Essays on the Economics of the Smartphone and Application Industry

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Min Jung Kim

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Thomas J. Holmes

September, 2013

Acknowledgements

I am very grateful to my advisor Thomas Holmes for his constant support, encouragement and helpful feedback. I would also like to thank Kyoo il Kim, Joel Waldfogel, Amil Petrin and Patrick Bajari for their invaluable help and advice. This paper also benefited from comments by the participants in the University of Minnesota Applied Microeconomics Workshop. All remaining errors are my own.

Dedication

To my husband, Minsung Kang, who held me up over the years

Abstract

This dissertation is in two parts. In Chapter 1, I empirically quantify the consumer welfare gains created by free smartphone apps that have recently emerged as one of the most used digital goods. I use a unique dataset of smartphone apps and transform sales ranks into quantities to overcome a lack of data on quantity. In the process, I suggest a new methodology of utilizing Google search data, which can also be used in future research. The estimation results show that smartphone apps created \$5.7∼10.9 billion and \$13.5∼23.3 billion of annual consumer benefits in 2010 and 2011, respectively – which can be translated into \$134∼260 in 2010 and \$157∼271 in 2011 per smartphone user on average – and more than 90% of the welfare gain is from free apps.

In Chapter 2, I examine how the contributing effects of mobile applications on smartphone adoption differ across smartphone operating systems and the extent to which this difference is explained by the role of platforms, focusing on the case of Apple vs. Google. I estimate a model of consumer demand for both smartphones and compatible apps where I specify the benefit provided by apps as the sum of individual app utilities. It is first shown that the selection of different types of consumers onto platforms should be accounted for in estimating app demand and thereby measuring the complementary effects. I take a novel approach to constructing geographically disaggregated sales panel by using Google web search data, as a way of addressing the selection issue. After controlling for the user heterogeneity, the results still suggest that Apple provided more app benefits to users and Android's stronger sales over the sample period come entirely from advantages in the price-adjusted quality of hardware. The overall quality of apps in Google Play was not inferior to that of the App Store, but Google is estimated to have delivered lower utility for a given set of apps possibly due to its open strategy.

Contents

List of Tables

List of Figures

Chapter 1

There is an App for Everything: Measuring the Value of Free Smartphone Apps

1.1 Introduction

We might be familiar with the line from a famous commercial: "There is an app for everything." Until just a few years ago the mobile app industry was almost nonexistent, but it has seen a dramatic growth with the rising penetration of smartphones and tablets since the launch of Apple's App Store in July 2008. Consumers now can choose from almost a million apps and there have been over 100 billion app downloads worldwide.[1] According to Flurry, a mobile analytics firm, the time spent using apps is also constantly growing: from 43 minutes per day in June 2010 to 127 minutes per day in December 2012, while the time spent on the web is staggering at around 70 minutes per day. Thus the consumption of apps pervades our everyday life and they must be providing huge benefits to consumers.

This paper measures the consumer welfare gains that smartphone apps have brought to US consumers. As with many empirical studies on online commerce, there is an issue of a lack of data on quantity and I use publicly available data on sales ranks of smartphone apps, transforming them into quantities. This, however, still presents a data challenge because translating ranks into sales requires observations of actual quantities and the amount of the data needed is even larger in the study. I suggest a new way of utilizing Google search data and show that it generates reasonable outcomes that are in line with prior research and industry statistics. With the quantity data obtained in this way, I then calculate the consumer surplus from apps, especially free apps, which is important because more than 95% of total app downloads are based on free apps. My interpretation is that free apps are like paid apps with zero price because both free and paid apps are delivered in the same format. By estimating the demand curve for free and paid apps together and applying the resulting value of price responsiveness estimate, I can obtain the quantitative measure of the consumer benefits from free apps as well as paid apps.

For robustness, I consider two alternative demand models with varying assumption on the market structure and estimate the models using a unique dataset of individual apps sold on Apple's App Store and Google Play during 2010 and 2011 – when Apple and Google took over the dominant position in both the smartphone and app markets.[2]

[1] It is known that the US accounts for roughly 30% of global app downloads.

The estimation results show that smartphone apps created \$5.7∼10.9 billion and \$13.5∼23.3 billion of annual consumer benefits in 2010 and 2011, respectively, which can be translated into \$134∼260 in 2010 and \$157∼271 in 2011, on average, per smartphone user. More importantly, most of the welfare gain is estimated to be from free apps: 90∼99% of the consumer surplus depending on time and underlying assumptions. As a point of reference, if I compare the consumer surplus estimates for paid apps with the paid-app revenues, they are of about the same magnitude, which seems reasonable. The results are also comparable to what [1] calculated as the consumer surplus generated by the diffusion of broadband access, \$4.8 to \$6.7 billion a year between 1999 and 2006. When considering the time spent using apps vs. browsing the web mentioned above, my estimation results seem to be consistent with the earlier empirical study as well.

This paper first contributes to the growing field of studies that make use of sales rank data. With the growing interest in online commerce, more and more papers use publicly available data on sales rank due to the unavailability of data on actual quantity. [2] suggests a technique for transforming ranks into sales by assuming a power law distribution for sales, and estimates the price index and price elasticities for online books. Since [2], there have been many empirical papers using the same technique to examine various economic aspects of online markets (e.g., see [3], [4] and [5]).[3] As mentioned above, translating ranks into sales still requires actual quantities. Previous studies either obtain actual sales data from sellers (which is a rare case) or conduct an experiment that shifts sales.

I contribute to the literature by proposing a new methodology of using Google search data when collaborating with sellers or conducting a purchase experiment are (practically) impossible. In my case, I have to recover the relationship between ranks and sales over a long period for a rapidly growing industry, which implies the increased amount of actual quantity data needed compared to prior research. Also, an experiment that shifts sales is practically impossible because rank data are available only for top-ranked apps and top free apps sold at least thousands of copies daily during my sample

[2] While the combined market share of Apple's iOS and Google's Android increased to 86.3% in 2011 from 34.0% in 2009, Apple's App Store and Google Play are known to take almost 90% share of the app market.

[3] [3] estimates the economic value of increased product variety made available through online markets; [4] evaluates pricing strategy in the consumer software industry; and [5] examines the effect of consumer reviews on online book sales.

period. I, instead, utilize search data provided by 'Google Trends' that shows weekly search trends for a certain keyword, to assist the recovery process from ranks into quantities. In addition, I do nonlinear estimation of the power law relation to deal with the difference in the frequencies of daily rank data and weekly search data.

The paper also belongs to the empirical literature that examines the welfare effect of the introduction of a new product. See, for example, [6], [7], [8], [9] and [3]. All these papers calculate the compensating variation when the new good in question is made practically unavailable to consumers, after estimating its demand curve. However, there are only a few papers that measure the welfare gain from the Internet or digital goods because most of them have no price. [1] evaluates the consumer surplus created by the diffusion of broadband by constructing a demand curve, but other studies thus take a different approach: assigning a value to the leisure time spent on the product. [10] specifies the consumer utility as a Cobb-Douglas style function of both the quantity purchased and time spent on the product and concludes that consumer gains from the Internet can be more than \$3,000 per year for the median person as of January 2005. [11] also uses a similar model and values the consumer surplus from the Internet at \$564 billion in 2011 or \$2,600 per user.

Erik Brynjolfsson, professor of MIT, stated at Techonomy 2012: "And we are missing more and more of the economy, more and more of what matters in the economy (by ignoring free goods)." Considering all the ways the Internet and digital goods have changed people's lives and the economy, measuring the value generated by them is important and this paper is one of such attempts. To my knowledge, the study is the first to quantify the gains from the introduction of free smartphone apps and this is possible due to extensive data collection for individual apps.

The paper is organized as follows. Section 2.2 describes the data and how I transform ranks into sales. Section 2.3 presents the app demand model and the way of measuring consumer welfare. The estimation procedure and results are reported in section 1.4. Section 2.6 concludes.

1.2 Data

1.2.1 App Data Collection

First, I collect rank and characteristics data of best-selling apps in the US market because it is not feasible to track all apps due to the large number of titles available and sales ranks are publicly available for a limited number of top apps. For iPhone users, apps can be downloaded directly to a target device or downloaded onto a personal computer via software called iTunes that is provided by Apple and works essentially like a web browser. From the underlying URLs that iTunes uses for browsing the App Store, I was able to obtain detailed characteristic and rank information on top 100 free and paid iPhone apps. In the case of Google Play (formerly known as Android Market), neither number of downloads nor rank is publicly available for individual apps until early 2011. I thus collect data on Android apps from the website AndroLib.com, one of the most trusted and visited websites by users and developers before the introduction of web-based Google Play, until June 2011. AndroLib shows lists of 150 free and paid apps that are most viewed on its site each day, and I gathered the ranks and characteristics of the apps on these lists.[4] For the remaining period, the data come from the website of Google Play.

The resulting dataset includes daily observations on title, seller, price, release date, recent update date, customer rating and additional characteristics for top 100 free/paid iPhone apps and Android apps for 14 months in 2010 and 2011.[5] In addition, histories of ranks up to 300 are also retrieved from www.appannie.com – for 2010 and 2011 in the case of iPhone apps and for the fourth quarter of 2011 in the case of Android apps. With the burgeoning interest in apps, there are now numerous research companies or web sites that provide data on app stores but they are of lower frequency or cover only ranks. Given that developers can change the details, especially prices, of their apps costlessly at any time, analyzing high-frequency panel data may allow more useful insights. Moreover, ranks are (mostly) determined by the number of downloads or page

[4] This is based on the following two assumptions: (i) Android users browsing apps at AndroLib.com are representative and (ii) the actual number of downloads for each app is proportional to the count of app page views.

[5] Rank, title and seller information is collected for the whole two years but price and additional characteristics are available only for the 14 months listed in table 1.1.

views for the past day, so I can get the most of information by collecting daily data.

Table 1.1 reports summary statistics for top 100 free and paid apps in the App Store and Google Play. On each app platform there are about 4,200 apps ever listed in the top 100 chart for each day of the sample period and paid apps tend to remain longer in the chart than free apps. The average prices of the top iPhone and Android paid apps are \$1.8 and \$4.0, respectively. Thus, popular Android apps tend to have higher and more dispersed prices than popular iPhone apps. Android apps are also both younger and has been updated more recently on average than iPhone apps. Paid apps get higher average customer rating than free apps in the App Store but this is reversed in Google Play. Lastly, table 1.1 also shows the difference in the category distribution of the top 100 apps. In the App Store about 70% of the apps are categorized as Games or Entertainment, but these categories take only about 45% in Google Play. Instead, Android apps in the Utilities and Productivity categories show stronger performance than iPhone apps in the same categoreis.

I will translate the observed ranking of each top app into quantity downloaded by assuming a power law distribution for app sales and the next subsection discusses its procedure and estimation results. As will be explained, some actual quantity data is still required to estimate the power law relationship. I was able to obtain time series data of the actual download quantities for a few apps from individual app developers confidentially and table 1.2 displays their summary statistics.[6] They are 2 iPhone free apps, 1 iPhone paid app, 1 Android free app and 1 Android paid app. Note that each app covers a different subperiod of my rank data and does not cover all 14 months.

Lastly, to define the size of the market for each app platform, I use data on smartphone quantities sold. They are obtained from Gartner, an information technology research firm, and consist of monthly unit sales by operating system in the United States from 2008 to 2011.[7] See section 2.2.1 for the summary statistics of smartphone quantities.

---

[6] I greatly thank the app developers for providing me with their app sales data.

[7] Concretely speaking, the original Gartner data is quarterly and I combine Google search data to construct monthly quantities as explained in section 2.2.2.

1.2.2 Google Search Data and Conversion of Ranks into Sales

Following [2], I transform the rank of each top app into quantity by assuming a power law (Pareto) distribution for app sales. Under the assumption, the relationship between ranks and app download quantities can be approximated by $\ln(\text{Quantity}) = \beta_1 - \beta_2 \ln(\text{Rank})$ and at least a few observations of actual download quantities are needed to estimate the relationship. It would be ideal to obtain actual sales data from sellers (see e.g., [3]) but sellers are extremely reluctant to release specific sales data. Many previous studies thus take an alternative approach of conducting a purchase experiment. For example, [2] purchased relatively large amounts of one low-ranked book that they knew sales information of, and compared the sales and ranks before and after the experiment to infer the relationship.

In my case, the Pareto relationship cannot be assumed to remain the same over time, given that the app industry is a rapidly growing one and I recover app downloads data over the two years. I thus allow β_1 to vary monthly while fixing β_2 over time:

$$\ln(\text{Quantity}_t) = \beta_{1\tilde{t}} - \beta_2 \ln(\text{Rank}_t), \tag{1.1}$$

where t and $\tilde{t}$ denote day and corresponding month, respectively. Note the increased need for the actual observations of quantities – 15 parameters to establish the relationship for 14 months. This is in contrast with previous studies that examine only a short period (e.g., three weeks taken over six months for a relatively stable online book market in [3] and two weeks in [4]). Even though I obtained actual download quantities for the five apps shown in table 1.2, they are not enough to recover β_2 and monthly varying β_1 in equation (1.1). Also, an experiment that shifts sales is practically impossible because rank data are available for only top-ranked apps and at least 1,300 copies were being sold daily for a free app ranked at 100 during my sample period.

To satisfy the increased need for data, I take a novel approach of combining web search data for app titles that are obtained from 'Google Trends' (www.google.com/trends). Google Trends is a service that analyzes a portion of Google web searches and provides weekly trends of the search volume of a given search term. I collect the Google search volumes for apps (by using an app title as a keyword) that I have rank data for and satisfy the following conditions: (i) app title can be viewed as an appropriate search keyword, (ii) app title has not been changed over time and is unique, (iii) app is exclusive (i.e., sold only on one platform) and (iv) app price is lower than two dollars.[8] Summary statistics for apps that have Google search data are shown in table 1.3. The portion of the apps with Google search data is about 1∼4% of all apps that I have rank data for and the daily sales rank for these apps shows a lot of variation.

By comparing the Google search volumes and actual quantities for the few apps in table 1.2, I find that they have a tight linear relationship of the form:

$$\text{Quantity}_w = a + b\,\text{Search}_w, \tag{1.2}$$

where w denotes week.[9] I estimate equation (1.2) for each type (i.e., free and paid) of each app platform (i.e., iPhone and Android) and the values of the R-squared of this regression are all around 0.7 as reported in table 1.4.[10] By extending this estimated relationship to other apps that have Google search data, I can compute the quantity proxies for all of them and now I have enough quantity data to estimate equation (1.1) with time-varying β_1.

One more difficulty arises from the difference in data frequencies: Google search volumes and thereby quantity proxies (derived as in the previous paragraph) are weekly while ranks are daily. Equation (1.1) can be rewritten as:

$$\text{Quantity}_{d_w} = \exp \beta_{1\tilde{t}}\,(\text{Rank}_{d_w})^{-\beta_2}, \tag{1.3}$$

where d_w and $\tilde{t}$ denote day of the week (from Sunday to Saturday) in week w and corresponding month, respectively, and then weekly quantity can be expressed as:

$$\text{Quantity}_w = \sum_{d_w} \text{Quantity}_{d_w} = \sum_{d_w} \exp \beta_{1\tilde{t}}\,(\text{Rank}_{d_w})^{-\beta_2} = \beta'_{1\tilde{t}} \sum_{d_w} (\text{Rank}_{d_w})^{-\beta_2}, \tag{1.4}$$

where $\beta'_{1\tilde{t}}$ and β_2 are the parameters to be estimated. Thus I do nonlinear estimation of equation (1.4), separately for Android free/paid apps and iPhone free/paid apps. With

[8] (i) refers to the case in which top search results are related to the app when its app title is used as a search keyword in Google website. (iv) is to deal with the possibility that the ratio between search volume and actual download vary with app prices.

[9] Note that actual download quantities for the five apps were not enough to get estimates for β_2 and monthly varying β_1 in equation (1.1) but they are enough to analyze the relationship between Google search volumes and quantities in equation (1.2).

[10] The values of the parameter estimates and any comparison of them is meaningless because Google search volumes are normalized differently for each case.

the Pareto parameter estimates in hand, the implied download quantities for all top apps (i.e., apps for which I have rank information) can be obtained.[11]

The estimates for the shape parameter, β_2, are reported in table 1.5. All estimates are around 1: 0.931 for free and 0.994 for paid in the case of iPhone apps, and 1.015 for free and 1.037 for paid in the case of Android apps. The value of R-squared is pretty high in all regressions, suggesting a tight log-linear relationship between Quantity and Rank. Here, a lower value of β_2 means a flatter curve or longer tail. Therefore, iPhone apps have a flatter curve than Android apps while free apps tend to have a flatter curve than paid apps. The values of the β_2 estimates are somewhat higher than those in previous papers – 0.78∼0.952 for online books (see [2], [3] and [5]) and 0.828 for consumer software (see [4]) – meaning that the sales of smartphone apps are more concentrated on top ones. Also, my estimates are close to the slope parameter estimates obtained by [12], 0.944 for iPhone paid apps and 0.985 for Android paid apps.

Furthermore, the estimates lead to plausible quantity outcomes that are in line with industry statistics. Apple and Google rarely release the cumulative number of global app downloads, so I calculate the app download quantities in the US over a certain subperiod of my sample as the increase in the cumulative global app downloads multiplied by the US download share estimate of App Annie (www.appannie.com), and then compare them with the download quantities computed using the above estimates. For example, about 1.96 billion copies of iPhone apps (= 7 billion worldwide * 28% US download share) were downloaded in the US during October 2010∼June 2011 according to Apple and App Annie, and my estimate for this is 1.91 billion. Similarly, 525 million Android apps (= 1.5 billion worldwide * 35% US download share) were downloaded in the US during March∼May in 2011 according to Google and App Annie, and my estimate gives 564 million for this. In addition, the portion of the app downloads taken by top 100 paid apps (among all paid apps) is estimated to be 42.2% for iPhone and 52.3% for Android, which compare with Flurry's 55% estimate for overall smartphone apps. Thus the computed quantities are generally consistent with industry statistics.

[11] Only top 100 apps are used for demand estimation in section 1.4 due to the availability of app characteristics, but I use all available data – including apps ranked at up to 300 – for equation (1.4).

1.3 Model

1.3.1 App Demand

Consider a consumer i who has recently purchased smartphone j.[12] In each day t, she can purchase an app among K_{jt}, the set of app titles available on the associated app platform (also denoted as j). The conditional indirect utility from app k is:

$$v_{ijkt} = \beta_{j\tilde{t}} + x_{jkt}\beta - \alpha p_{jkt} + \eta_{jkt} + \epsilon_{ijkt} = \delta_{jkt} + \epsilon_{ijkt}, \tag{1.5}$$

where $\beta_{j\tilde{t}}$ is a platform-month specific fixed effect, x_{jkt} represents observable app characteristics and includes app-specific fixed effects, age, time since last update, customer rating, etc.[13] p_{jkt} and η_{jkt} are app price and unobserved (by the econometrician) quality, and ϵ_{ijkt} is a logit error. Let δ_{jkt} denote the market mean utility from app k at time t as usual.

I now consider two models with varying assumptions on the market structure.

Model 1. *Each app constitutes its own market and apps are not substitutes at all. Thus consumers make the purchase decision independently for each app.*

Model 2. *Apps are substitutes and consumers can choose only one app among K_{jt}. Substitutability may be stronger among apps within same category (i.e., nested logit model).*[14]

These are all approximations to the true model.[15] Model 1 allows consumers to purchase multiple apps at a time but puts restriction on the substitutability of apps. Given that there are a large number of apps, each with its own distinct features, this is not likely to be a strong assumption.[16] For a comparison, I remove this restriction in

[12] The exact meaning of "recently" will be defined in section 2.4.

[13] $\beta_{j\tilde{t}}$ captures the user group difference in overall likeness for apps and the difference in app platform efficiency among platforms and over time, in addition to any seasonal effects (seasonal effects cannot be separately identified because the dataset covers only 14 months). It cannot be identified for all j and t from shifts in each app's fixed effect, so $\beta_{android,1}$ is normalized to zero.

[14] *For a nested logit model,* $v_{ijkt} = \beta_{j\tilde{t}} + x_{jkt}\beta - \alpha p_{jkt} + \eta_{jkt} + \sigma\zeta_{ijgt} + (1-\sigma)\epsilon_{ijkt}$ *and* $\delta_{jkt} = \beta_{j\tilde{t}} + x_{jkt}\beta - \alpha p_{jkt} + \eta_{jkt}$*, where g denotes category.*

[15] An ideal model would be a multiple-discrete choice model but there are too many products in the choice set to estimate such a model.

[16] [13] makes the same assumption for videogames while [14] finds that videogames are not strong substitutes for one another. Also, [15] shows that the demand of digital music is inelastic.

model 2. This model, however, makes the assumption of purchasing only one app per month, which is clearly rejected by the data as shown in section 1.2.1. I will examine in section 1.4 how the estimates for price elasticity and consumer gain vary depending on the assumptions made. Lastly, for model 1, the outside good is defined for each app k as not purchasing the app and denoted as k_0. For model 2, the outside good means no app purchase.

1.3.2 Consumer Welfare

Following the literature, the consumer welfare gain from the introduction of apps can be obtained as the compensating variation (CV). The CV for consumer i is the change in income needed to compensate the consumer to willingly give up apps and it solves:

$$V_i(p^0, y_i) = V_i(p^1, y_i + CV_i), \tag{1.6}$$

where V_i is the (unconditional) indirect utility, p^0 is the vector of prices with apps available and p^1 is the price vector with app prices set high enough such that app demand is zero (e.g., see [7]). The measure of total consumer welfare is then given by:

$$CV = M \int CV_i \, dP(i), \tag{1.7}$$

where M is market size and $P(\cdot)$ is distribution function.[17]

As shown by [16], this term has a closed form expression for (nested) logit models. For model 1:

$$CV_{jt} = M_{jt} \int \frac{\sum_{k \in K_{jt}} \max\{v_{ijkt}, v_{ijk_0t}\}}{\alpha} \, dP_\epsilon(\epsilon) = M_{jt} \frac{\sum_{k \in K_{jt}} \ln[\exp(\delta_{jkt}) + 1]}{\alpha}, \tag{1.8}$$

and for model 2:

$$CV_{jt} = M_{jt} \int \frac{\max_{k \in K_{jt} \cup 0}\{v_{ijkt}\}}{\alpha} \, dP_\epsilon(\epsilon) = M_{jt} \frac{\ln[\sum_{k \in K_{jt}} \exp(\delta_{jkt}) + 1]}{\alpha}, \tag{1.9}$$

where M_{jt} will be defined according to the underlying assumption (see section 2.4).[18]

[17] Subscripts j and t have been suppressed so far for ease of notation.

[18] This paper ignores the effect of the introduction of apps on other industries due to data limitation. The most affected ones are likely to be the Internet and web-related products.

1.4 Estimation and Results

1.4.1 Estimation

I first discuss how I obtain the size of the market, M_{jt}, in equations (1.8) and (1.9).[19]

Consumers who hold smartphone j at time t are potential app buyers, but it is likely that they purchase apps intensively right after smartphone purchase and their interest in apps decreases over time. I thus assume that the probability of purchasing any app decreases exponentially by θ each month after smartphone purchase. For model 1, the market size for platform j's apps is:

$$M_{jt} = Q_{jt} + \sum_{s=1}^{23} Q_{j,t-s}\,\theta^{s}, \tag{1.10}$$

where Q_{jt} denotes the quantity sold for smartphone j in month t and this term is added for up to 23 months ago considering the average two-year purchase cycle of mobile phones. θ will be varied among 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ and 1. For model 2, I need to make an additional assumption on the number of apps a consumer could purchase each time and multiply it to the numbers obtained in equation (1.10). I set this number differently for each θ so that the market share of the outside good (i.e., no app purchase) stays positive and not too low.[20]

Next the issue of endogeneity between price, p_{jkt}, and unobserved quality, η_{jkt}, should be properly addressed to get consistent estimates for the demand models. Following [17], the sum of each characteristic across other apps is used as instruments for model 2. While these BLP-instruments are based on oligopoly pricing, model 1 assumes a monopolistic market. I thus employ a different set of pricing instruments for model 1 as in [13]. The first instrument for price p_{jkt} is constructed as follows: for all apps released before app k, I take their prices when they were of same age as app k at time t, and the average is used. This is intended to capture any trends in cost over app life cycle, which is not fully captured by age. In addition, the average current price of all apps on the other platform is also used to reflect any app-wide cost shocks. Lastly, for

[19] For ease of notation, let t denote month in this section. The size of the daily market is simply that of the monthly market divided by the number of days in a month.

[20] Specifically, the number of apps a consumer could purchase each time is set to be 90 for $\theta = 0$, 70 for $\theta = \frac{1}{4}$, 50 for $\theta = \frac{1}{2}$, 30 for $\theta = \frac{3}{4}$ and 10 for $\theta = 1$, respectively.

a nested logit model, I also use the sum of each characteristic across other apps within the same category to instrument for within-nest share.

1.4.2 Results

Parameter estimates for the app demand models are presented in tables 1.6∼1.8. I run an instrumental variables regression to account for the endogeneity of price and group share after inverting market shares to market mean utility levels (see [18] and [17]).

Starting with model 1 (see table 1.6), the coefficients for observable characteristics including price are mostly significant and have expected sign. While the absolute values of the price coefficient estimate are somewhat small, Android users tend to be more sensitive to app price than iPhone users. App age and time since last update have negative effects on the app utility even though they are small in magnitude. The coefficient for the dummy variable indicating that an app was introduced or updated within past three days is significantly positive for both platforms. This is because apps that are recently introduced or updated have a chance to be featured in front page sections for new apps.[21] The effect of customer rating is somewhat ambiguous depending on the value of θ but it is estimated to affect the utility positively in Google Play. Note that model 1 is sensitive to the assumption on θ. As θ increases (i.e., as consumer interest in apps decreases more slowly over time), the price elasticity estimate becomes smaller (especially for iPhone apps) - the mean price elasticity decreases from 1.917 to 1.264 - and the adjusted R-squared gets higher (i.e., more is explained by the observable app attributes).

On the other hand, all coefficients for model 2, except the constant term, are estimated to be the same regardless of the value of θ (see tables 1.7 and 1.8). Customer rating is not significant anymore but similar observations can be made for other variables. Table 1.7 is for model 2 without grouping apps (i.e., simple logit model). The mean of estimated price elasticities is 1.840, which falls within the range of its estimates for model 1. Next, table 1.8 is for model 2 with grouping apps by similar categories. That is, it assumes a nested logit model where Games and Entertainment categories are in one nest, Utilities and Productivity categories are in another, the other categories

[21] They are the New and Noteworthy section of the App Store and the Top New Paid/Free chart of Google Play.

are in another and the outside option (no app purchase) is in a nest by itself. Results do not change much with the move to a nested logit model – note that the estimate for σ is not very large (0.266) – except that the relative responsiveness of iPhone users to app price is estimated to be much smaller. The mean price elasticity is now a little larger at 1.903.

Given these demand estimates, I now use equations (1.8) and (1.9) to calculate the consumer surplus gain from apps. Because I do not have data for all months in 2010 and 2011, the consumer surplus was first calculated for the 14 months used in estimation and then interpolated/extrapolated for the other months to obtain annual-level estimates. Tables 1.9∼1.12 report the results.

Table 1.9 shows that model 1 creates a range of welfare estimates due to the sensitivity to θ. The annual welfare gain from Android apps jumped to \$7.0∼10.8 billion in 2011 from \$1.6∼1.9 billion in 2010. For iPhone apps, the annual gain reached \$5.5∼7.6 billion in 2010, which is comparable to the level of Android apps a year later, and \$10.4∼13.1 billion in 2011. As a point of comparison, I also report the revenue from paid apps in table 1.10, which is estimated using the computed quantity and price in my dataset.[22] The consumer surplus estimates for paid apps are of about the same magnitude as the values for paid-app revenue, which seems reasonable. Note that most of the consumer welfare gain is from free apps as can be seen in table 1.9: 98∼99% for Android and 90∼98% for the iPhone depending on time and the value of θ. There is a tendency that the share of free apps gets smaller as θ increases (i.e., as consumer interest in apps persists longer over time after smartphone purchase).

Tables 1.11 and 1.12 display the consumer surplus estimates for model 2. For model 2, it is not straightforward to calculate the consumer surplus from free apps and paid apps separately because both are in one choice set (see equation (1.9)). By comparing the total consumer surplus estimates of model 1 with those of model2, however, I can see if the results are robust. Table 1.11 is for the simple logit model. The estimates are within a tighter range (as the value of θ varies) and generally lower than those for model 1: \$1.4 billion in 2010 and \$5.5∼5.9 billion in 2011 from Android apps, and \$4.2∼5.2 billion in 2010 and \$8.0∼8.8 billion in 2011 from iPhone apps. However, the nested logit

[22] Because I have price information for only top 100 apps, I first calculate the total revenue from top 100 paid apps and divide it by the download share of those apps among all paid apps.

model leads to higher values of consumer surplus than model 1 as reported in table 1.12 even though it assumes a similar market structure as the logit model: \$1.8 billion in 2010 and \$6.9∼7.4 billion in 2011 from Android apps, and \$7.4∼9.1 billion in 2010 and \$14.1∼15.5 billion in 2011 from iPhone apps.

Putting the results together, Android and iPhone apps created at least \$5.7 billion and \$13.5 billion of consumer benefits in 2010 and 2011, respectively.[23] These numbers can be translated into \$134 in 2010 and \$157 in 2011, on average, per smartphone user. When looking at the maximum values, the consumer gains from apps amount up to \$10.9 billion in 2010 and \$23.3 billion in 2011 – or \$260 in 2010 and \$271 in 2011 per smartphone user. The estimation results are comparable to what [1] calculated as the consumer surplus generated by the diffusion of broadband access, \$4.8 to \$6.7 billion a year between 1999 and 2006. As mentioned in the introduction, consumers started to spend more time using apps than browsing the web by 2011: smartphone and tablet users spent an average of 94 minutes per day on apps as of December 2011 while they spent 72 minutes per day on the Internet according to Flurry. Thus, my estimation results seem to be reasonable and consistent with industry statistics and earlier empirical studies.

1.4.3 Discussion

This section discusses the way how I estimated the consumer welfare gain from free apps. My interpretation was that free apps are like paid apps with a zero price because both free and paid apps are delivered in the same format. I obtained the quantitative measure of the consumer benefits from free apps by estimating the demand curve for free and paid apps together. This, however, can be problematic because consumer may value free apps in a different way than they value paid apps. As a robustness check, I estimate the model 1 and model 2 again by using only paid apps and see how the results change. What matters in the consumer surplus calculation is the price coefficent, α (see equations (1.8) and (1.9)), so I compare the price coefficients obtained by estimating the models with only paid apps with the original results. Table 1.13 displays the results. The price coefficients for the cases using all apps are the same as those in tables 1.6∼1.8.

[23] Android users can purchase apps on other platforms like Amazon Appstore but the revenue/download share of them was ignorable during 2010 and 2011.

When estimating model 1 with only paid apps, the price coefficients decrease by more than tenfold, which implies that the consumer surplus estimates will increase by more than tenfold.[24] However, model 2 creates price coefficients that are of the same magnitude as the results obtained by using all apps, even though those for Android apps decrease to half. The consumer welfare estimates by model 2 should be roughly the same for iPhone apps and double for Android apps, respectively. Even though it may be true that free apps behave different from paid apps, quantifying the benefits from free apps in a different way would require information on (non-monetary) costs involved with downloading free apps. In the absence of such data, my results provide at least a good reference point.

1.5 Conclusion

This paper measures the economic value created by smartphone apps, especially free apps, using a unique dataset of Android and iPhone apps. While transforming sales ranks into quantities, I suggest a new methodology of utilizing Google search data to overcome a lack of data on actual quantities. This approach can be also very useful in future research that faces the same data issue. The estimation results show that smartphone apps created \$5.7∼10.9 billion and \$13.5∼ 23.3 billion of annual consumer benefits in 2010 and 2011, respectively – which can be translated into \$134∼260 in 2010 and \$157∼\$271 in 2011 per smartphone user on average – and more than 90% of the welfare gain is from free apps.

One limitation of the study is that the possibility of different utilization of each app (e.g., depending on price and category) is ignored because of the data issue. Also, only apps that can be downloaded from app platforms (mainly third-party apps) are considered. Apps that are already installed on the device upon purchase (factory-installed apps) may also create non-negligible economic value but they are not considered in the paper due to the lack of data. If data on the time spent or frequency using apps are available, consumer surplus may be derived from the utility model that is a function

[24] Of course, compensating variation estimated from a logit model might be overestimated due to a heavy dependence on the idiosyncratic logit taste term as [7] pointed out. Adding micro moments as in [7] can lead to more precise estimates but this would require additional information such as average characteristics of app buyers.

of both direct expenditures and time as in [10] and [11]. This can provide a good comparison to the welfare estimates of this paper or from other conventional demand models, and account for the both limitations better, which is left as a future study.

Table 1.1: Summary Statistics for App Characteristics

	iPhone (free)		iPhone (paid)		Android (free)		Android (paid)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
# Unique Titles on Top 100 Chart	2,908		1,311		2,639		1,652	
# Days Listed on Top 100 Chart	13.9	34.9	30.7	63.8	15.5	49.3	24.7	60.9
Price ($)	0.0	0.0	1.8	1.8	0.0	0.0	4.0	5.0
Age (days)	285.6	303.7	333.4	288.7	203.0	169.4	217.2	171.5
Since Update (days)	45.4	78.5	74.3	126.9	39.6	114.6	41.1	107.0
Customer Rating (1∼5)	3.5	0.7	4.0	0.5	4.2	0.7	4.0	0.9
Category (%)								
Games	51.5		58.5		33.3		43.2	
Entertainment	12.8		12.2		12.0		4.8	
Utilities	5.4		6.5		8.2		17.8	
Productivity	1.2		1.7		5.8		6.9	
Social	8.3		2.2		14.0		4.8	
Media	4.4		4.5		10.0		8.9	
Business	0.1		0.2		0.1		4.4	
Others	16.3		14.2		16.7		9.3	

Note: Summary statistics are for top 100 free and paid apps for 14 months during 2010-2011 (Apr∼Jun and Nov∼Dec of 2010, and Jan∼Jun and Oct∼Dec of 2011). The data of Android apps are collected from the website AndroLib.com except for Oct∼Dec of 2011.

Table 1.2: Summary Statistics for App Sample with Actual Quantities

	Platform (type)	Obs	Period	Quantity		Obs with Rank	Rank	
				Mean	Std		Mean	Std
App 1	iPhone (free)	83	2011.06.20∼2011.09.10	5332.7	1183.1	14	171.6	50.6
App 2	iPhone (free)	130	2011.07.20∼2011.11.26	53261.6	38288.5	14	27.9	24.0
App 3	iPhone (paid)	165	2011.06.01∼2011.11.12	319.9	61.9	137	198.5	44.5
App 4	Android (free)	68	2011.10.25∼2011.12.31	5725.8	590.0	63	160.5	14.1
App 5	Android (paid)	236	2010.05.10∼2010.12.25	61.8	30.8	77	57.1	18.2

Note: Number of observation is in number of days. The mean and standard deviation of ranks are taken only over the periods during which I have rank data.

Table 1.3: Summary Statistics for App Sample with Google Search Data

	iPhone (free)		iPhone (paid)		Android (free)		Android (paid)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Number of apps	100 (3.4%)		48 (3.7%)		57 (2.2%)		14 (0.9%)	
Rank	132.5	78.4	184.8	81.1	91.4	61	85.7	46.3

Source: Summary statistics are for 2010 and 2011. Numbers in parenthesis represent the portion of apps that have Google search data out of all apps that have rank data. The mean and standard deviation of ranks are taken only over the periods during which I have rank data.

Table 1.4: Estimated Linear Relationship between Quantities and Google Search Volumes

	iPhone		Android	
	free	paid	free	paid
a	1178.2***	1456.0***	266.5**	95.9***
	(412.4)	(170.4)	(83.4)	(25.8)
b	27.9***	24.3***	4.6***	2.4***
	(9.2)	(4.9)	(1.1)	(0.5)
Number of obs	30	23	9	33
Adj R-squared	0.732	0.745	0.663	0.714

Note: Parameter estimates for free apps are divided by a factor of 100. Numbers in parenthesis represent standard errors for the estimates. *** indicates significance at the 1% level.

Table 1.5: Estimates for Shape Parameter β_2

	iPhone		Android	
	free	paid	free	paid
β_2	0.931***	0.994***	1.015***	1.037***
	(0.214)	(0.206)	(0.080)	(0.385)
Number of obs	743	1,229	944	768
Adj R-squared	0.931	0.919	0.965	0.876

Note: Numbers in parenthesis represent standard errors for the estimates. *** indicates significance at the 1% level.

Table 1.6: App Demand Estimation Results - Model 1

	θ				
	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Price	-0.464***	-0.389***	-0.303***	-0.220***	-0.194***
	(0.044)	(0.040)	(0.036)	(0.033)	(0.032)
Price * $d_{Android}$	-0.150***	-0.179***	-0.201***	-0.238***	-0.247***
	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)
Age	-0.081***	-0.068***	-0.046***	-0.017***	-0.007***
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
Age sq	0.002***	0.002***	0.001***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Updated	-0.012**	-0.010**	-0.006	-0.003	-0.004**
	(0.006)	(0.005)	(0.004)	(0.002)	(0.002)
Updated sq	0.001***	0.000***	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$d_{recent\ update}$ (Android)	0.058***	0.053***	0.068***	0.082***	0.083***
	(0.011)	(0.009)	(0.007)	(0.006)	(0.006)
$d_{recent\ update}$ (iPhone)	0.357***	0.315***	0.200***	0.124***	0.082***
	(0.041)	(0.035)	(0.026)	(0.014)	(0.008)
Customer Rating	-0.466***	-0.288***	-0.189***	-0.078***	0.017
	(0.055)	(0.045)	(0.036)	(0.023)	(0.014)
Customer Rating * $d_{Android}$	0.503***	0.334***	0.234***	0.148***	0.049*
	(0.062)	(0.051)	(0.042)	(0.032)	(0.025)
Constant	-3.196***	-3.569***	-4.066***	-4.694***	-5.502***
	(0.164)	(0.149)	(0.133)	(0.126)	(0.118)
Mean Price Elasticity	1.917	1.742	1.513	1.328	1.264
Number of Obs	160,825	160,825	160,825	160,825	160,825
Adj R-squared	0.696	0.727	0.800	0.914	0.929

Note: All regressions include platform-month specific fixed effects ($\beta_{j\tilde{t}}$ in equation (2.8)) and app fixed effects. Numbers in parentheses represent standard errors. ***, ** and * indicate significance at the 1%, 5% and 10% levels. The mean of price elasticities is taken only over paid apps.

Table 1.7: App Demand Estimation Results - Model 2 without Nests

	θ				
	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Price	-0.300***	-0.300***	-0.300***	-0.300***	-0.300***
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
Price * $d_{Android}$	-0.335***	-0.335***	-0.335***	-0.335***	-0.335***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Age	-0.005**	-0.005**	-0.005**	-0.005**	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age sq	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Updated	-0.010***	-0.010***	-0.010***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Updated sq	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$d_{recent\ update}$ (Android)	0.077***	0.077***	0.077***	0.077***	0.077***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
$d_{recent\ update}$ (iPhone)	0.080***	0.080***	0.080***	0.080***	0.080***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Customer Rating	0.001	0.001	0.001	0.001	0.001
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Customer Rating * $d_{Android}$	0.036	0.036	0.036	0.036	0.036
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Constant	-7.270***	-7.282***	-7.319***	-7.358***	-7.013***
	(0.159)	(0.159)	(0.159)	(0.159)	(0.159)
Mean Price Elasticity	1.840	1.840	1.840	1.840	1.840
Number of Obs	160,825	160,825	160,825	160,825	160,825
Adj R-squared	0.943	0.943	0.942	0.938	0.930

Note: The results are for simple logit models. All regressions include platform-month specific fixed effects ($\beta_{j\tilde{t}}$ in equation (2.8)) and app fixed effects. Numbers in parentheses represent standard errors. ***, ** and * indicate significance at the 1%, 5% and 10% levels. The mean of price elasticities is taken only over paid apps.

Table 1.8: App Demand Estimation Results - Model 2 with Nests

	θ				
	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Price	-0.170***	-0.170***	-0.170***	-0.170***	-0.170***
	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)
Price * $d_{Android}$	-0.333***	-0.333***	-0.333***	-0.333***	-0.333***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Age	-0.004**	-0.004**	-0.004**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age sq	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Updated	-0.007***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Updated sq	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$d_{recent\ update}$ (Android)	0.056***	0.056***	0.056***	0.056***	0.056***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$d_{recent\ update}$ (iPhone)	0.056***	0.056***	0.056***	0.056***	0.056***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Customer Rating	0.000	0.000	0.000	0.000	0.000
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Customer Rating * $d_{Android}$	0.020	0.020	0.020	0.020	0.020
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Nest Share (σ)	0.266***	0.266***	0.266***	0.266***	0.266***
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Constant	-5.865***	-5.876***	-5.913***	-5.952***	-5.608***
	(0.230)	(0.230)	(0.230)	(0.230)	(0.230)
Mean Price Elasticity	1.903	1.903	1.903	1.903	1.903
Number of Obs	160,825	160,825	160,825	160,825	160,825
Adj R-squared	0.943	0.943	0.942	0.938	0.930

Note: The results are for nested logit models. All regressions include platform-month specific fixed effects ($\beta_{j\tilde{t}}$ in equation (2.8)) and app fixed effects. Numbers in parentheses represent standard errors. ***, ** and * indicate significance at the 1%, 5% and 10% levels. The mean of price elasticities is taken only over paid apps.

Table 1.9: Consumer Surplus from Apps - Model 1

			θ				
			0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Android	2010	free	1,661	1,615	1,729	1,831	1,850
			(98.4)	(98.2)	(98.2)	(98.1)	(98.1)
		paid	27	29	32	35	37
		total	1,688	1,644	1,761	1,867	1,886
	2011	free	10,696	10,478	8,175	6,999	6,874
			(99.1)	(99.0)	(98.5)	(98.1)	(98.1)
		paid	102	109	122	132	136
		total	10,798	10,587	8,297	7,131	7,011
iPhone	2010	free	6,994	7,208	7,233	6,767	4,984
			(96.6)	(96.1)	(95.2)	(93.1)	(89.9)
		paid	247	291	368	498	559
		total	7,241	7,499	7,601	7,265	5,544
	2011	free	11,555	12,396	12,783	10,046	9,876
			(98.1)	(97.9)	(97.4)	(95.5)	(94.9)
		paid	226	269	343	469	531
		total	11,781	12,664	13,127	10,515	10,406

Note: Numbers in parentheses represent the share of free apps (%) and other numbers are in million dollars.

Table 1.10: Sales Revenue from Paid Apps

	2010	2011
Android	39.6	146.9
iPhone	218.6	357.8

Note: All numbers are in million dollars.

Table 1.11: Consumer Surplus from Apps - Model 2 without Nests

		θ				
		0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Android	2010	1,443	1,441	1,438	1,434	1,472
	2011	5,878	5,877	5,838	5,696	5,482
iPhone	2010	4,935	5,167	5,156	4,821	4,181
	2011	8,652	8,686	8,763	8,746	8,017

Note: The results are for simple logit models. All numbers are in million dollars.

Table 1.12: Consumer Surplus from Apps - Model 2 with Nests

		θ				
		0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Android	2010	1,821	1,819	1,815	1,810	1,857
	2011	7,419	7,417	7,367	7,188	6,919
iPhone	2010	8,708	9,118	9,099	8,507	7,377
	2011	15,267	15,326	15,463	15,434	14,148

Note: The results are for nested logit models. All numbers are in million dollars.

Table 1.13: Comparison of Price Coefficients: Using All Apps vs. Only Paid Apps

		Model 1					Model 2	
		0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	Logit	Nested Logit
All apps	Price	-0.464	-0.389	-0.303	-0.220	-0.194	-0.300	-0.170
	Price * $d_{Android}$	-0.150	-0.179	-0.201	-0.238	-0.247	-0.335	-0.333
Paid apps only	Price	-0.026	-0.026	-0.025	-0.025	-0.025	-0.295	-0.249
	Price * $d_{Android}$	-0.021	-0.021	-0.022	-0.022	-0.022	-0.027	-0.027

Note: All price coefficients are significant at the 1% level. The price coefficients for the case using all apps are the same as those in tables 1.6∼1.8.

Chapter 2

The Interdependence between Smartphones and Applications: The Role of Platforms

2.1 Introduction

At the core of the rapidly growing smartphone industry lies the smartphone operating system (OS) that is basically what makes mobile phones "smart". Smartphones become even smarter and more valuable to consumers when used together with complementary mobile applications (hereafter, apps) and apps have been entirely or mainly distributed through the app distribution systems operated by OS firms. Therefore, the smartphone-app industry is a typical example of platform markets with indirect network effects, but the role of OS firms is especially important in determining the size of the effect. I define "app platform quality" as the quality of the app environment that is created by the role of platforms – to coordinate OS, devices, and apps so that they work together seamlessly and to make the app distribution system user-friendly – and affects consumer benefits from apps.

While Apple's iOS and Google's Android have emerged as the dominant smartphone operating systems in the US market, the two firms took very different approaches.[1]

Google has licensed Android for free to any smartphone manufacturers and thus benefited from a broader selection of devices and lower prices. On the other hand, Apple has integrated iOS and its proprietary hardware, the iPhone, and been more focused on creating a seamless user experience.[2] Apple thus kept tight control over its app distribution system, App Store, and none of the other app platforms came close to Apple's lead for a long time. One interesting fact is that iPhone users are still buying more apps than Android phone users even though Google's app distribution system, Google Play, has recently caught up with the App Store in terms of app variety.[3] Many developers selling same apps in both marketplaces saw the sales in the App Store outnumbering that in Google Play and it has been at the center of interest whether this is because Apple has a better app environment (i.e., the app platform quality of Apple

[1] As of the first quarter in 2012, 80.5 percent of smartphone owners in the US had a device that runs iOS or the Android OS according to Nielsen (http://blog.nielsen.com/nielsenwire/online_mobile/who-owns-smartphones-in-the-us/).

[2] See, e.g., the statement written by Apple in response to the FCC's questions : "Apple's goal is to provide our customers with the best possible user experience. We have been able to do this by designing the hardware and software in our products to work together seamlessly." (http://www.apple.com/hotnews/apple-answers-fcc-questions/)

[3] There were over 500,000 and 400,000 apps in the App Store and Google Play, respectively, at the beginning of 2012 while the other OS-native app platforms had less than 100,000 apps.

is higher), possibly due to different strategies of Apple and Google.[4]

In this paper I empirically examine how apps have contributed to smartphone adoption growth and also the extent to which this contribution is determined by app platform quality. I first develop a structural model of consumer demand for both smartphones and their compatible apps, where I specify the benefit provided by apps to smartphone owners as the sum of the expected utility from each app currently available on the app platform (c.f. [13]).[5] Thus I recover the differential impact of an individual app on smartphone demand, which is possible due to a new dataset containing observations on sales, prices, and characteristics of individual apps. Controlling for app quality is important given that the distribution of app sales is highly skewed and apps can be classified into many categories. If consumers receive different levels of utility from the same apps or apps of the same quality across platforms, one reason might be the app platform quality differences.

To infer differences in the overall quality of apps (app selection) and app platform quality from the fact that people on each platform are buying a different amount of apps, however, a selection problem should be addressed. It is because consumers who purchase smartphones earlier in time or choose a smartphone with a better app system are already predisposed to purchasing apps. They are self-selected into different platforms over time, considering varying device price and platform quality. Furthermore, I notice that additional selection may occur since preferences for apps in general are (more) correlated with preferences for a certain device.[6] That is, iOS and Android may attract different types of consumers and iPhone users happen to like apps more due to a stronger correlation between preferences for the iPhone and preferences for apps – e.g., richer and younger consumers prefer the iPhone over Android phones, and at the same time, they tend to like apps more.

At one extreme, it might be the case that iPhone users and Android users are different in their propensity to purchase apps regardless of the smartphones they hold. At the other extreme, there is no difference in the user groups, but the App Store might

[4] See, e.g., http://iphonedevsdk.com/forum/business-legal-app-store/87717-any-interest-in-ios-vs-android-sales-comparison.html.

[5] This is what [19] call an ideal measure for the effects of software on hardware sales and extensive data on sales, prices, and quality of sofware is needed for this.

[6] Note that this selection is not related with app/app platform quality.

have better apps or deliver higher utility for a given set of apps because apps work better with the iPhone or the App Store provides an easier search tool or payment system. To distinguish between the two, we can think of a hypothetical situation where Android users are given an iPhone or vice versa. If the formal is true, we would not see any change in the amount of app purchases. On the other hand, Android (resp. iPhone) users would increase (resp. decrease) the amount of app purchases to match that of iPhone (resp. Android) users when they have the iPhone (resp. an Android phone) if the latter is true.

To separate the selection effects of heterogeneous consumers across platforms in a static setting, I take a novel approach to constructing geographically disaggregated sales data by using Google web search data.[7] As will be seen, tastes for smartphones and tastes for overall apps seem to be largely determined by demographic differences.[8] Using the data on smartphone sales for various geographic markets in the US allows me to account for demographic-driven differences in tastes for smartphones and apps, thereby sorting out the selection effects.

The estimation results indicate that the user groups are actually different between iOS and Android, and the preferences for apps are more correlated with the preferences for the iPhone. It is, therefore, critical to control for the user heterogeneity in evaluating different app systems and measuring the effects of apps on smartphone adoption. In a counterfactual experiment of giving the iPhone to Android users, however, I observe that Android users purchase more apps in earlier months of the dataset period, which implies that Apple also has a better app system. During 2010 and 2011, Android manufacturers sold 80 percent more smartphone devices than Apple, and Android's stronger sales are estimated to come entirely from advantages in the price-adjusted quality of hardware. Apple kept advantages in the app benefits for most of the period even though the magnitude is small. Google Play was not inferior in its app selection but its app platform quality remained lower than that of the App Store, which might

[7] I consider a static model due to data limitation (my dataset for apps is a discontinuous panel covering 14 months) and ignore the dynamic aspects associated with the durability of smartphones. Consumers, however, can be viewed as treating the mobile phone decision as a "repeat" purchase made every 2 years (given the 2-year purchase cycle) and many researchers (e.g., [17] and [7]) have also assumed myopic consumers in durable goods markets as [20] pointed out.

[8] Appendix A shows some survey results on how the market shares of smartphone operating systems and the propensity of downloading apps vary across different demographic groups.

be caused by Google's open strategy.

This paper belongs to the empirical literature that examines indirect network effects in hardware/software markets (e.g., [21] ; [19]; [22]; [23]; [24] ; [13]). While most of the previous papers have used only the total number of complementary goods under the assumption of homogeneous goods, I use both hardware and software sales data and estimate differential impacts of complementary products on hardware demand as in [13]. However, I additionally control for the selection of consumers into different operating systems that stems from the fact that preferences for apps can be more correlated with preferences for one smartphone than with those for others. This additional selection effect may be more important in the study because a great portion of the consumer utility is likely to be derived from the smartphone itself (i.e., device itself has a stand-alone use – as a mobile phone – without apps), which is different from video game consoles. Tastes for smartphones and apps are largely driven by demographic differences, as already mentioned, so having disaggregated data is crucial.

As stated before, the geographically disaggregated data on smartphone sales is constructed by taking a novel approach. I utilize search data provided by "Google Trends" that shows weekly search trends for a certain keyword in metro areas of the US, in addition to aggregate national-level sales data. By analyzing the relationship between the search volume and the actual sales of each smartphone at the national level and extending it to the metro areas, I can obtain estimated sales numbers for many markets of varying demographics within the country. This study is the first to make use of Google search data in demand estimation, and suggests one way of overcoming data limitation issues when bringing demographics into analysis is important but only aggregate data is available.

This paper also contributes to a recent literature on the smartphone industry. [25] studied how exclusive contracts affect pricing and entry incentives, focusing on the exclusive contract between Apple and AT&T over the iPhone. [20] also examined the exclusive arrangement between Apple and AT&T and its impact on consumer welfare. In spite of the inherent complementary nature of apps to smartphones, they mostly ignored the app part. To my best knowledge, [26] is the only paper dealing with the smartphone market and the app market together but he basically followed the existing framework of the literature estimating indirect network effects, only using the total number of apps. My approach of discovering differential impacts of apps on smartphone adoption is possible because I collected detailed characteristic and rank information on best-selling apps in the App Store and Google Play over time, which allows me to capture app quality properly.

The remainder of the paper is organized as follows. Section 2.2 describes data sets and some summary statistics. Section 2.3 presents the demand model for both smartphones and apps. Section 2.4 discusses the estimation procedure and identification. Estimation results are reported in Section 2.5. Section 2.6 concludes.

2.2 Data

2.2.1 Demand and Characteristics Data

I combine several sources of data for smartphone. The data on smartphone quantities were obtained from Gartner, an information technology research firm, and consist of quarterly unit sales by operating system in the United States from 2008 to 2011. I use the Gartner data together with Google search data to construct geographically disaggregated quantities, which I explain in detail in section 2.2.2. Smartphone handset price comes from the website camelcamelcamel.com and is no-contract price at Amazon.com.[9] I also collect handset characteristics from www.phonearena.com. Because I do not observe quantities at the handset level, I take weighted averages of the prices and characteristics of popular handsets for each operating system with weight being the impression share from Millennial Media, a mobile advertising firm.[10] The resulting data set covers the 23 months from February 2010 to December 2011. Lastly, I use cross-sectional data on the average signal strength for each wireless carrier in each designated market area (DMA) (see [20] for data description) to account for the effect of signal quality on smartphone choice.[11],[12]

[9] The website camelcamelcamel.com is an Amazon price tracker that provides price history charts for products sold by Amazon.

[10] An (ad) impression is a measure of the number of times an ad is seen by a user. Millennial Media publishes monthly report showing the shares of top mobile handsets in terms of the impressions across its network.

[11] I thank the authors for sharing the signal strength data.

[12] Even though I do not model consumer decisions in network service providers, signal quality still matters because of the exclusive contract between Apple and AT&T for the iPhone until February 2011.

Smartphone descriptive statistics are reported in table 2.1. There are substantial variations in handset prices and characteristics over time even though they were aggregated up to the operating system level. Figure 2.1 shows that as with other high-tech products, average prices fall slightly over time while firms continually enhance the quality of hardware, so we can expect the composition of smartphone purchasers to change over time. Most observable device characteristics are comparable between Android phones and the iPhone but average prices remain higher in the case of the iPhone. Table 2.1 also shows that there are a lot of device options for Android due to its openness while the iPhone is essentially only one device.

Next, I need a measure of the benefit provided by compatible apps to smartphone users to model the effects of apps on smartphone adoption. I collect rank, price and characteristics information on best-selling apps in the US market because it is not feasible to track all apps due to the large number of titles available for each OS and app sales are known to be highly concentrated in top-ranked apps. The app dataset includes daily observations on title, seller, price, release date, customer rating and additional characteristics for top 100 free/paid iPhone apps and top 100 free/paid Android apps for 14 months in 2010 and 2011.[13] ,[14] I translate the observed ranking of each top app into quantity downloaded by assuming a power law distribution for app sales following [2]. Section 1.2.2 outlines the procedure and estimation results for deriving numbers of app downloads.

Table 2.2 reports summary statistics for top free and paid apps in the App Store and Google Play. On each app platform there are about 4,200 apps ever listed in the top 100 chart for each day of the sample period and paid apps tend to remain longer in the chart than free apps. While roughly 100 million copies are sold in both markets each month on average, the sales of paid apps is much smaller in Google Play (only one third of the App Store) possibly because of higher app prices.[15] One important thing to note is

[13] In the case of Google Play, neither number of downloads nor rank was publicly available for individual apps until early 2011. I thus collected data on top Android apps from the website AndroLib.com, one of the most trusted and visited web sites by users and developers before the introduction of web-based Google Play (formerly known as Android Market), for the early period. AndroLib shows lists of 150 free and paid apps that are most viewed on its site each day, and I gathered the ranks and characteristics of apps on these lists.

[14] Rank, title and seller information was collected for the whole two years but price and additional characteristics are available only for the 14 months listed in table 2.

[15] Top 100 apps are estimated to account for about 40$\sim$50 percent of the total app downloads in

that the average number of app downloads per device is much higher for the iPhone; app purchases equal 44.7 per device for the iPhone and 25.3 for Android, respectively, in the case of free apps; for paid apps this number is equals 0.5 for Android, which is less than 15 percent of 3.5 for the iPhone. The total number of app purchases per phone for each app platform converge over time and become similar in the fourth quarter of 2011, but the gap between the two persists over the sample period. This may be because iPhone apps are of higher quality than Android apps, but this does not seem to be (entirely) attributable to app quality differences when observing that the average customer rating for free apps is rather higher in Google Play. There are possibilities that Android users derive lower utility from a given set of apps because apps are not working well with the OS/devices or user groups are different in the propensity to purchase apps between the two platforms.

In order to make this point clearer, I present summary statistics for top apps that are also on the other app platform in table 2.3. Apps are matched between the App Store and Google Play based on title, seller and free/paid type, and as a result, less than 10 percent of the top apps turn out to be sold on both app platforms.[16] A pair of matched apps can be assumed to have the same quality but table 2.3 shows that the gap in the numbers of app purchases per phone is still present, supporting the possibility of app platform quality difference or user heterogeneity. In section 2.3, I thus build a model of consumer demand for both smartphone and app that can identify both effects separately. Lastly, sellers introduce apps to the App Store six months earlier on average than to Google Play.

2.2.2 Construction of Disaggregated Data using Google Trends

As I mentioned in the introduction, the selection problem should be addressed to properly infer the effects of app selection and app platform quality on smartphone purchase decisions. Because tastes for smartphones and tastes for apps in general are

both app platforms. See section 1.2.2 for a detailed discussion.

[16] Many apps have to be matched manually because there are many cases where developers enter both markets with completely different seller names.

likely to be greatly affected by demographic characteristics, I construct geographically disaggregated data on smartphone sales. To do this, I combine the national-level quantity data by Gartner with micro search data obtained from "Google Trends" (http://www.google.com/trends/). Google Trends is a service that analyzes a portion of Google web searches and provides weekly trends of the search volume of a given search term across different locations in the US (see figure 2.2).[17] By analyzing the relationship between the search volume and the actual sales of each smartphone at the national level and extending it to metro areas, I can get estimated sales numbers for many markets of varying demographics within the country.

There are two underlying assumptions behind this approach: (i) Google users are representative consumers and (ii) search volume for a particular product reflects the actual quantity sold. The former assumption can be justified by Google's representativeness as a search engine.[18] Regarding the latter, note that a substantial amount of people use web search engines to collect information on goods they intend to buy or web sites they intend to visit (i.e., web search can be thought of as a preparatory step to the actual purchase or visit). I therefore expect some constant relationship between search volume and actual quantity sold, which I explain in the next paragraph.

I collect the search trends for the keywords of "android", "blackberry" and "iphone" for each of the 210 metro areas, which mostly coincide with Nielsen's DMAs (DMA is a region where the population receive the same television station offerings).[19] Figures 2.3 and 2.4 compare the movements of the Google search index and the actual quantity sold (at the national level) for each operating system. There may be a time lag between information search and actual purchase, so I experiment with lags from

[17] Instead of the raw level of search volume, Google Trends provides a query index of how often a particular search term is entered relative to the total search volume in a given geographic region at a point in time. The query index numbers for each search term are normalized and presented on a scale from 0 to 100, with 100 being the highest point across regions and time. For a detailed description, see [27] and a research paper by Google Israel Labs, "On the Predictability of Search Trends" which can be accessed from the following post: http://googleresearch.blogspot.com/2009/08/on-predictability-of-search-trends.html.

[18] Google accounted for 65 percent market share of the US search market during the sample period (Source: comScore's press releases on U.S. Search Engine Rankings).

[19] Google classifies search queries into categories and sub-categories using an automated classification engine (natural language processing methods). I do not filter on categories because smartphones and apps are new phenomena and thus do not fit into a single category well. Only for "blackberry", searches were restricted to the category of "Internet and Telecom" because it can also refer to a kind of fruit.

one week to one month. In what follows, I proceed with one-week lag since it gives the best fit. As Figure 2.3 shows, their movements are surprisingly similar for android even though they become somewhat loose for blackberry and iphone. However, Google search index still captures the overall trend and peak points of the actual quantity quite well. Figure 2.4 replicates figure 2.3 in scatter plots: the search volume and the sales for each smartphone seem to have a strong proportional relationship. The solid line in each graph represents a line going through the origin and we can see that the proportional relationship between search and sales is pretty tight.

By assuming a proportional relationship between the search volume and the actual sales, I construct monthly data on smartphone market shares for 210 metro areas. Figure 2.5 displays cross-sectional distributions of the constructed market shares for each operating system in the final month of the dataset. The smartphone market shares show a lot of variation even though the smartphone products offered, their prices and characteristics do no change across metro areas. The figure is for December 2011 that is after Verizon and Sprint introduced the iPhone, so the differences in signal strength should not be the main driver of this variation. It implies that demographic characteristics might be important in smartphone choice, which will be reflected in the model.

2.3 Model

In each month t, consumer i makes decisions in two stages: first, which smartphone to purchase among $j \in (0, ..., J)$ and second, which apps to purchase among the set of compatible app titles available on the platform (denoted as K_{jt}) if she chooses to purchase a smartphone. A consumer gets more information on apps in the second stage by actually using the smartphone and searching for apps. In the first stage, she forms expectation over the unobserved information and considers the total expected benefit from apps currently available when making a decision.[20]

[20] I here assume that consumers behave myopically at the time of the smartphone decision and only consider the possibility of apps currently available.

2.3.1 Smartphone Demand

Consumers choose among $J + 1$ alternatives where $j = 0$ denotes the outside good consisting of all feature phones as well as no purchase. In a general random coefficients model, the utility that consumer i living in market area m gets from smartphone running OS j at time t can be specified as:

$$u_{ijmt} = \beta_{ij} + x_{jt}\beta - \alpha_i p_{jt} + \beta^{\Gamma}\Gamma_{jt}(\beta_i^{app}) + \gamma SS_{jm} + \xi_{jmt} + \epsilon_{ijmt}, \tag{2.1}$$

where β_{ij} and β_i^{app} represent individual-specific preferences for operating system j and for apps in general. x_{jt} and p_{jt} are average (sales-weighted) device characteristics - including processor speed, screen size, camera resolution, etc - and price for OS j in month t. $\Gamma_{jt}(\beta_i^{app})$ is the expected value of being able to purchase apps currently on the platform. SS_{jm} is the signal strength for OS j in market m.[21] ξ_{jmt} is the market mean (across consumers) of the unobserved (by the econometrician) component of utility.[22]

As is obvious from equation (2.1), the utility a consumer derives from hardware and software is assumed to be additively separable following the literature. $\Gamma_{jt}(\beta_i^{app})$ will be defined in section 2.3.2 after I introduce the model of app demand. The distribution of consumer preference parameters for the operating systems can be modeled as multivariate normal with a mean that is a function of demographic variables as in Nevo (2001):

$$\beta_{ij} = \beta_j + \pi_j D_i + \sigma_j \nu_{ij}, \tag{2.2}$$

where β_j is the mean of individual preferences for operating system j. D_i is demographic characteristics to be drawn from Census survey data and ν_{ij} is a random shock to be independently drawn from the standard normal distribution.

This paper estimates a nested logit model that is a special case of the general model above where random coefficients are only given to group-specific dummy variables. I group the products into $G + 1$ sets, $g = 0, 1, ..., G$: Android and iOS are in one nest, Blackberry is in another and the outside option is in a nest by itself. For smartphone j

[21] I set this equal to the average signal strength of AT&T until January 2011, that of AT&T and Verizon until October 2011 and that of AT&T, Verizon and Sprint after October 2011 for the iPhone to account for the exclusive contract between Apple and AT&T. For the other products, it is the average signal strength of four major carriers - Verizon, AT&T, Sprint and T-mobile.

[22] To be exact, ξ_{jmt} is a metro area-month specific deviation from the mean of the unobserved component of utility that is controlled by including brand-specific dummy variables (β_j in equation (2.2)).

in group g, the utility of consumer is:

$$u_{ijmt} = \tilde{\pi}_j \bar{D}_m + x_{jt}\beta - \alpha p_{jt} + \beta^{\Gamma}\Gamma_{jt}(\bar{D}_m) + \gamma SS_{jm} + \xi_{jmt} + \zeta_{igmt} + (1-\sigma)\epsilon_{ijmt}, \quad (2.3)$$

where ϵ_{ijmt} is an identically and independently distributed extreme value and ζ_{igmt} is common to all products in group g. I still allow the preferences for each operating system j and for apps in general to be affected by demographics, but by mean demographics of each market area m (denoted as $\bar{D}_m$). This is as if consumers who live in the same geographic market are assumed to have the same (mean) demographic characteristics and thereby the same preferences for smartphone operating systems and apps except for the idiosyncratic error terms, ζ_{igmt} and ϵ_{ijmt}.[23]

2.3.2 App Demand

As shown in section 2.2.1, the assumption of purchasing only one app is clearly rejected by the data and I deal with a very large number of app titles. I thus assume that a consumer makes the purchase decision independently for each app (c.f. Lee (2011)). I specify a general random coefficients model first. If buying app k at time t, consumer i owning smartphone j receives a utility of:

$$v_{ijkt} = \beta_i^{app} + \beta_{jt}^{app} + x_{jkt}\beta^{app} - \alpha^{app} p_{jkt} + \eta_{jkt} + \epsilon_{ijkt}, \quad (2.4)$$

where β_i^{app} is an individual-specific preference for apps in general and β_{jt}^{app} is the app platform quality of OS j over time.[24] x_{jkt} are observable app characteristics and

[23] Consumers may face costs of switching among smartphone operating systems and become locked into current ones (see e.g., [28] and [29]), which is not considered in this paper due to data limitation. Addressing this would require individual-level choice data or long enough market-level data (when compared to the product purchase cycle) to introduce dynamic considerations. My justification for the simplification is that the smartphone industry was in its early stage and therefore a large portion of the smartphone purchases were made by new smartphone users during my sample period. Given that mobile phones have a 2-year purchase cycle, the number of consumers who purchased smartphones two years ago should be a good proxy for replacement demand (i.e., demand from consumers who want to replace the currently owned smartphones). Figure 2.6 compares the combined sales of Android phones and the iPhone to those of 2-year prior. The ratio between the two remained at 16.4% on average during 2010-2011, which implies that more than 80% of the purchases are made by first-time smartphone users and thus the effects of switching costs should not be very large.

[24] β_{jt}^{app} also includes seasonal effects but they cannot be identified separately due to data limitation (the dataset covers only 14 months). However, seasonal effects do not seem to be large (see section 2.5.1 and figure 2.7) and I thus interpret the term as the app platform quality. It again represents how well OS, devices and apps work together and how efficient an app distribution system is.

include app-specific fixed effects, age, time since last update and customer rating. p_{jkt} and η_{jkt} are app price and unobserved quality. Note that the idiosyncratic shock for apps is allowed to have a different variance from that for smartphones through the term β^{Γ} in equation (2.1) (see Lee (2011) for a detailed explanation).

The distribution of the consumer taste parameter for overall apps is similarly modeled as normal with a mean that is a function of demographic variables:

$$\beta_i^{app} = \pi D_i + \sigma \nu_i, \qquad \nu_i \sim N(0,1). \tag{2.5}$$

Then the expected value of being able to purchase apps currently on the platform can be defined as:

$$\begin{aligned} \Gamma_{jt}(\beta_i^{app}) &= \sum_{k \in K_{jt}} \int_{\epsilon} \max\{v_{ijkt}, v_{ijk_0t}\} dP_{\epsilon} & (2.6)\\ &= \sum_{k \in K_{jt}} \ln(\exp(\beta_i^{app} + \beta_{jt}^{app} + x_{jkt}\beta^{app} - \alpha^{app} p_{jkt} + \eta_{jkt}) + 1), & (2.7) \end{aligned}$$

where K_{jt} denotes the set of compatible app titles available on platform j in month t and $v_{ijk_0t} = \epsilon_{ijk_0t}$ is the utility from an outside option of not purchasing app k. Thus, I specify the benefit provided by apps to smartphone owners as the sum of the expected utility from each app currently available on the app platform.

The two extreme cases explained in the introduction would be realized in the model as following: (i) if iPhone users and Android users are different in their app preferences, then the distributions of β_i^{app} should differ for $j = iphone$ and $j = android$ (but no differences in β_{jt}) and I expect $Corr(\beta_{i,iphone}, \beta_i^{app}) > Corr(\beta_{i,android}, \beta_i^{app})$ when individuals who have high values of β_i^{app} also have high values of a taste for the iPhone, and (ii) if there is no selection of consumers between the two platforms but Apple has higher app platform quality (i.e., there is something better about iPhone apps that make them work better), then the levels of β_{jt}^{app} should be higher for $j = iphone$ than for $j = android$.

In this paper I estimate a logit model but I allow the preferences for apps to be affected by mean demographics of users in each platform at a point in time (i.e., $\bar{D}_{jt}$). The form of the equation to be estimated is:

$$v_{ijkt} = \tilde{\beta}_{jt}^{app} + x_{jkt}\beta^{app} - \alpha^{app} p_{jkt} + \eta_{jkt} + \epsilon_{ijkt} = \tilde{\beta}_{jt}^{app} + \delta_{jkt}^{app} + \epsilon_{ijkt} \tag{2.8}$$

where δ_{jkt}^{app} is the compoment of the utility varying by app and any variation in the app benefits induced by introduction of new apps of different quality over time will be captured by the term (x_{jkt} includes app-specific fixed effects). $\tilde{\beta}_{jt}^{app}$ now represents the mixed effects of the user heterogeneity in overall likeness for apps (that changes across platforms and time) and the app platform quality. It will explain the variation in app utility common across all apps (i.e., variation not related with app selection). $\tilde{\beta}_{jt}^{app}$ cannot be separately identified for all j and t from shifts in each app's fixed effect, so $\tilde{\beta}_{android,1}^{app}$ is normalized to zero. For some app characteristics, β^{app} is allowed to be different between the App Store and Google Play.

Now there is an issue in including mean demographics directly in the equation because the effects of $\bar{D}_{jt}$ will be absorbed in $\tilde{\beta}_{jt}^{app}$ terms. To distinguish between the two effects of user heterogeneity and app platform quality, I do a auxiliary regression of the recovered $\tilde{\beta}_{jt}^{app}$ on mean demographic characteristics of users in each smartphone platform in each month as well as a dummy variable for the App Store:

$$\tilde{\beta}_{jt}^{app} = \tilde{\pi}\bar{D}_{jt} + \beta_{j}^{app} + \omega_{jt}, \tag{2.9}$$

where $\tilde{\pi}$ captures the change in the value of purchasing any app to the outside option (not purchasing the app) as a function of demographics. This is, therefore, a special case of the general model where preferences for overall apps are determined by mean demographic characteristics, and the expected benefit from apps currently available can be constructed for each demographic type as in equation 2.7.

2.4 Estimation

2.4.1 Computation

For the general random coefficients model, a computational fixed point routine of iterating between smartphone demand estimation and app demand estimation is needed to control for the selection (see e.g., [30] and [13]). To evaluate parameters in the smartphone demand model, the expected value of apps, $\Gamma_{jt}(\beta_i^{app})$, is needed, which means that the estimation of app demand should precede the estimation of smartphone demand. However I have to form demographic distributions of users in each platform at each point in time to estimate the app purchase decision, which can be done by updating the market distributions with probabilities of purchasing a particular smartphone across time:

$$Pr(i|j,t) = \sum_i \frac{Pr(j|i,t)Pr(i)}{Pr(j|t)}, \tag{2.10}$$

where $Pr(j|i,t)$ and $Pr(j|t)$ are consumer type i's purchase probability and the market share of smartphone j at time t, and $Pr(i)$ is the share of consumer type i in the population. The iteration is, therefore, required between estimating smartphone demand and estimating app demand.

I can start either from estimating smartphone demand by setting $\Gamma_{jt}(\beta_i^{app}) = 0$ or from estimating app demand by assuming that the demographic distributions of users are not different from the market distributions (i.e., no selection), and then repeat the procedure of updating the values for $\Gamma_{jt}(\beta_i^{app})$ and $\{Pr(i|j,t)\}$ until they converge. In this way, the user differences in app likeness between the platforms can be controlled, even in a static setting, to the extent app preferences are explained by demographic characteristics.

In this paper, average demographic characteristics are used instead of taking simulation draws from the demographic distributions for each metro area or of users in each platform across time. Instead of using the predicted purchase probabilities obtained from the smartphone demand model, I compute mean demographic characteristics of smartphone holders in each month by weighting the mean demographic characteristics of each metro area with observed probabilities of purchasing a particular smartphone:

$$\bar{D}_{jt} = \sum_m Pr(m|j,t)\bar{D}_m = \sum_m \frac{Pr(j|m,t)Pr(m)}{Pr(j|t)}\bar{D}_m \tag{2.11}$$

where $Pr(j|m,t)$ and $Pr(j|t)$ are market shares of smartphone j in geographic market m and in the US, respectively, at time t and Pr(m) is the population share of m. This term can be proved to give the right sign for demographic effects on app utility. This is an approximation but I can avoid the need to iterate, under the assumption that the data on smartphone shares are generated by consumers' (optimal) joint decision on hardware and apps.

I start with estimating the app demand model in equations (2.8) and (2.9). The size of the potential market for apps in each platform in each month is defined as the number of consumers who purchase the smartphone in that month. Here I thus assume that consumers purchase all wanted apps within one month after smartphone purchase. This is because of data limitation but it is known that most of the app purchases are made in the first few weeks. I then compute the expected value of apps for each demographic type and include the measures in the smartphone demand estimation. The market size for the smartphone adoption side equals the population of 10 years and over in each metro area.

2.4.2 Identification

Smartphone Equation. The parameters for product characteristics and price, β and α, are identified from time variation in market shares as such characteristics change because they are same across cross-sectional markets. Even though there is no variation in the choice sets, product characteristics and price do change a lot over time as new devices are introduced for each smartphone operating system. While cross-sectional variation in sales identify both the effects of demographics on OS preferences and of signal strength, i.e., $\tilde{\pi}_j$ and γ, there is an additional identifying power for γ from time variation in sales around when new network providers are added for the iPhone. β^{Γ}, the parameter for the expected app utility, uses both time and cross-sectional variation in sales as its value for each demographic type changes over time (because of new app introductions or improvements in the app platform quality).

App Equation. App characteristics parameters (β^{app} and α^{app}), of course, are identified time variation in sales of each app. The identification of $\tilde{\beta}^{app}_{jt}$, the mixed effects of the user heterogeneity in overall likeness for apps and the app platform quality, comes from the comparison of same apps sold on both platforms over time. Lastly the impacts of demographics on app preferences, i.e., $\tilde{\pi}$, is identified by comparing $\tilde{\beta}^{app}_{jt}$ and mean demographics, $\bar{D}_{jt}$, across platforms and time. As explained in the previous section, the time-varying mean demographics for each platform are available because I constructed sales panel (by metro area and month) for smartphones.

2.5 Results

2.5.1 Estimation Results

Parameter estimates for the app demand model are presented in table 2.4. I run an instrumental variables regression to account for the endogeneity of price after inverting market shares to market mean utility levels (see [18] and [17]). The coefficients for observable characteristics including price are all significant and have expected sign. App age and time since last update have negative effects on the app utility even though they are small in magnitude. The coefficient for the dummy variable indicating that an app was introduced or updated within past three days is significantly positive for both platforms but it is three times larger in the App Store than in Google Play. This is because apps that are recently introduced or updated have a chance to be featured in front page sections for new apps.[25] Customer rating is estimated to affect the utility positively only in Google Play. While the absolute values of the price coefficient estimates are somewhat small, Android users tend to be much more sensitive to app price than iPhone users.

I report the estimates of $\tilde{\beta}_{jt}^{app}$ in table 2.5. Their values for the App Store remain higher than those for Google Play during the entire period even though the gap between the two platforms is narrowing over time. This means that iPhone users receive higher utility from an app of same quality than Android users do, which may be because iPhone users are more likely to love apps in general than Android users (i.e., user selection) or Apple manage its app platform better (i.e., higher app platform quality). To distinguish between the two, I regress the recovered $\tilde{\beta}_{jt}^{app}$ terms on average demographic characteristics of users in each smartphone platform in each month as well as a dummy variable for the App Store as in equation (2.9).

The estimation results are reported in table 2.6. Unsurprisingly, the coefficient for the App Store dummy is significant and positive. A more interesting result is that user income and age affect the utility from apps greatly; richer and younger consumers like apps more. The implied elasticities for the apps with an average market share in the final month (19.2%) are 18.7 and -5.2 for income and age, respectively. Their

[25] They are the New and Noteworthy section of the App Store and the Top New Paid/Free chart of Google Play.

minimum values are close to zero and occur for some apps ranked number one that are purchased by nearly all users. $\tilde{\beta}_{jt}^{app}$ estimates and mean user incomes for each platform are compared in figure 2.7 and it is evident that they move very closely over time.[26] These results indicate that iPhone users buy more apps partly because they tend to like apps more regardless of the smartphone they hold and the selection problem should be accounted for to properly measure the impacts of apps on smartphone choice.

Given the above results, the expected benefit from apps currently available can be constructed for each demographic type (i.e., $\Gamma_{jt}(D_m)$) as in equation 2.7 for Android and the iPhone and it is set to 0 for Blackberry.[27]

Table 2.7 displays the results for multiple specifications of smartphone demand. Columns (i)-(iv) estimate logit models instrumenting for price. The regression in column (i) includes only smartphone characteristics and signal strength, and then I introduce demographics into the regression in column (ii). The effect of adding demographics, which is possible due to the use of the disaggregated data, on the price coefficient is significant: its absolute value increases by more than tenfold. The implied price elasticity at the mean smartphone price is around 2 in column (ii). Remember that the coefficients on demographic variables capture the change in the value of each smartphone to the outside option (feature phones or currently owned phone) as a function of demographics. The results suggest that the value of smartphones increases with income, while its effect is most pronounced for the iPhone. Consumer age does not have significant effects on the smartphone utility in specification (ii). Column (iii) additionally includes the expected utility from apps as a regressor, which is calculated from the estimation results for app demand described above, but the selection effect is not considered yet (i.e., every consumer type receives the same app utility). Contrary to the expectation, the coefficient for the expected app utility is negative.

Column (iv) uses the expected app utility that is defined separately for each demographic type (i.e., $\Gamma_{jt}(D_m)$) and thereby accounts for the selection problem. The effect of apps on smartphone adoption is now estimated to be significantly positive as

[26] As mentioned in section 2.3.2, $\tilde{\beta}_{jt}^{app}$ also includes seasonal effects but they cannot be identified separately due to data limitation (only 14 months). However, seasonal effects do not seem to be large when comparing the estimated values of $\tilde{\beta}_{jt}^{app}$ for the months observed twice.

[27] This restriction is not a serious limitation given that Blackberry phones are mainly purchased for business purposes and the number of apps purchased per Blackberry device is known to be about the same to that of feature phones.

expected. Also, the coefficients on age terms turn significant and its effects have are bell-shaped. Column (v) presents the results of a nested logit model using the same regressors as in column (iv). By moving from the logit models to the nested model, I see a considerable improvement in fitting and the age effects become more significant. The results of specification (v) are used in the following analysis. As a comparison with the previous approach, column (vi) additionally includes the total number of available apps that is the variable used in most previous papers to measure complementary effects. The inclusion of the term hardly changes the estimates of the other coefficients and its coefficient is not statistically significant. In other words, the total number of apps does not have any explanatory power once the expected utility from apps is accounted for.

2.5.2 Discussion

In this section I use the estimation results and show: (i) how much of the gap in app purchases between Apple and Google is explained by user heterogeneity versus app system difference, (ii) the relative importance of smartphone quality and apps in smartphone platform competition, and (iii) the break down of the app advantages into app platform quality and app selection.

First I do a counterfactual experiment of giving the iPhone to Android users and examine whether and how they change the amount of app purchases. Table 2.8 displays the results. (A) and (C) report the actual numbers of app purchases per phone for Android and iPhone users, respectively, and (B) is for the hypothetical situation. The next two columns show the differences between (A) and (B) and between (B) and (C). When comparing (A) and (B), I keep the demographics of Android users but swap the app platforms, so the difference between the two should reveal the differences in the app platform quality and app selection. On the other hand, the gaps between (B) and (C) are purely caused by user group differences between Android and the iPhone. As table 2.8 shows, the effects of selection on app purchases are very large: iPhone users consist of consumers who tend to like apps more. Android users would have purchased around 30 less apps in 2010 and 10 less apps in 2011 if they had the iPhone. This underlines the need to account for the selection issues when estimating app demand and measuring apps' effects on smartphone decision. Note also that in earlier periods there seems to be something better about Apple apps that makes people want to buy them even after controlling for the selection (compare (A) and (B)).

Next I want to determine the relative importance of apps versus the (price-adjusted) quality of smartphones in driving smartphone sales. One way to do this is to look at how much of the log-odds ratio of the sales of Android phones to the iPhones is explained by the app benefit as opposed to the smartphone price and quality differences between the two platforms (see [31] and [19]). The log-odds ratio can be written as:

$$\begin{aligned} log(\frac{s_{Amt}}{s_{Imt}}) &= \delta_{Amt} - \delta_{Imt} \\ &= [(\beta_A - \beta_I)D_m + (x_{At} - x_{It})\beta - \alpha(p_{At} - p_{It}) + \gamma(SS_{Am} - SS_{Im}) + (\xi_{Amt} - \xi_{Imt})] \\ &\quad +[\beta^{\Gamma}(\Gamma_{At}(D_m) - \Gamma_{It}(D_m))], \end{aligned} \tag{2.12}$$

where $j = A$ and $j = I$ denote Android and iOS, respectively. The first bracket in the last equation of (2.12) can be thought of as the price and quality advantage of Android over iOS, and the second bracket as the advantage from apps. Note that the selection effect of different consumers onto both platforms is eliminated in equation (2.12) because I compare the utility levels from two products for the same consumer type. Figure 2.8 displays the percentage compositions of the two terms over time for the country. During the sample period, the price-adjusted smartphone quality explains most of the relative sales of Android phones over the iPhones. However, Apple keeps advantages in the app benefits during most of the period even though the magnitude is small. The relative Android sales are potentially decreased by around 4.5% in the beginning month due to inferior app benefits. I expect that the portion of the log-odds ratio explained by the app benefits was higher before 2010.

I further examine the differences in the complementary effects of apps by classifying them into the variances induced by app selection (captured by δ^{app}_{jkt} in equation (2.8)) versus the variances induced by app platform quality (captured by $\tilde{\beta}^{app}_{jt}$ after sorting out the demographic effects in equation (2.9)). As figure 2.9 shows, Google was as successful as the App Store in introducing high quality apps during the sample period. The app platform quality of the App Store, however, remained higher than that of Google Play. This coincides with the fact that Android users and developers are reported to experience problems in using and developing apps because of fragmentation among different devices and different OS versions, which arises from Google's open strategy. More data are needed to analyze it, but the results shed some light on the comparison of open and closed strategies in platform markets.

2.6 Conclusion

This paper examines the complementary effects of apps on smartphone adoption, focusing on the case of Apple's iOS versus Google's Android, by estimating smartphone demand and app demand together. To explain why these effects differ across smartphone operating systems, I address the selection problem of heterogeneous types of consumers onto platforms and identify the differences in app selection and app platform quality. This is all possible because of the extensive data work of collecting information on individual apps and combining smartphone sales data with Google search data.

By running a hypothetical experiment of giving the iPhone to Android users, I first show that the effects of the selection on the amount of app purchases are quite large: Android users would still purchase much less apps than iPhone users even if they had the iPhone. It is, therefore, important to control for the user heterogeneity in evaluating different app systems and measuring the effects of apps on smartphone adoption. After controlling for the user selection, the results suggest Apple provided more app benefits to the users and Android's stronger sales entirely came from advantages in the price-adjusted quality of hardware. Apple kept advantages in the app benefits mostly due to the higher app platform quality. These results show some evidences on the benefits and costs of open versus closed strategies in platform markets.

Table 2.1: Smartphone Summary Statistics

	Android		Blackberry		iPhone	
	Mean	Std	Mean	Std	Mean	Std
Quantity (M)	3.5	1.5	1.2	0.4	1.9	1.2
Price ($)	456.8	80.9	334.5	31.6	648.4	7.0
# Devices	176.1	104.7	58.6	6.2	3.0	0.0
Screen Size (inches)	3.6	0.2	2.6	0.1	3.5	0.0
Touch Screen (1∼3)	2.7	0.3	1.2	0.1	3.0	0.0
Processor Speed (MHz)	770.2	113.7	413.1	79.4	762.1	106.4
RAM (MB)	341.8	62.0	252.4	26.5	368.1	79.5
Camera Resol (Mpixels)	5.1	0.7	3.2	0.1	4.1	1.1
Talk Time (hrs)	7.0	0.6	5.4	0.1	12.7	0.7
Stand-by Time (hrs)	286.0	37.3	365.4	11.3	292.6	19.0

Note: Summary statistics are for the 23-month period between February 2010 and December 2011. Price is no-contract price at Amazon.com. Touch Screen; 1: not capacitive/multi-touch, 2: capacitive, 3: capacitive and multi-touch.

Table 2.2: App Summary Statistics

	iPhone (free)		iPhone (paid)		Android (free)		Android (paid)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
# Unique Titles on Top 100 Chart	2,908		1,311		2,639		1,652	
# Days Listed on Top 100 Chart	13.9	34.9	30.7	63.8	15.5	49.3	24.7	60.9
Monthly Total App Downloads (M)	89.0	39.0	6.0	1.0	100.0	61.0	2.1	1.2
Total App Downloads/Device	44.7	14.2	3.5	1.6	25.3	7.4	0.5	0.1
Price ($)	0.0	0.0	1.8	1.8	0.0	0.0	4.0	5.0
Age (days)	285.6	303.7	333.4	288.7	203.0	169.4	217.2	171.5
Since Update (days)	45.4	78.5	74.3	126.9	39.6	114.6	41.1	107.0
Customer Rating (1∼5)	3.5	0.7	4.0	0.5	4.2	0.7	4.0	0.9
Category (%)								
Games	51.5		58.5		33.3		43.2	
Entertainment	12.8		12.2		12.0		4.8	
Utilities	5.4		6.5		8.2		17.8	
Productivity	1.2		1.7		5.8		6.9	
Social	8.3		2.2		14.0		4.8	
Media	4.4		4.5		10.0		8.9	
Business	0.1		0.2		0.1		4.4	
Others	16.3		14.2		16.7		9.3	

Note: Summary statistics are for top 100 free and paid apps for 14 months during 2010-2011 (Apr∼Jun and Nov∼Dec of 2010, and Jan∼Jun and Oct∼Dec of 2011). The data on Android apps are collected from the website AndroLib.com except for Oct∼Dec of 2011.

Table 2.3: Summary Statistics for Apps Sold in Both Markets

	iPhone (free)		iPhone (paid)		Android (free)		Android (paid)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
# Unique Titles on Top 100 Chart	241		122		180		178	
# Days Listed on Top 100 Chart	46.3	80.0	67.9	100.9	32.8	54.4	36.6	66.0
Monthly Total App Downloads (M)	23.0	11.0	1.2	0.3	22.0	33.0	0.3	0.4
Total App Downloads/Device	10.8	3.0	0.7	0.3	4.5	5.7	0.1	0.1
Price ($)	0.0	0.0	2.0	2.2	0.0	0.0	4.6	6.6
Age (days)	403.4	328.7	424.5	318.7	213.0	196.4	221.9	190.0
Since update (Days)	50.9	88.3	84.7	139.3	110.0	189.0	69.8	139.5
Customer Rating (1∼5)	3.7	0.5	4.0	0.4	4.3	0.4	4.1	0.6

Note: Summary statistics are for top 100 free and paid apps that are also sold on the other app platform for 14 months during 2010-2011 (Apr∼Jun and Nov∼Dec of 2010, and Jan∼Jun and Oct∼Dec of 2011). The data on Android apps are collected from the website AndroLib.com except for Oct∼Dec of 2011.

Table 2.4: App Demand Estimation Results (1)

Variable	Estimate	S.E.
Price (Android)	-0.529***	(0.035)
Price (iPhone)	-0.328***	(0.037)
Age	-0.037***	(0.004)
Age sq	0.001***	(0.000)
Updated	-0.008**	(0.003)
Updated sq	0.000**	(0.000)
$d_{recent\ update}$ (Android)	0.074***	(0.007)
$d_{recent\ update}$ (iPhone)	0.199***	(0.020)
Customer Rating (Android)	0.051**	(0.025)
Customer Rating (iPhone)	-0.199***	(0.028)
Constant	-3.264***	(0.135)
Number of obs	160,825	
Adj R-squared	0.859	

Note: The regression includes app fixed effects. ***, ** and * indicate significance at the 1%, 5% and 10% levels.

Table 2.5: App Demand Estimation Results(2)

Variable	Android		iPhone	
	Estimate	S.E	Estimate	S.E
d_{Apr10}	0.000***	-	3.984***	(0.168)
d_{May10}	-0.172***	(0.020)	3.977***	(0.168)
d_{Jun10}	-0.377***	(0.021)	3.110***	(0.168)
d_{Nov10}	-0.019	(0.029)	2.828***	(0.172)
d_{Dec10}	-0.074**	(0.031)	3.429***	(0.173)
d_{Jan11}	0.145***	(0.034)	2.739***	(0.174)
d_{Feb11}	0.380***	(0.036)	2.853***	(0.177)
d_{Mar11}	0.488***	(0.039)	2.856***	(0.178)
d_{Apr11}	0.454***	(0.041)	2.488***	(0.176)
d_{May11}	0.558***	(0.043)	2.775***	(0.177)
d_{Jun11}	0.620***	(0.046)	2.475***	(0.179)
d_{Oct11}	1.104***	(0.059)	1.735***	(0.184)
d_{Nov11}	1.229***	(0.057)	1.638***	(0.186)
d_{Dec11}	1.148***	(0.060)	1.990***	(0.188)

Note: The regression includes app fixed effects. ***, ** and * indicate significance at the 1%, 5% and 10% levels.

Table 2.6: Estimation Results for $\tilde{\beta}_{jt}^{app}$

Variable	Estimate	S.E.
d_{iphone}	1.347***	(0.360)
log(Income)	23.193**	(9.578)
Age	-6.469**	(3.145)
Constant	-19.164	(12.590)
Number of obs	28	
Adj R-squared	0.892	

Note: ***, ** and * indicate significance at the 1%, 5% and 10% levels.

Table 2.7: Smartphone Demand Estimation Results

	IV Logit				IV Nested Logit	
Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Price ($100)	-0.052***	-0.673***	-0.482***	-0.670***	-0.642***	-0.646***
	(0.015)	(0.069)	(0.072)	(0.069)	(0.063)	(0.065)
Screen size (inches)	1.154***	1.452***	3.136***	1.445***	0.509**	0.469*
	(0.143)	(0.282)	(0.433)	(0.282)	(0.214)	(0.303)
RAM (100 MB)	-0.118*	0.020	-0.067	0.019	0.141*	0.170
	(0.074)	(0.088)	(0.093)	(0.088)	(0.083)	(0.176)
Processor speed (100 MHz)	-0.202***	-0.077*	-0.332***	-0.076*	-0.010	-0.032
	(0.045)	(0.044)	(0.053)	(0.044)	(0.044)	(0.127)
Camera resolution (Mpixels)	0.555***	0.681***	0.642***	0.681***	0.518***	0.511***
	(0.030)	(0.051)	(0.048)	(0.051)	(0.044)	(0.064)
Signal strength	0.108***	0.068*	0.068*	0.071*	0.070**	0.070**
	(0.038)	(0.039)	(0.038)	(0.039)	(0.032)	(0.032)
Mean Income						
$d_{iphone} * log(Income)$		40.079***	40.066***	40.744***	32.148***	32.374***
		(10.979)	(10.915)	(11.039)	(8.132)	(8.469)
$d_{android} * log(Income)$		39.891***	39.848***	40.561***	31.985***	32.215***
		(10.977)	(10.913)	(11.038)	(8.130)	(8.475)
$d_{blackberry} * log(Income)$		39.987***	39.990***	40.752***	32.033***	32.264***
		(10.989)	(10.926)	(11.059)	(8.150)	(8.493)
$\{log(Income)\}^2$		-1.791***	-1.790***	-1.825***	-1.446***	-1.456***
		(0.491)	(0.488)	(0.494)	(0.364)	(0.379)
Mean Age						
$d_{iphone} * Age$		-0.011	-0.011	0.255*	0.302**	0.301**
		(0.077)	(0.076)	(0.138)	(0.131)	(0.131)
$d_{android} * Age$		-0.002	-0.002	0.263*	0.303**	0.302**
		(0.078)	(0.078)	(0.139)	(0.130)	(0.131)
$d_{blackberry} * Age$		-0.012	-0.012	0.226*	0.271**	0.270**
		(0.076)	(0.075)	(0.128)	(0.121)	(0.122)
Age^2		0.000	0.000	-0.003*	-0.004**	-0.004**
		(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
App utility						
Common			-2.212***			
			(0.280)			
Demo-type specific				0.012***	0.010***	0.010***
				(0.005)	(0.004)	(0.004)
log(Number of apps)						0.052
						(0.272)
Number of obs	7056	7056	7056	7056	7056	7056
R-squared	0.242	0.264	0.270	0.265	0.540	0.537

Note: Numbers in parenthesis represent standard errors for the estimates. ***, ** and * indicate significance at the 1%, 5% and 10% levels.

Table 2.8: Counterfactual Experiment

	Android (A)	Counterfactual (B)	iPhone (C)	B-A	C-B	C-A
2010.2Q	18.3	28.4	58.6	10.1	30.2	40.4
2010.4Q	18.2	16.1	43.5	-2.1	27.4	25.3
2011.1Q	27.1	34.8	42.0	7.7	7.2	14.9
2011.2Q	25.9	30.1	40.9	4.2	10.8	15.0
2011.4Q	35.2	22.8	33.5	-12.4	10.6	-1.7

Note: (A) and (C) columns report the actual numbers of app purchases per phone for Android and iPhone users, respectively. (B) column is for a hypothetical situation where Android users are given the iPhone.

Figure 2.1: Variation in Smartphone Characteristics

(a) Price

(b) Processor Speed

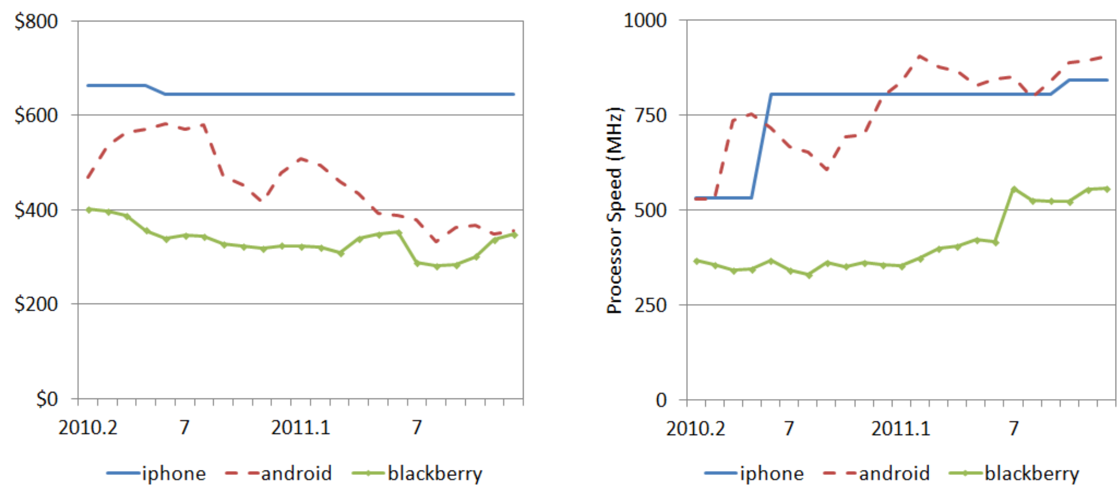


Figure 2.2: Google Trends (keyword = iphone)

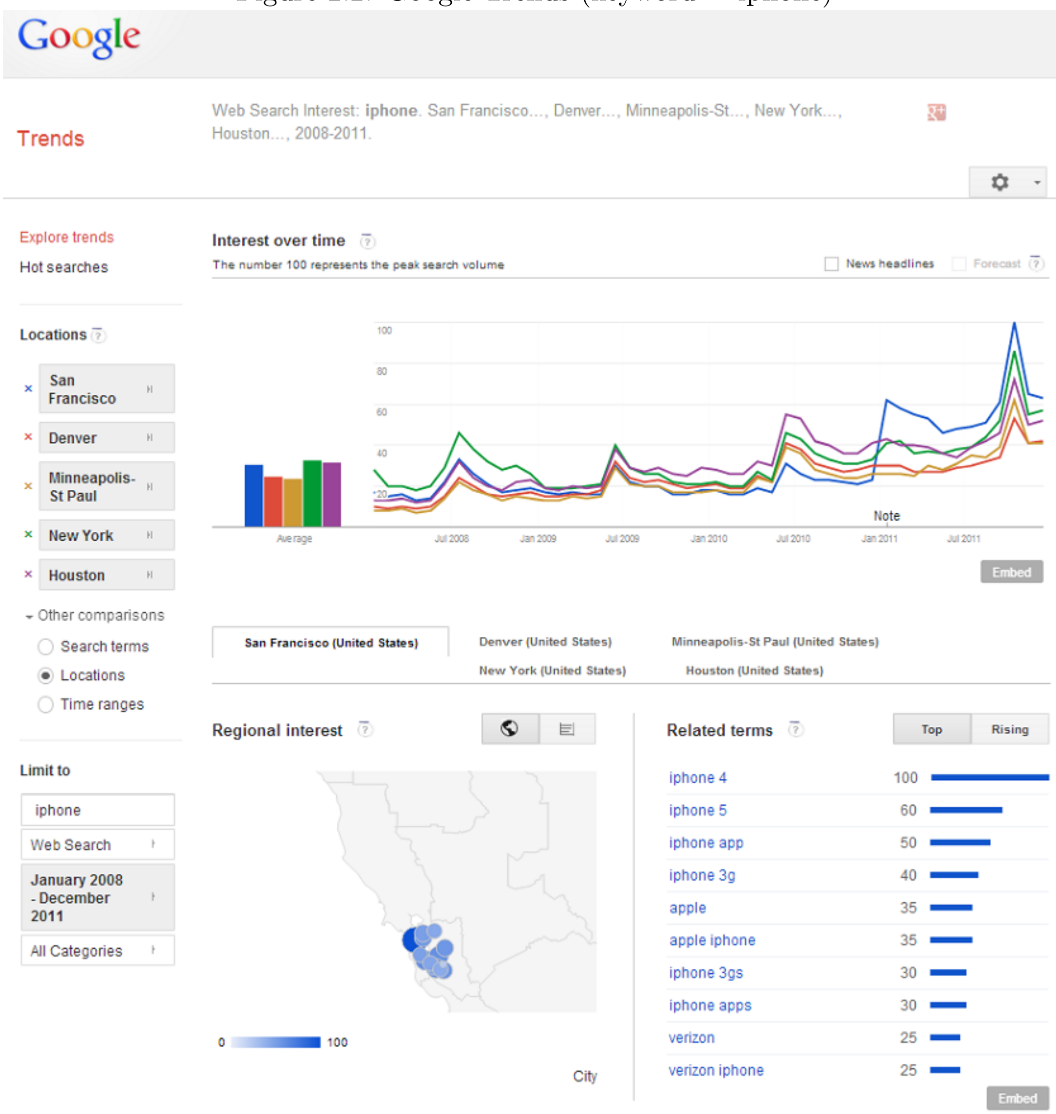
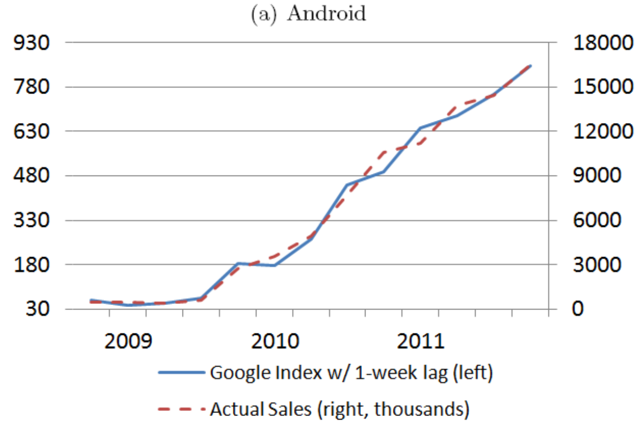
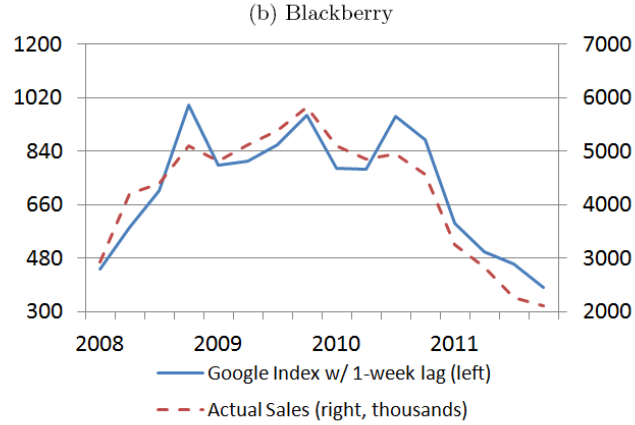


Figure 2.3: Google Search Volume vs. Actual Quantity (1)

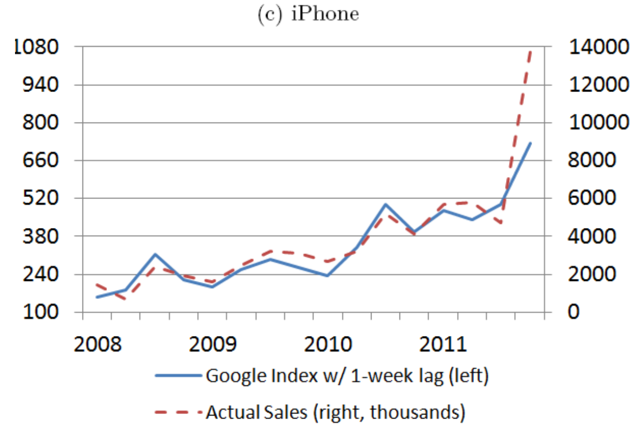
(a) Android



(b) Blackberry



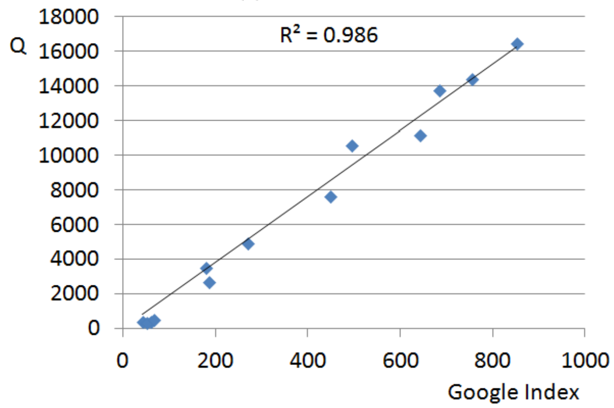
(c) iPhone



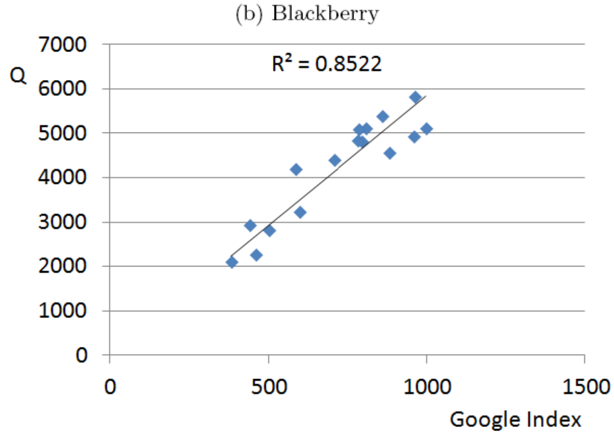
Note: Google search index is not comparable across smartphones since it is normalized differently for each search term.

Figure 2.4: Google Search Volume vs. Actual Quantity (2)

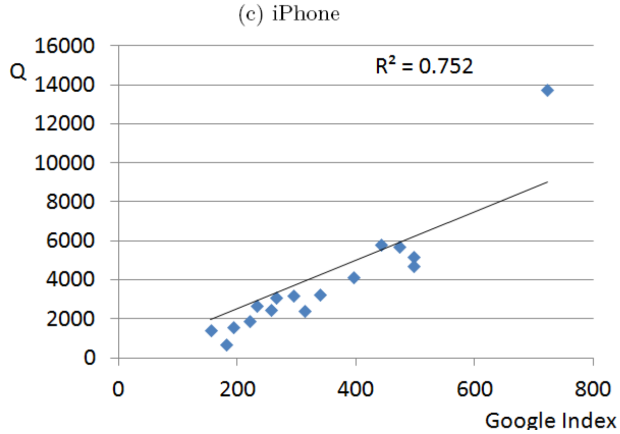
(a) Android



(b) Blackberry



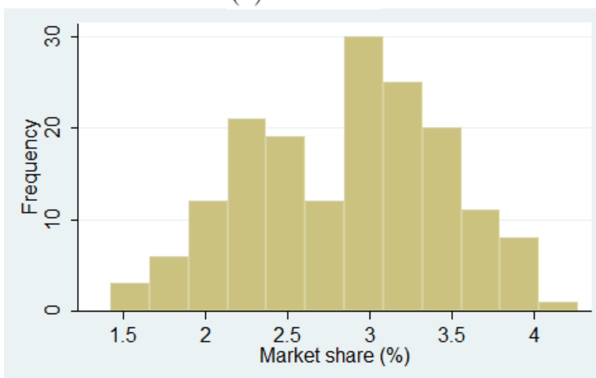
(c) iPhone



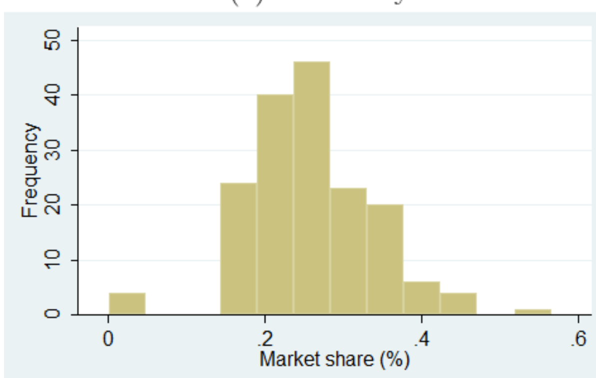
Note: Google search index is not comparable across smartphones since it is normalized differently for each search term.

Figure 2.5: Cross-sectional Variation in OS Shares

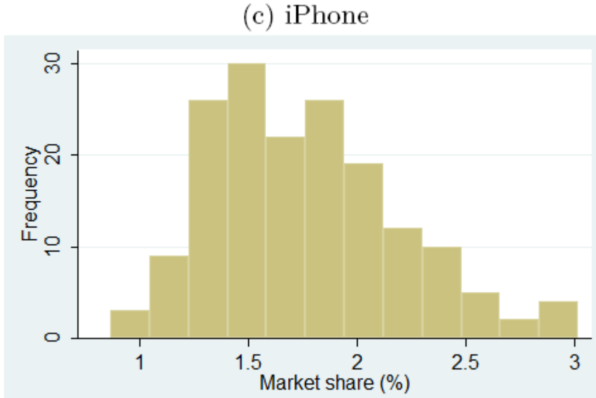
(a) Android



(b) Blackberry

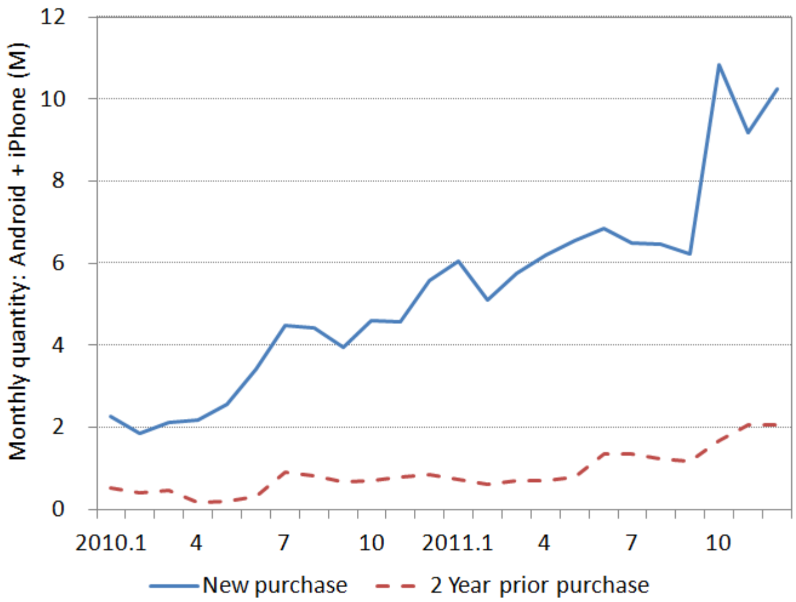


(c) iPhone



Note: The distributions are for December 2011.

Figure 2.6: Comparison of Current Sales and 2-Year Prior Sales



Note: The number of consumers who purchased smartphones two years ago should be a good proxy for replacement demand. The ratio between the two remained at 16.4% on average during 2010-2011.

Figure 2.7: $\tilde{\beta}_{jt}^{app}$ and Demographics

(a) $\tilde{\beta}_{jt}^{app}$ (b) $log(Income)_{jt}$

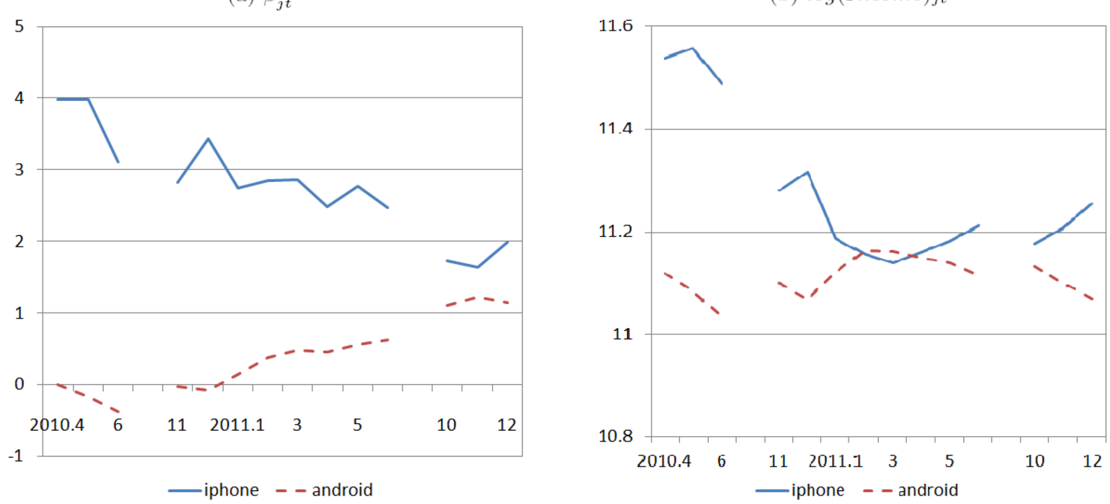


Figure 2.8: Log-Odds Ratio: Smartphone Quality vs. App Benefit

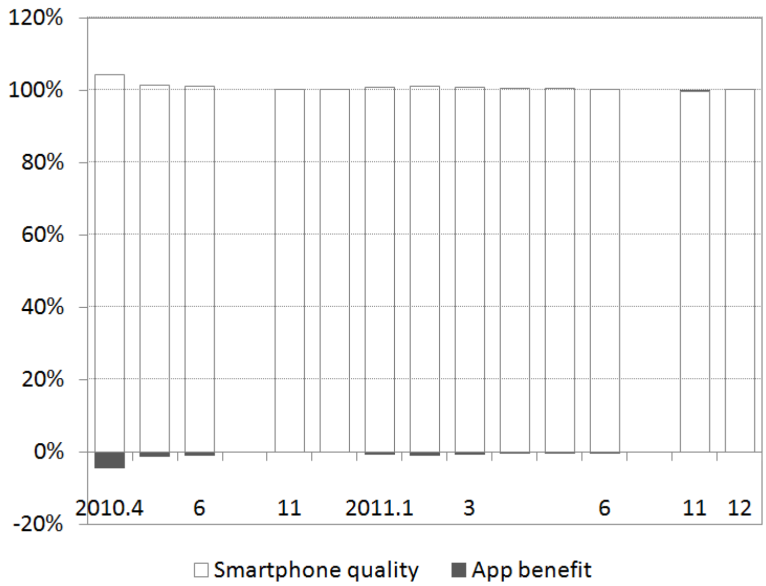
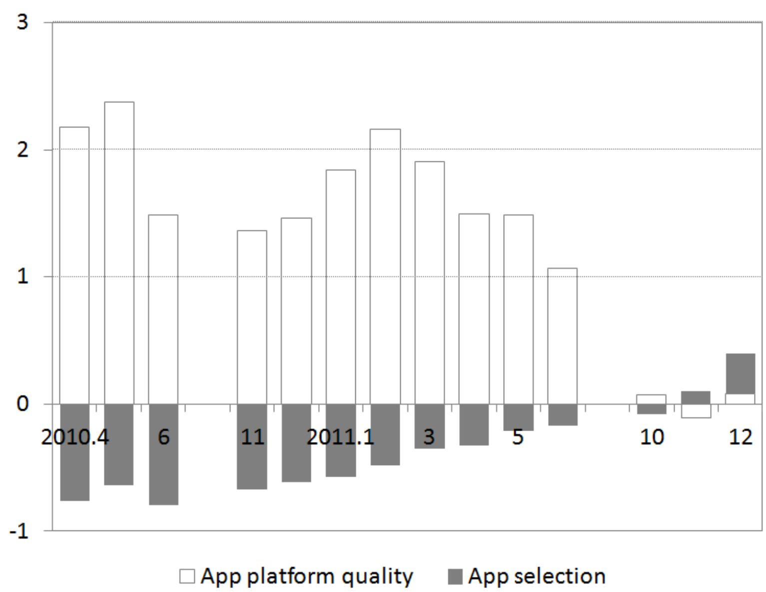


Figure 2.9: Composition of the Differences in the App Benefit (iPhone - Android)



References

[1] Shane Greenstein and Ryan C. McDevitt. The broadband bonus: Accounting for broadband internet's impact on u.s. gdp. Working Paper 14758, National Bureau of Economic Research, February 2009.

[2] Judith Chevalier and Austan Goolsbee. Measuring prices and price competition online: Amazon.com and barnesandnoble.com. *Quantitative Marketing and Economics*, 1:203–222, 2003.

[3] Erik Brynjolfsson, Michael D. Smith, and Yu (Jeffrey) Hu. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49:2003, 2003.

[4] Anindya Ghose and Arun Sundararajan. Evaluating pricing strategy using ecommerce data: Evidence and estimation challenges. *Statistical Science*, 21:131–142, 2006.

[5] Judith A. Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, August 2006.

[6] Jerry A. Hausman. Valuing the effect of regulation on new services in telecommunications. *Brookings Papers on Economic Activity*, 28(1997 Microeconomics):1–54, 1997.

[7] Amil Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729, August 2002.

[8] Austan Goolsbee and Amil Petrin. The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica*, 72(2):351–381, March 2004.

[9] Aviv Nevo. New products, quality changes, and welfare measures computed from estimated demand systems. *The Review of Economics and Statistics*, 85(2):266–275, 2003.

[10] Austan Goolsbee and Peter J. Klenow. Valuing consumer products by the time spent using them: An application to the internet. *American Economic Review*, 96(2):108–113, September 2006.

[11] Erik Brynjolfsson and JooHee Oh. The attention economy: Measuring the value of free digital services on the internet. In *ICIS*, 2012.

[12] Rajiv Garg and Rahul Telang. Inferring app demand from publicly available data. *MIS Quarterly (Forthcoming)*, May 1, 2012.

[13] Robin S. Lee. Vertical integration and exclusivity in platform and two-sided markets. Working Paper, 2012.

[14] Harikesh S. Nair. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics*, 5(3):239–292, 2007.

[15] Brett Danaher, Yan Huang, Michael D Smith, and Rahul Telang. An empirical analysis of digital music bundling strategies, 2011. Working Paper.

[16] Kenneth A. Small and Harvey Rosen. Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–30, 1981.

[17] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–90, July 1995.

[18] Steven T. Berry. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262, Summer 1994.

[19] Harikesh Nair, Pradeep K. Chintagunta, and Jean-Pierre H. Dube. Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics*, 2(1):23–58, 03 2004.

[20] Ting Zhu, Hongju Liu, and Pradeep K. Chintagunta. Wireless carriers exclusive handset arrangements: An empirical look at the iphone. Working Papers 11-35, NET Institute, 2011.

[21] Neil Gandal, Michael Kende, and Rafael Rob. The dynamics of technological adoption in hardware/software systems: The case of compact disc players. *RAND Journal of Economics*, 31(1):43–61, 2000.

[22] Matthew T. Clements and Hiroshi Ohashi. Indirect network effects and the product cycle: Video games in the u.s., 1994-2002. *Journal of Industrial Economics*, 53(4):515–542, 2005.

[23] Kenneth S. Corts and Mara Lederman. Software exclusivity and the scope of indirect network effects in the u.s. home video game market. *International Journal of Industrial Organization*, 27(2):121–136, March 2009.

[24] Jean-Pierre H. Dube, Gunter J. Hitsch, and Pradeep K. Chintagunta. Tipping and concentration in markets with indirect network effects. *Marketing Science*, 29(2):216–249, 2010.

[25] Michael Sinkinson. Pricing and entry incentives with exclusive contracts: Evidence from smartphones. Working Paper, 2011.

[26] Yutec Sun. The value of branding in two-sided platforms. Working Paper, 2012.

[27] Hyunyoung Choi and Hal Varian. Predicting the present with google trends. *The Economic Record*, 88(s1):2–9, 06 2012.

[28] Paul Klemperer. Markets with consumer switching costs. *The Quarterly Journal of Economics*, 102(2):375–394, May 1987.

[29] Joseph Farrell and Paul Klemperer. Coordination and Lock-In: Competition with switching costs and network effects. In M. Armstrong and R. Porter, editors, *Handbook of Industrial Organization*, volume 3, pages 1967–2072. Elsevier, 2007.

[30] Gregory S. Crawford and Ali Yurukoglu. The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–85, April 2012.

[31] Hiroshi Ohashi. The role of network effects in the us vcr market, 1978-1986. *Journal of Economics & Management Strategy*, 12(4):447–494, December 2003.

Appendix A

Smartphone/App Preferences and Demographics

There are many survey results that report how demographic characteristics of users vary across different smartphone operating systems or how the market shares of each OS vary across different demographic groups. For example, figure A.1 displays how the preferences for each OS are related to income and age.[1] iPhone users tend to be richer than Android users (40% of iPhone users have income of over $100,000 while just 28% of Android users fall within the same demographics), and Android and iPhone are more popular among younger consumers.

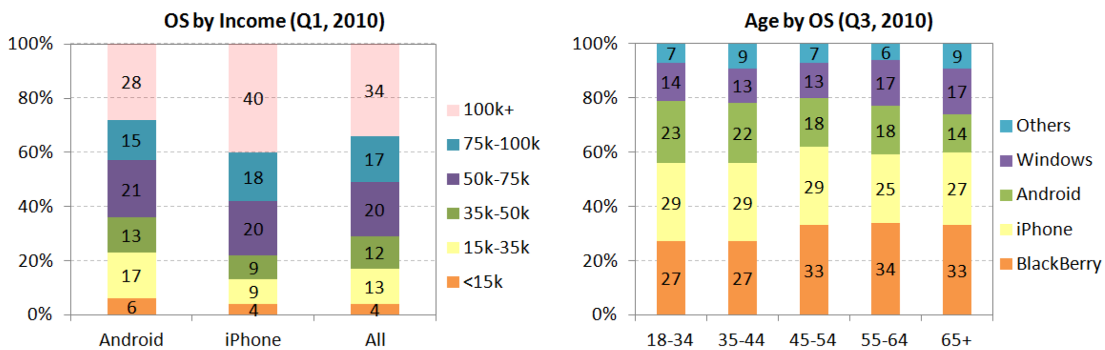
In addition, table A.1 show the percentage of cell phone owners who download apps for different income and age groups.[2] The propensity of downloading apps increases as consumers get richer but it sharply decreases as consumers get older.

[1] Available at http://blog.nielsen.com/nielsenwire/online_mobile/iphone-vs-android and http://blog.nielsen.com/nielsenwire/online_mobile/mobile-snapshot-smartphones-now-28-of-u-s-cellphone-market

[2] Available at http://www.pewinternet.org/Reports/2010/The-Rise-of-Apps-Culture.aspx and http://pewinternet.org/Reports/2011/Apps-update.aspx

Figure A.1: Preferences for OS and Demographics

(a) Income (b) Age



Source: The Nielsen Company

Table A.1: Percentage of cell phone owners who download apps

	Income			Age		
	<50k	50k-75k	75k+	18-29	30-49	50+
May 2010	27	29	38	52	31	11
Aug 2011	33	38	55	60	46	15

Source: Pew Research Center