiproduct search and the joint search effect. *The American Economic Review 104*(9), 2918–2939.

7 Appendix A: List of Categories

Table 9: Google Play Game Categories Before and After March 2014

Before	After
Casual	Casual
Racing	Racing
Sports	Sports
Arcade & Action	Action
	Adventure
	Arcade
	Board
Cards & Casino	Card
	Casino
	Education
	Family
	Music
	Live Wallpaper
Brain & Puzzle	Puzzle
	Role Playing
	Simulation
	Strategy
	Trivia
	Word

Table 10: Google Play Non-Game Categories

Books & Reference	Libraries & Demo	Productivity
Business	Lifestyle	Shopping
Comics	Media & Video	Social
Communications	Medical	Sports
Education	Music & Audio	Tools
Entertainment	News & Magazines	Transportation
Finance	Personalization	Travel & Local
Health & Fitness	Photography	Weather

8 Appendix B

8.1 Data Management - Categories for Reduced Form Analysis

In order to examine entry in this market at the category level, I need to calculate the number of unique apps that appear in each month in each category. As the category composition changed in March 2014, the categories are not directly comparable before and after March 2014. However, since I can track apps over time, I can construct "counterfactual categories" for the pre-March 2014 data. For example, even though the "Family Games" category did not exist before March 2014, I can see all the apps that switched into this category from the old categories ("Aracde & Action", "Card & Casino", "Board & Puzzle"). Assuming that the apps that switched into "Family Games" would have always entered into this category, I can then calculate how many apps would have entered into "Family Games" before March 2014.[65]

If an app entered and exited the market before March 2014, it does not have a post-March 2014 category. If I ignore these apps, this creates a measurement error that would be worse the further I go back in time. I resolve this problem by exploiting the detailed textual descriptions of the apps found in the data. I use a Random Forest machine learning algorithm that first maps the descriptions of the classified post-March 2014 apps into categories, and then goes back and applies this mapping to the descriptions of apps that entered and exited the market before March 2014.

More formally, after removing "stopwords" (e.g., "and", "or") I convert app descriptions into vectors of words and then into term frequency-inverse document fre-

[65]This is not a "true" counterfactual since it does not take into account the strategic decision-making of apps. See below for more.

quencies. This method allows me to assign the highest weight to words appear frequently in a particular description relative to the average description. I use the April 2014 apps as the training set for the Random Forest classifier (I obtain similar results with other classifiers such as KNN). I then apply the classifier to apps in every month prior to March 2014. This is very similar to how Liu, Nekipelov, and Park (2014) map Google Play categories into Apple iTunes categories.

These "counterfactual" categories do not take into account the strategic decisions of apps, but are a purely mechanical allocation based on the apps' descriptions. It is entirely possible that if strategic decisions were taken into account, the categories would be somewhat different. The entry model in Section 5 tackles some of these issues.

8.2 Data Management - Predicting App Downloads

The raw data includes a range of downloads that an app has accrued over its lifetime - for example [1-5] downloads, or [1,000-5,000] downloads, or [1 million - 5 million] downloads. The full list of download ranges is below. Since this range is observable monthly or weekly, it is conceptually straight forward to define "per-period downloads" as the difference in lifetime downloads - for example, it can be the difference in the lower bounds of lifetime downloads, or in the average of the lifetime downloads.

An issue with this measure is that the size of the range increases with the number of downloads. For example, the ranges start at 4 downloads ([1-5], [5-10]), but then increase to 40 ([10-50]), and eventually to 400 million ([100 million - 500 million]). This introduces two sources of measurement error, which become worse for more successful apps: (1) this measure will overstate the per period downloads for successful apps that move from one bandwidth to another. For example, if a firm has a bandwidth of 100 thousand to 500 thousand downloads, and then they move up to 500 thousand to 1 million downloads in the next period, it could mean that they sold 500 thousand units this period, or 3 units. (2) this measure will understate the per-month downloads for successful apps that do not move bandwidths, since moving bandwidths is harder the more successful you are. For example, an app in the [100 million - 500 million] download bandwidth can sell millions of units every month while still remaining in the same bandwidth.

To recover the weekly or monthly downloads of apps in the Google Play market, I rely on two features of the data. First, the bandwidth of lifetime downloads for new entrants is equal to the bandwidth on per-week downloads: I know that a new app in the 10 thousand to 50 thousand range sold between 10 thousand and 50 thousand units. Second, I observe the weekly category rankings which reflect the 500 most-

Table 11: List of Cumulative Download Ranges

Lower Bound	Upper Bound
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 million
1 million	5 million
5 million	10 million
10 million	50 million
50 million	100 million
100 million	500 million
500 million	1 billion

downloaded apps in each category roughly over the past week.[66] If I look at the market at a weekly frequency, then I know the rankings and downloads of new apps (see Appendix B for summary statistics regarding these apps).[67] This allows me to use these apps to predict the downloads of other apps in the market, after making an assumption about the relationship between app ranks and downloads.

Several past studies of online markets with best-seller lists assume that the rank-downloads relationship is defined by a Pareto Distribution (Chevalier and Goolsbee 2003, Chevalier and Mayzlin 2006, Garg and Telang 2012).[68] That is, there is a

[66]It is not known how the lists are determined, but Google releases (AdWeek.com) as well as anecdotal industry evidence (Quora) suggest that they reflect the downloads of apps over the previous several days.

[67]I can assign the lower bound of the bandwidth as the number of weekly downloads, the upper part of the bandwidth, or the average of the bandwidth. In the rest of the analysis of this paper I will assign the lower end of the bandwidth, although the results do not change if I assign the other measures.

[68]It is possible that the Pareto distribution is not entirely correct for predicting downloads in this market. In particular, Eeckhout (2004) and Gabaix (2016) show that the Pareto distribution

negative exponential distribution where an app at rank n has exponentially fewer downloads than the app at rank $n+1$. I can use this assumption and fit a Pareto distribution (which consists of one parameter) for every week and category around the observations that I have for new apps. I do this by running an OLS regression of the logarithm of the rank of new app j in category c at week t on the logarithm of the downloads for every category and week:

$$ln(\text{Downloads}_{jct}) = \delta_c + \delta_{month} + \beta ln(\text{Rank}_{jct}) + \mu_{jct}$$

where the δs are category and month dummies, and where μ_{jct} is a random mean zero variable representing measurement error. The β is a slope coefficient.[69] As the dependent variable *Downloads* I use the lower bound of the bandwidth (minimum downloads in a week).[70] After running the regression (see results tables below), and knowing the number of apps in each category in each period, I can predict the downloads of all apps in the market (e.g., if a category has 2,000 apps, I can generate a prediction for the downloads of the 2,000th app). Only the top 500 ranks are observed. To generate rankings for the unranked apps, I sort them based on their number of cumulative lifetime downloads and their age, and then break up any ties by randomizing.[71]

This prediction algorithm depends, in part, on variation in the rankings of apps over time. New apps should be able to enter into the rankings at different points in the distribution for me to estimate the Pareto relationship accurately. This seems to be true in the data. While there is a large proportion of apps that do not change their rankings from week to week, there are also many apps that move at least two spots on a weekly basis. A graph showing the distribution in weekly changes in app rankings is below.

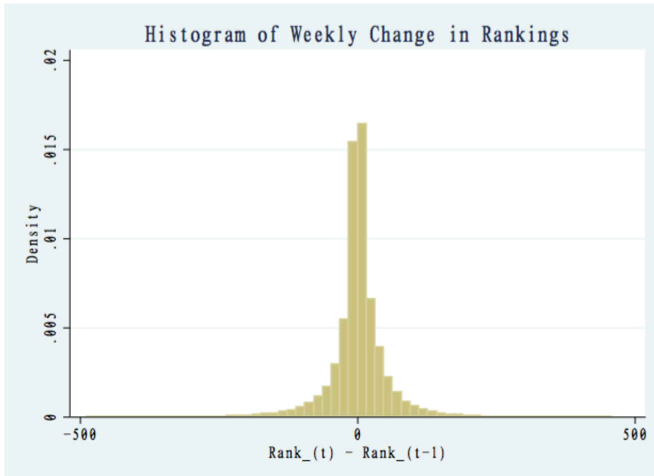
accurately predicts the rank-size relationship for the upper tail of the distribution but not for the lower tail. There, the relationship is more accurately characterized by the log-normal distribution. However, since the Pareto distribution has only one parameter, whereas the log-normal distribution has two, the Pareto distribution is simpler to estimate.

[69] I have also experimented with slope coefficients that vary by category and month, and the results do not change qualitatively.

[70] Results using the upper bound or an average clearly overstate the number of downloads - for example, each of the top 50 apps are predicted to have over 10 million downloads weekly. However, the main results still hold qualitatively.

[71] To check that randomization does not in itself generate any of the results, I've rerun the analysis several times with different randomized seeds. The results do not change.

Figure 10



8.2.1 Summary Statistics of New Apps

Table 12: **Summary Stats of New Apps at Weekly Level**

Variable	Mean	Std. Dev.	Min	Max	N Obs
Games					
Download Lower Bound	27,802	245,351	0	1 million	10,863
Rank	256	147	1	500	10,863
Non-Games					
Download Lower Bound	13,982	414,378	0	5 million	17,375
Rank	278	140	1	500	17,375

Table 13: **Regression Results on ln(Downloads)**

	(1) Games	(2) Non-Games
ln(Rank)	-1.168***	-0.926***
	(0.079)	(0.055)
Month FE	YES	YES
Category FE	YES	YES
ln(Rank) interactions with Month FE	YES	YES
Observations	10,020	16,738
R-squared	0.490	0.530

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

9 Appendix C: Additional Reduced Form Results

9.1 Weekly Game and Non-Game App Entry

9.2 Test of the Timing of the Entry Treatment Effect

With monthly data I can test whether the treatment effect indeed occurred in March 2014 rather than in an earlier period (say, in mid 2013) but is still absorbed by the treatment dummy. To do that, I allow the treatment effect to vary by month by introducing interactions between monthly date dummies and the treatment dummy, effectively estimating a potential treatment effect for every month (relative to some early baseline period). The estimating equation then changes slightly to become:

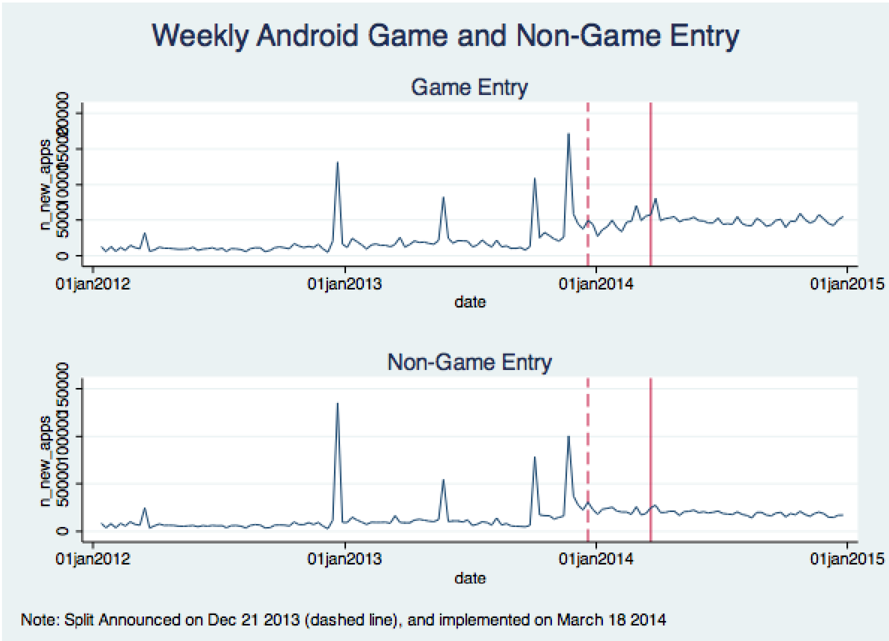
$$Y_{ct} = \sum_t \tau_t \ (Game_c \times \delta_t) + \delta_c + \delta_t + \epsilon_{ct} \tag{27}$$

where the τ_ts now capture period specific treatment effects, relative to a baseline period - March 2012, the first period of the sample.[72]

Figure 8 above shows the period specific treatment effects. It suggests that the main treatment effect hits exactly after the increase in the number of categories. There are two periods of pre-split positive effects (one of those statistically significant

[72] I can use different baselines, which give similar results.

Figure 11



at the 95 percent level), but these can be explained since the split was announced in December 2013, meaning that some firms may have entered the market early to be better positioned for when the split takes place.

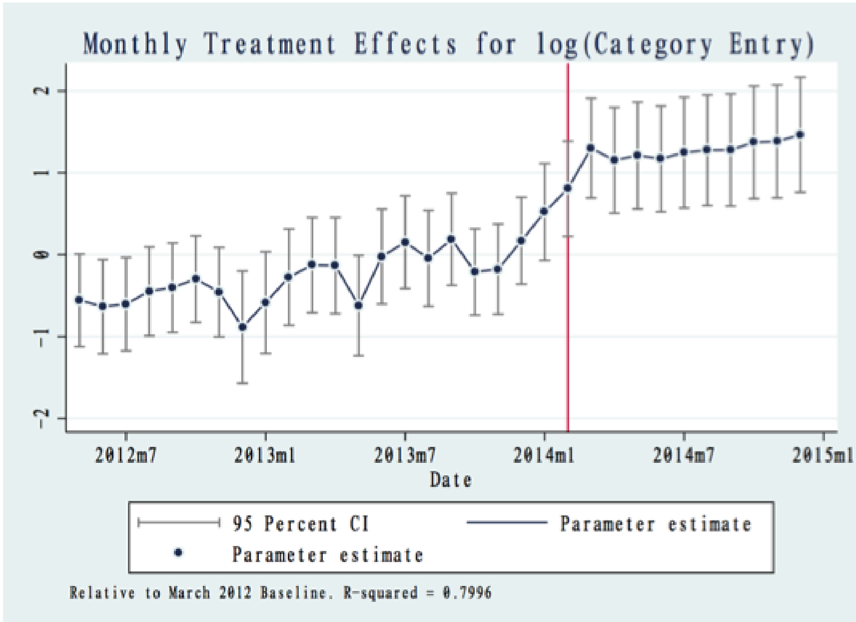
9.3 Category Specific Entry Treatment Coefficient Estimates

9.4 Quality: Mean Size of the App (in MB)

Table 6 shows regression results from the same type of regressions as Table 4, but using the mean size of the apps in MB as the dependent variable. The size of the app is representative of the quality of the app as described above. Although Column (1) at the game/non-game level suggests that average quality for games increased, Columns (2) and (3), which are at the category level, suggest the opposite. They show that, on average, new game apps experienced a decline in average app size relative to new non-game apps following the split in the game categories, representing a relative decline in quality.

The (game) category specific heterogeneous treatment effects - presented in Figure 9 below - shed some light on the difference between the aggregate and the category specific regression coefficients above. In particular, there is a large dispersion in

Figure 12



the coefficients across the categories. Additionally, while most of the coefficients are negative, there are a few categories that have positive treatment effects, and which could have driven the overall results (the categories with the most negative coefficients are relatively small).

The relationship between the coefficient estimates at the category level and category size is similar to the 1 star share relationship - the correlation is roughly positive. It suggests that smaller game categories experienced more substantial drops in the quality of new apps as compared to the larger game categories. Similarly, the right panel shows the relationship between average category app size in the pre-split period and the treatment coefficients. As in the ratings regression, it suggests that the highest quality categories - i.e. the categories with the highest average app size - experienced the biggest drop in quality after the split (had the largest negative coefficients).

9.4.1 Category Specific Size Treatment Coefficient Estimates

Table 14: **Category Specific Treatment Effects on log(N Entrants)**

Category Name	τ_g	Category Name	τ_g
Arcade	0.040	Educational	2.383***
	(0.106)		(0.137)
Card	-0.405**	Strategy	1.985***
	(0.163)		(0.123)
Casual	0.011	Trivia	2.241***
	(0.100)		(0.132)
Puzzle	-0.209*	Word	1.474***
	(0.113)		(0.132)
Action	2.061***	Adventure	2.702***
	(0.112)		(0.143)
Board	1.883***	Family	2.923***
	(0.116)		(0.168)
Casino	1.536***	Music	2.532***
	(0.151)		(0.204)
Racing	0.395***	Role Playing	1.804***
	(0.117)		(0.119)
Sports Games	0.010	Simulation	3.140***
	(0.143)		(0.160)
Observations	1,469		
R-squared	0.829		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

9.5 Placebo Tests for Quality

While the reduced form results are suggestive, some of the category specific results, particularly those that rely on category size, could be driven by mean reversion or the size of the categories rather than the treatment effect. To disprove this hypothesis I run a placebo test.

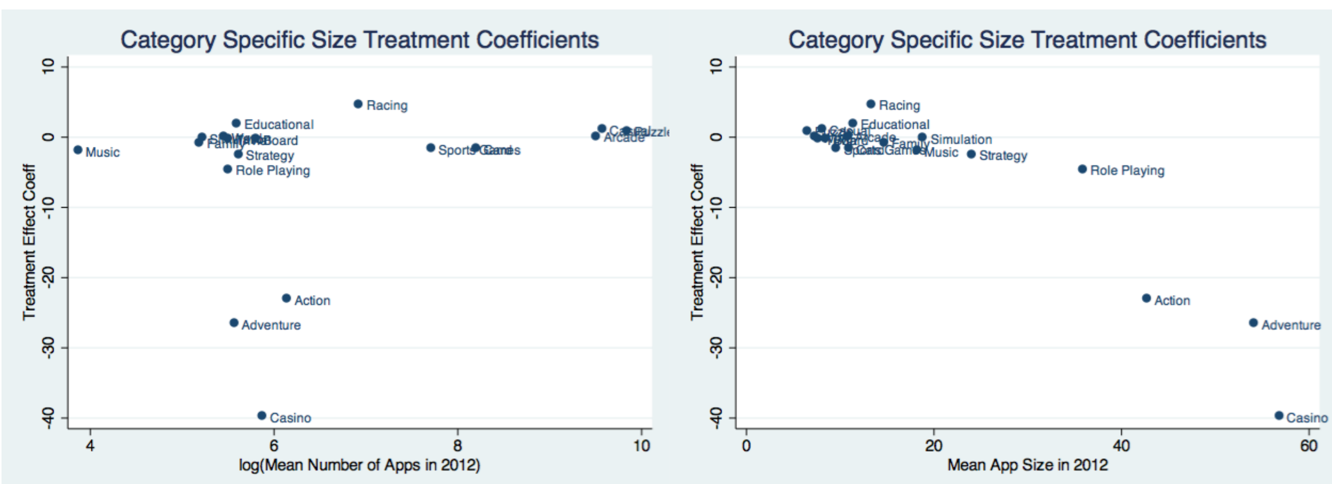
I split my three year sample into two - a subsample of observations for January 2012 to December 2013, and a subsample of observations from January 2013 to December 2014. The split in the categories and the reduction in search costs happens

Table 15: **Regression Estimates for Mean App Size (in MB)**

VARIABLES	(1) Category Groups	(2) Separate Cats	(3) w/ Cat FE
Games × Post Split Period	2.064***	-5.299***	-5.295***
	(0.601)	(1.234)	(1.074)
Games (Treated Group)	6.325***	16.160***	
	(0.411)	(1.101)	
Time FE	YES	YES	YES
Category FE		NO	YES
Observations	70	1,468	1,468
R-squared	0.949	0.265	0.554

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 13



during the second subsample (in March 2014), although the difference in size between different Game app categories are persistent throughout the two subsamples.

Consequently, I run a category specific regression (analogous Column 2 in Tables 4 and 5) for both subsample. In the 2012/3 subsample, I set up a placebo treatment to start in March 2013 (exactly one year before the true treatment starts). I

Table 16: **Category Specific Treatment Effects on App Size (in MB)**

Category Name	τ_g	Category Name	τ_g
Arcade	-3.58***	Educational	-4.68***
	(1.21)		(1.36)
Card	-0.825	Strategy	-6.18***
	(2.82)		(2.0)
Casual	-3.02***	Trivia	-1.27
	(0.853)		(1.75)
Puzzle	-1.55	Word	-4.66**
	(1.08)		(1.82)
Action	-27.2***	Adventure	-30.4***
	(5.4)		(8.5)
Board	-4.84***	Family	-7.72***
	(1.52)		(1.14)
Casino	-48.7***	Music	-2.67
	(7.1)		(4.42)
Racing	-9.94***	Role Playing	-11.84**
	(1.36)		(4.4)
Sports Games	-3.8*	Simulation	-6.22***
	(2.0)		(1.66)
Observations	1,469		
R-squared	0.560		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

expect that the placebo treatment will generate zero treatment effects, whereas the true treatment in the second part of the sample should generate non-zero treatment effects.

The following table shows the regression results for the share of 1 star ratings, and for the mean size of the app for the two periods. Columns (1) and (3) show the 2012/3 regressions, whereas Columns (2) and (4) show the 2013/4 regressions.

Both placebo treatment effects in Columns (1) and (3) are not statistically significant. The placebo treatment for the 1 star rating regression is a full order of magnitude lower than the "true" treatment, and the placebo treatment for the mean

Table 17: **Placebo Test Regressions Results**

	Share of 1 Star Ratings		Mean App Size (MB)	
	(1)	(2)	(3)	(4)
	2012/3	2013/4	2012/3	2013/4
Placebo Treatment	0.002		2.077	
	(0.004)		(1.780)	
True Treatment		0.028***		-5.562***
		(0.003)		(1.022)
Time FE	YES	YES	YES	YES
Category FE	YES	YES	YES	YES
Observations	881	966	881	966
R-squared	0.350	0.468	0.602	0.646

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

size regression is going in the opposite direction of the "true" treatment effect (both are not statistically significant).

I also run the category specific treatment effects for the placebo period for the share of 1 star ratings and for the mean size of the apps. These results are pictured in the two figures below. For the placebo regression, the category specific treatment effects are plotted against the mean number of apps in the category in 2012, and for the "true" regression, the category specific treatment effects are plotted against the mean number of apps in the category in 2013. The ordering of the categories does not change between 2012 and 2013, meaning that the relationship between the treatment coefficients and category size should be comparable between the two sets of regressions.

In the 1 star share regressions, the placebo category treatment coefficients have essentially a zero correlation with the size of the categories (in terms of the number of apps). This, together with their very small magnitude, suggests that the placebo treatment essentially had no effect on the share of 1 star reviews, as expected. The magnitudes are much larger (often twice as big) in the "true" regression, and there is a clearly negative correlation between category size and the treatment effects - as expected.

The mean size regressions are similar. In the placebo regressions, there is a weakly

Figure 14

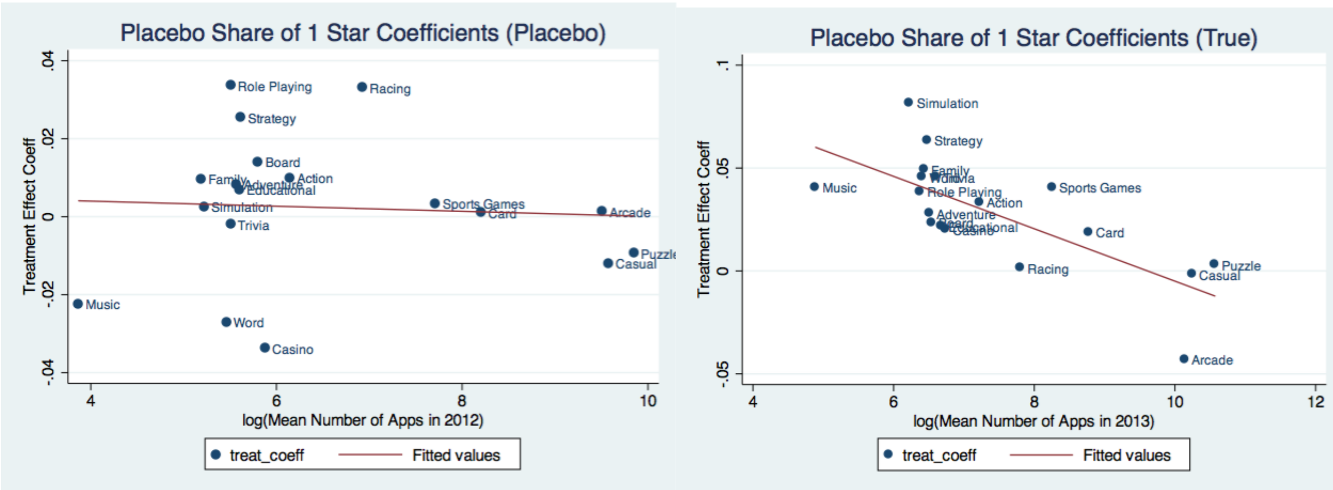
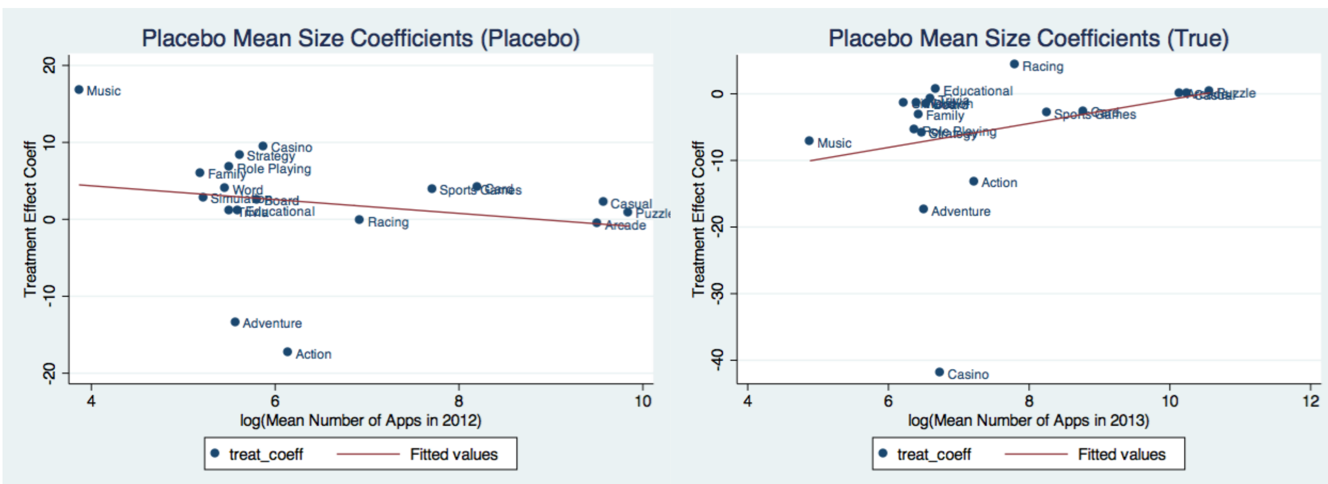


Figure 15

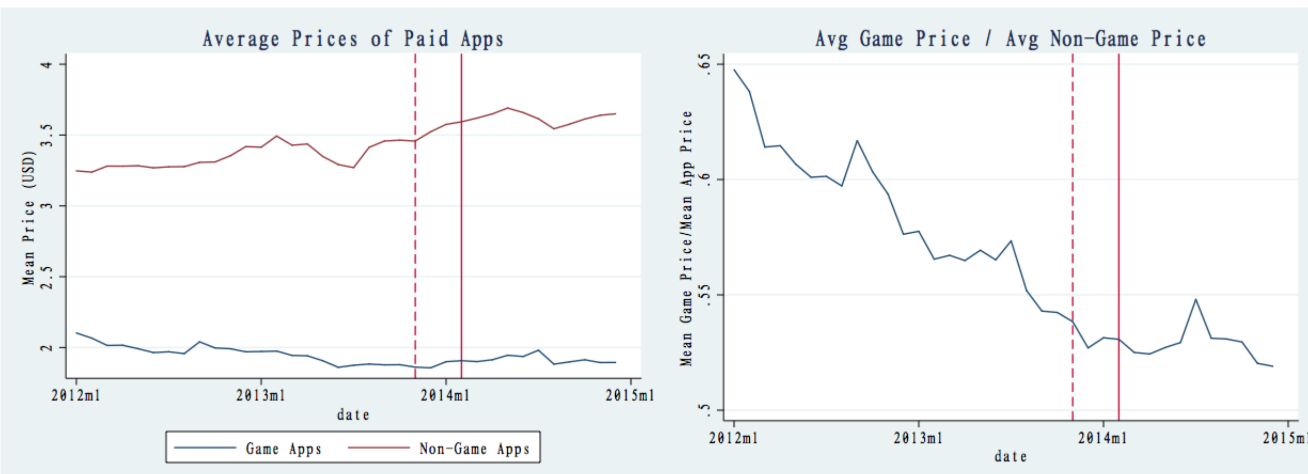


negative relationship between category size and the placebo treatment effects, which are mostly positive in magnitude and not statistically significant. This suggests that in the pre-split period, the smallest categories experienced a larger increase in mean size (increase in quality) relative to the larger categories. This is consistent with the story that apps in the small categories (or types) had to be "better" to be successful. By contrast, the relationship between the treatment coefficients and category size is positive in the "true" period, and the coefficients themselves are mostly negative for the small categories. Again, this is consistent with the story of the split in categories either reducing competition or otherwise reducing the incentives of developers in small categories to produce "high quality" apps.

9.6 Prices

The average prices for all paid apps and for new paid apps do not follow parallel trends for games and non-games.[73] Average prices for all games fall as compared to non-games, driven by the absolute decline in game prices and the absolute increase in non-game prices over the sample period (Figure 10). The absolute decline in game pricing can be a reflection of the increasing importance of in-app advertising and in-app purchases in the app economy. Developers view more expensive paid apps to be less profitable than less expensive paid apps or free apps, where developers recover their fixed costs more slowly over time. After the announcement of the split, the average prices of games start increasing again, meaning that the ratio of average game to non-game prices stabilizes in the post-split period.

Figure 16

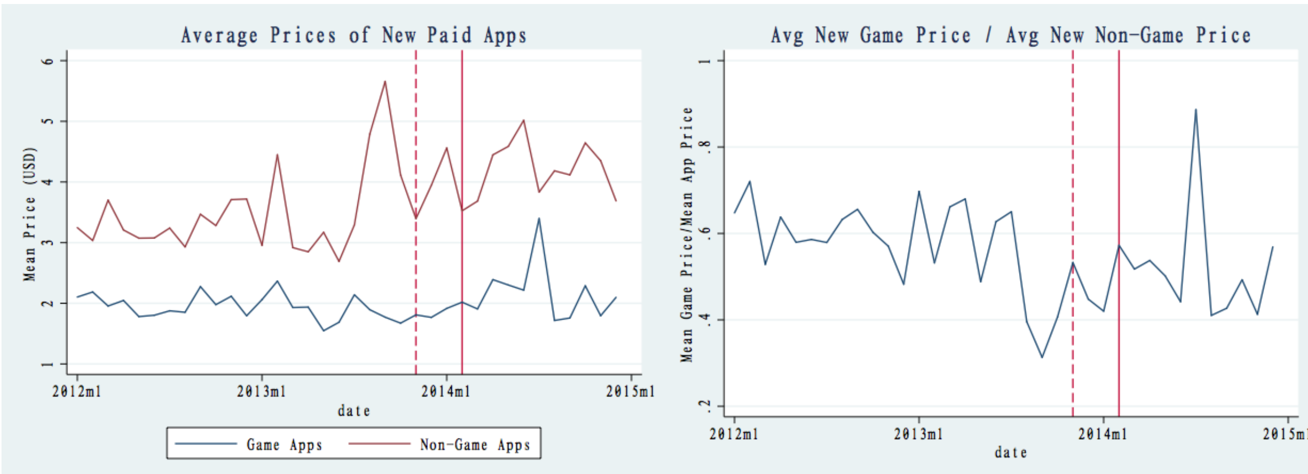


Lower search costs from the game category split could be the cause of the price changes. In particular, the average prices of new games (in the left panel of Figure 11) show a spike right after the split takes place. Similar to the effects on quality, this could be the result of increased demand or of higher valuation consumers being able

[73]The ratio of free to paid apps is constant between games and non-games. It is continuously falling for both throughout the sample.

to discover more preferred game-apps more easily (Bar-Isaac, Caruana and Cunat 2012).

Figure 17



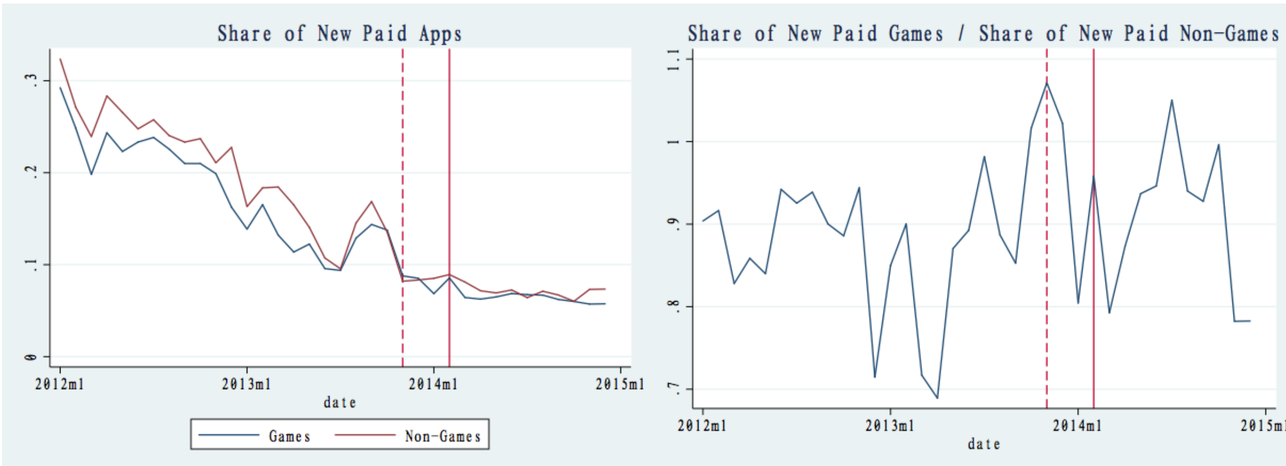
I also test differences in the share of new paid games and non-game apps that enter into the market. Changes in the revenue streams of paid and non-paid apps (e.g., the increasing prevalence of in-app purchases) may result in changes in the number of entrants into the market. These changes could then be what I capture with my entry difference-in-difference estimates.

However, the data does not support that hypothesis. The left panel of the following figure shows that there are substantial changes in the share of paid apps that are entering into the market over time. The share of new paid products falls from over 30% to less than 10%. However, this pattern is consistent for both games and for non-game apps. The right panel of the figure shows that there are no changes in the patterns of free and paid product entry between games and non-games after the split.

9.7 Downloads

The previous suggest that the increase in production (and reduction in quality) of game apps versus non-game apps was driven primarily by the increase in the number of game categories. The specific mechanism that would drive this change in developer incentives is the reduction in search costs for consumers and the improved visibility (and potentially easier marketing) for the products. If consumer search costs are

Figure 18



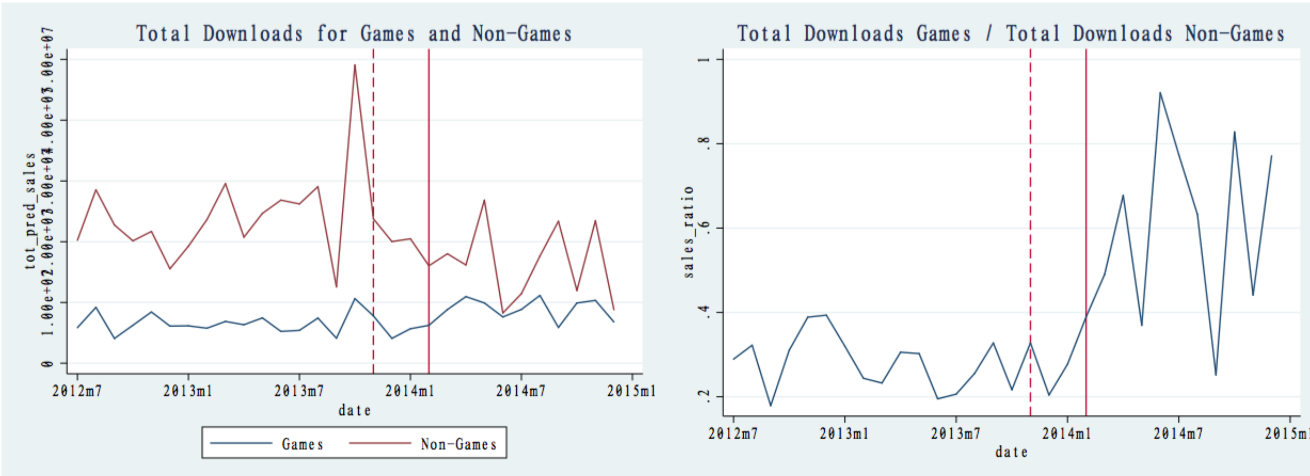
lowered and consumers can more easily find their preferred products, then I would expect downloads to increase for games vs non-games in the post-split period.

This is indeed the case. The figure below shows the total number of downloads for games and non-games. The figure suggests that the aggregate number of downloads for games was relatively stable and trending together with the aggregate number of downloads for non-games. However, at the point of the split, the aggregate number of downloads for games started increasing, whereas the aggregate number of downloads for non-games stayed constant and then trended downwards slightly.[74]

If there was no market expansion, it could be possible to argue that the reduction in search costs and the increase in entry resulted in excessive entry and higher deadweight loss (Mankiw and Whinston 1986, Berry and Waldfogel 1999). However, since there was market expansion, it seems that the market was initially not covered, and it is harder to claim the loss of welfare due to excess entry.

[74] There is a slight pre-trend in the period prior to the split. This could be the result of higher quality games entering prior to the split occurring in order to position themselves to get a good ranking at the time of the split.

Figure 19



10 Appendix D

10.1 Linear Demand Estimates

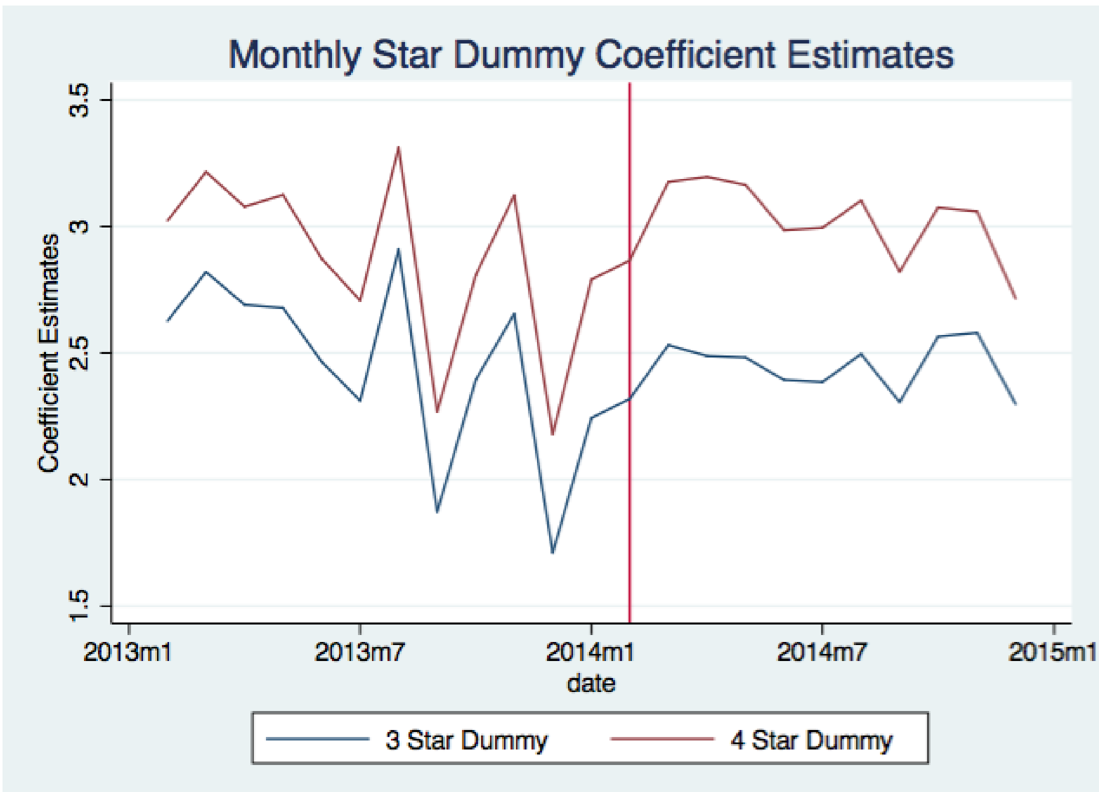
10.2 Additional Monthly Coefficient Estimates

Table 18: **Additional Demand Estimates**

	Games		
	(1)	(2)	(3)
ln(N Apps in Category) - γ	0.684***	0.604***	0.637***
	(0.004)	(0.002)	(0.004)
$\gamma \times$ Post Split	-0.318***	-0.294***	-0.305***
	(0.002)	(0.002)	(0.003)
$\gamma \times$ New App	-0.127***		
	(0.004)		
$\gamma \times$ New App $\times$ Post Split	0.036***		
	(0.001)		
$\gamma \times$ Paid App		-0.037***	
		(0.009)	
$\gamma \times$ Paid App $\times$ Post Split		0.018***	
		(0.001)	
Price	-1.226***	-0.813***	-1.207***
	(0.09)	(0.034)	(0.089)
ln(Size)	0.068***	0.050***	0.056***
	(0.001)	(0.001)	(0.001)
N Screenshots	0.022***	0.018***	0.019***
	(0.0003)	(0.0002)	(0.0003)
Video Preview Dummy	0.261***	0.289***	0.290***
	(0.004)	(0.003)	(0.004)
Paid App	-0.313*	-1.268***	-0.275
	(0.173)	(0.141)	(0.168)
New App	-0.911***		
	(0.02)		
Average Star Rating FE	YES	YES	YES
Date FE	YES	YES	YES
Observations	4,790,522	4,790,522	4,790,522

All columns show the results for free and paid game apps.
Columns (3) uses a linear specification of demand.
Average star dummies removed for ease of reading.
Price IVs include the average chars. of all other apps in the same category.
Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 20


Monthly Star Dummy Coefficient Estimates
Coefficient Estimates
3.5
3
2.5
2
1.5
2013m1
2013m7
2014m1
2014m7
2015m1
date
3 Star Dummy
4 Star Dummy