ABSTRACT
A fundamental change brought forth by the advent of the mobile internet has been the widespread adoption of mobile phone based applications (apps). Mobile apps are now being used worldwide to perform a variety of tasks - access social networks, read ebooks, play games, listen to music, watch videos and so on. As consumers increasingly use mobile apps, it is important to estimate consumer demand for mobile apps and quantify the value created by these new apps and services. We build a structural model of user demand for mobile applications. We use a panel dataset consisting of top 300 ranked applications’ sales rank, prices, and characteristics data from the two leading app stores – Apple App Store and Google Android Market. We address the price endogeneity issue intrinsic in demand estimation by building a random coefficient logit demand model in a similar vein to the BLP (1995) method. Our results show that demand increases with the file size of apps and the length of app description, but decreases with the age of apps. We incorporate consumer heterogeneity in the model and find that older consumers tend to be more sensitive to the prices of apps than younger consumers.

Categories and Subject Descriptors
J.4 [Social and Behavioral Sciences]: Economics.

General Terms
Management, Economics.

Keywords
Mobile internet, mobile apps, demand estimation

1. INTRODUCTION
A fundamental change brought forth by the advent of the mobile internet has been the widespread adoption of mobile phone based applications (apps). Mobile apps are now being used worldwide to perform a variety of tasks - access social networks, read ebooks, play games, listen to music, watch videos and so on. According to a new Nielsen report, already one in four US adults have smart phones that are more powerful than the computers initially used to send men to the moon. Nielsen predicts that by the end of 2011, the majority of mobile subscribers in the US will have smart phones and will be spending most of their time on apps.

Mobile applications have existed for years, but clunky user interfaces on devices and hard-to-use application stores made it difficult for consumers to download applications. And then came Apple in 2008 with its App Store for the iPhone. The iPhone, which was easy to use, coupled with the App Store that allowed users to search and download apps from iTunes, turned mobile applications into an overnight success. Today, the App Store is considered the largest and most successful mobile application storefront out there. According to IDC (2010), in 2010 more than 500,000 applications were downloaded 10.9 billion times and in 2014 global downloads are projected to reach 76.9 billion downloads worth approximately US$35 billion.

As consumers increasingly use mobile apps, it is important to estimate consumer demand for mobile apps and quantify the value created by these new apps and services. Knowledge of heterogeneous consumer demand (i.e., how sensitive to a dollar increase or decrease in app prices) in mobile apps markets can help app developers and managers continuously improve their app features, better monetize their apps, and increase profits. For example, there are so called premium “glossing apps” which are becoming popular. TomTom Navigation app is $99.99 and Golfs: Golf GPS app is $29.99. Such glossing apps are about 4-times more expensive than most popular apps (Distimo 2011). This example demonstrates that some consumers are willing to pay more for additional features or/and high quality apps. Moreover, given that mobile carriers collect one-off or subscription fees for paid-apps, it can be critical for them to obtain precise estimates of user demand. This is because such demand-side information can be used to determine whether to develop mobile apps in-house or to outsource app developers, or both.

From a broader perspective, mobile technologies facilitate the delivery of many new products and services across mobile platforms. As these platforms develop and mature, it will be important to quantify their value for customers, firms and society. While much of the attention in academic research and in the press has been on examining user behavioral differences in the mobile channel versus traditional channels, we believe that important benefits lie in new products and services made available through these mobile platforms. While prior work has studied the impact of the mobile web on users’ multimedia content creation and consumption behavior (for example, Ghose and Han 2010, 2011), the value of new apps and services made available through mobile platforms has remained unquantified. This paper contributes by producing the first study that estimates a structural model of user demand in a mobile app setting.

In demand estimation, prices in general are correlated with the error term. Hence, the estimate of the price (sensitivity) will be biased due to the unobserved product characteristics. This is because prices are a function of marginal cost and a markup term, and moreover, the markup term is a function of the unobserved product characteristics, which is also included in the error term in the demand equation (Nevo 2010). We address the price endogeneity issue intrinsic in demand estimation by building a random coefficient logit demand model in a similar vein to the BLP (1995) method.

We use a panel dataset consisting of top 300 ranked app’s sales rank, prices, and characteristics data from the two leading app stores – Apple App Store and Google Android Market. Our results show that demand increases with the file size of apps and the age
of apps, but decreases with the length of app description. We incorporate consumer heterogeneity in the model and find that older consumers tend to be more sensitive to the prices of apps than younger.

2. PRIOR LITERATURE
In this section, we discuss multiple streams of relevant literatures such as user behavior in mobile media and user demand estimation from the introduction of new goods.

First, our paper builds on and relates to the literatures on user behavior in mobile media. A stream of relevant literature has discussed users’ usage patterns of voice calls and short message service (SMS) in the mobile phone setting. For example, Danahar (2002) and Iyengar et al. (2006) study how many phone call minutes are consumed under different pricing packages. Kim et al. (2010) examine to what extent the usage of mobile phone voice service can substitute short message service. In addition, our study is related to the emerging stream of literatures on user behavior in mobile internet. An emerging stream of relevant work has investigated the economic and social impact of user-generated multimedia content in the mobile internet by mapping the interdependence between content generation and usage (Ghose and Han 2011), modeling how consumers learn about different kinds of content (Ghose and Han 2010), documenting differences in search costs and location effects on mobile phones vs. PCs (Ghose et al. 2011), and examining the impact of network characteristics on social contagion on the mobile internet (Ghose, et. al. 2012).

Second, a long literature documents the models of demand estimation. One model that has made a significant contribution to the field is the random-coefficients discrete-choice model of demand (Berry et al. 1995, henceforth BLP). The BLP method is superior to the logit model because it can be estimated using only market-level price and quantity data and it deals with the endogeneity of prices (Nevo 2000). To our knowledge, no previous study has examined a structural model of user demand in a mobile app setting. Our paper aims to fill this gap in the literatures.

3. DATA DESCRIPTION
We use a panel dataset consisting of top 300 ranked applications’ sales rank, prices, and characteristics data from the two leading app stores – Apple App Store and Google Android Market. We collected the data from South Korea market between October 6, 2011 and December 14, 2011 (76 days).

Our dataset includes app-level, daily data on apps’ prices and sales quantity from the app store and app characteristics. Since neither app stores provide information on the download of apps, we first calibrate the relationship between ranks and sales using an additional panel data in which we have information on ranks and actual download of apps from a mobile carrier’s app store. Then we predict download of apps in both Apple App Store and Google Android Market, which are used in actual demand estimation. In addition, we have aggregate-level information on user demographics such as age and gender for each of two app stores. Further, we have category-level app market share data from app stores other than two leading ones. We use such information to compute the total market size. Table 1 shows the summary statistics of the key variables used in our model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>App price (US$)</td>
<td>3.42</td>
<td>9.91</td>
<td>0</td>
<td>449.99</td>
</tr>
<tr>
<td>App size (Mega Bytes)</td>
<td>67.34</td>
<td>209.31</td>
<td>1</td>
<td>1910.4</td>
</tr>
<tr>
<td>App age (Days)</td>
<td>311.59</td>
<td>193.31</td>
<td>1</td>
<td>1248</td>
</tr>
<tr>
<td>App description (100 chars)</td>
<td>22.17</td>
<td>12.548</td>
<td>0.26</td>
<td>96.68</td>
</tr>
</tbody>
</table>

4. MODEL
In this section, we discuss how we build our random coefficient structural model and describe how we estimate it.

In our model, the utility for consumer i from choosing app j in market t can be represented as:

$$ u_i = X_i \beta + a_i \xi + \epsilon_i , $$

where $X_i$ is a vector of observable characteristics of app j in market t and $\beta$ is a vector of the random coefficients (i.e., taste parameters) associated with those app characteristics. The observed app characteristics in our sample include:

• app size (mega bytes),
• app age (time elapsed since app launch),
• app categories, and
• length of textual app description by the app developer (number of characters).

$P_t$ is the price of app j in market t and $a_i$ is a random coefficient that captures consumers’ heterogeneous tastes towards price. $\xi_i$ represents the unobserved (by researchers) characteristics of app j. Unobserved app characteristics, for example, can include the impact of unobserved marketing promotion or systematic shocks to demand. Lastly, $\epsilon_i$ is a mean-zero stochastic term.

Consumer i chooses app j that yields maximal utility. Market shares are obtained from aggregating over consumers. We define a “market” as the combination of an “app store-day.” Correspondingly, the market share for each app is calculated based on the number of apps purchased for that app in that market divided by the “total size of that market.” It is important to note that consumers in our data can access only two types of app stores: the Apple’s App Store and the Android market. With regard to market size, we define the total number of apps purchased in a certain market based on data both from these two leading app stores and from other app stores (i.e., third-party app stores, mobile device-based app stores). Hence, the outside good is defined as “no purchase from the two leading app stores but from other app others during a given day.”

Our model captures taste heterogeneity of users by incorporating observed and unobserved individual characteristics. Formally, this is modeled as:

$$ \left( \frac{\hat{\alpha}}{\hat{\beta}} \right) = \frac{\hat{\alpha}}{\hat{\beta}} + \Pi D + \Sigma \nu, $$

The vector ($\hat{\alpha}, \hat{\beta}$), which is referred to as the mean utility of price and app characteristics, is common to all consumers. It measures the average weight placed by the consumers. In addition, $D$ is a vector of demographic variables that include user age and gender. $\Pi$ is a matrix of coefficients that measure how the taste
characteristics (e.g., whether an app is provided by a third-party developer, app size, app age, app categories) vary with demographics. Further, \( v_i \) captures the unobserved characteristics. Similar to the literature in demand estimation (Berry et al. 1995, Nevo 2001), we assume that \( v_i \) has a standard normal distribution, and the vector \( \Sigma \) allows for each element of \( v_i \) to have a different standard deviation (Nevo 2000). Our goal is then to estimate the mean utilities vector \( \mu \), \( \beta \), \( \Pi \) matrix of coefficients, and the standard deviations in vector \( \Sigma \).

Combining equations (1) and (2), our model is as follows:

\[
\mu_t = \delta_t \left( X_t, P_t, z_t, \bar{u}_t \right) + \mu_t \left( \alpha_t, \beta, D_t, v_t \parallel \Pi, \Sigma \right) + \epsilon_t ,
\]

where \( \delta_t = X_t \beta + \alpha_t + \epsilon_t \) representing the mean utility of app \( j \) in market \( t \) and \( \mu_t = X_t (\Pi \beta + \Sigma v_t) \) representing a mean-zero deviation from the mean utility level. Given that we have price, constant, 4 app characteristics, and 4 app category dummies, \( [P_t, X_t] \) is a 1 x 10 row vector. We also include control variables such as app dummies and cumulative number of user reviews on apps.

5. Empirical Analysis and Results

In this section, we present our key results on demand estimation. The results of the estimate are in Table 2. Estimates of the mean utility levels for each app characteristic are presented in the first column. Our results show that demand increases with the file size of apps and the age of apps (time since release), but decreases with the length of app description. Then, in the second column, the standard deviation captures the effects of heterogeneity around the mean utility levels due to the unobserved demographics. Lastly, the last two columns present the effect of demographics on the mean utility levels. For example, the estimate of interaction between app price and user age (i.e., -0.006) suggests that while the average consumer is sensitive to the price of apps, older consumers tend to be less price sensitive than younger consumers.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (Std. dev.)</th>
<th>Interactions with Demographic Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Age (Male)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.282** (0.076)</td>
<td>-0.006** (0.003)</td>
</tr>
<tr>
<td></td>
<td>0.073** (0.034)</td>
<td>-0.072** (0.032)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.334*** (0.019)</td>
<td>-0.302 (0.229)</td>
</tr>
<tr>
<td></td>
<td>0.084 (0.239)</td>
<td>-0.071 (0.402)</td>
</tr>
<tr>
<td>App size</td>
<td>0.339*** (0.020)</td>
<td>0.894** (0.445)</td>
</tr>
<tr>
<td>App age</td>
<td>1.224*** (0.018)</td>
<td>0.942 (0.710)</td>
</tr>
<tr>
<td>App description length</td>
<td>-1.425*** (0.021)</td>
<td>0.724 (0.572)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>23,145</td>
<td></td>
</tr>
<tr>
<td>GMM Objective Value</td>
<td>1.46</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Asymptotically robust standard errors are given in parentheses. ** denotes significant at 0.05, *** denotes significant at 0.01.

6. Conclusions

Mobile applications are a rapidly growing segment of the global mobile market. Various factors that have been contributing to that growth include advancements in network technologies, the lowering of mobile data usage cost, the growing adoption of smart phones around the world, and a continuous increase in application usability. As consumers increasingly use mobile apps, it is important to understand the underlying drivers of user demand for mobile apps. In this paper, we estimate a structural model of user demand in a mobile app setting.

Our results show that for the average consumer demand increases with the file size of apps and the age of apps, but decreases with the length of app description. In addition, older consumers tend to be more price sensitive than younger consumers.

Data availability issues suggest that some caution is warranted in the demand estimation. For example, we do not have information about consumers’ mobile internet rate plans (i.e., fixed monthly fees with unlimited internet access, usage-based fee), thereby transmission charges which incur when consumers download apps using their phones. Furthermore, the mobile app store in our data is based on apps built upon the Apple App Store and the Google Android Market in South Korea and it is possible that the magnitudes of the consumer demand will vary across platforms and countries. Finally, companies can determine their app prices and app characteristics, hence such variables can be econometrically endogenous in empirical demand systems. Future research can examine endogeneity in the supply side.

Notwithstanding these limitations, our analysis documents estimates for demand function for mobile apps. To the extent that prices and product characteristics of mobile apps affect market outcomes, the increasing size of the mobile app store may have profound implications for the future direction of mobile commerce.

7. REFERENCES


