An Explorative Study on Sales Distribution in M-commerce

Yongjin Park
KAIST College of Business
livediary@business.kaist.ac.kr

Youngsok Bang
CUHK Business School
bangyoungsok@baf.cuhk.edu.hk

Jae-Hyeon Ahn
KAIST College of Business
jahn@business.kaist.ac.kr

Abstract

Despite the proliferation of studies on the sales distribution in e-commerce, little research has been conducted on the sales distribution in the m-commerce channel. This study empirically examines the sales distribution of various product categories in the mobile channel, using the large transaction data from a leading e-marketplace in Korea. Overall, transactions in the mobile channel are more concentrated to head products compared to the PC channel sales, but the pattern is inconsistent across product categories. Transactions in product categories of high average price (e.g., computers) and low purchase frequency rate (e.g., health care products) are less concentrated to head products in the mobile channel than the PC channel. The revenue distribution, however, shows the opposite. Head products generate relatively less revenue in the mobile channel than the PC channel. We provide explanations why the mixing results appear across product categories and between the distribution types.

1. Introduction

The demand curve of the online market is characterized by the “long tail” where hard-to-find products in traditional retail markets have higher shares and generate substantial profits [1]. Without binding constraints on product storage and display space, online markets carry more products and provide more options to buy for consumers. Consumers discover niche products that better suit their needs using powerful search tools available in online markets. Online retailers get a selection advantage over brick-and-mortar retailers and significantly enhance the consumer surplus with their product variety [5][6].

As the demand shift towards niche products in online markets casts significant meanings in digital economy with online retailers’ different revenue structure and increased consumer surplus, many researchers empirically investigated the long tail phenomenon in various markets including book [5][20], clothing [4], music [8], and video [9] and DVD [16][23]. However, prior research on the demand curve in online markets is more or less limited to the comparison of sales distribution between the brick-and-mortar channel and traditional online channel or between two online channels where the mobile channel was not available. Little is known about how the demand curve is shaped in the presence of mobile channels, which have become popular purchase channels but have unique properties differentiated from the traditional PC channel.

Internet-enabled mobile devices such as smartphones allow consumers to access online markets and search for products anytime and anywhere. On one hand, ubiquitous access and search possibilities afforded by mobile channels release the transportation cost of finding a fixed place that has the Internet connection and a PC for online shopping, which could promote more searches in online markets. For example, a consumer could continue product search by hopping from the PC to the smartphone when s/he needs to go outside. This would reshape the demand curve in online market by fattening the tail. On the other hand, small screens and limited usability of mobile devices put constraints on extensive product search in online markets and could entice consumers to featured products, resulting in fattening the head. Considering both possibilities, the impact of mobile channels on demand curve remains the empirical question.

Further, as most prior studies focus on the demand curve of a single product category, there is little guidance how the sales distribution in the mobile channel would differ across product categories. The results of previous studies report largely different degrees of online sales dispersion across product categories. The Gini coefficient of the sales distribution of online markets reports ranging from 0.35 to 0.93 depending on the product category (Table 1).
Table 1. Gini coefficients in prior research

<table>
<thead>
<tr>
<th>Paper</th>
<th>Context</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>[22]</td>
<td>Movie ratings in Netflix</td>
<td>0.8067, 0.8315, 0.8193, 0.8155, 0.8339, 0.8491 (from 2000 to 2005)</td>
</tr>
<tr>
<td>[4]</td>
<td>Women’s clothing</td>
<td>0.49 (Internet), 0.53 (catalog)</td>
</tr>
<tr>
<td>[13]</td>
<td>VOD sales in Germany</td>
<td>0.637 (mean for 111 weak from March 2005), 0.553 (min), 0.732 (max)</td>
</tr>
<tr>
<td>[18]</td>
<td>Amazon book sales</td>
<td>0.73 (overall mean), 0.76 (science book), 0.51 (computer and Internet)</td>
</tr>
<tr>
<td>[25]</td>
<td>Total downloads on the Google Play app market</td>
<td>0.93 (paid app), 0.88 (free app)</td>
</tr>
<tr>
<td>[14]</td>
<td>Mobile App usage from Android users in Korea</td>
<td>0.74 (overall mean), 0.83 (communications), 0.79 (social media), 0.74 (entertainment), 0.64 (news), 0.35 (game)</td>
</tr>
<tr>
<td>[16]</td>
<td>Movie DVD sales in U.S.</td>
<td>0.67 (before the broadcast on TV), 0.59 (after the broadcast on TV)</td>
</tr>
<tr>
<td>[23]</td>
<td>Movie DVD rental in U.S.</td>
<td>0.806 (January 2001), 0.843 (July 2005)</td>
</tr>
<tr>
<td>[19]</td>
<td>Internet news</td>
<td>0.816, 0.762 (NYT articles in the pre-and the post-paywall period), 0.653, 0.630 (LAT articles before and after NYT’s paywall rollout)</td>
</tr>
</tbody>
</table>

2. Comparison of sales distribution between the PC and mobile channels

2.1. Data

We use a large transaction dataset from the database of a leading e-marketplace in Korea that had initially provided the PC channel only and launched the mobile channel later. The dataset contains a random sample of 60,000 users and their entire transactions during two years (from June 2009 to May 2011). The e-marketplace launched its mobile channel on June 1st, 2010, and our data period covers one year before and after the mobile channel introduction by the e-marketplace.

To clearly compare the sales distribution between the PC and mobile channels, we exclude 27,238 users from the initial dataset who signed-up the e-marketplace during our observation period. The incoming of new users could distort the tail length, as their joining might be associated with the intention to purchase products at the time of signing-up [13]. Our main sample contains 32,762 users and their over 1.5 million transaction records including 45,477 transactions conducted through the mobile channel (Table 2). Among 37 product categories in the e-marketplace, “women’s clothing” category has the largest number of transactions both in the PC and mobile channels. The mobile channel transaction accounts for 4.7% of the whole transactions.

Table 2. PC and mobile channel transactions

<table>
<thead>
<tr>
<th>Channel</th>
<th>June 2009 ~ May 2010</th>
<th>June 2010 ~ May 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>677,995</td>
<td>915,780 (95.269%)</td>
</tr>
<tr>
<td>Mobile</td>
<td>-</td>
<td>45,477 (4.731%)</td>
</tr>
</tbody>
</table>

2.2. Sales distribution in the mobile channel

The mobile channel adopters and non-adopters might be heterogeneous regarding demographic composition or purchasing habits, which could be the cause of different shapes of the sales distribution between the PC and mobile channels. For example, mobile channel adopters might be younger and more innovative than non-adopters. To tease out the pure channel effect on the sales distribution, we first employ the propensity score matching (PSM) to match orders in the PC and mobile channels [4]. We use users’ demographic variables (gender, age) and their order summary measures (recency, frequency, and monetary value) before the mobile channel introduction by the e-marketplace to generate the propensity score of each order. Studies in the direct marketing literature show
that recency, frequency, and monetary value are sufficient statistics to summarize purchase histories of each customer, which provide theoretical justification for the use of these measures [10].

The PC and mobile samples are matched well as none of demographic variables and order summary measures are significantly different after matching (Table 3), which addresses the potential selection bias issue in comparing the sales distribution between channels [4].

### Table 3. Propensity score matching results

<table>
<thead>
<tr>
<th>Matching variables</th>
<th>Mobile sample</th>
<th>PC sample</th>
<th>Difference</th>
<th>Matched PC sample (probit)</th>
<th>Matched PC sample (logit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Ratio</td>
<td>0.358</td>
<td>0.350</td>
<td>0.008***</td>
<td>0.357</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Age</td>
<td>30.034</td>
<td>31.790</td>
<td>1.756***</td>
<td>30.005</td>
<td>29.990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.037)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Recency</td>
<td>3.307</td>
<td>3.803</td>
<td>0.496***</td>
<td>3.309</td>
<td>3.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Frequency</td>
<td>3.365</td>
<td>3.921</td>
<td>0.556***</td>
<td>3.368</td>
<td>3.369</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Monetary</td>
<td>3.325</td>
<td>3.848</td>
<td>0.523***</td>
<td>3.327</td>
<td>3.327</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Sample size</td>
<td>38,626</td>
<td>862,371</td>
<td></td>
<td>38,626</td>
<td>38,626</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.

Then, we generate the Lorenz curve and calculate the Gini coefficient of the sales distribution in the PC and mobile channels to compare the demand shape between the two channels. Lorenz curve depicts the sales proportion of the bottom x% of products. Higher Lorenz curve (closer to the 45° equality line) represents the longer tail of the sales distribution. The Gini coefficient measures the ratio of the area in between 45° equality line and the Lorenz curve to the area under the equality line. A greater Gini coefficient represents more concentration of sales on head products.

When we sort products by demand, the Lorenz curve of the mobile channel locates below the curve of the PC channel, meaning more niche products were sold in the PC channel than the mobile channel (Figure 1). The Gini coefficient of the sales distribution in the mobile channel (0.494) is also higher than the PC channel (0.438), confirming that the PC channel has the longer tail than the mobile channel.

However, when we sort products by revenue, the results are opposite. The Lorenz curve of the mobile channel locates above the curve of the PC channel, showing more revenue was generated from head products in the PC channel (Figure 2). The Gini coefficient of the sales distribution in the mobile channel (0.594) is also lower than the coefficient in the PC channel (0.696).

For more rigorous analysis, we examine the log-linear relationship between sales and sales rank of product \( j \) [4][5] (Eq 1). Lower rank means higher sales. The \( \beta_1 \) estimated from the sales distribution of the mobile channel would have greater (smaller) absolute value than the PC channel if it is more (less) negatively skewed. To statistically test the difference in size of \( \beta_1 \) between the two channels, we incorporate the dummy for mobile transactions (Mobile) and run a simple regression to check the significance of \( \beta_3 \) after pooling the PC and mobile samples (Eq 2).

\[
\ln(Sales_j) = \beta_0 + \beta_1 \ln(Sales Rank_j) + \varepsilon_j \quad \text{(Eq 1)}
\]

\[
\ln(Sales_j) = \beta_0 + \beta_1 \ln(Sales Rank_j) + \beta_2 Mobile_j + \beta_3 Mobile_j \times \ln(Sales Rank_j) + \varepsilon_j \quad \text{(Eq 2)}
\]
The analysis results show that the sales rank is more strongly associated with the demand in the mobile channel than the PC channel and the difference between the two channels is significant (Table 4), meaning head products made relatively more sales in the mobile channel compared to the PC channel. In the case of revenue, on the other hand, the sales rank is more strongly associated in the PC channel than the mobile channel, and the difference is significant (Table 5). It means that head products relatively generated more revenue in the PC channel than in the mobile channel. Quantile regression analyses show the consistent results (Table 4 and 5).

In sum, demand is more concentrated to popular products in the mobile channel than the PC channel, whereas revenue is the opposite.

2.3. Comparison of sales distribution by product categories

Considering the large variance in Gini coefficients reported in previous studies (Table 1), we perform a category-level comparison of the sales distribution between the PC and mobile channels (Table 6). In nine categories, the demand distribution in the mobile channel reports lower Gini coefficients than the PC channel. Specifically, transactions in product categories of high average price (e.g., computer, audio/video supply) and low purchase frequency rate (e.g., health care products) were less concentrated to head products in the mobile channel than the PC channel. When it comes to the revenue distribution, on the other hand, niche products relatively generated more sales in the mobile channel than the PC channel for most product categories.
mobile and PC channels (Figure 3). Most categories of the demand (revenue) distribution between the bedding/curtain/carpet, sports clothing & accessories, audio/office users preferred the PC channel to the mobile channel compared to the PC channel. This suggests generated more sales but less revenue in the mobile channel especially in those categories. Three categories locate in Q1 where head products generated more sales and revenue in the mobile channel. The contrasting results across product categories and between the demand and revenue distributions naturally lead us to the question of why. We open the discussion and provide the possible explanation about our results in the next section.

### 3. Search cost and sales distribution

The long tail phenomenon in online markets is mainly driven by two forces—the increased product variety on the supply side [6][23][26] and advanced search tools available online on the demand side [4][13]. Unlike brick-and-mortar retailers, online retailers are free from the constraints of product storage and display space, which enable them to provide the increased product variety to the market. Advanced online search tools lower users’ search cost to find products that better fit their needs, making niche products viable in the market.

Both PC and mobile channels are electronic channels that are free from physical constraints. Therefore, there is no difference between the two channels on the supply side. When it comes to the demand side, however, the two channels would incur different search cost. The search cost in the PC channel consists of two parts, 1) the transportation cost of finding the place that has the fixed Internet and a PC [2] and 2) the information processing cost of finding and comparing alternatives to make a purchase decision [21]. Small screens and limited usability of mobile devices hamper extensive information processing and incur higher information processing cost. However, the ubiquity of mobile channels allows a better accessibility to online markets and reduces the transportation cost. Considering the both, product search in the mobile channel could incur either higher or lower cost than the PC channel.

Information-intensity of products refers to the amount of information required to make a purchase decision [3]. Searching for information-intensive products would incur much higher cost in the mobile channel than in the PC channel, as it involves higher information processing cost than the PC channel [3]. Consequently, more searches and longer demand distributions are expected in the PC channel than in the mobile channel for information-intensive products. On the other hand, when searching for products with low information-intensity, more searches and longer demand distributions are expected in the mobile channel, as the transportation cost accounts for most search cost incurred.

Both product price and purchase frequency rate could affect the information-intensity of products. Anecdotal evidence shows that consumers spend more...
time and effort for search when purchasing expensive products than inexpensive products. Prior studies also show that product price is the major indicator of involvement [17] and drives more search behaviors [7] [15]. Consumers might have pre-knowledge for the products which they are frequently buying, and would conduct less search for those products [24].

4. Conclusion

Despite the proliferation of studies on the sales distribution in e-commerce, little research has been conducted on the sales distribution in the m-commerce channel. Furthermore, since most prior studies focus on the demand curve of a single product category, there is little guidance how the sales distribution in the mobile channel would differ across product categories. This study aims to expand this stream of research by investigating the sales distribution in the mobile channel across multiple product categories.

Overall, transactions in the mobile channel are more concentrated to head products compared to the PC channel sales, but the pattern is inconsistent across product categories. Transactions in product categories of high average price (e.g., computers) and low purchase frequency rate (e.g., health care products) are less concentrated to head products in the mobile channel than the PC channel. From the search cost perspective, we provide explanations why the sales distribution in a certain category has a longer (or shorter) tail in the mobile channel.

In case of the revenue distribution, on the other hand, head products contribute less revenue in the mobile channel than the PC channel. This implies that users tend to purchase expensive products through the PC channel than the mobile channel given the product category.

This study provides several theoretical implications. First, to the best of our knowledge, this is one of the first studies that investigate the sales distribution in the m-commerce channel and compare the sales distribution in the mobile channel with the PC channel across product categories. From the search cost perspective, we explain why the sales distribution of certain product categories is more skewed in the mobile channel than the PC channel. Prior studies presume that the mobile channel has higher search cost than the traditional PC channel due to smaller screen size and limited usability [11][12]. However, our results imply that the search cost can be different depending on product category, and the PC channel has fatter head than the mobile channel in some product categories.

5. References