Specification Document

Complementarity and social aspect of App-Store:

Empirical Study of an African App-store

Author: Meisam Hejazi Nia

Advisor: Dr. Norris Bruce

Date: 07/02/2014

**Abstract**

Fueled by widespread smartphone adoption and increasing mobile internet access the mobile app market has experienced significant growth during the last 3 years. The mobile app market reached 64 billion download in 2012 up from 8.2 billion downloads in 2010 and its total revenue in 2012 was $18 billion up from $5.2 billion in 2010. In this market three types of agents operate: (1) platform owners (2) app developers and (3) consumers. For the platform owner that seeks the diffusion of its platform, it is important to understand which app-category is more relevant to drive the diffusion of its platform, and what app-categories are substitutes. In addition from policy stand point platform owner may be interested to know the effect of paid options and the assortment size on consumer’s decision. As mobile phones are ubiquitous, individual’s preference for an app can influence the preference of others in many ways, ranging from psychological benefits of social identification and inclusion to the benefit of network externalities. In other word individual decision to adopt a product may be driven by aggregate process of population adoption. In this study we develop a hierarchical Bayesian spatial non-linear state space structural model[[1]](#footnote-1) to study the individual preference’s dependence on aggregate population adoption and level of substitution and complementarity of different app-categories and app choices. We test our model using 11 month unique individual panel data of consumer purchases in a newly launched app store in Africa[[2]](#footnote-2). We run counterfactual analysis to find implications of different platform owner’s incentives for the diffusion of different app categories in the app store, optimal assortment size, and optimal structure for app-store.

Keywords: Mobile App, Big Data, Interdependence of Consumer Preference, Consumer misperception, Bounded Rationality, Social Learning, Diffusion Model, App Category Complementarity, Choice Model, Structural Model, Hierarchical Mixture Model, State Space Model, Factor Model, Metrapolis Hasting, Extended Kalman Filter

**Background**

This study draws upon several streams within the literature that have investigated: *i*) Interdependence of consumer preference; *ii*) Product Category Complementarity; iii) Mobile app store dynamics; and *iv*) new product diffusion models in general. Given the breadth of these areas across multiple disciplines, what follows is only a brief review of these relevant streams.

**Interdependence of consumer preference**

Quantitative models of consumer purchase behavior often do not recognize that consumer choices may be driven by the aggregate diffusion of awareness and preferences. Economic models of choice typically assume that an individual’s latent utility is a function of brand and attribute preferences, not the preferences of others, yet in the case of free experience goods, when there is no pricing mechanism, a new model which accounts for aggregate diffusion of awareness and preferences may explain the choice of consumers better.

Preferences and choice behavior are influenced by a consumer’s own tastes and awareness, and also the taste and awareness of others. People who live in the same geographical area often become aware of an innovation sooner and their preference are more coherent, resulting in the choices that are interdependent. Interdependence may be driven by social concern, by endorsements from respected individual that increase a brand’s credibility, or by learning the preference of others through diffusion of information. Many studies tried to address this issue including Yang and Allenby (2003), which uses cross sectional data to model preference dependanc. Stephen and Toubia (2010) studies value implications of social network between sellers in a large online social commerce marketplace. Lehmmes and Croux (2006) apply bagging and boosting classification techniques to a customer database of an anonymous U.S. wireless telecommunication company, and find that both methods significantly improve accuracy in predicting churn.

Bell and Song (2007) examine customer trials at Netgrocer.com and drawing on studies in marketing and economics conjecture that exposure spatially to proximate others (through direct social interaction or observation), can influence decisions of those who have yet to try, find that trials arise from utility-maximizing behavior and they find that neighborhood effect is significantly positive and economically meaningful. Hartmann et al. provides a selective review of how insights from other discipline can improve and inform modeling choices of social interactions. Aral and Walker (2011) present a method that uses in vivo randomized experimentation to identify influence and susceptibility in networks while avoiding the biases inherent in the traditional estimates of social contagion to identify social influence in the networks.

Nair et al. (2010) quantify the impact of social interaction and peer effects in the context of physicians’ prescription choices, using detailed individual-level prescription data, along with self-reported social network information, and find that physicians prescription behavior is significantly influenced by the behavior of research-active specialists, or “opinion leaders,” in the physician’s reference group. Bradlow et al. (2005) reviews literature on generalizing the notion of a map to include demographic and psychometric representations, to argue for spatial models to capture a variety of effect that impact firm or consumer decision behavior. Hartmann (2010) develops a model for demand estimation in the context of social interactions. It models the decision of a group of customers as an equilibrium outcome of an empirical discrete game, in which group members must be satisfied with chosen outcome. Yang et al. (2005) estimate the preference independency in television program viewership between spouses using simultaneous equation model, and show that the models that ignore interdependency of preference yield biased estimates of consumers sensitivity to observed attribute preferences.

Narayan et al. (2011) studies how multi-attribute product choices are affected by peer influence, and propose two-stage conjoint-based approach to examine three behavior mechanisms of peer influence, finding that when faced with information on peer choices, consumers update their attribute preferences in Bayesian manner. Kurt et al. (2011) investigate the interactive influence of the presence of an accompanying friend and consumer’s agency communion orientation on the consumer’s spending behavior, finding that shopping with friend can be expensive to agency-oriented consumer (e.g. males) but not for communion-oriented consumer (e.g. females). Chung and Rao (2012) present a general consumer preference model for experience products that allows for situations in which it is not easy to consider some qualitative attributes of a product or when there are too many attributes relative to the available amount of preference data, by capturing the effects of unobserved product attributes with the residuals of reference consumer for the same product. Choi et al. (2010) develop a Bayesian spatio-temporal model to study two imitation effects in the evolution of demand at an Internet retailer, allowing imitation behavior to be reflected both in geographic proximity and in demographic proximity, finding that the proximity effect is especially strong in the early phases of demand evolution, whears the similarity effect becomes more important with time.

**App Category Complementarity**

Sriram et al. (2007) presents a framework of durable good purchasing behavior in related technology product categories, and using a household level information on category-level firmst-time adoption decisions in three categories—personal computer, digital cameras, and printers--- it finds strong complementary relationship between the three categories. Liu et al. (2010) studies complementarities and the demand for home broadband internet services, with an eye on the demand side explanation rather than supply side for large share of phone companies in the DSL broadband market, and it finds evidence of strong complementarities between the consumption of broadband and its related categories. Yang and Ching (2013) develop a structural consumer life-cycle model to investigate consumer’s adoption and usage decision of ATM cards, and using data on ATM cards adoption in Itly, it finds that one could significantly overestimate adoption costs for the elderly when ignoring their shorter life span.

Ma et al. (2012) models household’s contemporaneous brand choice outcome in complementary product categories, explicitly capturing cross-category dependencies in brand choice outcome of household, and using scanner panel data on cake mix and frosting categories, it finds that complementarity accounts for the vast majority of the estimated cross category effects in demand. Karaca-Mandic (2011) estimates complementarities between DVD player adoption and availability of content on DVD, and using household level quasi-panel data on adoption decision of households, onthly data on availability of movies for rental on DVD at local regions and on the number of movies released on DVD nationally, it finds significant complementarity.

Lee et al. (2013) propose a direct utility model with a latent decision sequence for measuring asymmetric complementarity effects to capture differential response to cross-category purchases and inventories, and using scanner panel data on milk and cereal purchases, it investigates cross-price elasticities and spillover effects of merchandising variables. Niraj et al. (2008) models more than one purchase decision of the household simultaneously, through two-stage bivariate model of incidence and quantity outcome in multiple categories, and it finds that cross-category promotional spillovers are asymmetric between the two product categories of bacon and eggs. Wedel and Zhang (2004) develop a sales model that can be used to assess cross-category price effects that are specific to individual items (stockkeeping units) from store-level sales data.

Song and Chintagunta (2006) study the cross-category effects of marketing activities using aggregate store-level scanner data on four product categories--- liquid laundary detergents, powdered laundry detergents, and liquid fabric softeners, and sheet fabric softeners—, and it finds evidence for a complementary relationship between liquid softeners and both forms of detergents. It also finds that the magnitude of cross-category elasticities are brand specific, i.e., different brands in a category have a different price impact on the demand for a brand in another category.

Oestreichter-Singer and Sundrararajan (2008) tests the conjectures that the explicitly visibility of “product networks” can alter demand spillover across their constituent items on the data about copurchase networks and demand levels associated with more than 250,00 interconnected books offered on Amazon.com, finding that on average the explicit visibility of a copurchase relationship can lead to up to an average threefold amplification of the influence that complementary products have on each others’ demand level. Bezawada et al. (2009) investigate the effects of aisle and display placement on the sales affinities between categories, by developing a spatial model of brand sales that allows for asymmetric store-specific affinity effects between two or more categories.

**Mobile app store dynamics**

Ghose and Han (2014) build a structural econometric model to quantify the vibrand platform competition between mobile (smartphone and tablet) apps on the Apple iOS and Google Android platforms and estimate consumer preference toward different mobile app characteristics, finding positive effect of in-app purchase option wherein user can complete transactions within the app, and negative effect of in-app advertisement option. Carare (2012) uses daily data on the ranking by sales of the top 100 apps sold through Apple’s App Store, and it finds that lagged bestseller rank information has causal effect on mobile apps’ demand. Garg and Telang (2013) present an innovative method to use public data to infer the rank-demand relationship for the paid apps on Apple’s iTune App store, and it finds that the top-ranked paid app for iPhone generates 150 times more downloads compare to the paid app ranked at 200. Liu et al. (2012) examines the effect of the freemium strategy on Google play, and by analyzing a large panel dataset, it finds that the freemium strategy is positively associated with increased sales volume and revenue of the paid apps.

Ghose et al. (2011) explores how internet browsing behavior varies between mobile phones and personal computers, and arguing that smaller screen sizes on mobile phone increase the cost to the user of browsing for information, and wider range of offline locations for mobile internet usage that causes importance of local activity, it finds significant ranking effect and more geographic match relevance on mobile internet. Ghose and Han (2011b) develops and estimates a dynamic structural model of user behavior and learning with regard to content generation and usage activities in mobile multimedia environments, in which user learn about two different categories of web content: social network and community and mobile portal, and it allows for both direct experience and indirect experience from neighbors, finding that learning based on direct experience is more accurate (has less variability) than learning based on indirect experience. Ghose and Han (2011a) quantify how user mobile internet usage relates to unique characteristics of the mobile internet, focusing on examining how the mobile-phone-based content generation behavior of users relates to content usage behavior, analyzing whether there is positive or negative interdependence. Kim et al. (2008) develop a structural model to examine user demand for voice service and SMS, and it measures the own- and cross-price elasticity of these services, finding cross price elasticity significant due to complementarity and substitution.

**New product diffusion models**

Chaterjee and Eliashberg (1990) develop a model of innovation diffusion using micromodeling approach that explicitly consider the determinants of adoption at the individual level in a decision analytic framework, incorporating heterogeneity in the population with respect to initial perceptions, preference characteristics, and responsiveness to information. More recently, Young (2009) examines three broad classes of diffusion models—contagion, social influence, and social learning—and shows how to incorporate heterogeneity into each at a high level of generality without losing analytical tractability. Lelarge (2012) studies a model of diffusion where the individuals behavior is the result of strategic choice, using simple coordination game with binary choice, and it describes the condition of a new action to become widespread in a random network.

Iyengar and Van den Bulte (2011) studies how opinion leadership and social contagion within social networks affect the adoption of new product, and find evidence of contagion operating over network ties, even after controlling for marketing efforts and arbitrary system wide changes. Moon and Russel (2008) develops a product recommendation mechanism based on principal of customer preference similarity stemming from prior purchase behavior is a key element in predicting customer product purchase. Song and Chitagunta (2003) develops an empirical model for the adoption process of a new durable product that accounts for consumer heterogeneity as well as consumers’ forward-looking behavior, and it decomposes the effect of the entry of Sony into primary demand expansion and switching from other brands, finding that there exist several segments in the market with different preferences for the brand and different price sensitivities leading to differences in adoption timing and brand choice across segments.

Dover et al. (2012) shows how networks modify the diffusion curve by affecting its symmetry, and it demonstrates that network’s degree distribution has a significant impact on the contagion properties of the subsequent adoption process, proposing an agent based method for uncovering the degree distribution of the adopters network underlying the dissemination process. Trusov et al. (2013) proposes an approach for identifying systematic conditions that are stable across diffusions, transferable to new product introduction within a given network, using Facebook applications data, , and it shows that incorporation of such systematic condition improves prelaunch forecast. For more complete review of models please review Peres et al. (2010).

This study is different from above mentioned studies in that it uses a unique long time series of a panel of consumers over a course of 11 month over free mobile apps, which allows us to identify to identify the diffusion of awareness and preferences of consumers in different geographical locations. In addition multiple app categories over a long time allow us to identify the relationship between different app categories. In addition, to the best of our knowledge we are the first to model the evolution of social preference in consumer choice.

**Model**

We observe choice information for a set of individuals. In contrast to prior studies such as Carare (2011) do not account for consumers’ decision within the app category, we start our model with app category. More specifically, we assume consumers go through two stages of decision making: First consumer selects the product category that she is interested in among the available product categories, where  denotes outside good option, and then given the app category in the second stage consumer selects among available apps in the app category, where  denotes outside good option or any app other than the apps under this study. In this specification we assume that if a consumer selects an app category, she will definitely download one of the apps. This modeling decision may be justified by referring to sunk opportunity cost of consumer for decision making, and tiny cost of download. More formally we can define the probability of purchase of app in the category as:



Consumer information processing theory and household production model, assume that consumer considers perceived risk of a purchase decision, and she is more prone to select a product that she has more familiarity with. These studies such as Ratchford (1982) and Kok and Xu (2011) suggest that consumer decision may follow nested logit structure. Our different formulation of app choice probability by conditioning on category, allows us to have a more structural model which allows us to study search behavior of consumer over different app categories on app store. Our two stage modeling approach is close to the specification of Kok and Xu (2011) on consumer’s decision among product types in the first stage and among the brands for a given product type in the second stage. Figure below shows consumer’s decision to download an app. Another interpretation of our model is that first consumer faces a universal set which she uses a screening rule to filter. We assume the screening rule is selecting the product category. Other screening rules could be to select one of the attributes as Gilbrid and Allenby (2004) suggest.

We specify the utility of consumer purchase among app categories in the app store in the following form:

 (1)

|  |  |
| --- | --- |
| Variable | Definition |
|  | Utility of consumer i, from purchasing in category j at time t, given location |
|  | Consumer i’s perception of goodwill of app category j at time t, given location l |
|  | State variable that summarizes the history number of app’s that consumer i has downloaded in category j until now (to account for state dependency) |
|  | Variance of price of mobile app’s in app category j at time t. According to search theory, it indicates more uncertainty, so it should have negative effect. |
|  | Mean of file size of mobile app’s in app category j at time t (first for the limitation of mobile phone memory size, second as the proxy for complexity or quality of mobile app) |
|  | Mean of the tenure of mobile app’s in app category j (proxy for maturity of app’s in the category) |
|  | Total number of free mobile app’s available in app category j at time t, while search theory suggests lower is better, product differentiation suggests more is better. |
|  | Total number of paid mobile app’s available in app category j at time t. |
|  | Parameters to estimate |

We accounted for state dependence in our model by incorporating the history of consumer downloads of apps in categories j from inception until time t (). If consumer downloads an app from app category j in previous occasion, t-1, then  , otherwise if the consumer selects outside option, then  remains unchanged: . This specification induces a first-order Markov process on the choices. Furthermore, our incorporation of goodwill of app category j in location l observed by consumer i at time t is an alternative approach to the approach used by Yang and Allenby (2003) to incorporate interdependence of awareness and preference (perhaps due to observational learning in offline world as mobile phones are ubiquitous) of consumers.

Assuming that random utility term, , is type 1 extreme value distribution, consumer’s i in location l app category choices j at time t are given by a multinomial logit model as follows:

 (2)

Here, we assume that the mean utility of outside good is zero, .

We assume that consumers have misperception about the real goodwill of each mobile app category. Erdem and Keane (1996) assumes that misperception error of consumer can be modeled with normal distribution, and many structural modelers follow the same approach e.g. Zhao et al. (2013), so we find the assumption of normal misperception error innocuous. More formally, we assume that goodwill of a mobile app category can be modeled as:

(3)

Where  and  are parameters to estimate, and  is common factor across individuals. In order to identify the model we set .

Further, in order to reduce the size of our estimation, we use factor model for the good-will of app categories across different locations in the following form:

(4)

Where and  are parameters to estimate, and  is common factor across locations. Where in order to identify the model we set .

Bass diffusion model (1969) suggests specific hazard rate functional form for that the diffusion of adoption. Here we may be able to consider that goodwill of the app also follows the same form, and given the model of Naik, Prasad and Sethi (2008) and Simon and Horskey (1983) of competition to find the effect of complementarity or substitution effect, we adopt the following form for the diffusion of each app category:

 (5)

|  |  |
| --- | --- |
| Variable | Definition |
|  | cumulative goodwill of each app category j at day t |
|  | market size of the app category j |
|  | error term of diffusion of category j at time t |
|  | Coefficient of substitution or complementarity of app category k and j |
|  | External market diffusion factor for app category j |
|  | Internal market diffusion factor for app category j |
|  | Variance parameter to estimate |

In summary equation (4) is our observation equation and equation (5) is our state equation in app category state space choice model.

We account for heterogeneity in individual and category parameters, through a random effect model. In order to avoid misspecification of the distribution of consumer heterogeneity in preferences, and allow for multimodal densities of preferences across groups (or “segments”) of consumers, we adopt the method proposed by Dube et al. (2010) to use a flexible semi-parametric heterogeneity specification consisting of a mixture of multivariate normal distributions, to allow for thick tail, skewness, and multimodality. We denote vector of non-state consumer-level parameters by  and vector of non-state category level parameters by . We accommodate consumer and category heterogeneity by assuming that  and are drawn from two separate distribution common across consumers and common across categories. We employ mixture of normal as the first stage prior, to specify an informative prior, avoiding the problem of over fitting. The first stage consists of a mixture of  multivariate normal distributions respectively, and the second stage consists of priors on the parameters of the mixture of normal:

 (6)

 (7)

 (8)

 (9)

Where and represents the hyper-parameters of the priors on the mixing probabilities and the parameters governing each mixture component. To obtain a truly non-parametric estimate using the mixture of normal model requires that the number of mixture components  increase with the sample size. We adopt the approach proposed by Dube et al. (2010) to fit models with successively large numbers of components and to gauge the adequacy of the number of components by examining the fitted density associated with each number of components.

Consumer searches apps of their favor given the app category, so we specify from location l, consumer i’s utility from purchasing mobile app a in category j, at time t, according formally as:

 (10)

|  |  |
| --- | --- |
| Variable | Definition |
|  | Utility of consumer i, from purchasing app a in category j at time t , given location |
|  | Consumer i’s perception of goodwill of app a in app category j at time t, given location l |
|  | Dummy variable indicating whether app a is featured at time t |
|  | Tenure of mobile app a at time t(number of days from creation) |
|  | State variable that summarizes the history of consumer i download of app a in category j until now (to account for state dependency) |
|  | File size of mobile app a in app category j at time t (first for the limitation of mobile phone memory size, second as the proxy for complexity or quality of mobile app) |
|  | Parameters to estimate |

We accounted for state dependence in our model by incorporating the history of consumer I downloads of apps a in categories j from inception until time t (). If consumer downloads an app from app a in category j in previous occasion, t-1, then  , otherwise if the consumer selects outside option, then  remains unchanged: . This specification induces a first-order Markov process on the choices. Like app category portion of our model, our incorporation of goodwill of app a in location l observed by consumer i at time t is an alternative approach to the approach used by Yang and Allenby (2003) to incorporate interdependence of awareness and preference (perhaps due to observational learning in offline world as mobile phones are ubiquitous) of consumers.

Assuming that random utility term, , is type 1 extreme value distribution, consumer’s i in location l app choices a in app category j at time t are given by a multinomial logit model as follows:

 (11)

Here, we assume that the mean utility of outside good (app’s that we do not consider) is zero, .

Like app category model, here we also assume that consumers have misperception about the real goodwill of each mobile app. Formally, we assume that goodwill of a mobile app can be modeled as:

(12)

Where  and  are parameters to estimate, and  is common factor across individuals. In order to identify the model we set .

Further, in order to reduce the size of our estimation, we use factor model for the good-will of apps across different locations in the following form:

(13)

Where  and are parameters to estimate, and  is common factor across locations. Where in order to identify the model we set .

Similar to app category diffusion and to account for complementarity and substitution effect between different app’s in the category we specify app diffusion in the following form:

 (14)

|  |  |
| --- | --- |
| Variable | Definition |
|  | cumulative goodwill of each app a at day t |
|  | Market size of mobile app a in category j at time t |
|  | error term of diffusion of app a at time t |
|  | Coefficient of substitution or complementarity of app k and a |
|  | External market diffusion factor for app a |
|  | Internal market diffusion factor for app a |
|  | Variance parameter to estimate |

In summary equation (13) is our observation equation and equation (14) is our state equation in app state space choice model.

We account for heterogeneity in individual and app level parameters similar to the approach we used for app category. We denote vector of non-state consumer-level parameters by  and vector of non-state category level parameters by . We accommodate consumer and category heterogeneity by assuming that  and are drawn from two separate distribution common across consumers and common across apps. We employ mixture of normal as the first stage prior, to specify an informative prior, avoiding the problem of over fitting. The first stage consists of a mixture of multivariate normal distributions respectively, and the second stage consists of priors on the parameters of the mixture of normal:

 (15)

 (16)

 (17)

 (18)

Whererepresents the hyper-parameters of the priors on the mixing probabilities and the parameters governing each mixture component. To obtain a truly non-parametric estimate using the mixture of normal model requires that the number of mixture components  increase with the sample size. Just like category level parameters, we adopt the approach proposed by Dube et al. (2010) to fit models with successively large numbers of components and to gauge the adequacy of the number of components by examining the fitted density associated with each number of components.

This closes our model specification. The Direct Acyclic Graph (DAG) of conditional distributions is presented in the following figure:



























































Mobile app

Mobile app category

**Extension**

To extend this research we plan to add supply side down the road, to understand parameters of choices of app-developers that determine her entrance into specific app-category. Another admissible extension is to model the diffusion of app and app category in groups of users with similar purchase behavior. In other word, at each point in time we can cluster users based on their purchase behavior and study how each mobile app and app category diffuses in each of those clusters. Another extension to this study is to model consumer learning in the category, as one may posit that consumers who purchase from a category may have less misperception about the category of product evolution, as more purchase may be correlated with more active search behavior. Another extension could be to find interaction between diffusion of different locations, identifying opinion leadership across locations. Another extension could be to consider that consumers are forward looking and form step-ahead expectation about the goodwill of a mobile app and the mobile app category.

**Data**

The data were collected by an African telecom operator on downloads of mobile apps’ on the app store of its global partner. The app-store is launched within 11 month prior to our study, so we have 11 month daily downloads of mobile apps on the app store. Apps are categories into different categories, and each app is categorized into one category. We selected 17 different app categories, and within each category we selected app’s that have significantly greater number of downloads given their category during the period of study, total of 68 apps. This choice allows us to assume that the app’s are in an oligopoly market. We obscure the identity of some apps for the purpose of confidentiality. Following table shows the list of the categories that we select and their corresponding apps:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| index | Category | Downloads in the category | Number of apps in our sample | Name of the apps  (total downloads) |
| 1 | Action/Strategy | 9,803 | 3 | Danfo Reloaded II (166)  RICO (311)  LasGidi Run (8,829) |
| 2 | Books and References | 1,179 | 6 | aYoruba (52)  aHousa (41)  E-Country (39)  MobileDiary (36)  Students and Workers diary (30)  NewsHunt (109) |
| 3 | eBook | 4,125 | 5 | Country’s Constitution (153)  Open Heavens 2013 (138)  Okadabooks (638)  Country’s Constitution (358)  Open Heavens 2014 (2740) |
| 4 | Games | 1,619 | 7 | Gangstar City (22)  Taxi Driver (21)  MobiLotto (24)  Stonehunt (149)  Hill Climb Racint 2-Free (58)  Z Runner Free (25)  Moto Fever (53) |
| 5 | Health/Diet/Fitness | 1,900 | 3 | Doctor (886)  Home Workout Program (207)  CareCheck (89) |
| 6 | Humor/Jokes | 1,443 | 4 | Chat with Annabel (534)  Akpors Comedy (345)  Country Comedians (209)  N-MEME for Android (180) |
| 7 | Internet/WAP | 1,770 | 3 | UC Browser (90)  One Browser (90)  Operator’s Info APP (1246) |
| 8 | Logic/Puzzle/Trivia References | 1,367 | 5 | Huzzle Puzzle Free (32)  Gidigames (36)  Current Affairs Quiz (588)  Word Search Swype-Puzzle Game (81)  Keke Marwa Slice (94) |
| 9 | Movie/Trailers | 9,275 | 3 | Afrinolly (65)  Nollywood Movies and News (70)  Dobox (9,018) |
| 10 | Radios/Players | 2,446 | 2 | Operator’s Songstar (2,371)  Eah Radio (42) |
| 11 | Reference & Dictionaries | 1,498 | 2 | Fact About Life\_and\_Physics\_Dictonary (1,266)  Explore Country (134) |
| 12 | Social Networks & Messaging | 1,454 | 6 | Fring (164)  MO-Call Mobile VoIP (50)  Jobberman (65)  WiFi hacker (53)  Zeeders Social for Blackberry (51)  ChatZone (41) |
| 13 | TV/Shows | 2,485 | 2 | Live Mobile TV (2,255)  Online TV (92) |
| 14 | University | 2,974 | 7 | Exam Keypoints (157)  Calculator (273)  Grade Point Calculator (299)  JAMB CBT (416)  Onpoint: gpa calc & timetable (381)  Fusion (318)  Country’s Educational Grading System (122) |
| 15 | Utilities/Synchronization | 1,525 | 6 | PixMix-Photo Sharing (102)  Traffic Reporter (82)  Trace (70)  Birthday Greetings (75)  Kedu (163)  Nepasituation (174) |
| 16 | Video & TV | 12,904 | 2 | Youtube Downloader Video (1,529)  SuperTube (10,941) |
| 17 | Social Network |  | 2 | Whatapp Messenger (174)  Viber (111) |

We investigate geographic source of dependence of awareness and preference of consumers through good will of the mobile app. Geographic neighborhood is created by physical proximity, through mapping IP address of each consumer to the corresponding city. For this purpose we wrote a crawler and scraper in Perl to scrape the location of each corresponding IP in an IP-location mapping website[[3]](#footnote-3). We consider 28 locations based on the density of download. Following table summarizes these locations:

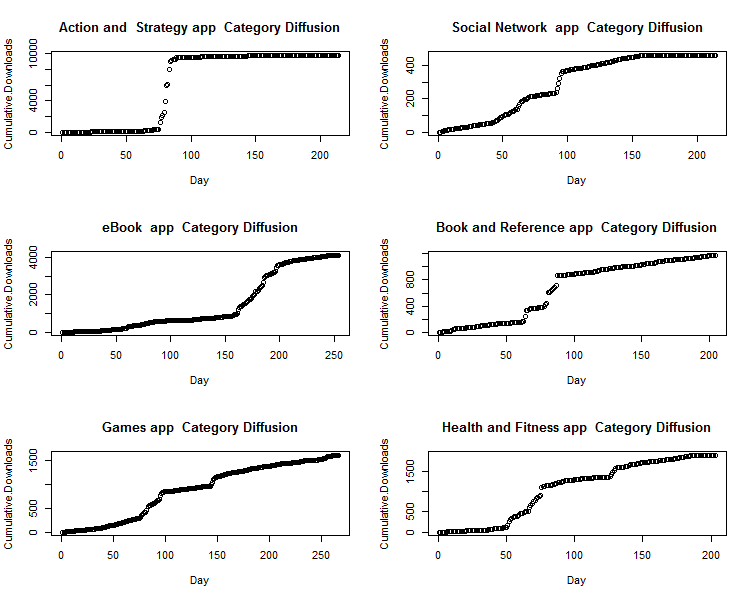
|  |  |
| --- | --- |
| Continents | Number of Towns within in our sample |
| Africa | 11 |
| United States | 8 |
| Europe | 7 |
| China | 1 |

Our data also included longitude and latitude of each IP address, yet as mobile phones are nomadic and ubiquitous, we find the choice of city a more reasonable one. Social nature of mobile phone and mobile apps is another reason which may reinforce that our assumption is innocuous. As mobile phone has become inseparable part of humanity, to the point that not only they use it when they are alone in bus, before sleeping, or even in the class, but also they use it in parties, in office, and in the leisure time, and in social events. Usage in social events makes mobile apps visible, so it creates opportunity of learning in many social events.

The data consists of 18,297 consumers, live in those 27 towns, and with 57,767 unique downloads daily for the course of 11 month in 2013 and 2014. Therefore, in terms of size of the data, we can refer to it as a Big data. Following table shows the basic statistics of data at mean of category level:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category Data Summary | Mean | Variance | Min | Max |
| Number of available apps in the Category | 137.958 | 289.571 | 12.336 | 1227.617 |
| Average Tenure of App’s in the category (Days) | 297.723 | 65.017 | 169.845 | 413.913 |
| Number of available free apps in the category | 121.803 | 257.428 | 7.976 | 1090.103 |
| Average day’s that an app is featured in the category | 0.358 | 0.528 | 0.000 | 1.487 |
| Average file size of app’s in the category | 2537388.812 | 1834221.933 | 647681.122 | 7254383.845 |
| Variance of prices of app’s in the category | 3.184 | 9.258 | 0.000 | 38.673 |

In the following figure we draw diffusion curve of cumulative adoption of each sample of 6 app categories:



For each user we have created time, and number of downloads. This allows us to distinguish between heavy and light users, so we can probably use it to explain heterogeneity of users. For each download we know source of download, whether the mobile app is featured and the device name. For each location we can collect whether and population demographic to explain heterogeneity between locations. For each app category and app we can find the target demographic, so it may be helpful to explain heterogeneity in terms of these data.

**Identification and Estimation**

In order to identify our choice model we use logit model, which has a fix scale. We do not include any dummy variable for preference for specific apps, as we are not sure that applies to our context. To set the location of the utility we normalize the utility of outside good to zero, as explained before. We identify individual level parameters using large panel of individuals we have in our sample. We identify the goodwill of mobile app category and mobile app through the diffusion structure we impose. We can identify location, mobile app category, and mobile app parameters through shrinkage and variation of different location, app categories, and mobile apps in our data, and finally we identify state space parameters through variation through long time series in our panel data.

We use hybrid of Metropolis-Hasting(MH), Gibbs sampling(GS), and Extended Kalman Filter (EKF) forward filtering backward sampling techniques for our estimation. Our algorithm consists of three stages, and we parallelized the algorithm for MH and GS of individual choices, and EKF and GS of app category and mobile apps. Our algorithm consists of “eleven” stages:

(E1)

(E2)

and are matrixes consisting of I rows, each with the  and  parameters for each consumer, and  are the vector choice observation for consumer I, and and  are the matrix of covariates. The first stage of the MCMC in (E1) is a set of I parallel Metrapolist algorithms tuned to each consumer MNL likelihood on Metropolis proposal distributions. We code the tune to be done automatically without any “pre-sampling of draws on the basis of a fractional likelihood that combines the consumers’ likelihood fractionally with the pooled MNL likelihood, as Rossi, Allenby and McCulloch (2005), chapter 5 specifies. This procedure avoids the problem of undefined likelihood for tuning purposes. In order to optimize the first stage of MCMC in (E1), we used Metropolis hasting draw around the mode.

The second stage of (E1) and the first stage (E2) are standard unconstrained Gibbs Sampler for a mixture of normal. The “label-switching” problem for identification in mixture of normal is not present in our application, as we are interested only in the posterior distribution of quantity which is label-invariant, i.e. mixture of normal density itself. For drawing from multinomial distribution we have following conditional probabilities:

 (E3)

 (E4)

Where  are discrete random latent variables with outcome probabilities and . This conditioning allows us to draw from mixture of normal distribution. Similar to equation (E3) and (E4) we have conditions to draw from mixture of normal distribution for other category, app, and individual level parameters we specified in DAG above for fifth, eight and eleventh stage of (E1), and fourth, seventh and tenth stage of (E2).

Third and sixth stage of (E1) and second and fifth stage of (E2) are standard unconstrained Gibbs Sampler for a factor model to shrink observations across individuals given time t. Fourth and seventh stage of (E1) and third and sixth stage of (E2) are standard unconstrained Gibbs Sampler for mean and variance of shrinkage of observations across time given individual i. Ninth stage of (E1) and eighth stage of (E2) are Extended Kalman Filter to draw unobserved state of shrunk goodwill of category and app, by linearizing Bass diffusion model (BDM), and finally tenth stage of (E1) and ninth stage of (E2) are standard unconstrained Gibbs Sampler for mean and variance of state equation of goodwill of category and app.

MCMC that we apply here is a hybrid method with a customized Metrapolist chain for the draw of the consumer, category, and app level parameters coupled with a standard Gibbs sampler for a mixture of normal conditional on the draws of consumer, category and app’s parameters. In other word, once the collection of consumer, category, and app parameters are drawn, the MCMC algorithm treats these as “data” and conducts Bayesian inference for a mixture of normal.

The prior used are:



We assessed the prior hyperparameters to provide proper but diffuse distributions:



The Drichlet prior on  is conjugate to the multinomial distribution, and can be interpreted as the size of a prior sample of data for which the classification”observations” is known. The number of observation of each “type” or mixture component is given by the appropriate element of . This prior assumes that each type is equally likely and that there is only a very small amount of information in the prior equal to a sample size of .5. As the number of normal components increases, we do not change how informative the prior is, through scale of elements of the a vector by . As our model become more complicated, we parallelized the computation of many stages as much as possible to use multi-core of server for our estimation[[4]](#footnote-4).

**Counterfactual analysis**

We analyze “what if” scenarios to find optimal incentive policy of platform owner for different product categories to increase the diffusion of the platform. This incentive policy could be in the form of lower revenue share for specific category. In addition, for platform owner it may be relevant to know what is the optimal assortment size to generate more revenue in a category, so we simulate different scenarios to find optimal policy. Third paid apps may cannibalize the free apps, or vise versa, so it may be interesting to know what is the optimal combination of free and paid apps that result in higher diffusion level. Another counterfactual could study how removing one app category affect consumer’s choice. Platform owner may be interested to know the effect of specialization or differentiation, and the optimal structure of the app store. To study these issues we use indirect utility of consumer, by numerically integrating the utility of individual consumer across population to get the expected market size.

**Conclusion**

We model consumer’s decision to download from mobile app store, focusing on complementarity and substitution of app categories and social dependency of awareness and preferences of consumers. We used a unique panel data to estimate our model. Our model explains the level of substitution and complementarity of the mobile app categories, and it investigates the level of social dependency of preference and awareness of consumers in a geographical area. We run counterfactual to understand how incentivizing one category can impact the total diffusion of the platform, and to know what is the optimal assortment size and optimal category hierarchy on the app store. Our analysis also allows us to find the impact of paid apps on the diffusion of free apps. We believe our study has various managerial implications.

**References**

Bell, David R., and Sangyoung Song. "Neighborhood effects and trial on the Internet: Evidence from online grocery retailing." Quantitative Marketing and Economics 5.4 (2007): 361-400.

Bezawada, R., Balachander, S., Kannan, P. K., & Shankar, V. (2009). Cross-category effects of aisle and display placements: A spatial modeling approach and insights. *Journal of Marketing*, *73*(3), 99-117.

Bradlow, Eric T., et al. "Spatial models in marketing." *Marketing Letters* 16.3-4 (2005): 267-278.

Carare, Octavian. "THE IMPACT OF BESTSELLER RANK ON DEMAND: EVIDENCE FROM THE APP MARKET\*." *International Economic Review* 53.3 (2012): 717-742.

Chatterjee, R. A., & Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Science*,*36*(9), 1057-1079.

Jaihak Chung, Vithala R. Rao (*2012*) A General Consumer Preference Model for Experience Products: Application to Internet Recommendation Services. Journal of Marketing Research: June 2012, Vol. 49, No. 3, pp. 289-305.

Choi, Jeonghye, Sam K. Hui, and David R. Bell. "Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer." *Journal of Marketing Research* 47.1 (2010): 75-89.

Dover, Yaniv, Jacob Goldenberg, and Daniel Shapira. "Network traces on penetration: Uncovering degree distribution from adoption data." *Marketing Science* 31.4 (2012): 689-712.

Dubé, J. P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*,*41*(3), 417-445.

Erdem, Tülin, and Michael P. Keane. "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets." *Marketing science* 15.1 (1996): 1-20.

Oestreicher-Singer, Gal, and Arun Sundararajan. "The visible hand? Demand effects of recommendation networks in electronic markets." *Management Science* 58.11 (2012): 1963-1981.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han. "How is the mobile Internet different? Search costs and local activities." *Information Systems Research*24.3 (2012): 613-631.

Ghose, Anindya, and Sang Pil Han. "Estimating Demand for Mobile Applications in the New Economy." *Management Science* (2014).

Ghose, Anindya, and Sang Pil Han. "An empirical analysis of user content generation and usage behavior on the mobile Internet." *Management Science*57.9 (2011): 1671-1691.

Ghose, Anindya, and Sang Pil Han. "A dynamic structural model of user learning on the mobile Internet." *Available at SSRN 1485049* (2011).

Ghose, Anindya, and Sang Pil Han. "A dynamic structural model of user learning on the mobile Internet." *Available at SSRN 1485049* (2011).

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*(3), 391-406.

Hartmann, Wesley R. "Demand estimation with social interactions and the implications for targeted marketing." *Marketing Science* 29.4 (2010): 585-601.

Hartmann, Wesley R., et al. "Modeling social interactions: Identification, empirical methods and policy implications." *Marketing letters* 19.3-4 (2008): 287-304.

Horsky, Dan, and Leonard S. Simon. "Advertising and the diffusion of new products." *Marketing Science* 2.1 (1983): 1-17.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente. "Opinion leadership and social contagion in new product diffusion." *Marketing Science* 30.2 (2011): 195-212.

Didem Kurt, J. Jeffrey Inman, Jennifer J. Argo (*2011*) The Influence of Friends on Consumer Spending: The Role of Agency– Communion Orientation and Self-Monitoring. Journal of Marketing Research: August 2011, Vol. 48, No. 4, pp. 741-754.

Karaca-Mandic, Pinar. "Role of complementarities in technology adoption: The case of DVD players." *Quantitative Marketing and Economics* 9.2 (2011): 179-210.

Kim, Y., Telang, R., Vogt, W. B., & Krishnan, R. (2010). An empirical analysis of mobile voice service and SMS: a structural model. *Management Science*,*56*(2), 234-252.

Lee, Sanghak, Jaehwan Kim, and Greg M. Allenby. "A direct utility model for asymmetric complements." *Marketing Science* 32.3 (2013): 454-470.

Lelarge, Marc. "Diffusion and cascading behavior in random networks." *Games and Economic Behavior* 75.2 (2012): 752-775.

Aurélie Lemmens and Christophe Croux (2006) Bagging and Boosting Classification Trees to Predict Churn. Journal of Marketing Research: May 2006, Vol. 43, No. 2, pp. 276-286.

Liu, Charles Z., Yoris A. Au, and Hoon Seok Choi. "An Empirical Study of the Freemium Strategy for Mobile Apps: Evidence from the Google Play Market." (2012).

Liu, Hongju, Pradeep K. Chintagunta, and Ting Zhu. "Complementarities and the demand for home broadband internet services." *Marketing Science* 29.4 (2010): 701-720.

Ma, Yu, P. B. Seetharaman, and Chakravarthi Narasimhan. "Modeling dependencies in brand choice outcomes across complementary categories."*Journal of Retailing* 88.1 (2012): 47-62.

Manski, Charles F. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60.3 (1993): 531-542.

Garg, Rajiv, and Rahul Telang. "Inferring app demand from publicly available data." *MIS Quarterly* 37.4 (2013): 1253-1264.

Naik, Prasad A., Ashutosh Prasad, and Suresh P. Sethi. "Building brand awareness in dynamic oligopoly markets." *Management Science* 54.1 (2008): 129-138.

Harikesh S. Nair, Puneet Manchanda, Tulikaa Bhatia (*2010*) Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. Journal of Marketing Research: October 2010, Vol. 47, No. 5, pp. 883-895.

Narayan, Vishal, Vithala R. Rao, and Carolyne Saunders. "How peer influence affects attribute preferences: A Bayesian updating mechanism." *Marketing Science* 30.2 (2011): 368-384.

Niraj, Rakesh, V. Padmanabhan, and P. B. Seetharaman. "Research Note-A Cross-Category Model of Households' Incidence and Quantity Decisions."*Marketing Science* 27.2 (2008): 225-235.

Peres, Renana, Eitan Muller, and Vijay Mahajan. "Innovation diffusion and new product growth models: A critical review and research directions." *International Journal of Research in Marketing* 27.2 (2010): 91-106.

Rossi, P. E., & Allenby, G. M. R. McCulloch (2005). *Bayesian Statistics and Marketing*.

Yang, Sha, Vishal Narayan, and Henry Assael. "Estimating the interdependence of television program viewership between spouses: A Bayesian simultaneous equation model." *Marketing Science* 25.4 (2006): 336-349.

Moon, Sangkil, and Gary J. Russell. "Predicting product purchase from inferred customer similarity: An autologistic model approach." *Management Science*54.1 (2008): 71-82.

Yang, Sha, and Greg M. Allenby. "Modeling interdependent consumer preferences." Journal of Marketing Research 40.3 (2003): 282-294.

Song, I., & Chintagunta, P. K. (2003). A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quantitative Marketing and Economics*, *1*(4), 371-407.

Song, Inseong, and Pradeep K. Chintagunta. "Measuring cross-category price effects with aggregate store data." *Management Science* 52.10 (2006): 1594-1609.

Sriram, S., Pradeep K. Chintagunta, and Manoj K. Agarwal. "Investigating consumer purchase behavior in related technology product categories."*Marketing Science* 29.2 (2010): 291-314.

Andrew T. Stephen, Olivier Toubia (2010) Deriving Value from Social Commerce Networks. Journal of Marketing Research: April 2010, Vol. 47, No. 2, pp. 215-228.

Michael Trusov, William Rand, and Yogesh V. Joshi (*2013*) Improving Prelaunch Diffusion Forecasts: Using Synthetic Networks as Simulated Priors. Journal of Marketing Research: December 2013, Vol. 50, No. 6, pp. 675-690.

Wedel, Michel, and Jie Zhang. "Analyzing brand competition across subcategories." *Journal of Marketing Research* 41.4 (2004): 448-456.

Yang, Botao, and Andrew T. Ching. "Dynamics of consumer adoption of financial innovation: The case of ATM cards." *Management Science* 60.4 (2013): 903-922.

Young, H. Peyton. "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning." *The American economic review* (2009): 1899-1924.

Zhao, Yi, et al. "Modeling consumer learning from online product reviews."*Marketing Science* 32.1 (2013): 153-169.

1. I would like to thank Dr. Soysal and Dr. Ratchford for sending me to Duke workshop on structural modeling, and teach me related techniques. Also I would like to thank Dr. Bruce for teaching me how to conduct a quality empirical research in marketing, and code Kalman Filter. [↑](#footnote-ref-1)
2. I would like to thank Dr. Ying Xie and Dr. Elisabeth Honka for their revision of the proposal I prepared and send out to the company. I also would like to thank Dr. Rao for making us to write a lot during his seminar, allowing us to organize our mind, and positioning our research paper. [↑](#footnote-ref-2)
3. http://www.iplocation.net/index.php [↑](#footnote-ref-3)
4. If it took too long, then perhaps we could use Texas System Super computers for estimation [↑](#footnote-ref-4)