

THE NEGATIVE IMPACT OF MOBILE DEVICES ON NICHE PRODUCT CONSUMPTION

Research-in-Progress

Anindya Ghose

New York University
Stern School of Business
aghose@stern.nyu.edu

Sung-Hyuk Park

New York University
Stern School of Business
spark@stern.nyu.edu

Abstract

Internet based information technology has provided the paradigm shift from Pareto principle to long tail phenomenon. However, the advent of small size mobile devices with higher search costs raises a negative question on this paradigm shift. The aim of this study is investigating the negative impact of mobile devices on niche product consumption. We collect 10 million user-device level transaction data with user profiles and content characteristics from a nation-wide e-book company in East Asia. We analyze the large size data with Pareto curve estimation and econometric modeling. As key findings, we found that smart phone users' product sales are more concentrated than those of users with PCs or smart pads. Our empirical results support that mobile commerce markets do not follow long tail phenomenon, but follow "Pareto Principle" in terms of sales diversity because smart phone users have less willingness to purchase unpopular products than smart pad users.

Keywords: Pareto principle, mobile commerce, long tail, niche product, search costs

Introduction

Information technology and internet markets have allowed people to consume niche products, thereby creating longer tails for internet businesses in terms of sales distribution. According to Brynjolfsson et al.'s study (2011 and 2003), internet's long tail reflects the lower search costs on internet commerce services and this allows consumers have much higher benefit from access to increased product variety in the long tail. However, in the world of mobile commerce, it is hard to keep the benefit because mobile devices (e.g., smart phones) have small size screens and clunky user interfaces so that the mobile users have higher search costs than PC users (Ghose et al. 2013; Ghose and Han, 2012; Hwang et al., 2012). Since mobile device users are different from PC users, it is getting important to understand users' mobile-specific behavior to define successful sales strategies. Even for the electronic commerce companies which have their own story of success in existing PC-based businesses, it is still challengeable to know how mobile users are different from their traditional users. Thus, we raise an important research question to show whether existing PC-based business rules are seamlessly applied to the world of mobile or not.

One of the most significant contributions of the internet based technology is that it allows people to have lower search costs (or advanced shopping environment) so that it brings a paradigm shift from the long lasting Pareto principle (known as the 80/20 rule) to the long-tail sales distribution. However, although existing studies have successfully suggested a new era of the niche products consumption in the online-based electronic commerce markets (Brynjolfsson et al., 2011 and 2003), no study has answered to the following questions: (1) whether the internet's long tail phenomenon is replicated in the mobile market? (2) if mobile user behavior is different from PC users', particularly, how consumption behavior differs between smart phone users and smart pad users? For example, because of the smallest screen size and the worst inconvenient user interface, smart phones make their owners have higher search costs than other device (i.e., PC or smart pads) owners. Thus, we can expect that users have the smallest collective share of niche (or unpopular) products when they use smart phones.

To answer our research questions, we conduct sales diversity analyses including econometric model analyses using data from a nation-wide e-book company in East Asia. Our data set is novel because it has large size transactions (more than 10 million data records between June 2012 to December 2012), and user-device level data which allow us to examine each device-specific user behavior. Specifically, our data set allows us to examine a negative/positive impact of having a smart phone (or a smart pad) on consuming products at market/individual level.

For the market level data analysis, we measure Gini coefficients of sales diversity to describe the concentration of market shares for each of smart phone, smart pad, and PC transaction data sets. Interestingly, we show that smart phone users have the lowest sales diversity among others, and smart pad users have lower sales diversity than PC users. In other words, we provide an evidence to examine a negative relationship between search costs and sales diversity (i.e., as search costs increase sales diversity decreases, reversely). In addition, we estimate the impact of smart phones and pads on niche product consumption. Rank data allow us to measure the effort required due to popularity. Higher ranking effects means it is more popular than other products, so it needs relatively less search costs to access the product. Through individual-device level panel data analyses, we show that users have less likelihood to purchase niche products when they have mobile devices such as smart phones.

Our analysis yields two main results. First of all, we find the negative and statistically significant relationship between search costs and niche products consumption. Mobile devices which have lower search costs than PCs allow users have more concentrated sales distribution (i.e., not long tail phenomenon, but Pareto principle). Secondly, smart phone users who have the smallest search costs and the most inconvenient user interface show the most concentrated sales distribution than others. This means that although the smart device is the most advanced technology, it allows users to relapse into the world of Pareto principle where markets are dominated by a small number of best-selling products.

The rest of this paper is organized as follows. Section 2 presents literature reviews, and section 3 gives data descriptions with summary statistics. Section 4 and 5 explain Pareto curve estimation and econometric modeling results. The last section summarizes our key findings with managerial implications.

Related Works

Niche vs. Riche

This paper aims to understand how mobile device users' product consumption behavior is different from online devices users' behavior in terms of sales diversity. Since Brynjolfsson et al. (2003) have introduced the concept of long tail phenomenon on the internet based electronic commerce, many articles have found evidences of a paradigm shift from the Pareto principle to the long tail phenomenon. The key message is that the product selection is greater on the internet than offline, because internet based information technologies such as search engines and recommender systems provide more convenient shopping environment to users with lower search costs (Brynjolfsson et al., 2011). For example, more users in nowadays spend their money to purchase not only best-selling products, but also niche products (Chellappa et al., 2007; Oberholzer-Gee, 2007). In addition, according to the Brynjolfsson et al. (2011)'s recent study, "Goodbye Pareto Principle, Hello Long Tail," consumers' usage of internet search and recommender systems allow them to have lower search costs, and promote the consumption of niche products. Also, in the Fleder and Hosanagar (2009)'s study, the authors showed that the impact of search tools (e.g. content recommender systems) on sales diversity was positive in terms of individual level diversity as a result of choice model simulations.

In summary, internet have played a key role to allow online consumers purchase more diverse products including unpopular products which have no chances to be displayed in offline stores. However, from our best knowledge, no one has explained the answer of the next research questions whether the shift toward niche products (i.e., long tail) in the online business is duplicated in the mobile world, or not.

Mobile Device Users' Behavior

When consumers are fully informed about a product, screen size does not matter (Balasubramanian 1998). However, according to the well-known human-computer interaction (HCI) theories, smaller size screens increase users search costs such as interaction efforts, and reading or viewing time (Sweeney and Crestani, 2006). In addition, higher search costs discourage users from consuming new or unpopular products, so that these users have relatively low probability to purchase niche products than others who have lower search costs (Mahmood et al., 2000; Brynjolfsson et al. 2011).

Due to the problem that the size of screen matters, one can expect that mobile device users show different behavior from online users'. Even when users generate new contents (e.g., posting on Facebook pages or uploading video clips) or use existing contents mobile, they show different content creation and consumption behavior (Ghose et al. 2013). For example, small screen device users watch less content in their main page than traditional PC users, and the small screen users may consume less information in terms of both information quantity and quality. Interestingly, the impact of mobile devices varies according to the size of screen. In other words, screen size matters and the difference in the screen size matters, as well. In Hwang et al. (2012)'s empirical study, they found the e-book consumers with larger screen mobile device (e.g., smart pads or tablet PC) had less price sensitivity, and higher probability to purchase niche products than small screen mobile device (e.g., smart phones) users. Therefore, in our setting, we examine how consumers' sales diversity varies when they use different types of devices such as PCs, smart pads, and smart phones, respectively.

Data Description

We analyze a large size digital content consumption data collected from one of the largest e-book companies in East Asia. This company provides an e-book business platform as Amazon.com does. The company's users download free e-books or purchase paid e-books by using their various kinds of devices such as PC, smart pad (e.g., iPad and Galaxy Tab, etc.), and smart phone devices as Amazon.com's users purchase e-books through their devices such as smart phone, smart pad, PC and Kindle device. The usage of multiple devices allows us to observe the users' device-specific content consumption behavior. In addition, an important strength of our data is that we have a large number of e-books (i.e. more than 20,000) with different categories including Novel, Business and Finance, Self Improving, Hobby, Health

and Living, and History among others. Specifically, our data set consists of users' profile including demographic and device registration information, detail e-book profiles such as price, author, publisher, release date, genre, rank, and description (textual information), and large size e-book purchase transactions. Therefore we believe that our data have the advantage of allowing us to validate the impact of different devices on content consumption behavior.

The observation period was six months from June 2012 to December 2012. Although the time period is not very long, the size of the raw log data was larger (raw data have about 10 million data records) than other data sets have been used in previous academic researches. An observation in our panel data consists of particular user-device-book-time(daily) level data records. Since our data set contains novel information such as monetary value (or conversions), sales rank, and sales amount, we believe that our data set provides reliable analysis results. Table 1 presents the summary statistics of the variables in our data set.

Measure of Sales Diversity

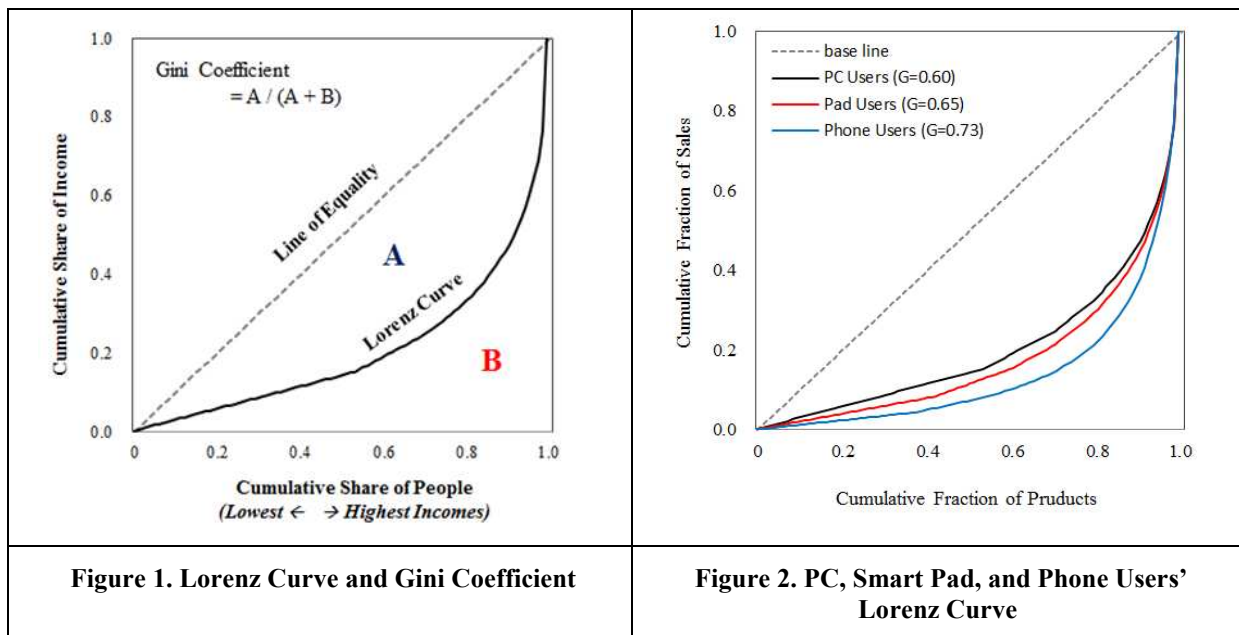
Gini coefficient is a widely accepted measure of distributional inequality (Sen, 1976), and it has been applied to measure wealth inequality for nations all over the world. For example, in Figure 1, the line with a 45 degree slope represents that a nation's wealth is equally distributed by people. The Gini coefficient is simply defined as $G = 1 - 2 \int_0^1 L(x) dx = \frac{A}{A+B}$, where A denotes the area lies between the equality line and the Lorenz curve (Lorenz, 1905), $L(x)$, and B denotes the area under the Lorenz curve (i.e., $\int_0^1 L(x) dx$). In Fleder and Hosangar (2009)'s study, the authors measured the Gini coefficient of sales

Table 1. Summary Statistics					
Variable	Description	Mean	Std. dev.	Min	Max
<i>Book-Device Level Data</i>					
Sales_Phone	Sales ratio through smart phones	0.3026	0.4594	0	1
Sales_Pad	Sales ratio through smart pads	0.5066	0.5000	0	1
Sales_PC	Sales ratio through PCs	0.1908	0.3930	0	1
<i>Individual Level Data: User Profile</i>					
Age	Users' age	34.62	7.50	15	70
Gender	Users' gender(1:male, 0:female)	0.7459	0.4354	0	1
Tenure	Days since subscribed (experience)	400.09	269.38	77	1,109
<i>E-book Characteristics</i>					
Rank	Sales Rank	4,837	7,213	1	41,494
Price	Sales price (USD)	4.28	3.20	0.09	52
Content Size	File size (Megabytes)	6.73	10.08	0.1	75.95
Book Age	Days since content released	172.07	263.22	1	1,322
Description	Length of content description	4,849.56	5,318.26	34	38,644
Review	Number of review comments	1.36	6.03	0	485
Rating	Average score of user rating	4.61	0.64	0	5

Note. We sampled 4,503 individuals who have three devices: PC, smart pad and phone. Each book is available on all three device types.

diversity to describe the concentration of market shares. In their study, the Lorenz curve, $L(x)$, denotes the percentage the company's sales amount generated by the lowest 100x% of products (during a specific time period). When most customers prefer to purchase popular products (or best-sellers), the company expects high Gini coefficient as a sales diversity score. The Gini coefficient can be theoretically ranged from 0 to 1, and as the Gini coefficient decreases it indicates a more equal distribution. Since a low Gini is related to the long-tail phenomenon and a higher Gini is related to the Pareto principle, measuring the Gini coefficient allows us to identify whether a company has a concentrated sales distribution or not.

As results of macro level data analysis, we plot three different Lorenz curves to examine that different device users show different content consumption patterns (e.g., long-tail vs. Pareto distribution) in terms of sales diversity. Figure 2 illustrates three Lorenz curves based on the company's PC, smart pad, and smart phone users' sales transactions, respectively. Interestingly, smart phone users show the most convex curve with the highest Gini coefficient among others. In addition, the reported $G=0.73$ represents that a small number of popular products account for most of sales provided by the smart phone users. Thus, smart phone users' content consumption behavior relatively does not follow the long-tail distribution. In addition, the order of the three Gini coefficients is summarized as PC users ($G=0.60$) < pad users ($G=0.65$) < phone users ($G=0.73$), and this implies PC users have the most diverse sales products, and smart pad users shows higher diversity as the second best. Our findings are consistent with prior study on multiple device users' content consumption behavior. According to Hwang et al.(2012), having a new smart pad device positively influences on the consumption of niche products because the wider device allows users to have lower search costs. Since the screen size of smart phones is smaller, and smart phones have lower search costs than other devices, we believe that the smart phone users prefer to purchase more popular content which usually displayed (or recommended) in the company's main service page. Therefore, the smart phone users record the highest Gini coefficient.



Econometric Modeling

Lorenz curves and Gini coefficients show the difference of sales distribution among three device channels. However, these results are not based on statistical significance. Therefore, we construct an econometric model to measure the effects of E-book Rank and Mobile Device Type on content consumption. Based on the well-known Pareto distribution model (Chevalier and Goolsbee, 2003; Brynjolfsson et al., 2003), we choose two aspects of digital content consumption: device type and rank. With control variables such as user profiles (e.g., gender and age) and book characteristics (e.g., price, filesize, and description length),

we estimate the user level probability of purchase using the following difference-in-difference equation. Specifically, we assume that the sales of the j-th content are:

$$Sales_j = \beta_0 + \beta_1(Rank_j^{-\hat{\lambda}}) + \beta_2Pad + \beta_3Phone + \beta_4Pad \times (Rank_j^{-\hat{\lambda}}) + \beta_5Phone \times (Rank_j^{-\hat{\lambda}}) + \sum_h \beta_{6,h} BookCharacteristics_{j,h} + \sum_k \beta_{7,k} UserProfiles_{j,k} + \varepsilon_j, \text{ where } -\hat{\lambda} \text{ is the estimated shape parameter of Pareto distribution.}$$

To estimate the regression coefficients, we assume that our dependent variable follows the negative binomial distribution, and this assumption is consistent with the most related prior study (Brynjolfsson et al. 2011). In addition, our control variables consist of user profiles and product characteristics to control consumer, and content heterogeneity, respectively.

Table 2. Negative Binomial Regression Results: Coefficients and Standard Errors

Variables	Model 1 (full model)	Model 2 (no controls)	Model 3 (product controls)	Model 4 (user controls)	Model 5 (no phone data)	Model 6 (no pad data)
Constant	1.658** (0.099)	1.676** (0.017)	1.477** (0.067)	2.065** (0.072)	1.611** (0.122)	1.705** (0.122)
$Rank_j^{-\hat{\lambda}}$	74.843** (5.857)	164.822** (6.440)	74.599** (5.840)	164.544** (6.445)	74.616** (6.055)	73.615** (5.975)
Pad	0.296** (0.023)	0.253** (0.023)	0.292** (0.023)	0.263** (0.024)	0.329** (0.023)	
Phone	0.777** (0.024)	0.703** (0.023)	0.787** (0.023)	0.680** (0.024)		0.724** (0.024)
$Pad \times Rank_j^{-\hat{\lambda}}$	53.965** (8.623)	61.757** (9.726)	54.256** (8.626)	60.950** (9.715)	54.659** (8.649)	
$Phone \times Rank_j^{-\hat{\lambda}}$	134.187** (9.821)	154.03** (10.678)	133.94** (9.818)	154.09** (10.674)		131.897** (9.800)
Price _j	-0.026** (0.003)		-0.026** (0.003)		-0.015** (0.004)	-0.024** (0.004)
Content Size _j	-0.010** (0.001)		-0.010** (0.001)		-0.003** (0.001)	-0.015** (0.001)
Book Age _j	-0.001** (0.000)		-0.001** (0.000)		-0.001** (0.000)	-0.001** (0.000)
Description _j	0.000** (0.000)		0.000** (0.000)		0.000** (0.000)	0.000** (0.000)
Review _j	0.058** (0.002)		-0.059** (0.002)		0.056** (0.002)	0.060** (0.002)
Rating _j	0.049* (0.013)		0.050* (0.013)		0.062** (0.018)	0.050** (0.016)
Age _j	-0.004* (0.002)			-0.013** (0.002)	-0.011** (0.002)	-0.003 (0.003)
Gender _j	-0.036 (0.023)			-0.038 (0.022)	0.212** (0.029)	-0.151** (0.028)
Tenure _j	0.000 (0.000)			0.000** (0.000)	0.000** (0.000)	0.000** (0.000)
Pseudo R ²	0.692	0.682	0.691	0.684	0.622	0.706
Number of Obs.	57,274	57,274	57,274	57,274	34,560	37,725

Note. * significant at the 5% level; ** significant at the 1% level. Base is PC users. Standard errors are in parentheses.

For all model, we set $-\hat{\lambda} = -1.013$ which is equal to the Pareto curve estimation results from the following basic mode: $\ln(Sales_j) = \beta_0 + \beta_1 \ln(Rank_j) + \beta_2 Pad + \beta_3 Phone + \beta_4 Pad \times \ln(Rank_j) + \beta_5 Phone \times \ln(Rank_j) + \varepsilon_j$.

As summarized in Table 2, Model 1, the full model which consists of all variables, reports the most comprehensive result. As book rank decreases (toward more popular content), $Rank_j^{-\lambda}$ increases accordingly. Since the estimated regression coefficient of $Rank_j^{-\lambda}$ is positive, it supports that $Rank$ is negatively correlated with $Sales$ which means that sales quantity decreases when content has less popularity in general. The result is consistent with our Pareto curve estimation results. Interestingly, the cross term between Pad and $Rank$ and the cross term between $Phone$ and $Rank$ are positive, as well. The results imply that the marginal impact of $Rank_j^{-\lambda}$ on $Sales$ increases from 74.843 to $74.834 + 53.965 (= \beta_1 + \beta_4)$ when a user has a smart pad, and it also increases from 74.843 to $74.843 + 134.187 (= \beta_1 + \beta_5)$ when a user has a smart phone. These results are statically significant under 99% confidence level. Therefore we can say that smart phone users have much larger marginal effect of $Rank_j^{-\lambda}$ than PC or smart pad users have. In model 1, we control both book-specific characteristics such as sales price, content size, book age, length of description, number of reviews, and rating scores, and user profiles such as average age, gender, and tenure (prior experience). As reported in Table 2, the impact of one unit increase in price is negative and significant. Similarly, the regression coefficient of content size is also negative which means that when other conditions are equal, large size content negatively influences on the sales of the content.

In Model 2, 3, and 4, we check the robustness of the main effects (e.g., $Rank_j^{-\lambda}$ and cross terms) reported in Model 1. Each model shows alternative specifications which have less number of control variables than the full model. Firstly, in Model 2, the main effects do not change when all controls are omitted. Interestingly, the regression coefficient of $Rank_j^{-\lambda}$ is larger than the one from Model 1. Secondly, in Model 3, the results show that the main effects also do not change when user profile variables are omitted. Thirdly, when we exclude product characteristic variables, the results are consistent with prior regression results. Similarly, in Model 5 and 6, we exclude phone and pad data, respectively. As summarized in Table 2, the regression coefficients show that our results are robust when phone or pad variables are omitted. In short, all Models consistently show that smart phone users have larger marginal effect of $Rank_j^{-\lambda}$ on $Sales$ than PC or smart pad users, and this is consistent with our findings illustrated in Figure 2.

Conclusion

It has been introduced that product variety increase social welfare (Ghose et al 2006; Brynjolfsson et al. 2003). In our setting, if mobile users have more concentrated sales distributions than online (PC) users, content providers which have less popular or unpopular content cannot have opportunities to sell their products to users through mobile channels. In addition, the market which has less diverse products cannot provide higher social welfare and even cannot survive longer than the market which has more diverse products. Therefore we believe that mobile users relapse into the world of Pareto principle where they have difficulties to search various kinds of products as they purchase something in off-line stores.

Through empirical data analyses including a sales diversity diagnosis (i.e., Gini coefficient calculation), a market level Pareto curve estimation, and basic and advanced econometric analyses, we show that our results are always consistent with each other, and with the findings from prior studies. Our results show that sales quantity (or demand) decrease when users have mobile devices such as smart phones and pads. We suggest an empirical evidence to support that smart phone users have the lowest likelihood to purchase niche products than other device users, and smart pads users have lower probability to purchase niche products than PC users, as well. Although it is identified that the smart phone is not an ideal device to purchase niche products, we believe that companies can help their smart phone users to purchase more by providing recommendation and product search services. As suggested by Brynjolfsson et al. (2011) such internet-based information technologies are effective to help users easily search and purchase niche products. Therefore, mobile contents providers and mobile commerce companies should invest their asset to design and implement good recommender systems and search tools so that their potential customers can effectively decrease their search costs when they use small screen mobile devices.

On the other hand, it has been suggested that sales diversity is positively associated with customer retention rates in terms of customer relationship management (Park and Han 2013). Therefore marketing managers should put their efforts to create sales diversity based marketing strategies and action plans so that their customers purchase not only best-selling contents but also niche products and have higher

retention rate than before. We believe that this strategic approach is still valid to the smart phone users and can be a useful and practical blueprint to keep the high level diversity in mobile content market.

Our main contributions are summarized into two parts. First, we state the issue that although the mobile devices are more advanced and have been newly released than traditional PCs, the mobile devices provide worse shopping environment to purchase niche products due to the smaller screen size and more inconvenient user interface. Through several empirical analyses, we show that mobile purchases have a shorter tail than PC purchases. According to Ghose et al. (2013), search costs are higher on mobile due to a smaller screen size, and higher search costs lead to higher concentration of sales of the most popular products (Brynjolfsson et al., 2011). We believe that this is the first work to empirically demonstrate the link between search costs (i.e., device type) and sales concentration, and our findings contribute to the understating of mobile device users' unique contents consumption behavior. Second, we also provide useful managerial implications to field experts. To overcome such mobile obstacle, mobile commerce companies should implement recommender systems or search tools so that such information technologies help their users easily browse various kinds of products and purchase them through their mobile devices.

This study has some limitations that we conclude our results by using an e-book data set provided by one company in East Asia, which means that we do not have other electronic commerce data sets required to generalize our key findings. In addition, we could not fully control the effect of endogeneity such as e-book company's marketing promotions which leads to bias in estimation of regression coefficients. For the future research projects, we have a plan to collect firm level cross sectional data sets from multiple electronic commerce markets to provide more general findings, and to conduct advanced analyses to understand the mediating relationship of search costs between mobile devices and sales concentration.

References

- Adomavicius, G., and Kwon, Y. 2009. "Toward more diverse recommendations: Item re-ranking methods for recommender systems," *In 19th Workshop on Information Technologies and Systems*.
- Brynjolfsson, E., Hu, Y. J., and Smith, M. D. 2011. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," *Management Science* (57:8), pp. 1373-1386.
- Brynjolfsson, E., Hu, Y. J., and Smith, M. D. 2003. "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science* (49:11), pp. 1580-1596.
- Chevalier, J., Goolsbee, A. 2003. "Measuring prices and price competition online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics* (1:2), pp. 203-222.
- Chevalier, J., and Mayzlin, D. 2006. "The effect of work of mouth on sales: Online book reviews," *Journal of Marketing Research* (43:3), pp. 345-354.
- Fleder, D., Hosanagar, K. 2009. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity," *Management Science* (55:5), pp. 697-712.
- Ghose, A., Smith, M. D., Telang, R. 2006. "Internet exchanges for used books: An empirical analysis of product cannibalization and welfare impact," *Information Systems Research* (17:1), pp. 3-19.
- Ghose, A., Goldfarb, A., and Han, S. 2013. "How is the mobile internet different? Search costs and local activities," *Information Systems Research*, Forthcoming.
- Ghose, A., and Han, S. 2012b. "Estimating Demand for Mobile Applications in the new Mobile Economy," *In Proceeding of International Conference on Information Systems (ICIS), Orlando, Florida, USA*.
- Hwang, K., Park, S. and Han, I. 2012. "Multi-Screen Strategy for Selling Mobile Content to Customers," *In Proceeding of International Conference on Information Systems (ICIS), Orlando, Florida, USA*.
- Lorenz, M.O. 1905. Methods of measuring the concentration of wealth. *Publications American Statistics Association*. (9:70), pp. 209-219.
- Mahmood, M. A., Burn, J. M., Gemoets, L. A., and Jacquez, C. 2000. "Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature," *International Journal of Human-Computer studies* (52), pp. 751-771.
- Park, S., Han, S. 2013. "From Accuracy to Diversity in Product Recommendations: Relationship between Diversity and Customer Retention," *International Journal of Electronic Commerce (forthcoming)*.
- Sen, A. 1976. "Poverty: An ordinal approach to measurement," *Econometrica* (44:2), pp. 219-231.
- Sweeney, S., and Crestani, F. 2006. "Effective Search Results Summary Size and Device Screen Size: Is There a Relationship?" *Information Processing and Management* (42:4), pp. 1056-1074.