The Dissertation Committee for Sandeep Arora certifies that this is the approved version of the following dissertation:

TO FREE, OR NOT TO FREE: THE IMPACT OF FREE VERSIONS, AVERAGE USER RATINGS, AND APP CHARACTERISTICS ON THE ADOPTION SPEED OF PAID MOBILE APPS

Committee:

Vijay Mahajan, Supervisor

Frenkel ter Hofstede, Co-Supervisor

Prabhudev Konana

Jason Duan

Ty Henderson

TO FREE, OR NOT TO FREE: THE IMPACT OF FREE VERSIONS, AVERAGE USER RATINGS, AND APP CHARACTERISTICS ON THE ADOPTION SPEED OF PAID MOBILE APPS

by

Sandeep Arora, B.Tech.; PGD

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2014

Acknowledgements

I am deeply grateful for the training and guidance that I received from my advisors, Vijay Mahajan and Frenkel ter Hofstede. Without their unconditional support, I would not have been able to accomplish what I have so far.

I also thank my committee members Jason Duan, Prabhudev Konana, and Ty Henderson for their suggestions, insights, and encouragement.

TO FREE, OR NOT TO FREE: THE IMPACT OF FREE VERSIONS, AVERAGE USER RATINGS, AND APP CHARACTERISTICS ON THE ADOPTION SPEED OF PAID MOBILE APPS

Sandeep Arora, Ph.D.

The University of Texas at Austin, 2014

Supervisors: Vijay Mahajan and Frenkel ter Hofstede

The mobile application (App) industry has grown tremendously over the past five years, primarily fueled by small App development businesses. Lacking advertising budgets, these relatively unknown, small businesses often offer free versions of their paid Apps to reduce customer uncertainty about App quality and get noticed in the crowded App industry. In this research I investigate the implications of offering free versions on the adoption speed of paid Apps by building on the existing marketing and information systems literature on sampling and versioning. Using a unique dataset of 2.82 million observations from 4,180 Apps and accounting for endogeneity, I find that while the strategy of offering free versions of paid Apps is popular, it impacts the adoption speed of paid Apps negatively. I also find that the presence of free versions has a larger negative impact on the adoption speed of Apps bought for fun and pleasure (hedonic Apps) and in the later life stages of paid Apps. I expect that the results of my study will enable

App developers to make informed decisions about offering free versions of paid Apps and prompt academicians to produce more work focusing on this industry.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION

Background

The App industry has grown from zero revenue and jobs in 2008, to $25 billion and 466,000 jobs by 2011 and it continues to grow rapidly (Mandel 2012). With more than 1600 new Apps being added every day to the existing 2 million, analysts expect that the impact of the App industry will grow to $75 billion by 2015 (Jahns, Mikalajunaite, and Meehan 2011). In addition to its direct impact on the economy, the App industry has a significant impact on the $219 billion smartphone industry (Wortham and Wingfield 2012). Moreover, recent projections show that mobile devices will become the preferred mode of accessing the Internet by 2015 (Weber 2013). Since Apps account for 80% of mobile Internet access time (Accenture 2012), the App industry holds significance for most businesses.

Many small businesses have ventured into App development because the App industry has low entry barriers and provides easy access to a large customer base. In fact, more than 78 percent of prominent App developers are small businesses (Godfrey, Reed, and Herndon 2012). The proliferation of small business entrants in the App development space is the primary reason for the rapid growth of the App industry. But this growth has also created some challenges for App developers and customers. App developers with their slim marketing budgets struggle to get their Apps noticed in a crowded industry where multiple Apps are launched every day (Abecassis 2012). Also, these Apps are distributed through App stores (such as Google Play store, iTunes App store, etc.), which do not have stringent quality control systems for screening new Apps. As a result, a large number of Apps in App stores have repetitive content, contain malware, or infringe on customer privacy (D'Orazio 2013; Perez 2013). Hence there is a high degree of uncertainty among customers regarding App quality. Finally, the revenue distribution in the App industry is skewed. For example, 4 Apps on Google Play store and 7 Apps on iTunes App store (out of 700,000 Apps in each store) account for 10% of the revenue of these stores (Spriensma 2012), and more than 60% of Apps fail to break even[1].

To encourage App discovery and reduce customer uncertainty, a popular strategy adopted by App developers is to offer free Apps- 79% of Apps on Google Play store and 59% of Apps offered on iTunes App Store are free[2][3]. A large proportion of these free Apps are offered as samples of full featured paid Apps. App developers adopting the strategy of offering free versions of paid Apps hope that a significant proportion of customers who install the free version will buy its paid version and may also recommend the App[4]. Despite the popularity of this strategy, there is a dearth of guidance available for App developers about its effectiveness as the App industry has received limited attention by researchers.

Marketing scholars have produced a rich stream of work on sampling, which is the practice of offering free samples of paid products. However, sampling literature focuses on consumer goods which differ from information goods (such as Apps) in their characteristics (Jones and Mendelson 2011) and sampling practices (Wang and Zhang 2009). Hence I argue that the knowledge generated by sampling literature, while relevant for consumer goods, may offer limited guidance to managers looking to offer free versions of information goods. Another stream of literature close to my area of investigation focuses on the practice of offering multiple versions of information goods, or versioning. In some recent work done on versioning, operations and information systems scholars investigate the implications of offering free versions

[1] http://app-promo.com/wake-up-call-infographic/ accessed on June 10, 2013.
[2] http://www.appbrain.com/stats/free-and-paid-android-applications accessed on June 10, 2013.
[3] http://148apps.biz/app-store-metrics/?mpage=appprice accessed on June 10, 2013.
[4] In a survey of 110 App developers, that I conducted in February 2014, 82% listed 'promotion of paid App' as a reason for offering free version of their paid App.

of paid computer software. Since Apps are small, specialized software, I expect that the findings of versioning literature apply in the context of Apps. Hence I build my framework with inputs from versioning literature. However, based on my literature review (Table 1), I see the following opportunities to contribute to both versioning and sampling literature:

----- Insert Table 1 about here -----

First, both sampling and versioning literature lack large scale empirical studies as most of the existing research focuses on constructing analytical models to serve as decision tools for managers looking to offer free versions. Secondly, a free version is only one of the many sources of information about the quality of a paid product (Shapiro 1983). Existing literature does not explore the role of the free version in the presence of other information sources of paid product quality. Thirdly, unlike consumer goods, most developers of information goods offer free versions throughout the life of their corresponding paid products. However, current literature does not investigate how the impact of offering free versions on the performance of paid products changes across their life. Finally, despite its increased size and significance, there is a lack of research focusing on the App industry.

I attempt to plug these gaps in the existing literature by conducting a large scale empirical study to investigate the impact of offering free versions on the performance of paid Apps. I also investigate how, apart from the free version, two additional information sources of paid App quality- average user rating of a paid App (henceforth referred to as 'paid App average user rating') and developer reputation- impact paid App performance. I further explore how the impact of these information sources on paid App performance change across the different life stages and characteristics of a paid App (Figure 1a). More specifically, I address the following questions in this research:

- What is the impact of offering a free version on the performance of its paid App?
- In what conditions is it beneficial for App developers to offer free version of a paid App?
- How does the impact of offering a free version, paid App average user rating, and developer reputation on the performance of a paid App change across its characteristics and life stages?

----- Insert Figure 1a about here -----

I use a dataset containing 2.82 million observations from 4,180 Apps offered on Google Play store for my analysis. I compile this dataset from three different sources as there is no single source that offers data required for my analysis. Hence, I introduce a unique dataset which has not been analyzed by marketing scholars till now. I use these data to model the impact of offering a free version, developer reputation, and paid App average user rating on the adoption speed of paid Apps. I estimate complementary log-log proportional hazard models at different life stages of paid Apps to investigate the change in the impact of these information sources on the adoption speed of paid Apps across their life. I include a random intercept in my model to account for App specific unobserved heterogeneity. I also account for endogeneity in my analysis using Coarsened Exact Matching (CEM) technique (Iacus, King, and Porro 2011, 2012). My analysis provides several interesting results.

My results show that despite the popularity of the practice of offering free versions of paid Apps, it has a negative impact on the adoption speed of paid Apps. I further find that this negative impact changes across the life and characteristics of paid Apps. The presence of a free version has a larger negative impact on the adoption speed of Apps bought for fun and pleasure (hedonic Apps) and in the later life stages of a paid App. I also find that high paid App average user rating and developer reputation have positive impacts on paid App adoption speed, but these impacts are smaller for hedonic Apps. Finally, I observe that these positive impacts of paid App average user rating and developer reputation on paid App adoption speed move in opposite directions across the life of a paid App. I find that early in the life of a paid App, developer reputation has a larger positive impact and paid App average user rating has a smaller positive impact on its adoption speed. However, towards the later life stages of a paid App, the positive impact of developer reputation becomes smaller whereas the positive impact of paid App average user rating on its adoption speed increases.

My findings have managerial and theoretical implications. As per my knowledge, this is a pioneering work focusing on the fast growing App industry and I hope that it would attract the attention of other academicians to produce work in this area. My work also contributes to sampling literature, which is devoid of any large scale empirical work. My data suggest that App developers are not very strategic in offering free versions and that many developers are simply following the current norm. Hence I anticipate that the results of my research will enable App development businesses to refine their free version strategies.

Organization of the Dissertation

The remainder of the dissertation is organized as follows. In chapter two, I present the theory that I use to construct my hypotheses and then formally state the hypotheses. As the data for this project were compiled from multiple sources and this is the first research using such data, I dedicate the entire chapter three to data collection and collation. In chapter four, I describe the methodology used to account for endogeneity and to test my hypotheses. The results of my analyses are summarized in chapter five. I conclude this dissertation with a discussion of its contributions to marketing theory and managerial practice, its limitations, and opportunities for further research in chapter six.

CHAPTER TWO: THEORY AND HYPOTHESES

With more than 1600 new Apps released by small businesses every day, App developers face two challenges: how to boost App discovery and reduce customer uncertainty about App quality. Classic work done on signaling theory deals with issues of high and low quality actors in the market. Signaling theorists suggest that producers have complete information about the quality of their products but prospective customers have limited information about product quality, which creates information asymmetry in the market (Akerlof 1970; Spence 1973). Producers can bridge this information asymmetry by advertising, leveraging their reputation, offering free trials or warranties, or by relying on third party product evaluations (including word-of-mouth) (Shapiro 1983). However, most of these modes for disclosing product quality information are missing for Apps. While the number of Apps has grown exponentially in the last five years, most of these Apps have not been evaluated by third parties, thus limiting external information sources of App quality. Most Apps are developed by small, relatively unknown developers; hence customers cannot rely on developer reputation to discern the quality of such Apps. Most App developers do not have advertising budgets to increase product awareness and provide information about App quality to prospective customers. The information sources of App quality are even smaller for newly-released Apps, as no user ratings are available for them. In the absence of other information sources of App quality, a large number of developers offer free versions of their paid Apps to raise product awareness and reduce uncertainty about the quality of their paid Apps.

Alternative Monetization Strategies

A discerning reader might argue that developers offering free versions of paid Apps may be pursuing alternative monetization strategies, namely – in-App purchase (IAP) and/ or in-App advertising. IAP is an extremely popular monetization strategy for a class of Apps which are categorized as ‘freemium’. While customers could install a ‘freemium’ App for free, they have to pay to use certain advanced features of the App. These ‘freemium’ Apps only have free versions (and no accompanying paid versions) available in an App store. My analysis focuses on paid Apps with (and without) distinct free versions available in an App store, so ‘freemium’ Apps are excluded from my analysis. I checked for the presence of IAP in the free version of paid Apps in my data and as expected, only 6% free versions contain IAP. In-App advertising is another monetization strategy used by App developers. However, unlike IAP, it is not possible to identify presence of in-App advertising without decompiling the source files of Apps or manually examining all Apps, which is a time consuming process. Hence, in February 2014, I surveyed 110 App developers and asked them about the reasons behind offering free versions of paid Apps. 82% of App developers listed ‘promotion of paid Apps’ whereas only 17% listed ‘advertising revenue’ as a reason for offering free versions of paid Apps. In light of these results, I am confident that a majority of free versions of paid Apps are offered as a trial of paid Apps and not for generating revenue through IAP or in-App advertising. I now hypothesize about the impact of offering free versions of paid Apps.

Impact of Offering Free Version

Offering free versions of products to raise awareness and reduce customer’s uncertainty about their quality is not a novel concept. Producers of consumer goods have been offering free samples of their products (sampling) since long and this continues to be a popular practice (Bawa and Shoemaker 2004). Firms could use sampling to change product image (Hamm, Perry, and Wynn 1969), generate word of mouth (Holmes and Lett 1977), and boost immediate sales (Heiman et al. 2001; Heiman and Muller 1996). However, scholars show that the results of sampling may not always be desirable for a producer. Jain, Mahajan, and Muller (1995) show that sampling might have a negative impact on product profitability if the proportion of potential customers who are offered samples is above an optimum level. Bawa and Shoemaker (2004) show that offering samples will cannibalize the sales of products with low repurchase rate.

However, the entire sampling literature focuses on consumer goods, which are different from information goods (such as Apps) on the following parameters:

- *Repurchase rate:* Consumer goods are typically purchased multiple times in the lifetime of a customer. Evidence from existing literature on sampling of consumer goods suggests that offering samples cannibalizes the sales of products with low repurchase rate, but the benefits of sampling increase with increase in the repurchase rate of products (Bawa and Shoemaker 2004). However, most information goods are only purchased once, and thus have unit demand. The role of samples for goods which are only purchased once, thereby having a zero repurchase rate, is an unexplored area in existing marketing literature. I note that the sparse marketing literature on demonstration of consumer durables (Heiman and Muller 1996) focuses on products which have low repurchase rates. However as these demonstrations are time bound, this practice is similar to offering time limited trials of information goods, an area I discuss later.

- *Marginal cost:* The marginal cost of producing and distributing free versions of information goods is zero, whereas samples of consumer goods cost money to produce and distribute. Hence, the profitability impact of offering a free version may be different for consumer and information goods.
- *Availability of free version:*
    - *Breadth:* Free versions of information goods are available to all customers, whereas samples of consumer goods are targeted towards a subset of customers. Existing marketing literature provides guidance on the benefits of targeted sampling and the optimum level of sampling (Jain, Mahajan, and Muller 1995). In the context of Apps, however, App developers do not have control over which and how many customers download a free version.
    - *Length:* Free versions of information goods are generally offered throughout the life of their paid versions, which is rare for sampling of consumer goods. The impact of free version presence on the performance of a paid App might change across its life, which is an unexplored area in existing literature.
- *Longevity of free version:* Unlike consumer goods, information goods are non-consumables. Hence the free version of an App, once downloaded by a customer, may continue to exist indefinitely (unless it is a time limited trial, an option I discuss later). Scholars show that the benefits of offering samples increase with the number of purchase occasions since the time of sample delivery (Bawa and Shoemaker 2004). In the context of Apps, the impact of free version presence on the performance of a paid App is unclear as a free version is not perishable.

- *Quality of free version:* Free versions of information goods differ from the full, paid product in quality as they have less features than the paid product (unless it is a full-featured, time limited trial), whereas samples of consumer goods differ from the full product in quantity. Scholars show that samples reduce customer uncertainty by revealing the true quality of a product. However, the role played by free versions which are inferior in quality to their paid versions is an area that has not been addressed in sampling literature.

I argue that the above differences between the characteristics and sampling practices of consumer and information goods imply that the findings of sampling literature, while relevant for consumer goods, may offer limited guidance to managers looking to offer free versions of information goods.

Scholars in operations and information systems investigate the impact of offering free versions of information goods on the performance of paid versions as a special case in versioning literature. Versioning has been shown to be a sub-optimal strategy for information goods as they have negligible marginal cost (Bhargava and Choudhary 2001; Jones and Mendelson 2011). Scholars show that if the marginal costs of low and high quality versions offered by a monopolist are the same, the benefits of offering a low quality version of a product are dominated by the resulting cannibalization of its high quality version. Despite this existing academic evidence about the undesirable effects of versioning, producers of information goods such as computer software continue to offer free versions of their paid products (Whinston, Stahl, and Choi 1997).

Operations and information systems scholars investigate this anomaly between academic wisdom and market practices to explore conditions in which it is profitable for firms to offer free versions of information goods. They note that free versions of information goods primarily occur in two forms: free version having fewer features than the paid version, but with no time restriction (feature limited free version); and free version having the same features as the paid version, but with time restriction (time limited trial). These scholars argue that the choice of feature limited free version or time limited trial will depend on whether an information good gains value when more individuals use it (positive network effects) (Cheng and Liu 2012; Faugère and Tayi 2007). They propose that offering a feature limited free version will be more profitable in the presence of positive network effects, as a firm will be able to capture network effects from the users of both free and paid versions. In the absence of positive network effects, the primary role of free versions is to reduce customer uncertainty about the quality of information goods. Since a time limited trial has all the features of the paid product, it will be more effective in allaying customer uncertainty about product quality than a feature limited free version, without cannibalizing the sales of the paid version due to the time restrictions associated with it. Hence, for information goods that do not have positive network effects, offering a time limited trial will be more profitable rather than offering a feature limited free version (Cheng and Liu 2012). In the App industry, most Apps (other than the ones categorized under the sub-category 'social') do not have positive network effects. This is unlike computer software products, for which positive network is extremely important (Cheng and Tang 2010). Despite the fact that most Apps lack strong network effects, 99% of free versions offered in conjunction with paid Apps are feature limited. In such a scenario, I expect that offering a feature limited free version will have a negative impact on its paid App performance.

Dey and Lahiri (2013) explore the role of customer uncertainty about a product's usefulness in determining optimal versioning strategy for software. They show that even when the marginal costs of all versions are zero, versioning expands the market when uncertainty is low. They further show that when customer uncertainty about a product's usefulness is high, versioning does not expand the market, but producers could charge hefty price premiums for upgrading customers, making versioning desirable. A common assumption that information systems scholars make in versioning literature is that computer software are offered by monopolists. The argument given for this assumption is that the nature of computer software allows their producers to act as monopolists, even if there are multiple players in the market (Dey, Lahiri, and Liu 2013). In consequence, these monopolists may be able to increase the prices of their products without worrying about competition, thus charging hefty price premiums from prospective customers.

In the App industry, a majority of Apps have been developed by small businesses with unknown reputations (Godfrey, Reed, and Herndon 2012). This, coupled with lack of stringent vetting of Apps by App stores, has resulted in the presence of a large number of low quality Apps in App stores (D'Orazio 2013; Perez 2013). Hence there is high customer uncertainty about App quality. However, unlike the computer software industry, there is a high degree of competition in the App industry where a large number of undifferentiated Apps are sold through a small number of App stores. Consequently, App developers are unable to charge hefty price premiums on their paid Apps. Hence, I expect that offering free versions of paid Apps will have a negative impact on the performance of paid Apps.

I use the duration taken by paid Apps to reach certain installation or adoption levels as a measure of their performance. I describe my performance measure in detail in the 'Data and Measures' section of this dissertation. Also, since 99% of the free versions in my sample are feature limited free versions, all my hypotheses should be read in the context of feature limited free versions. For brevity, I use the term ‘free version’ instead of ‘feature limited free version’ from here on. Hence I hypothesize that,

H_1 Paid Apps offered in conjunction with free versions will have lower adoption speed than paid Apps without free versions

Impact of Developer Reputation

In addition to free versions, prospective customers could acquire additional information about the quality of paid Apps based on the reputation of the developers of these Apps. Marketing scholars have shown that strong brands help in faster market penetration (Aaker 1991), command greater brand loyalty (Russell and Kamakura 1994), and could boost firm performance (Srivastava, Shervani, and Fahey 1998). I argue that developer reputation will be especially important in the App industry because of the large number of unknown developers and the newness of the industry. The significance of developer reputation will also be high as many traditional sources of product quality information such as advertising and third-party evaluations are not available to customers in App stores. Hence I hypothesize that,

H_2 Paid Apps developed by reputed developers will have a higher adoption speed than paid Apps developed by unknown developers

Impact of Paid App Average User Rating

Scholars show that word-of-mouth or its online counterpart, customer rating, is an important source of information about product quality and has a significant impact on product performance (Duan, Gu, and Whinston 2008). The rise of the Internet has boosted the significance of word-of-mouth as customers now have easy access to product reviews and ratings before they choose a product. Word-of-mouth gains even greater significance when products are distributed through digital channels, as prospective customers are designedly exposed to product reviews and ratings before they buy a product. Scholars have shown that online ratings of products influence product sales (Chevalier and Mayzlin 2006), revenues (Liu 2006), and firm value (Villanueva, Yoo, and Hanssens 2008). Consistent with the evidence about the impact of customer ratings of a product on its performance, I expect that high paid App average user rating will have a positive impact on its adoption speed.

H_3 Paid Apps with higher average user rating will have a higher adoption speed than paid Apps with lower average user rating

Impact across Life Stages of Paid App

A unique characteristic about the App industry is that once App developers launch free versions of their paid Apps, they generally continue to offer them throughout the life of the paid Apps. I expect that the impact of a free version, developer reputation, and paid App average user rating on its adoption speed will change across the life of a paid App.

In the early life stage of a paid App, the reliability of the information conveyed by customer reviews will be low as the number of customer reviews will be small. This life stage of a paid App will also be characterized by high uncertainty about its quality as there are not too many sources of information about its quality. Consequently, a large number of customers will rely on the free version and developer reputation to gain additional information and reduce uncertainty about the quality of a paid App. Hence the negative impact of a free version on paid App adoption speed will be partially offset by its uncertainty reduction role. Also high developer reputation will have a strong positive impact on the adoption speed of a paid App.

However, as more customers install and review a paid App, the number of its ratings will increase and these ratings will become a reliable source of information for determining its quality. Hence in the later life stages of a paid App, the positive impact of high paid App average user rating on its adoption speed will become stronger. An increase in the number of ratings of a paid App will also reduce the uncertainty about its quality. Thus the positive impact of developer reputation on paid App adoption speed will diminish due to this reduced uncertainty about paid App quality and due to an increase in the significance of other sources of information about paid App quality. Also, the uncertainty reduction role of the free version will diminish in the later life stages of a paid App which will be reflected in an increase in the cannibalizing effect of the free version.

Additionally, the type of customers entering the market will also change across the life of a paid App. Scholars have shown that customers who buy a product early in its life are opinion leaders, affluent, and less price-sensitive, whereas late adopters have been shown to be highly influenced by other customers and to be more price sensitive (Brancheau and Wetherbe 1990; Nagle 1987). Hence, I expect that a lower proportion of customers will upgrade from the free version to the paid App in the later stages in the life of a paid App than in its earlier life stages. This will result in a stronger negative impact of free version presence on the adoption speed of a paid App in its later life stages. I further expect that the positive impact of high paid App average user rating on its adoption speed will become stronger in the later stages of its life when late adopters, who have been shown to be swayed by external influence, enter the market. Hence I hypothesize that,

H_4 The positive impact of high developer reputation on the adoption speed of paid Apps will be larger for early adoption levels than for later adoption levels

H_5 The positive impact of high paid App average user rating on the adoption speed of paid Apps will be smaller for early adoption levels than for later adoption levels

H_6 The negative impact of free version presence on the adoption speed of paid Apps will be smaller for early adoption levels than for later adoption levels

In addition to the life stage of a paid App, the impact of a free version on the performance of the paid App will vary across its characteristics, which I discuss below.

Impact across Characteristics of Paid App

Apps offered on Google Play store are broadly categorized into Games and Applications. Games are Apps that are pleasure oriented and are purchased for their fun and enjoyment value, and hence are primarily hedonic. Applications are Apps that are goal oriented and are purchased for their practical and functional attributes, and hence are primarily utilitarian[5]. In this research I focus on this characteristic of paid Apps, i.e., whether a paid App is primarily hedonic (Game) or utilitarian (Application).

Most of the existing work on versioning of information goods assumes that after exposure to a free version, customers update their perceptions about product quality of the paid version in an identical manner. However in some recent work scholars recognize that due to their individual preferences, not all customers will end up with the same impressions about the quality of a paid version after trying its free version (Dey, Lahiri, and Liu 2013). I argue that customer assessment of the quality of a product could be divided into the objective assessment of product quality and subjective assessment of product fit to individual taste. I further posit that even

[5] I note that these hedonic and utilitarian dimensions are not mutually exclusive but they may not be equally salient for all product categories. However, since most product categories are evaluated more positively on one dimension than the other, the overall attitude toward a category will be primarily hedonic or utilitarian (Batra and Ahtola 1991).

though all products will be assessed on objective and subjective dimensions, the significance of these dimensions will vary across products. Scholars show that the value derived from hedonic goods is more subjective and personal than the value derived from utilitarian goods (Hirschman and Holbrook 1982).

Hence I expect that, given their hedonic nature, Apps in the Games category will be evaluated more subjectively based on their fit to individual customer's tastes. Apps in the Applications category, however, will be evaluated more objectively based on their product attributes and quality. I argue that even though prospective customers can get information regarding product attributes and quality from sources such as producer's reputation, sample, word-of-mouth, and third-party evaluations, evaluation of a product's fit to their tastes can only be made from direct exposure to a product[6]. Hence, even though high developer reputation and paid App average user rating will have positive impacts on paid App adoption speed, I hypothesize that this effect will be smaller for paid Games than for paid Applications, as these information sources do not provide direct exposure to the App.

H_7 The positive impact of high developer reputation on the adoption speed of paid Apps will be smaller for paid Games than for paid Applications

H_8 The positive impact of high paid App average user rating on the adoption speed of paid Apps will be smaller for paid Games than for paid Applications

The impact of free version presence on paid App adoption speed across Games and Applications is more complex as there are two counter-forces acting. On the one hand, a free version provides direct exposure to an App, which will help in reducing uncertainty regarding the quality of an App and will let prospective customers determine if an App fits their tastes.

[6] This assumption is confirmed by my data where I see that the variance of rating of paid Games is significantly higher than that of paid Applications when no free versions are offered. However, once free versions are offered for the same paid Apps, the variance of rating of Games becomes significantly lower than that of Applications.

This should in turn boost the performance of paid Games more than paid Applications, as the significance of ‘fit to taste’ is higher for hedonic products than for utilitarian products. On the other hand, the presence of a free version may delay the purchase of the paid App as a proportion of customers may choose to first try the free version and then upgrade to the paid App. I expect that due to the discretionary nature of hedonic products (Okada 2005) and hedonic elements associated with impulse purchases (Rook 1987), a delay between the exposure to a hedonic product (stimulus) and decision to buy (transaction) will have a negative impact on its sales. I expect that this will be reflected in a greater negative impact of free version presence on the adoption speed of paid Games than that of paid Applications.

Scholars also show that compared to utilitarian products, purchases of hedonic products are more difficult to justify (Dhar and Wertenbroch 2000).There is further evidence from existing literature that purchasing hedonic products is associated with a feeling of guilt (Lascu 1991). Hence consumers resort to self-rationing to prevent overconsumption of hedonic products by maintaining smaller inventories and buying smaller package sizes (Wertenbroch 1998). I expect that free versions of Games will reduce customer guilt as getting a product for free will help customers to justify their purchases. Also, as these free versions have fewer features than their paid Apps, they will help customers to ration the time they spend on playing Games. This will be reflected in a greater negative impact of free version presence on the adoption speed of paid Games than that of paid Applications.

Since ex-ante I am not sure if the positive uncertainty reduction impact or the negative cannibalization impact of free versions of paid Games will dominate, I hypothesize the net effect to be in both directions and let my data reveal which one of these hypotheses is supported.

H_{9a} The negative impact of free version presence on the adoption speed of paid Apps will be smaller for paid Games than for paid Applications

H_{9b} The negative impact of free version presence on the adoption speed of paid Apps will be larger for paid Games than for paid Applications

I summarize my hypotheses in the detailed framework given in Figure 2.

----- Insert Figure 2 about here -----

CHAPTER THREE: DATA AND MEASURES

I test my hypotheses using a dataset of more than 2.82 million observations for 4,180 paid Apps sold on Google Play store, which is one of the largest App stores in the world. Even though Google Play store, like other App stores does not reveal historical information about Apps sold on it, I was able to identify and collect this information from three different secondary data sources. I describe the data collection process and characteristics of my data in the following sections.

Data Collection

Each App sold on Google Play store has a dedicated webpage that provides App-specific contemporaneous information such as its description, developer name, current price, user ratings on a 5 point scale, average user rating, user reviews, current version, date of the last update, App category, and a range indicating the number of times the App has been installed. Google Play store is a great source for obtaining contemporaneous information about Apps, but historical information about Apps, which I require for my analysis, is not available in the store. Hence for collecting historical data about Apps, I explored web services which have been collecting daily data from Google Play store. I found three such web services but none of them individually had all the data required for testing my hypotheses. Hence, I used web-scraping to collect historical data about Apps from all three sources. I later combined the data collected from different sources using a Relational Database Management System (RDBMS) to compile a dataset of more than 460,000 Apps (> 90% of total Apps available in Google Play store at that time), which has all the variables required for my analysis.

I collected my data on 13th November 2012, and for my current analysis, ignored all the Apps which were launched after 10th January 2012[7]. I did this to ensure that I have enough time-varying data for each App and that Apps in my dataset had enough time to reach at least 500 installations, which is the first adoption level in my data. Since I collected data from secondary sources using web-scraping, I carefully reviewed my data for any inaccuracies or missing information. Around 79% of Apps in my database are free, which is similar to the estimate of the proportion of free Apps (79%) suggested by other available sources. I excluded the Apps which did not have all the information required for my analysis. For my current project, I identified a dataset with 7 million observations for 11,738 paid Apps for which I have data on the following variables:

Measures

Performance of paid Apps: Google Play store does not reveal the actual number of installations (equivalent to unit sales of Apps), revenues, or profitability of Apps sold on it. However, the webpage of each App on Google Play store displays a range indicating the number of times an App has been installed. Also, Google Play store only displays the current range of installs of an App and does not provide any information on the date on which that range was reached. However, my data sources have been collecting daily data from Google Play store. Hence I was able to collect the dates on which an App moves from one install range to the next. In this project, I focus on five such install ranges: 500- 1,000; 1,000- 5,000; 5,000- 10,000; 10,000- 50,000; and 50,000- 100,000. I assume that the number of installations of an App on the date when it reaches the 500- 1,000 range are 500; when it reaches the range 1,000- 5,000 are 1,000; and so on. I refer to these installation levels as 'adoption levels' and use the time taken by

[7] Hence the Apps in my data were launched between 21st February, 2009 and 10th January 2012.

Apps to reach these adoption levels as an indicator of their performance. I do not include adoption levels below 500 and above 50,000 due to the scarcity of data for those adoption levels.

Life stages of paid Apps: The existing measures of different stages in the product life cycle (PLC) such as takeoff and slowdown, which have been used for consumer goods, may not apply to Apps. Golder and Tellis (1997) define takeoff as the percentage increase in sales of a product that changes with base sales levels of the product. They also mention that to obtain a credible signal of takeoff, base sales must be at least 50,000 units as a large percentage increase from a smaller base level of sales is not a reliable indicator of takeoff. However, in my dataset, only 19 Apps (out of 4,180) reach the 50,000 adoption level. This may indicate that most Apps have still not taken off or that the PLC of Apps is different from consumer goods, but that is not the focus of this research. Since the existing classification of stages in a product's life cycle does not seem to apply to Apps, I use the five adoption levels in my data as the different life stages of a paid App. I would like to point out that I do not intend to equate these adoption levels to the PLC stages defined in the current literature, but use them as indicators of early and later stages in the life of a paid App.

Paid App average user rating $\left(rat_{it}\right)$: After customers install an App from Google Play store, they can rate and review that App. Customers can rate an App on a 5 point scale with 1 being the lowest rating. In addition to rating an App, customers can also write a text review for the App. The webpage of each App displays information about the total number of ratings of the App, number of ratings on different scale points (1 to 5), and the average user rating based on individual ratings. This webpage also displays the text of the latest 480 user reviews and the ratings accompanying these reviews. One of my data sources provides historical user reviews and corresponding ratings of Apps on Google Play store. Hence, I only observe historical ratings of users who have provided both, a text review and a numerical rating. I assume that the ratings which I see in my data are good indicators of the overall rating of Apps. I test this assumption by calculating the correlation between the paid App average user ratings visible on Google Play store on the day I collected my data with their respective ratings that I calculated using data available to me. The correlation between these two ratings is around 0.8 which supports my assumption that the paid App average user ratings in my data are a good approximation of the paid App average user ratings visible to customers.

Free version presence (fv_{it}) : Free versions of paid Apps are not identified explicitly on Google Play store. The webpage of each paid App has a section which displays up to four more Apps developed by the same App developer, if the App developer has multiple Apps present on Google Play store. In a majority of cases, if a free version of the paid App is offered on Google Play store, it is also featured in this section. Customers could also click on the name of the App developer to navigate to a webpage containing the complete list of Apps sold by that App developer to find the free version, if it exists. I adopted the following methodology to automatically identify free versions of the paid Apps in my database. I compiled a database of 448,065 free Apps and their developers from my three data sources. Then I coded and executed a search algorithm to match paid Apps in my data to these free Apps based on a match on developer name. After populating these matches in my data, I wrote a script to compare the name and unique identifiers of paid Apps to the corresponding characteristics of free versions based on a list of keywords. If a combination of a paid App and a free version matched on these keywords, I tagged that paid App as having a free version on Google Play store. I evaluated the accuracy of this method of identifying free versions using a random sample of 2460 paid Apps. I manually identified the presence of free versions of these paid Apps and then compared the result of my manual search with the free versions identified by my search algorithm. My algorithm identified the presence (or absence) of free versions accurately in 93% cases, with 4.6% false negatives and 2.4% false positives. Given these results, I am confident of the performance of my search algorithm and use it to identify free versions for the 11,738 paid Apps in my database. I identify paid Apps with free versions by using a free version presence dummy variable, which is set to '1' for paid Apps which have free versions present on Google Play store.

Developer reputation $\left(top_{it}\right)$: Most of the popular Apps on Google Play store have been developed by businesses which did not exist or were unheard of before the rise of the App industry. In May 2011, Google started assigning Top Developer status to producers of high quality and popular Apps. Since then, 166 such developers have had a 'Top developer' badge displayed next to their name. I use this top developer recognition as a proxy of high developer reputation for my analysis. I identify these top developers by using a top developer dummy variable which is set to '1' for observations on or beyond 11th May 2011 if an App developer is a top developer.

Category $\left(cat_{i}\right)$: Apps on Google Play store are broadly divided into Apps bought for fun and pleasure (Games) and goal oriented and functional Apps (Applications). I use a category dummy variable which I set to '1' for Apps which are categorized as Games. As mentioned earlier, the evidence from information systems literature suggests that the network effects associated with information goods play a crucial role in determining the impact of a free version on the performance of its paid version. Hence, I identified Apps classified under the sub-category 'social' as they are, by design, expected to have strong positive network effects. However, as less than 1% of Apps in my data fall under this sub-category, I did not include this variable in my final analysis.

Price $\left(price_{it}\right)$ **and number of days since last update**: I expect the price of an App to impact the number of its installations; hence I include it as a control variable in my model. Google Play only reveals the current price and versions of Apps. However, one of my data sources provides historical prices of all the Apps in my database. As some of these prices were not in US dollars, I used the currency exchange rate provided by Citibank N.A. on 6th February, 2013 to convert all App prices into US dollars. My data source also has the dates on which Apps were updated by their developers. As I expected that frequently updated Apps will have higher installs, I included 'number of days since last update' as a control variable in my initial models. However, as this variable did not have a significant impact on paid App adoption speed and its exclusion did not change my model estimates significantly, I dropped this variable in my final analysis.

Free version quality $\left(fvqual_{i}\right)$: Scholars propose that the impact of free version presence on the performance of paid version of information goods also depends on the proportion of full features available in the free version (Wallenberg 2009). If the proportion of features available in the free version is low then customers may not like the free version which will decrease their likelihood of purchasing the paid version. However, if the proportion of features available in the free version is high, then a large number of customers will find the free version as an appropriate substitute of the paid version, thus cannibalizing its sales. Hence, the impact of the proportion of features available in the free version on the likelihood of purchase of the paid App is expected to have an inverted-U shape. As identifying the proportion of full features available in free versions is a tedious, manual task, I use free version quality as its proxy. However, as I do not have third-party evaluations of App quality for most Apps in my database, I take the average user rating of free versions on the date of my data collection as a measure of their quality. I include both, linear and quadratic terms of free version quality as controls in my model.

Other control variables: Given the increasing privacy concerns about customer data being collected by mobile devices, I include the number of permissions $(perm_i)$ required by a paid App as a control variable in my model. Finally, as variance of ratings of a product has been shown to impact its sales (Sun 2012), I also include the variance of ratings of paid Apps $\left(\mathrm{var}_{it}\right)$ as a control variable in my model.

My data contain 11,738 paid Apps out of which 11% are Games and 19% have free versions present on Google Play store. Almost 99% of these free versions are feature limited. I also observe that developers of Games offer free versions in much higher proportion (43%) than developers of Applications (16%). Finally, 1% of paid Apps in my data have been developed by App developers with high reputation and these developers offer free versions of their paid Apps in lower proportion (8.3%) than relatively unknown developers (18.6%). A look at the growth rates of Apps (Figure 3) reveals a pattern similar to the familiar S-shaped diffusion curve (e.g., Mahajan, Muller, and Bass 1990).

---- Insert Figure 3 about here -----

CHAPTER FOUR: METHODOLOGY

My methodology is guided by the nature of data available to me and the underlying process that generates these data. To explain the rationale behind my model selection, I describe the data generating process below.

Data Generating Process

When a customer views a webpage belonging to a paid App, he will, in addition to the description of the App, see the following time varying and time constant information regarding that App.

Time varying information: Time varying information of an App includes:

- A range indicating the number of times the App has been installed. As discussed, I use time taken by paid Apps to reach different adoption levels, which are the lower ends of these ranges, as the dependent variable in my analysis;
- Paid App average user rating $\left(rat_{it}\right)$;
- Number of times that the App has been rated 1, 2, 3, 4, or 5 on a five point scale. I capture this spread of ratings by calculating the variance of ratings $\left(\text{var}_{it}\right)$;
- Total number of times the App has been rated $\left(n_{it}\right)$. Only the customers who have installed an App are able to rate an App. Since the number of ratings of the App will be highly correlated with the actual number of installs of the App, I do not include this variable in my models;

- Reputation of the developer, indicated by the presence (or absence) of the top developer batch (top_{it});
- Price of the paid App $(price_{it})$; and,
- Link to the free version (fv_{it}) of the paid App, if it exists.

Time constant information: Time constant information of an App includes:

- Category of App (Game or Application) (cat_i).

In addition to these variables, a customer can also see the text of the latest 480 reviews for the paid App, date of its last update, and the number of permissions required by the App to run. If a free version of the paid App is present (i.e., $fv_{it} = 1$) a customer can also view similar information for the free version. In my current model, I do not focus on the time-varying information of a free version but concentrate on the quality of free version $(fvqual_i)$ and its quadratic term $(fvqualsq_i)$.

This is the information available to a customer when he makes one of the following three choices: to install the paid App; or to install the free version (if it is available); or to do nothing. However, as some of the above-mentioned variables change over time, I expect that their impact on a customer's decision to choose one of these three alternatives will also change. For example, if a customer views the webpage of a paid App immediately after its release ($t = 0$), the values of these variables will be $rat_{i0} = \text{var}_{i0} = 0$. Hence, at this stage in the life of the paid App, in addition to the information provided by its description, developer reputation, and its price, the only way a customer could obtain additional information about its quality is by installing its free version, if it exists. However, when a customer visits the webpage of the same paid App after it has been installed and rated by other customers, i.e., later in the life stage of the paid App, these variables will have non-zero values and will provide additional information about the quality of the paid App.

Model Selection

In my data, I have the dates on which paid Apps achieve the following adoption levels: 500; 1,000; 5,000; 10,000; and 50,000. I refer to the duration that a paid App takes for moving from one adoption level to the next as a 'spell'. As I estimate separate models for each adoption level in my data, an App can only have a maximum of one 'spell' in each of my models.

I note that the change in the number of installations of Apps is a continuous process but I observe this at discrete time periods, i.e., once a day. Another characteristic of my data is that Apps enter my dataset at different times, i.e., on the day of their launch. Finally, for each model that I estimate, the spells for some Apps will be right censored as the observation period ends before those Apps have had the chance to reach an adoption level. Given these characteristics of my data, I build my model based on the framework proposed by Gönül and Ter Hofstede (2006). I model the likelihood of an App achieving an adoption level using discrete-time proportional hazard function. In each of my models, the hazard function is the probability of a paid App i, reaching the adoption level in time period t, given that it has not reached the adoption level till that time. I use the proportional hazard specification, where the hazard of an App i reaching the adoption level at time D is:

$$h(\mathrm{d}_{it}, \mathrm{X}_{it}, \mathrm{Z}_i) = \lim_{\Delta t \to 0} \frac{P[\mathrm{d}_{it} \le D < \mathrm{d}_{it} + \Delta t \mid D > \mathrm{d}_{it}]}{\Delta t} = h_0(\mathrm{d}_{it}) \exp(\beta_i' X_{it} + \alpha' \mathrm{Z}_i) \qquad (1)$$

where D is the stochastic representation of duration.

This proportional hazard specification has two components:

1) $h_0(\mathrm{d}_{it})$ - This part captures the dependence of the likelihood of achieving an adoption level on the duration since the start of the spell (d_{it}).

2) $\exp(\beta_i' X_{it} + \alpha' \mathrm{Z}_i)$ - This component is the proportionality term which contains all the time-varying $(\beta_i' X_{it})$ and time constant $(\alpha' \mathrm{Z}_i)$ factors that could impact the likelihood of achieving an adoption level.

As mentioned above, even though the underlying process generating my data is continuous, I only observe my data in discrete time periods i.e., daily. Hence my data is interval censored. For such interval censored data, it can be shown that the discrete version of the hazard takes the complementary log-log functional form (Jenkins 2005). So the discrete specification of the proportional hazard reduces to,

$$h(\mathrm{d}_{it}, \mathrm{X}_{it}, \mathrm{Z}_i) = P[\mathrm{D} = \mathrm{d}_{it} \mid D > d_{it} - 1] = 1 - \exp[-\exp(\beta' \mathrm{X}_{it} + \alpha' \mathrm{Z}_i + \xi_{it})] \qquad (2)$$

where $\xi_{it} = \log\left[\int_{d_{it}-1}^{d_{it}} h_0(u_{it}) du_{it}\right]$ corresponds to the baseline hazard $h_0(d_{it})$, $\beta' \mathrm{X}_{it}$ captures the impact of time varying characteristics and $\alpha' \mathrm{Z}_i$ captures the impact of time constant characteristics of Apps in equation 1. The likelihood of an App i reaching an adoption level in period t, is given by $(h(\mathrm{d}_{it}, \mathrm{X}_{it}, \mathrm{Z}_i))^{y_{it}} (1 - h(\mathrm{d}_{it}, \mathrm{X}_{it}, \mathrm{Z}_i))^{1-y_{it}}$, where $y_{it} = 1$ if the App reaches an adoption level in period t, else $y_{it} = 0$.

I use the following functional form to capture the effects of the time since the beginning of a spell,

$$\xi_{it} = \alpha_0 + \alpha_1 d_{it} + \zeta_i$$

Here α_0 is the intercept and $\alpha_1 d_{it}$ captures the linear impact of duration[8] since the start of the spell on the likelihood of an App achieving the adoption level. I also include a random component, ζ_i to account for unobserved heterogeneity. This random component is distributed normally $(\zeta_i \sim N(0,\psi))$ and is independent from the covariates. Finally, ξ_{it} has a standard extreme-value type-1 or Gumbel distribution, given the covariates and the random intercept ζ_i.

For the functional specification of the proportionality term in the hazard function, I consider the role of free version presence (fv_{it}), paid App average user rating (rat_{it}), variance of paid App (var_{it}), reputation of App developer (top_{it}), category of App (cat_i), linear and quadratic specifications of variable indicating the quality of free version $(fvqual_i, fvqualsq_i)$, price of paid App $(price_{it})$, and the number of permissions $(perm_i)$ required by paid App. There are observations in my data when no ratings of paid Apps exist as customers have not reviewed those Apps till then. To account for the effect of no ratings of a paid App, I also include a dummy variable in my model (k_{it}) which takes the value '1' when no user ratings of paid Apps are available. I estimate two models at each of the five adoption levels in my data- one containing only the main effects of the variables of interests (and control variables) and the second model which includes the hypothesized interactions in addition to the variables from the main effects model.

In the models with interactions, I include interactions of the App category dummy (cat_i) with the free version presence dummy (fv_{it}), developer reputation dummy (top_{it}), and paid App

[8] I also included the quadratic duration term in my models and it did not affect my results significantly. Hence I did not include this term in my final models.

average user rating (rat_{it}). Hence the functional form of the proportionality term in the hazard function looks as follows:

Main effect model:

$$\beta' X_{it} + \alpha' Z_i = \beta_1 fv_{it} + \beta_2 top_{it} + \beta_3 rat_{it} + \beta_4 price_{it} + \beta_5 cat_i + \beta_6 \operatorname{var}_{it} + \beta_7 fvqual_i + \beta_8 fvqualsq_i + \beta_9 perm_i + \beta_{10} k_{it}$$

Model with interactions effects:

$$\beta' X_{it} + \alpha' Z_i = \beta_1 fv_{it} + \beta_2 top_{it} + \beta_3 rat_{it} + \beta_4 cat_i * rat_{it} + \beta_5 cat_i * top_{it} + \beta_6 fv_{it} * cat_i + \beta_7 price_{it} + \beta_8 cat_i + \beta_9 \operatorname{var}_{it} + \beta_{10} fvqual_i + \beta_{11} fvqualsq_i + \beta_{12} perm_i + \beta_{13} k_{it}$$

Endogeneity

In my analysis I compare the performance of paid Apps with and without free versions and attribute any difference in their performance to the presence of free versions. However, to attribute this difference in performance to free version presence, I have to rule out endogeneity affecting my results. Offering a free version is a decision taken by the App Developer. Hence there is the possibility that the characteristics of paid Apps with free versions are very different from the characteristics of paid Apps without free versions. Thus any difference in the performance of these two groups of paid Apps might be due to the difference in their characteristics, and not because of free version presence.

Using the instrumental variable approach to account for endogeneity does not seem to be straightforward in my case as my dependent variable and key independent variable are discrete (Wooldridge 2002), and instruments are not readily available. Hence, I use the Coarsened Exact Matching (CEM) technique (Iacus, King, and Porro 2011, 2012) which has found recent application in economics (Azoulay, Graff Zivin, and Wang 2010) and management literature

(Singh and Agrawal 2011) but has not been used by marketing scholars till now. CEM involves selecting a set of observed covariates on which one intends to balance the treatment and control groups. The next step is to perform ‘coarsening’ of the available data on these covariates and create a large number of strata to cover the entire support of the joint distribution of the covariates. Then each observation is allocated to one unique stratum and only those treatment groups are selected which have at least one control group in the same stratum. The ‘pruned’ (but uncoarsened data) is then used for the final analysis. CEM simulates a randomized experiment by creating control and treatment groups which differ only in their treatments, as all the other observed covariates are matched/ balanced.

I balanced my data using CEM on those observed characteristics of paid Apps which could influence an App developers’ decision to offer free version of a paid App. These characteristics are: average rating, number of ratings, price, variance of rating, days since launch, category, and the calendar quarter of the year. This matching process ensured that each paid App with a free version in the data has one paid App without free version with characteristics similar to the paid App with free version on the day its free version was launched. I argue that, after this matching, the only difference between the groups of paid Apps with and without free versions is the presence of free version. Hence any difference in the performance of these two groups could be attributed to the presence of free version.

The details of the matched dataset are given in Table 2 and Figure 4. This dataset has 4,180 paid Apps out of which 50% have free versions and 22% of paid Apps are Games. The number of Apps achieving different adoption levels is given in Figure 4. The correlation chart given in Table 2 does not reveal any significant multicollinearity problems. The correlations between all but three variables- free version presence, quality of free version, and the squared term of the quality of free version- are less than 0.2. Given that free version quality is a time constant variable that only takes non-zero values for observations where free versions exist (i.e., free version presence = 1), these high correlations are expected and can be safely ignored (Allison 2012). Also, other than these three variables, no other variable has a variance inflation factor above 4, and none of the conditional indexes associated with the eigenvalues of the variable matrix exceed 15. These tests imply the absence of significant multicollinearity problems (Johnston 1991).

---- Insert Table 2 and Figure 4 about here -----

Estimation

I use maximum likelihood estimation to estimate my models. The first challenge in my estimation was to obtain the marginal joint probability of the responses for each App (i) not conditioning on the random intercept ζ_i, but still on the covariates $\mathrm{X}_{it}, \mathrm{Z}_i$. For obtaining this marginal joint probability, I have to first integrate out the random intercept (ζ_i). However, as this integral does not have a closed form expression, I approximate it by using the mean and variance adaptive Gauss-Hermite integration method. This approximation can be viewed as replacing the continuous density of ζ_i with a discrete distribution with R possible values of ζ_i having probabilities $\Pr(\zeta_i = e_r)$ (Rabe-Hesketh and Skrondal 2012). I use an adaptive quadrature approximation as it has been shown to perform satisfactorily even when the function being integrated has a sharp peak, though ordinary quadrature approximation performs poorly in such cases (Rabe-Hesketh, Skrondal, and Pickles 2002, 2005). As adaptive quadrature approximation does not work when there are too few quadrature points (fewer than five), I use 12 quadrature points[9] for my analysis. Finally, the marginal likelihood is the joint probability of all the responses for all Apps. As I assume that the responses of all Apps are mutually independent, this marginal likelihood is just the product of marginal joint probabilities of the responses of all Apps. For complementary log-log models (and other generalized models), the marginal likelihood does not have a closed form, hence I use the Newton-Raphson algorithm to maximize the likelihood with maximum iterations set to 16000.

I first estimate duration models with only the main effects of the three information sources of paid App quality and control variables. The results of these models capture the impact of free version presence, reputation of App developer, and paid App average user rating on its adoption speed. I estimate these main effect models across all five adoption levels. A comparison of the coefficient estimates across adoption levels also reveals how the impact of information sources of paid App quality on its adoption speed changes across the life of a paid App. In the second part of my analysis, I include the hypothesized interactions to my main effects models and estimate these models across all five adoption levels. I present the results of these models in Table 3 and 4 and discuss them in detail in the following section.

[9] I also used 16 and 20 quadrature points for my estimation but these changes do not affect my results significantly.

CHAPTER FIVE: RESULTS

Before turning to the discussion of the hypothesized effects, I consider the effects of control variables (Table 3). Given the low price of most paid Apps, price does not have a significant impact on their adoption speed. The results given in Table 3 also reveal that the adoption speed of Games is higher than that of Applications. This is consistent with the broader market appeal of Games, which are bought for fun and pleasure, as compared to Applications which are goal oriented and are bought for accomplishing specific tasks. My results also show that the quality of a free version has a positive and significant impact on paid App adoption speed. But the quadratic term for free version quality has a negative and significant impact on paid App adoption speed. Hence, as expected, free version quality has an inverted U shape effect on paid App adoption speed. Interestingly, my results suggest that this inverted U shape effect disappears towards the later stages in the life of a paid App, when the coefficients of both free version quality and its quadratic term become positive. This suggests that the role of a free version changes across the life stages of a paid App, an area which has not been explored by scholars till now. I was also surprised to see that despite all the attention given by mass media to customer privacy, an increase in the number of permissions required by a paid App has a positive impact on its adoption speed. One reason for this effect could be that higher quality Apps may require greater number of permissions to run due to the richness of their features and such Apps also have higher adoption speed.

Impact of Information Sources of Paid App Quality on Paid App Adoption Speed

My results given in Table 3 indicate that free version presence has a negative impact on the adoption speed of paid Apps across all adoption levels. These results support my hypothesis (H_1) that, due to a mismatch between the characteristics of Apps and the nature of free versions being offered, the presence of a free version will have a negative impact on the adoption speed of its paid App. Consistent with the evidence from existing literature and my hypotheses, I find that both, high developer reputation and paid App average user rating have positive impacts on the adoption speed of paid Apps, thus supporting H_2 and H_3.

----- Insert Table 3 about here -----

Impact across Life Stages of Paid App

The results (Table 3) also indicate that the impacts of the three information sources of paid App quality on its adoption speed change across the life of the paid App. As hypothesized, I find that the negative impact of free version presence on the adoption speed of a paid App becomes more negative towards the later stages in the life of the paid App. The coefficient estimate of free version presence moves from -0.201 at the 500 adoption level to -0.92 at the 10,000 adoption level. This indicates support for my hypothesis (H_6) that the cannibalization effect of free version presence on paid App installs will become stronger in the later life stages of paid Apps. I also find the impacts of developer reputation and paid App average user rating on paid App adoption speed move in opposite directions across the life of a paid App. High developer reputation has a strong positive impact on paid App adoption speed in the early life stages (b= 1.256 at the 500 adoption level) of a paid App but this impact disappears in the later life stages (b= -0.061 at the 10,000 adoption level). However, the positive impact of high paid App average user rating on its adoption speed becomes stronger in the later life stages (b= 0.435 at the 10,000 adoption level) of a paid App as compared to its early life stages (b= 0.141 at the 500 adoption level). These results indicate support for hypotheses H_4 and H_5. I note that most coefficient estimates are not significant at the 50,000 adoption level, which is likely due to the small sample size. The number of paid Apps reaching this adoption level is only 19.

Impact across Category of Paid App

My results given in Table 4 indicate that the impacts of the three information sources of paid App quality on its adoption speed change across paid App category. The negative impact of free version presence on the adoption speed of paid Apps is stronger for paid Games than for paid Applications. This indicates that the delay in the purchase of paid Apps caused due to the presence of free versions has a greater negative impact on the adoption speed of paid Games than paid Applications due to the discretionary and impulsive nature of their purchase. This supports my hypothesis H_{9a}. My results indicate partial support for hypotheses H_7 and H_8. I find that, as hypothesized, indirect information sources of paid App quality- developer reputation and paid App average user rating- have a smaller positive impact on adoption speed of paid Games than that of paid Applications. Even though the sign of these effects are in the hypothesized direction across most adoption levels, these effects are not significant across all adoption levels indicating partial support for hypotheses H_7 and H_8.

----- Insert Table 4 about here -----

To summarize, I find strong support for six and partial support for two of my hypotheses. My results also support my hypothesis H_{9a} that the presence of free versions will have a greater negative impact on the adoption speed of paid Games than that of paid Applications.

CHAPTER SIX: DISCUSSION AND IMPLICATIONS

The App industry has grown in size and significance over the past five years but has attracted limited attention of marketing scholars till now. I present this pioneering work focusing on this fast growing industry by investigating the popular practice of offering free versions of paid Apps. I find that despite its popularity, the current practice of offering feature limited free versions of paid Apps is not an optimal strategy. My results support my hypotheses that offering a feature limited free version of a paid App will have a negative impact on paid App adoption speed. In addition to presence of a free version, I also investigate the impact of two other information sources of paid App quality on its adoption speed. I find that both, high developer reputation and paid App average user rating have positive impacts on paid App adoption speed. I also find that the impact of the information sources of paid App quality on its adoption speed varies across the life stage and category of a paid App. In the early life stages of a paid App, there is lack of information about its quality, hence developer reputation has a strong positive impact and free version presence has a smaller negative impact on its adoption speed. However, as additional information sources of paid App quality such as paid App average user rating become more reliable towards the later stages in the life of a paid App, the positive impact of developer reputation on paid App adoption speed diminishes. Similarly, the negative impact of free version presence on paid App adoption speed becomes more negative towards the later life stages of the paid App. Finally, my results reveal that sources that provide information about paid App quality through indirect exposure to the App- such as developer reputation and paid App average user rating- have a smaller positive impact on the adoption speed of paid Apps which are primarily hedonic in nature. I further find that the delay in the purchase of a paid App caused by the presence of a free version has a greater negative impact on adoption speed of hedonic Apps as their purchase is generally discretionary and impulsive.

This project has implications for practitioners and academicians. As mentioned earlier, despite the growth in the App industry, there is limited guidance available to App developers. In my various discussions with App developers during this project, I found that due to the zero development and marginal costs of free versions, App developers offer free versions of paid Apps freely. However, my results suggest that the current practice of offering free versions of paid Apps is a myopic and sub-optimal strategy as it impacts the adoption speed of paid Apps negatively. I also find that even though the negative impact of free version presence on paid App adoption speed increases in the later life stages, App developers continue to offer free versions across all life stages of paid Apps. My data also suggest that developers of Games offer free versions in much greater proportion than developers of Applications, even though my results indicate that the cannibalization effect of free version presence is greater for paid Games than for paid Applications. I hope that my findings will prompt App developers to reconsider their current strategies. In light of my results, App developers may look at offering time limited free versions rather than feature limited free versions of paid Apps. It may also be beneficial for App developers to stop offering free versions in the later life stages of paid Apps, when the uncertainty about the quality of paid Apps is low. Finally, as free versions have greater negative impact on the adoption speed of paid Games, developers of Games may consider alternative modes of promoting their Apps and reducing customer uncertainty about the quality of their Apps.

This research contributes to sampling and versioning literature which lack large scale empirical studies. I also contribute to the existing literature by taking a more comprehensive view of the implications of offering free versions by investigating the impact of offering a free version on the performance of its paid App across its life stages and characteristics. I show that the impact of offering free version, developer reputation, and paid App average user rating on its adoption speed changes across the life and characteristics of the paid App. I also introduce a new dataset and research area in marketing literature and hope that this research could spark the interest of other marketing scholars to produce work focusing on this important and fast growing industry.

Limitations and Future Research

As my research focuses on an area which has not been explored by marketing scholars till now, I had to compile a new dataset for my analysis. The lack of availability of historical data of Apps from the primary source added to the complexity of this task. Since I had to collect my data from secondary sources, I had to use approximate measures of some variables in my models. I do not have data about the profitability and revenue generated by paid Apps. Hence, I use the adoption speed of paid Apps as a measure of their performance in my analysis. However, lower adoption speed may not always correspond to lower profitability or revenues of a paid App. Hence, future research in this area could use alternative performance measures. The data source that I used to collect data about ratings of paid Apps does not have data on all customer ratings of paid Apps. This data source only provides ratings which are accompanied by text reviews. As customers on Google Play store have an option of rating an App without writing a text review, the ratings in my data are a subset of the ratings visible to prospective customers. Even though I verified that the ratings in my data are good indicators of the actual ratings of paid Apps (i.e. ratings visible to customers), this is still a limitation of my data. Further, I used a search algorithm to automatically identify free versions of paid Apps, as identifying free versions of paid Apps is a time consuming, manual process. I verified the accuracy of my search algorithm by comparing its results with manual identification of free version presence for 2460 Apps. Even though my algorithm was able to correctly identify free version presence (or absence) in more than 90% of the cases, the accuracy of my search algorithm is not 100%. Hence my results may have been affected by wrong classification of free version presence of paid Apps. My efforts to contact Google Play store to collect data directly from them were not fruitful, but I hope that future researchers would be able to get data directly from Google Play store or identify better secondary sources of data. In this project I show that the impact of a free version on the adoption speed of its paid App changes across the life of the paid App. However, I do not identify the life stage of a paid App when an App developer should stop offering the free version. Future research in this area could develop analytical models to help App developers identify life stages of their paid Apps at which they could stop offering free versions and thus maximize the sales of their paid Apps. Finally, I do not explicitly incorporate the effect of competition in my analysis. As the decision to offer free versions and their impact on paid App performance may be influenced by the nature and intensity of competition, future research could incorporate the role of competition as well. I account for endogeneity by matching the observed characteristics of paid Apps with and without free versions. However, given the limitations in my data, my technique does not account for missing variables that affect my results. I hope that as data availability in this area improves, scholars will be able to use more traditional approaches such as instrumental variables to account for observed and unobserved variables affecting the results.

The App industry offers fantastic research opportunities due to its unique characteristics and data availability. I hope that future scholars would leverage on these opportunities afforded by the App industry to produce research with wide ranging impact.

TABLES

Table 1

LITERATURE REVIEW

	Authors	**Methodology/ Product**	**Relevant findings**
Sampling literature	Hamm et al. 1969/ Bettinger, Dawson, and Wales 1979	Experiment / Consumer goods	Free samples help in changing the image of a product.
	Holmes and Lett 1977	Experiment / Consumer goods	Free samples create favorable brand attitude and generate word of mouth.
	Marks and Kamins 1988	Experiment / Consumer goods	Sampling (vs. advertising) leads to higher belief and attitudinal confidence.
	Jain, Mahajan, and Muller 1995	Analytical / Consumer goods	Targeted and optimal level of sampling for products with high coefficient of imitation accelerates adoptions and leads to higher peak sales.
	Heiman et al. 2001	Analytical / Consumer goods	Sampling helps in boosting immediate sales and creates goodwill.
	Bawa and Shoemaker 2004	Experiment / Consumer goods	Impact of sampling **(**Acceleration-Cannibalization-Expansion) will depend on mix of consumer segments. If repeat purchase following trial is nonexistent, then cannibalization will dominate.
	Lehmann and Esteban-Bravo 2006	Analytical / Consumer goods	Targeted giveaways help when there is a small group of 'slow to adopt on their own customers' who influence other customers. Giveaways decrease sales in the first period and increase them in subsequent periods.
Versioning literature	Faugere and Tayi 2007	Analytical / Information goods	Model for finding the optimal combination of trial time and proportion of features to be included in free samples.
	Cheng and Tang 2010	Analytical / Information goods	It is profitable for a firm to offer feature-limited trial (vs. segmenting the market by offering two versions of different quality) when network intensity is high.
	Cheng and Liu 2012	Analytical / Information goods	It is profitable for a firm to offer time-limited (vs. feature-limited) trial when the network effect is moderate and when customers highly underestimate the quality of the software.
	Dey and Lahiri 2013	Analytical / Information goods	Versioning is profitable when uncertainty about product quality is low, as it expands the market. Versioning is also profitable when uncertainty is high and firms are able to charge high price premium for upgrading customers.

Table 2

DESCRIPTIVE STATISTICS AND CORRELATIONS

	Variable	**Mean (SD)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	**Free version presence**	0.430 (0.50)									
2	**Top developer**	0.012(0.11)	0.013 ***								
3	**Paid App average user rating**	3.600(1.47)	0.091 ***	0.020 ***							
4	**Price**	2.956(7.14)	0.007 ***	0.112 ***	-0.026 ***						
5	**Category**	0.219(0.41)	-0.019 ***	0.076 ***	0.017 ***	-0.079 ***					
6	**Paid App rating variance**	0.988(1.15)	0.031 ***	0.052 ***	-0.051 ***	0.030 ***	0.005 ***				
7	**Free version Quality**	1.746(1.87)	0.812 ***	0.024 ***	0.088 ***	0.002 ***	-0.015 ***	-0.025 ***			
8	**Free version Quality Square**	6.548(7.85)	0.717 ***	0.030 ***	0.107 ***	-0.002 ***	-0.023 ***	-0.054 ***	0.972 ***		
9	**Paid App permissions**	3.904(3.90)	0.107 ***	0.025 ***	-0.022 ***	0.166 ***	-0.108 ***	0.126 ***	0.122 ***	0.106 ***	
10	**Duration**	271.899(223.60)	0.019 ***	-0.010 ***	0.056***	-0.018 ***	0.036 ***	0.063 ***	-0.089 ***	-0.088 ***	-0.099 ***

* p =< 0.1, **p =< 0.05, ***p =< 0.01

Notes: Standard errors in parentheses.

Table 3

RESULTS- MAIN EFFECTS HAZARD MODELS

Independent Variables	500	1000	5000	10000	50000
Free version presence	-0.201*** (0.07)	-0.82*** (0.09)	-0.482*** (0.18)	-0.92*** (0.27)	3.467 (9.09)
Top developer	1.256*** (0.27)	-0.15 (0.26)	0.983*** (0.28)	-0.061 (0.28)	-0.118 (1.3)
Paid App average user rating	0.141*** (0.03)	0.241*** (0.04)	0.585*** (0.12)	0.435**(0.21)	2.398 (2.31)
Price	0.003 (0)	-0.003 (0)	-0.016 (0.01)	-0.007 (0.02)	-0.473 (0.3)
Category	-0.053 (0.06)	0.24*** (0.07)	0.499*** (0.13)	0.606*** (0.19)	0.855 (0.96)
Paid App rating variance	0.357*** (0.02)	0.077** (0.03)	0.301*** (0.08)	-0.151 (0.15)	0.441 (1.39)
Free version quality	0.364*** (0.06)	0.39*** (0.07)	0.242 (0.16)	0.152 (0.26)	-4.544 (5.25)
Free version quality square	-0.054*** (0.01)	-0.039** (0.02)	-0.009 (0.03)	0.025 (0.06)	0.837 (0.76)
Paid App permissions	0.051*** (0.01)	0.031*** (0.01)	0.062*** (0.01)	0.038**(0.02)	0.209* (0.12)
Duration	-0.003*** (0)	-0.002*** (0)	0.001* (0)	0.004*** (0)	0.012*** (0)
Paid App rating absent	-21.256 (2701.04)	NA	NA	NA	NA
Intercept	-7.201*** (0.16)	-6.574*** (0.22)	-10.49*** (0.67)	-8.35*** (1.09)	-25.451** (11.94)
Log likelihood	-15717.674	-9717.229	-3901.105	-1816.674	-164.345

* p =< 0.1, **p =< 0.05, ***p =< 0.01

Notes: The table shows parameter estimates, with standard errors in parentheses.

Table 4

RESULTS- INTERACTION EFFECTS HAZARD MODELS

Independent Variables	**500**	**1000**	**5000**	**10000**	**50000**
Free version presence	-0.116 (0.08)	-0.739*** (0.09)	-0.235 (0.19)	-0.709** (0.28)	2.728 (9.31)
Top Developer	1.431*** (0.39)	0.039 (0.42)	2.101*** (0.39)	0.3 (0.34)	1.161 (1.56)
Paid App average user rating	0.18*** (0.03)	0.253*** (0.05)	0.646*** (0.13)	0.228 (0.22)	0.739 (2.16)
Paid App average user rating * Category	-0.159*** (0.06)	-0.044 (0.08)	-0.142 (0.16)	0.426 (0.26)	3.301 (2.5)
Top developer * Category	-0.246 (0.54)	-0.338 (0.55)	-1.929*** (0.54)	-0.86 (0.57)	-3.526 (2.52)
Free version presence * Category	-0.378*** (0.12)	-0.358*** (0.13)	-0.629*** (0.23)	-0.611* (0.32)	1.829 (1.81)
Price	0.003 (0)	-0.003 (0)	-0.022**(0.01)	-0.004 (0.02)	-0.483 (0.3)
Category	0.734*** (0.23)	0.571* (0.32)	1.446** (0.65)	-0.716 (1.05)	-13.427 (10.5)
Paid App rating variance	0.357*** (0.02)	0.074** (0.03)	0.303*** (0.08)	-0.182 (0.14)	0.023 (1.36)
Free version quality	0.372*** (0.06)	0.389*** (0.07)	0.215 (0.16)	0.208 (0.26)	-4.426 (5.38)
Free version quality square	-0.057*** (0.01)	-0.039** (0.02)	-0.008 (0.03)	0.012 (0.06)	0.831 (0.78)
Paid App permissions	0.051*** (0.01)	0.031*** (0.01)	0.064*** (0.01)	0.034** (0.02)	0.213* (0.12)
Duration	-0.003*** (0)	-0.002*** (0)	0.001* (0)	0.004*** (0)	0.011*** (0)
Paid App rating absent	-22.21 (4420.85)	NA	NA	NA	NA
Intercept	-7.398*** (0.18)	-6.662*** (0.24)	-10.848*** (0.73)	-7.586*** (1.12)	-16.447 (11.19)
Log likelihood	-15707.535	-9712.814	-3889.789	-1812.306	-161.961

* p =< 0.1, **p =< 0.05, ***p =< 0.01

Notes: The table shows parameter estimates, with standard errors in parentheses

FIGURES

Figure 1

IMPACT OF INFORMATION SOURCES OF PAID APP QUALITY ON PAID APP PERFORMANCE- CONCEPTUAL MODEL

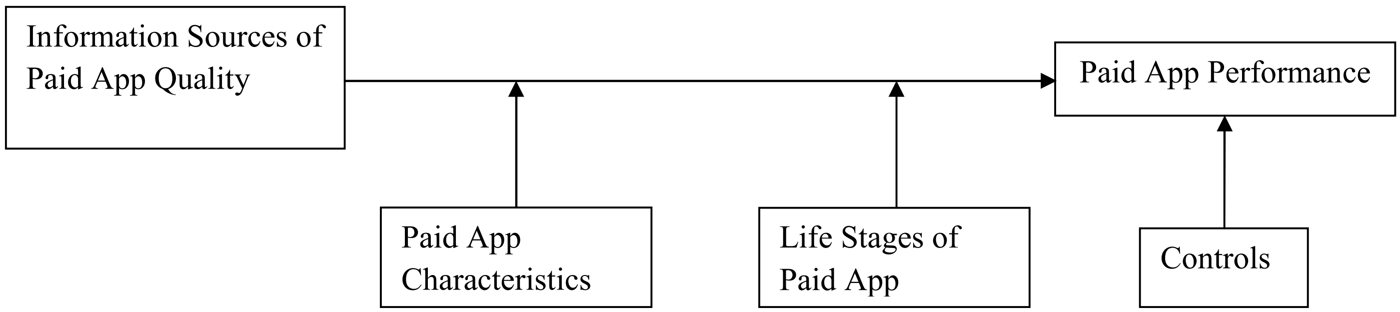


Figure 2

IMPACT OF INOFRMATION SOURCES OF PAID APP QUALITY ON PAID APP ADOPTION SPEED - ACROSS APP CATEGORY AND LIFE STAGES

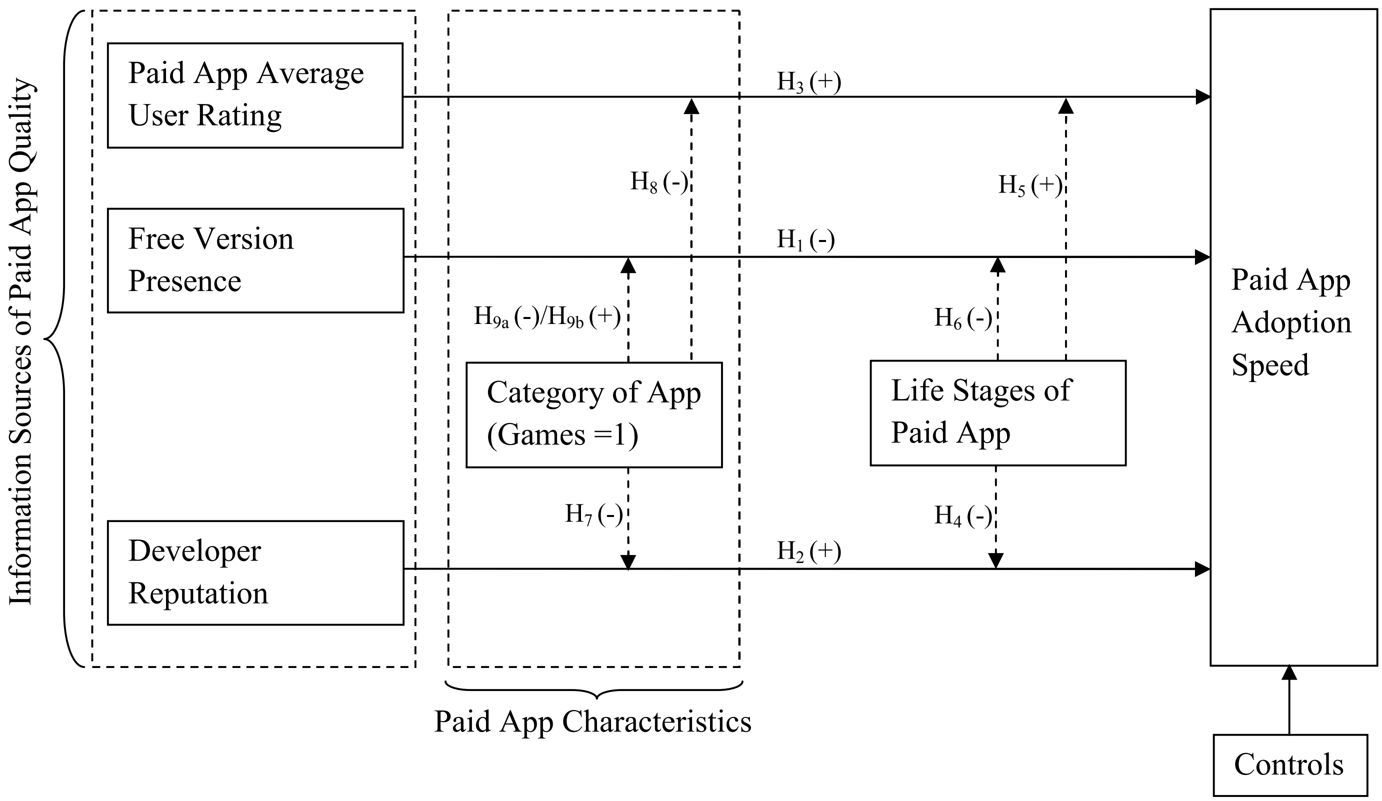


Figure 3

AVERAGE GROWTH RATE OF PAID APPS

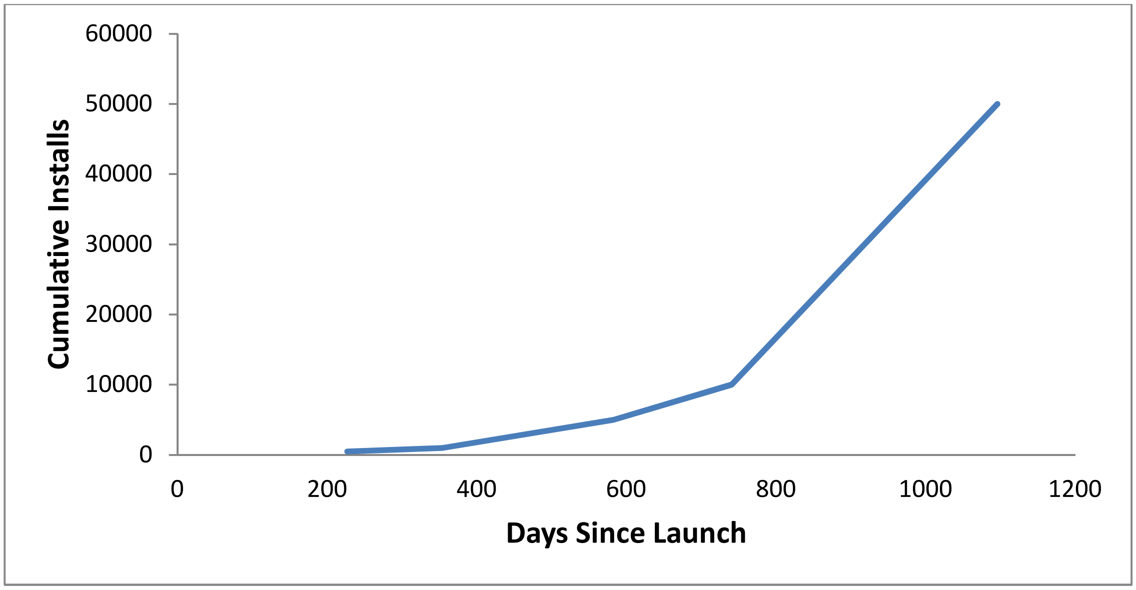
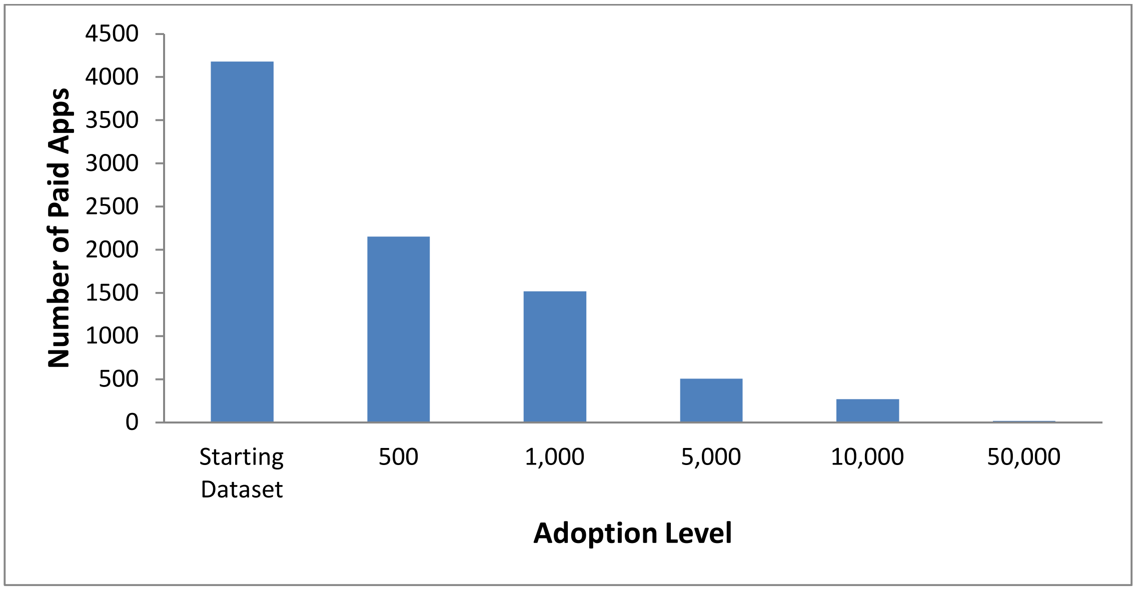


Figure 4

NUMBER OF PAID APPS ACROSS ADOPTION LEVELS



BIBLIOGRAPHY

Aaker, D.A. (1991), *Managing brand equity*, New York, NY: Free Press.

Abecassis, Arie (2012), "Can Mobile App Discovery Be Fixed?," *Mashable*, (accessed May 17, 2013), [available at http://mashable.com/2012/06/14/mobile-app-discovery/].

Accenture (2012), "Mobile Web Watch Survey 2012: Mobile Internet—Spawning New Growth Opportunities in the Convergence Era," Accenture.

Akerlof, George A. (1970), "The market for 'lemons': Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, 488–500.

Allison, Paul (2012), "When Can You Safely Ignore Multicollinearity?," (accessed August 23, 2013), [available at http://www.statisticalhorizons.com/multicollinearity].

Azoulay, Pierre, Joshua Graff Zivin, and Jialan Wang (2010), "Superstar extinction," *Quarterly Journal of Economics*, 25, 549–89.

Bawa, Kapil, and Robert Shoemaker (2004), "The Effects of Free Sample Promotions on Incremental Brand Sales," *Marketing Science*, 23, 345–63.

Bettinger, C.O., L.E. Dawson, and H.G. Wales (1979), "The Impact of Free-Sample Advertising," *Journal of Advertising Research*, 19, 35–39.

Bhargava, H.K., and V. Choudhary (2001), "Information goods and vertical differentiation," *Journal of Management Information Systems*, 18, 89–106.

Brancheau, James C., and James C. Wetherbe (1990), "The adoption of spreadsheet software: testing innovation diffusion theory in the context of end-user computing," *Information Systems Research*, 1(2), 115–43.

Cheng, H.K., and Q.C. Tang (2010), "Free trial or no free trial: Optimal software product design with network effects," *European Journal of Operational Research*, 205, 437–47.

Cheng, Hsing Kenneth, and Yipeng Liu (2012), “Optimal software free trial strategy: The impact of network externalities and consumer uncertainty,” *Information Systems Research*, 23(2), 488–504.

Chevalier, Judith A., and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research*, 43, 345–54.

D’Orazio, Dante (2013), “BlackBerry puts quantity over quality, lets one developer own a third of its app store,” *The Verge*, (accessed September 4, 2013), [available at http://www.theverge.com/2013/8/22/4648730/blackberry-lets-developer-submit-a-third-of-apps].

Dey, Debabrata, and Atanu Lahiri (2013), “The Zero-Day DLC Strategy: A Case for Versioning to Facilitate Product Sampling,” SSRN Scholarly Paper, Rochester, NY: Social Science Research Network.

Dey, Debabrata, Atanu Lahiri, and Dengpan Liu (2013), “Consumer Learning and Time-locked Trials of Software Products,” *Journal of Management Information Systems*, (Forthcoming).

Dhar, Ravi, and Klaus Wertenbroch (2000), “Consumer Choice Between Hedonic and Utilitarian Goods,” *Journal of Marketing Research*, 37, 60–71.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), “Do online reviews matter? — An empirical investigation of panel data,” *Decision Support Systems*, 45, 1007–16.

Faugère, Christophe, and Giri Kumar Tayi (2007), “Designing free software samples: a game theoretic approach,” *Information Technology and Management*, 8(4), 263–78.

Godfrey, Jonathan, Morgan W Reed, and E. Whitley Herndon (2012), “Apps Across America,” Association for Competitive Technology (ACT).

Golder, Peter N., and Gerard J. Tellis (1997), “Will it ever fly? Modeling the takeoff of really new consumer durables,” *Marketing Science*, 16(3), 256–70.

Gönül, Füsun F., and Frenkel Ter Hofstede (2006), “How to compute optimal catalog mailing decisions,” *Marketing Science*, 25(1), 65–74.

Hamm, B.C., M. Perry, and H.F. Wynn (1969), “The effect of a free sample on image and attitude,” *Journal of Advertising Research*, 9, 35–37.

Heiman, A., B. McWilliams, Z. Shen, and D. Zilberman (2001), “Learning and forgetting: Modeling optimal product sampling over time,” *Management Science*, 532–46.

Heiman, Amir, and Eitan Muller (1996), “Using demonstration to increase new product acceptance: Controlling demonstration time,” *Journal of Marketing Research*, 422–30.

Hirschman, Elizabeth C., and Morris B. Holbrook (1982), “Hedonic Consumption: Emerging Concepts, Methods and Propositions,” *Journal of Marketing*, 46, 92–101.

Holmes, J.H., and J.D. Lett (1977), “Product sampling and word of mouth,” *Journal of Advertising*, 17(5), 35–40.

Iacus, Stefano M., Gary King, and Giuseppe Porro (2011), “Multivariate matching methods that are monotonic imbalance bounding,” *Journal of the American Statistical Association*, 106(493), 345–61.

——— (2012), “Causal inference without balance checking: Coarsened exact matching,” *Political Analysis*, 20(1), 1–24.

Jahns, Ralf-Gordon, Egle Mikalajunaite, and Daniel Meehan (2011), “The Market for Mobile Application Development Services (2010-2015),” research2guidace.

Jain, D., V. Mahajan, and E. Muller (1995), "An approach for determining optimal product sampling for the diffusion of a new product," *Journal of Product Innovation Management*, 12, 124–35.

Jenkins, Stephen P. (2005), "Survival analysis," Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Johnston, John (1991), *Econometrics Methods*, McGraw-Hill.

Jones, R., and H. Mendelson (2011), "Information goods vs. industrial goods: Cost structure and competition," *Management Science*, 57, 164–76.

Lascu, Dana N. (1991), "Consumer guilt: examining the potential of a new marketing construct," *Advances in Consumer Research*, 18(1), 290–95.

Lehmann, Donald R., and Mercedes Esteban-Bravo (2006), "When Giving Some Away Makes Sense to Jump-Start the Diffusion Process," *Marketing Letters*, 17, 243–54.

Liu, Y. (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70, 74–89.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54, 1–26.

Mandel, M. (2012), "Where the Jobs Are: The App Economy," TechNet.

Marks, L.J., and M.A. Kamins (1988), "The use of product sampling and advertising: Effects of sequence of exposure and degree of advertising claim exaggeration on consumers' belief strength, belief confidence, and attitudes," *Journal of Marketing Research*, 25(August), 266–81.

Nagle, Thomas T. (1987), *The Strategy & Tactics of Pricing*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

Okada, Erica Mina (2005), "Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods," *Journal of Marketing Research*, 42, 43–53.

Perez, Sarah (2013), "Nearly 60K Low-Quality Apps Booted from Google Play Store in February, Points to Increased Spam-Fighting," *TechCrunch*, (accessed May 17, 2013), [available at http://techcrunch.com/2013/04/08/nearly-60k-low-quality-apps-booted-from-google-play-store-in-february-points-to-increased-spam-fighting/].

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2002), "Reliable estimation of generalized linear mixed models using adaptive quadrature," *Stata Journal*, 2(1), 1–21.

Rabe-Hesketh, Sophia, and Anders Skrondal (2012), *Multilevel and Longitudinal Modeling Using Stata, Volume II*, College Station, Texas: Stata Press.

Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles (2005), "Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects," *Journal of Econometrics*, 128(2), 301–23.

Rook, Dennis W. (1987), "The Buying Impulse," *Journal of Consumer Research*, 14(2), 189–99.

Russell, Gary J., and Wagner A. Kamakura (1994), "Understanding brand competition using micro and macro scanner data," *Journal of Marketing Research*, 289–303.

Shapiro, Carl (1983), "Optimal Pricing of Experience Goods," *The Bell Journal of Economics*, 14, 497–507.

Singh, Jasjit, and Ajay Agrawal (2011), "Recruiting for ideas: How firms exploit the prior inventions of new hires," *Management Science*, 57(1), 129–50.

Spence, A. Michael (1973), "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–74.

Sprïensma, Gert Jan (2012), "2012 Year In Review," Distimo.

Srivastava, R.K., T.A. Shervani, and L. Fahey (1998), “Market-based assets and shareholder value: A framework for analysis,” *Journal of Marketing*, 62, 2–18.

Sun, Monic (2012), “How Does the Variance of Product Ratings Matter?,” *Management Science*, 58(4), 696–707.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), “The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth,” *Journal of Marketing Research*, 45, 48–59.

Wallenberg, F. (2009), “Judging a Book by its Cover—Online Previews and Book Sales,” UC Berkeley School of Information Report, .

Wang, C.A., and X.M. Zhang (2009), “Sampling of information goods,” *Decision Support Systems*, 48, 14–22.

Weber, Lauren (2013), “How Your Smartphone Could Get You a Job,” *Wall Street Journal*.

Wertenbroch, Klaus (1998), “Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice,” *Marketing Science*, 317–37.

Whinston, Andrew B., Dale O. Stahl, and Soon-Yong Choi (1997), *The Economics of Electronic Commerce*, MacMillan Publishing Company.

Wortham, Jenna, and Nick Wingfield (2012), “To Fill Out Its App Store, Microsoft Wields Its Checkbook,” *The New York Times*.