Master’s Thesis of Mingi Goo

Heterogeneous Rank Effects in Online Marketplace

온라인 커머스에서 랭킹 효과의 이질성 분석

February 2023

College of Business Administration
Seoul National University
Marketing Major

Heterogeneous Rank Effects in Online Marketplace

Mingi Goo

Submitting a master's thesis of
Business

October 2022

College of Business Administration
Seoul National University
Marketing Major

Mingi Goo

Confirming the master's thesis written by
Mingi Goo
December 2022

Chair	주 우 진	(Seal)
Vice Chair	김 병 도	(Seal)
Examiner	김 준 범	(Seal)

Abstract

This paper studies the rank effect heterogeneity in the online marketplace and suggests a practical implication for marketing managers to set the optimal digital marketing strategies. Because of the increasing economy of online marketplaces, the position or rank effect is a crucial issue in the marketing literature. The latest literature has focused on the effects of sponsored search results on search engine advertising, though it is known that organic results are more critical than search ads. This research is novel to focus on the effect of organic results in the online marketplace. For analysis on the unit of product level, this paper constructs the rank index through weighted average by keyword search volumes. In the model, the rank effect was specified by the interaction of product-level and category-level averaged variables with the rank index, with the covariates of product-level time-variant variables and two-way fixed effects. Some products were selected randomly to escape the curse of dimensionality. The estimation result suggests that product sales increased in rank and the number of Q&A and reviews. Meanwhile, categories with high price dispersion experienced a lower rank effect, and categories with information asymmetry experienced a lower rank effect. The overall characteristics of the category, such as average price, product attributes, and competition intensity, do not have a significant rank effect. In conclusion, I suggest that marketing managers implement search engine optimization in online marketplaces if their products are in the category with a higher rank effect. This paper finally took a snapshot of the online marketplace by exploiting a vast dataset and extending the marketing literature to the new area. Future research considering hierarchical modeling and endogeneity can investigate more robust and rigorous causality.

Keyword: Rank effect, Online Marketplace, Product category, Digital Marketing, Search Engine Optimization, Marketing Strategy

Student Number: 2021-29714

Table of Contents

Chapter 1. Introduction

Online commerce is rapidly growing. Especially during the COVID-19 pandemic, online commerce has grown fast and massive. In 2021, the two online commerce giants of South Korea, Naver, and Coupang, announced that their sales increased by 35% and 54% compared to 2020 and recorded 1.47 trillion Korean Won and 18.4 billion dollars, respectively.[1]

Consumers can access the assessment, information, and recommendation of others in online marketplaces much more quickly than in traditional marketplaces as references for their purchases. If a product is positioned at the top of a website, a consumer might consider it a good product, so she is more likely to purchase it. In this sense, it is significant for sellers to expose their products at the top of a website. This kind of marketing is already a popular concept in digital marketing; if the product is located at the top position of the search engine results page via paid advertising, it is called search engine marketing (SEM). If a seller or a company tries to win the top position in an organic search result, it is called search engine optimization (SEO).

So far in marketing literature, numerous works have analyzed the effect of SEM on search engines (Chan and Park 2015; Dou et al. 2010; Ghose, Ipeirotis, and Li 2014; Jerath, Ma, and Park 2014). Some

[1] https://www.fnnews.com/news/202203120850523475

of them have also studied rank effects in the online marketplace (Morozov et al. 2021a; Ursu 2018). Although they successfully figured out the heterogeneous position effect, they did not deal with the heterogeneity of the position effect across product categories. This paper generates new knowledge from this point of view.

It is an essential issue for marketing managers to consider a new marketing campaign for their products in an online marketplace. The managers would wonder whether the campaign should aim to improve the brand value or improve its position at the top. To answer this question, this paper estimated the position effect in the most purchased categories and identified how the characteristics of categories affect the position effect. If the product for which a marketing manager wants to implement a campaign has a sensitive position effect, she would be advised to execute the SEO. If not, she would make her brand valuable first in the long-term perspective.

This research took the empirics-first (EF) approach rather than the theory-first approach (Golder et al., 2022). Although the EF approach does not seem rigorous from the theory view, the EF approach is more suitable here because of the nature of this research, which studies highly empirical and managerial questions.

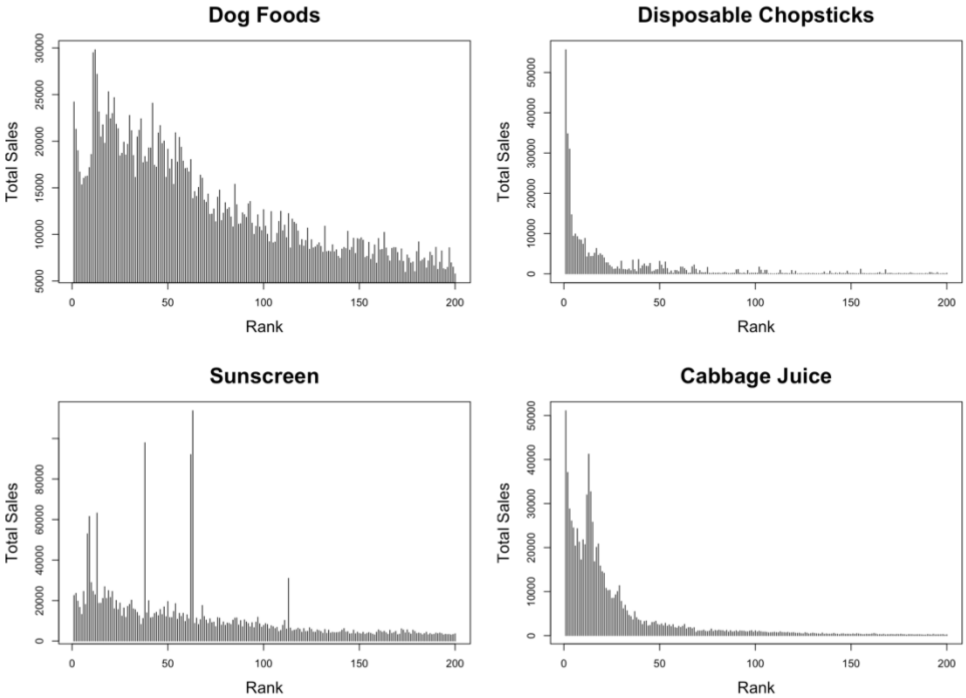

Figure 1. Casual observations of relationships between total sales and the rank of a few keywords.

Figure 1 shows relationships between rank and sales volume for a few keywords. For example, the graph in the upper left of figure 1 stands for the entire sales when the data was collected and summed in the same ranks from 1 to 200 of the keyword of dog foods. As the graphs show, the relationships between total sales and the ranks are remarkably different in the keywords. In the dog foods keyword, the total sales decrease in rank almost linearly. In the disposable chopsticks keyword, the total sales decrease significantly and exponentially in rank, while in the sunscreen keyword, it decreases slowly. In the cabbage juice keyword, it reduces sharper than sunscreen but more slowly compared to disposable chopsticks.

The primary research question of this paper starts from this graph: what is in the relationship between product or product category characteristics and rank effects? Which product category has a more sensitive rank effect? If this question could be answered, the marketing managers of online commerce can make decisions on their marketing strategy, not just from their intuition and experiences but from the data. To this end, this research used all available data in the online marketplace and connected them to the pieces of marketing literature.

The remaining part of this paper is organized as follows. The next section reviews the precedent literature, primarily on the position effect on websites and consumer behavior theories related to categorizing products. Section 3 introduces the data. Section 4 shows the models specifying heterogeneity across the position effect of categories. Section 5 shows the results of the model estimation and interpretation. Section 6 concludes.

Chapter 2. Literature Review

2.1. Position Effect

Numerous pieces of literature in the marketing field have already studied the position effect online from various views. Especially most of them investigate how the product characteristics affect the position effect. The most relevant literature to this research is how effects differ in position by product categories on mobile. Bart, Stephen, and Sarvary (2014) predominantly investigated what product-related conditions of mobile display advertising (MDA) are beneficial in influencing consumer attitudes and purchase intention. When it comes to product-related conditions, scholars adopted utilitarian consumption and hedonic consumption as one axis and involvement as the other axis. They finally concluded that in utilitarian and high-involvement settings, the effect of MDA is relatively significant, so the setting is more worthwhile to execute than the other settings.

Other relevant work to this research investigates the endogeneity of rank effect and the rank effect on click-through rates (CTR). Earlier, Ghose and Yang (2009) already found the rank effect on the search engine in the context of sponsored marketing. Ghose, Ipeirotis, and Li (2014) expanded this topic to the product search engine, an online travel agency (OTA) where consumers search for hotels. Ursu (2018) further investigated the effect of rankings on the OTA platform. There are some implications for these works. One is that ranking itself is an endogenous variable. Those position effect literature constructed models to deal with the endogeneity or implemented randomized experiments to tackle this issue. The other significant problem is that they dealt with click-through rates (CTRs), not with sales. As Ursu (2018) found out, the rank effect does influence CTRs but not sales.

Although many works have already investigated the heterogeneity of the position effect, they have been in the context of sponsored search on search engines. One mainstream is to identify the heterogeneity of rank effects across keywords, as the work of Rutz, Trusov, and Bucklin (2011). They proposed different strategies to bid or select a keyword to advertise for marketing managers in other purposes and situations. Ghose and Yang (2009) and Agarwal, Hosanagar, and Smith (2011) found that a lower position makes fewer click-through rates, but the conversion rates increase. Especially this effect gets more substantial for more specific keywords. Conclusively, they showed that the topmost position is not always profitable. Blake, Nosko, and Tadelis (2015), in sponsored results on eBay, one of the most dominant e-commerce of the United States, found a specific heterogeneity of the position effect across keywords. The brand keywords had no short-term benefit, while the non-brand keywords significantly affected consumers who rarely purchase. However, the frequent purchasers are not influenced by those ads, resulting in the total returns not being profitable.

Another stream is heterogeneity across advertisers. Jerath et al. (2011) investigated the position paradox in sponsored search that the superior firms with lower positions gain more clicks than the inferior ones with higher ranks. Given that the higher position needs more bids to win, the result may imply that the brand power exceeds the position effect in some contexts. Dou et al. (2010) also proposed a brand positioning strategy in information system literature: they found that the unknown brand can get a favorable reputation when its position is higher than the already-renowned brands in their research. Narayan and Kalyanam (2015) found that the position effect is more substantial when the advertiser is small, and consumers have rare experiences with the keyword. The position effect is weaker when the keyword hints at a specific brand or product information. Baye, De Los Santos, and Wildenbeest (2016) and Jeziorski and Moorthy (2018) further found in organic product search and search advertising that the prominence of advertisers could reinforce the position effects, so the prominent advertisers do not necessarily need the top position.

Even if these works successfully studied the position effect, their setting was in sponsored search, and only a few were in organic search. Also, they focused on the search engine and the CTR of the sites. This paper is different from previous research since it has a different context, and it studies product sales in organic search. The reason why organic search is essential is already found. Jerath, Ma, and Park (2014) found that most consumers' clicks are for organic search results rather than sponsored ones, although the sponsored ones are located on the upper side than the organic ones in SERPs (Search Engine Result Pages). Furthermore, they found that consumers who search less popular keywords tend to click more per search. Finally, they connected the consumer's search behavior with involvement: the involvement of a consumer is correlated inversely to the popularity of a keyword.

The nature of the position effect is highly connected to the search cost and product uncertainty. Since consumers' search starts from the top-ranked items (Granka et al. 2004), searching for lower-ranked items requires more search costs. Earlier marketing literature already showed that the assumption that the search attractiveness decreases as the search cost increases and product uncertainty increases is rational and modeled the concept quantitatively (Kim, Albuquerque, and Bronnenberg 2010). In this context, it is reasonable to speculate that consumers would be likely to search for a higher position due to the search cost increasing as they search for a lower position, and they would be possible to search for a lower position to decrease product uncertainty if the product they are searching has a high level of uncertainty.

2.2. Purchasing Behavior

An early approach in economics to categorize products is whether consumers can determine the product quality before purchasing. If so, the product is called search goods, or otherwise, it is called experience goods. (Nelson 1970, Nelson 1974) Internet, however, significantly increased the capability of consumers to search for a product and has changed their search behaviors. The scholars discovered that consumers who search for experience goods tend to spend more time on a page and be more affected by other consumers' feedback, whereas consumers who search for search goods tend to explore more pages (Huang, Lurie, and Mitra 2009).

Early marketing literature first considered the concepts of involvement and differences between brands (Assael 1987; Kotler and Armstrong 1988), which cause different product purchasing behavior of consumers. Involvement has been considered as an axis of characteristics of products or consumer purchasing behavior. Petty, Cacioppo, and Schumann (1983), Holbrook and Lehmann (1983), and Celsi and Olson (1988) investigated that involvement moderates the effect of advertisement. Consumers pay more attention to an ad in a high-involvement setting than a low one. Bloch (1983) made a connection between involvement and risk. Taylor and Joseph (1984) claimed that high- or medium-priced and durable goods have high involvement, while low-priced and non-durable goods have low involvement. Laurent and Kapferer (1985) broadened the nature of involvement by identifying its profile; it is related not solely to risk but also other constructs, including hedonic value.

Meanwhile, modern consumer behavior marketing literature considers hedonic and utilitarian characteristics or consumptions for categorizing products or consumers' behavior. Dhar and Wertenbroch (2000) investigated the different choice behavior between hedonic and utilitarian goods. Childers et al. (2001) also found that purchase motivation differs in online settings. Li et al. (2020) discovered that consumers employ different search paths in purchase characteristics; for utilitarian purchases, they use social media and product pages while searching for third-party reviews for hedonic purchases.

Chapter 3. Data

3.1. Data Source and Shopping Environment

Storelink provided the data for this paper. Storelink is a South Korean startup company to offer marketing solutions in online commerce. The provided data is collected from Naver Shopping, operated by a leading Korean search engine, Naver.com, similar to Amazon.com in the United States and Alibaba in China. The online service aggregates different seller items by the same categories or keywords. Once a consumer requests a query on the shopping site, it provides corresponding search results, in which sponsored products appear at the top of the list while organic results follow. Figure 2 shows an example of search results after requesting the keyword “sunscreen” in Korean, where the blue box indicates a sponsored product. Storelink collects data on the position of each product in the organic list, from the top to the 200^{th}, depending on each searched keyword, which is selected thanks to its dominance, popularity, or high demand from the marketing consulting clients. The data contains numerous product information, such as prices and product characteristics.

Sellers upload their products’ descriptions on Naver Shopping using a standardized format for different products so that consumers can consistently view essential information specific to each product type. As depicted in Figure 2, product characteristics such as

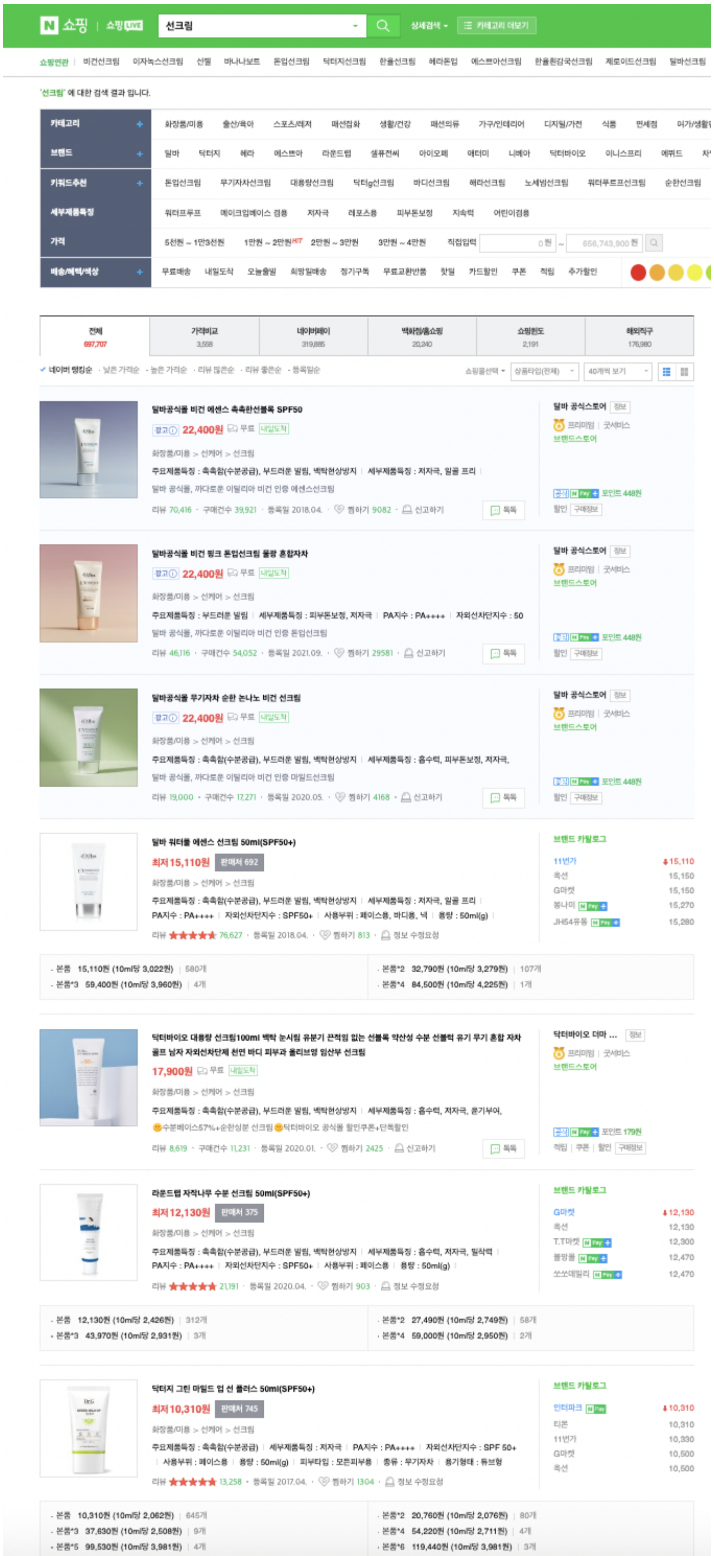

Figure 2. An example of SERPs, which resulted from a query "sunscreen."

moisturizing can be found. This research exploited this information by counting how many product characteristics are provided on the websites, and this variable will be called the number of attributes hereafter. Meanwhile, the shopping portal categorizes products into four depths. For example, a product from Figure 2 can be categorized into the sunscreen category (3rd depth) of the sun-related product category (2nd depth) of cosmetics (1st depth). Nevertheless, most of the products are classified into third depth. This research targets the heterogeneous rank effects across these third-depth categories.

Additionally, the cumulative number of product reviews has also been collected. These are kinds of electronic word of mouth in which consumers who have bought the product leave messages about the product. Although the review data used in this paper do not imply any further information, such as the positiveness or negativeness of the reviews, it still includes the number of reviews for a product cumulatively. The data also includes the number of questions and answers (Q&A). If consumers have questions about the product they are willing to buy, they can directly ask the sellers about the product to resolve the uncertainty about the product. The reviews are notes consumers leave after purchasing, whereas the questions and answers are notes consumers leave before purchasing. In addition, the promotion information is also gathered. This research considered the simple specification: if any price discount promotion is being executed for a product at the time, this will be 1; otherwise, 0. The sales are also collected. It indicates the total sales of the previous three days. Although the data is the total sales, not daily sales, this variable is used directly. Plus, missing values in sales data are recorded as "0," which makes it impossible to distinguish whether the product was not sold or the data was missed. The data with zero-value sales are filtered out to avoid misinterpretation.

Storelink also collects how many keywords are searched. Although this is not an accurate number, they calculated it using their algorithm; Naver.com provides the proportion of the number of keywords searched each day to the maximum number across a given period. Also, the shopping portal provides an exact value of the number of keywords searched in a day. Collecting these two pieces of information, the marketing company conducts reasonable computation about the keyword search volumes. However, it only collects some of the keywords containing some popular or marketing-targeting keywords. Also, the data for unpopular keywords with a volume below zero is not provided if the number is below 10. Therefore, it is assumed that the uncollected keywords are not frequently searched, and their volumes are substituted with 5, the average number of 0 to 10.

3.2. Data Selection and Description

I prepared a sample dataset for the empirical analysis to reduce the computation burden since the data contains observations of more than 200 million. First, the daily data is aggregated into a weekly basis. If an item is observed at least a day in a week, it can be included in the week's observation. After the aggregation, the data have 4.7 million observations of 420 thousand items from the 37^{th} week of 2021 to the 38^{th} week of 2022. Categories in which the average number of products across the observed period is less than 200 are dropped. Since the company collects 200 items on a keyword daily, it is rational that if a category contains pertinent information, one category will include at least 200 items since it can contain different results from different keywords. Namely, a category can cover products more than a keyword. The data of products observed less than three times were dropped for the panel analysis. This process selected 177,491 products from 189 categories, totaling 2,233,692 observations (unbalanced panel).

The rank variable increases when the position is lower. For ease of interpretation, minus rank $\hat{r}_{jkt,w}$ is used.

$$\hat{r}_{jkt,w} = -Rank_{jkt,w} \qquad -(1)$$

where $Rank_{jkt,w}$ stands for the rank value of product j from the category k and keyword w at time t. The rank variable is derived from various keywords. Table 1 shows an example of a part of the data.

Although all the data in table 1 indicates the same product of the same date, they are gathered from different keywords. As a result, a product's data can contain various rank values while others are equal. Direct use of this data might cause bias since duplicated data is used. The rank index r_{jkt} is considered in order to avoid the potential issue, defined by the weighted average of the searched volume of the keyword.

$$r_{jkt} = \frac{\sum_{w}^{W} \widehat{r_{jkt,w}} \times n_{j,w}}{\sum_{w}^{W} n_{j,w}} \qquad -(2)$$

where $n_{j,w}$ stands for the searched volume of the keyword w where the product j is shown.

Since the original data were highly skewed, all the variables, except for the promotion binary variable and the variable indicating time lag from the first product registration on the marketplace, were log-transformed. Table 2 shows the descriptive statistics of the raw data. The value of one is added for some variables when the minimum is zero for log transformation or to keep the values from getting negative.

Three-level variables were considered in this research, apart

Keyword	Product Name	Rank	Date	Price	Keyword Search Volume	...
Dog Summer Clothes	Cute Dog Clothes	1	2022-10-01	7400	12	...
Dog	Cute Dog Clothes	74	2022-10-01	7400	5698	...
Dog Clothes	Cute Dog Clothes	35	2022-10-01	7400	4513	...
...	...	...	...	...	...	...
Dog Hoody	Cute Dog Clothes	5	2022-10-01	7400	48	...
Pet Supplies	Cute Dog Clothes	42	2022-10-01	7400	39	...

Table 1. An example of raw data which shows a data of same product from different keywords.

from the rank index. First, the time-variant product-level variable vector X_{jkt} was used. The variables include (1) the product price, (2) the cumulative number of reviews, (3) the cumulative number of Q&A, (4) the indicator variable indicating whether the promotion is executed, and (5) the lagged time from the first registration of the product on the marketplace (called as time difference hereafter). Product-level average variable vector $\bar{Y}_{jk}$ is also considered, (1) the average price.

of the product across time, (2) the average number of reviews across time, (3) the average number of Q&A across time, and (4) the average time difference across time. That is, for a time-variant variable x_{jkt}, an element of X_{jkt}, gives the time-variant variable $\bar{y}_{jk}$, an element of $\bar{Y}_{jk}$, by $\bar{y}_{jk} = \frac{1}{T_{jk}} \sum_t^{T_{jk}} x_{jkt}$, where T_{jk} refers to the observed number of times of the product j from category k. The average variables $\bar{Y}_{jk}$ contain the time-invariant information of a product by doing so.

The category-level variable vector $\bar{Z}_k$ is considered as well. This vector includes (1) the average characteristics across products in category k, (2) the standard deviation of the prices in the category, (3) the number of brands in the category, (4) the average number of Q&As across products, (5) the average number of the reviews across

	mean	std	min	25%	50%	75%	max
$sales_{jkt}$	56.72	642.18	1.00	2.33	6.71	23.17	128398.00
$price_{jkt}$	30789.40	69225.44	10.00	8000.00	15900.00	29900.00	7421455.0
r_{jkt}	94.55	56.79	1.00	45.00	91.43	143.00	200.00
$promotion_{jkt}$	0.48	0.50	0.00	0.00	0.00	1.00	1.00
$Q\&A_{jkt}$	41.12	298.29	0.00	0.00	0.00	11.00	24616.50
$review_{jkt}$	1931.43	8110.98	0.00	81.00	347.00	1271.00	688306.00
$timediff_{jkt}$	591.17	634.70	0.00	166.00	405.00	786.00	6702.17

Table 2. Descriptive statistics of data.

products, (6) the average number of the attributes of the products in the category, and (7) the standard deviation of the number of the attributes of the products in the category. Please note that these variables are not directly derived from X_{jkt} or $\bar{Y}_{jk}$, which only have information about the selected products of the categories. Instead, $\bar{Z}_k$ includes the information of all products in category k. Note that for any variable x, $\mu_{x,k} = \frac{1}{\sum_{j=1}^{J} T_{jk}} \sum_{j=1}^{J} \sum_{t}^{T_{jk}} x_{jkt}$, namely, the average value across products in a category, and $\sigma_{x,k} = \sqrt{\frac{\sum_{j,k}^{J,k} \sum_{t}^{T_{jk}} (x_{jkt} - \mu_{x,kt})^2}{\sum_{j,k}^{J,k} T_j}}$, the standard deviation across products in a category, and $brand_k$ refers to the number of brands being sold in category k.

Chapter 4. Model

4.1. Model Development

The primary model is considered with product- and time-fixed effects. (Wooldridge 2002a).

$$log\, s_{jkt} = \beta_0 + \beta_j + \beta_t + \alpha_{jk}\, log\, r_{jkt} + \gamma X_{jkt} + e_{jkt} \qquad -(3)$$

where β_0 stands for the intercept, β_j stands for the product fixed effect, β_t stands for the time fixed effect, e_{jkt} is an error term following the iid distribution with mean zero. Following the specification of Wooldridge (2002a), each fixed effect contains $J-1$ product-fixed effect for a category k and $T-1$ time-fixed effect for a product j of category k for the identification. Remark that the rank effect α_{jk} is specified with heterogeneity across products. Like Ghose and Yang (2009), α_{jk} is specified with the mean value of products and categories:

$$\begin{aligned} log\, s_{jkt} = \beta_0 + \beta_{jk} + \beta_t + \left(\alpha_0 + \alpha_1 \bar{Y}_{jk} + \alpha_2 \bar{Z}_k\right) log\, r_{jkt} + \gamma X_{jkt} \\ + e_{jkt} \end{aligned} \qquad -(4)$$

in which the coefficients can be estimated with the ordinary least squares. Here, α_0 indicates the main effect of the rank, and α_1 and α_2 stand for the impact of products' characteristics on the rank effect and the effect of categories' characteristics on the rank effect, respectively. By identifying $\{\alpha\} = \{\alpha_0, \alpha_1, \alpha_2\}$, the model describes the market-average effect of the product-level and category-level on the rank effect.

Chapter 5. Results

5.1. Estimation

In this analysis, the large number of products causes the curse of dimensionality: to estimate all the fixed effects, the model has to estimate $m + \sum_{k=1}^{K} J_k + \mathrm{T}$ coefficients, where m means the number of independent variables. In this case, the total dimension is 18+ 177,491+ 52=177,561. Instead, fifty products from 150 categories were randomly selected to reduce further computation burden. Some products have been registered as categories of more than one. It can happen if a seller registers a product in different categories on different days. The final number of the products analyzed is 7,447 products, not 7,500.

		Model (1)	Model (2)	Model (3)	Model (4)
β_0	Intercept	6.3314***	6.6229***	6.5216***	6.619***
$\gamma\ (X_{jkt})$		(0.327)	(0.335)	(0.331)	(0.339)
	$\log price_{jkt}$	-0.3768***	-0.378***	-0.376***	-0.3797***
		(0.018)	(0.018)	(0.018)	(0.018)
	$promotion_{jkt}$	0.0751***	0.0734***	0.0678***	0.0665***
		(0.021)	(0.021)	(0.021)	(0.021)
	$\log Q\&A_{jkt}$	0.0304***	0.0311***	0.0301***	0.0306***
		(0.002)	(0.002)	(0.002)	(0.002)
	$\log review_{jkt}$	0.0332***	0.0316***	0.0339***	0.0321***
		(0.006)	(0.006)	(0.006)	(0.006)
	$timediff_{jkt}$	0.0000	0.0000	-0.0000	-0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
α_0	$\log r_{jkt}$	0.3058***	0.3967***	0.9417***	0.8267***
		(0.005)	(0.048)	(0.109)	(0.111)
$\{\alpha_1\}\ (\bar{Y}_{jk})$	$\mu_{price,jk}$		-0.0158***		0.001
			(0.004)		(0.005)
	$\mu_{Q\&A,jk}$		-0.041***		-0.0367***
			(0.004)		(0.004)
	$\mu_{review,jk}$		0.0274***		0.0269***
			(0.003)		(0.003)
	$\mu_{timediff,jk}$		-0.0001***		-0.0001***
			(0.0000)		(0.0000)
$\{\alpha_2\}\ (\bar{Z}_k)$	$\mu_{price,k}$			-0.0155	-0.008
				(0.016)	(0.018)
	$\sigma_{price,k}$			-0.0261**	-0.0248*
				(0.013)	(0.013)
	$brandcnt_k$			0.004	0.0034
				(0.006)	(0.006)
	$\mu_{Q\&A,k}$			-0.0105***	-0.0089***
				(0.003)	(0.003)
	$\mu_{review,k}$			-0.0158**	-0.0217***
				(0.008)	(0.008)
	$\mu_{attributes,k}$			-0.017	-0.0237
				(0.015)	(0.015)
	$\sigma_{attributes,k}$			-0.0392**	-0.0315*
				(0.016)	(0.016)
	Fixed Effects	Yes	Yes	Yes	Yes
	R^2	0.8390	0.8400	0.8400	0.8400
	$Adjusted R^2$	0.8240	0.8250	0.8240	0.8250
	F-statistics	55.1	55.22	55.17	55.26

Table 3. The estimation results of models.
Notes. Standard errors are in parenthesis.
$*** < .01$; $** \, p < .05$; $* \, p < .1$

5.2. Results

Table 3 shows the estimation results of the primary model, equation 4. Model (4) refers to the primary model, while model (1), model (2), and model (3) stand for the model without $\bar{Y}_{jk}$, $\bar{Z}_k$, and both $\bar{Y}_{jk}$ and $\bar{Z}_k$, respectively. Although all models show similar R^2, adjusted R^2, model (4) presents slightly higher than any other models do. Thus, I will mainly discuss and interpret the result of model (4).

The result shows that the category characteristics $\bar{Z}_k$ covers essential information on the rank effect. The coefficients of X_{jkt} were similar between fixed effect variables, but the main effect α_0 of rank r_{jkt} was underestimated in the models without $\bar{Z}_k$. In model (4) and model (3), α_0 was estimated as .83 and .94, respectively, whereas that of model (1) and (2) was .31 and .40. Additionally, $\mu_{price,jk}$ was insignificant in model (4) ($p>.6$). At the same time, it was significant in model (2) ($p<.01$). This may imply that the nature of $\mu_{price,jk}$ is from $\bar{Z}_k$ than itself.

The estimated γ, the coefficients of X_{jkt}, suggests that the estimates have face validity. The estimated price elasticity γ_{price} is −.38. Note that the coefficient estimated from the log-log model can be directly interpreted as the elasticity. The elasticity is in the reasonable interval analyzed in the meta-analysis (Bijmolt, Van Heerde, and Pieters, 2005). The number of reviews, a kind of word of mouth (WOM), was also positively related to product sales as the literature ($\gamma_{review} = .0269$; Chevalier and Mayzlin, 2006). The number of Q&A also positively affected product sales ($\gamma_{Q\&A} = .0306$).

There were some interesting findings in the category-level estimates. First, a category with significant price variance experienced a weak rank effect (Coefficient of $\sigma_{price,k} = -.0248$). Similarly, a category with a significant variance in the number of attributes experienced a weak rank effect (Coefficient of $\sigma_{attributes,k} = -.0315$). On the other hand, the average price and the average number of attributes do not impact the rank effect (p-value of coefficient of $\mu_{price,k} = .651$, p-value of coefficient of $\mu_{attributes,k} = .116$). This result means that the characteristics influencing the rank effect are not the average level of price and product attributes but the price dispersion and heterogeneity across products. It may be because consumers expect to find a more satisfying product or price by searching to lower ranks if the category has a considerable price variance and heterogeneity across products. One significant issue is that the p-value of the coefficient of $\mu_{attributes,k}$ is marginally higher than the 10% significance level. It implies that the variable is significant if the sample, model, or estimating method is changed. If this is the case, the potential result will imply that overall category characteristics — such as utilitarian-consumed vs. hedonic-consumed, information goods vs. search goods, and high-involvement goods vs. low-involvement goods — can also affect the rank effect.

On top of that, categories with more Q&As and reviews tended to have smaller rank effects (coefficient of $\mu_{Q\&A,k} = -.0089$, coefficient of $\mu_{review,k} = -.0217$). The category where consumers have more significant uncertainty about their purchasing products may have a weaker rank effect since they search for low-ranked products. This interpretation is plausible since the two variables can be interpreted as measuring the information asymmetry in a category. They resolve product uncertainty by providing information as word of mouth (Berger, 2014).

Additionally, the coefficients of an averaged number of reviews have different signs by its level (coefficient of $\mu_{review,jk} = .0269$, coefficient of $\mu_{review,k} = -.0217$). It can mean that categories in which consumers tend to make WOM have a weaker rank effect, but products with more reviews tend to have a more substantial rank effect in a given category. The lagged time was not significant in the level of product (p-value of coefficient of $timediff_{jkt}$ > .8), whereas it was significant in the level of average (coefficient of $\mu_{timediff,jk}$ = -.0001). Even if the effect seems extremely marginal, since the coefficient is not normalized, it cannot be concluded that it has no impact. It may hint that as the period a product was registered is older, the rank effect becomes weaker. Competition intensity across brands was not a significant predictor for the rank effect (p-value of coefficient of $brandcnt_k$ > .6).

Chapter 6. Discussion

6.1. Conclusion and Managerial Implication

In summary, the model estimates suggested that the rank effect was significantly related to product and category characteristics. While the category characteristics may influence the rank, the rank effect decreased in the heterogeneity and price dispersion. In contrast, overall characteristics and average price did not have sufficient evidence to affect the rank. Furthermore, a category with information asymmetry tended to attenuate the rank effect. This result proposes the possibility that different rank effects across product categories exist, and the heterogeneity, price dispersion, and information asymmetry make the rank effect different across product categories.

Marketing managers, who consider the profitability of performance marketing, such as SEM or SEO in online commerce, can be recommended to implement the SEO on the online marketplace if their products belong to a homogeneous, low price-dispersion category with sufficient information provided to consumers online. When it is ambiguous to identify the marketing strategy that is more profitable than other strategies, they can directly calculate the expected profit from the marketing campaign and speculate which one fits their purpose and circumstances more. If they conclude that SEO is not sufficiently profitable, they can instead consider price-discount promotions or establish long-term strategies such as brand marketing.

This research contributed to the marketing literature in three points. Firstly, this study exploited a new and unique dataset and produced a different result. Meanwhile, by taking the empirics-first approach, this research focused on investigating data patterns and connecting them to the existing theories. Second, this research took a snapshot of the online marketplace by studying numerous product categories simultaneously rather than focusing on a category. Although existing studies in the online marketplace have already proceeded with their research, they focused on a category such as a hotel industry (Ursu, 2018). In contrast, this paper analyzes various market categories and analyzes them. In this sense, this paper highlights studies on the general online marketplace.

The most important contribution of this research is to identify the heterogeneous rank effect in the online marketplace. The latest literature spotlighted the search, especially on the sponsored search in search engines. Blake, Nosko, and Tadelis (2015) studied the heterogeneous keyword effect in online commerce (eBay), but the rank effect was not focused on. This paper extended the latest topic of marketing literature to a new area.

6.2. Limitations and Directions for the Future Research

As discussed, this research only employed some of the data because of the curse of dimensionality. Although the sample is randomly selected, it would be more robust research after using all the data, and its generalizability will also be reinforced.

Furthermore, the model in this paper estimated the rank effect by interaction for simplicity. For rigor, the hierarchical method can estimate the same model (Rossi, Allenby, and McCulloch, 2003). If so, the rank effect will be a function of product and category characteristics.

One limitation of this research is that the endogeneity in the rank variable needs to be dealt with appropriately. Recent marketing literature has already found that the rank variable coefficient estimates suffer from endogeneity because the former variables significantly affect the present rank value. This nature of the rank variable triggers the simultaneity bias (Ghose and Yang, 2009). In this sense, the rank variable and the error term may be positively correlated. That is, $cor(r_{jkt}, e_{jkt}) > 0$. Thus, the rank effects in this study might be overestimated than actual values.

The past literature suggested various methodologies to manage the endogeneity for causal inferences. Ghose and Yang (2009) and De los Santos and Koulayev (2017) solved the problem by estimating simultaneous equations. Similarly, Baye, De los Santos, and Wildenbeest (2016) tackled the issue using the ranks of the same products or keywords from the other search engine as the instrumental variable. However, this research exploited the rank index, the weighted average of rank values on keyword search volume, so it is difficult to find appropriate instrumental variables. Instead, some researchers proposed different approaches. Rutz and Trusov (2011) dealt with the endogeneity issue with the latent instrumental variable approach proposed by Ebbes et al. (2005). Meantime, Narayanan and Kalyanam (2015) carried out their research on regression discontinuity design. These two approaches can be appropriate for the rank index approach. Other than that, randomizing experiment by Ursu (2018) can also be one of the possible choices. By doing these directions, future research is expected to provide more rigorous causality in rank effect and purpose empirical solutions for the contemporary marketing issues in the online marketplace.

Bibliography

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, location, location: An analysis of the profitability of position in online advertising markets," *Journal of Marketing Research*, 48 (6).

Bajari, Patrick, Jeremy T. Fox, and Stephen P. Ryan (2008), "Evaluating wireless carrier consolidation using semiparametric demand estimation," *Quantitative Marketing and Economics*, 6 (4).

Bart, Yakov, Andrew T. Stephen, and Miklos Sarvary (2014), "Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions," *Journal of Marketing Research*, 51 (3).

Baye, Michael R., Babur de los Santos, and Matthijs R. Wildenbeest (2016), "What's in a name? Measuring prominence and its impact on organic traffic from search engines," *Information Economics and Policy*, 34.

Berger, Jonah (2014), "Word of mouth and interpersonal communication: A review and directions for future research," *Journal of Consumer Psychology*.

Bijmolt, Tammo H.A., Harald J. van Heerde, and Rik G.M. Pieters (2005), "New empirical generalizations on the determinants of price elasticity," *Journal of Marketing Research*, 42 (2).

Blake, Thomas, Chris Nosko, and Steven Tadelis (2015), "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment," *Econometrica*, 83 (1).

Bloch, Peter H (1981), "An Exploration Into The Scaling Of Consumers' Involvement With A Product Class," *Advances in Consumer Research*, 8 (1).

Celsi, Richard L. and Jerry C. Olson (1988), "The Role of Involvement in Attention and Comprehension Processes," *Journal of Consumer Research*, 15 (2).

Chan, Tat Y. and Young Hoon Park (2015), "Consumer search activities and the value of ad positions in sponsored search advertising," *Marketing Science*, 34 (4).

Chevalier, Judith A. and Dina Mayzlin (2006), "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*.

Childers, Terry L., Christopher L. Carr, Joann Peck, and Stephen Carson (2001), "Hedonic and utilitarian motivations for online retail shopping behavior," *Journal of Retailing*, 77 (4).

Dhar, Ravi and Klaus Wertenbroch (2000), "Consumer choice between hedonic and utilitarian goods," *Journal of Marketing Research*, 37 (1).

Dou, Wenyu, Kai H. Lim, Chenting Su, Nan Zhou, and Nan Cui (2010), "Brand positioning strategy using search engine marketing," *MIS Quarterly: Management Information Systems*, 34 (SPEC. ISSUE 2).

Ghose, Anindya and Sha Yang (2009), "An empirical analysis of search engine advertising: Sponsored search in electronic markets," *Management Science*, 55 (10).

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li (2014), "Examining the impact of ranking on consumer behavior and search engine revenue," *Management Science*, 60 (7).

Golder, P. N., Dekimpe, M. G., An, J. T., van Heerde, H. J., Kim, D. S., & Alba, J. W. (2022), "Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation," *Journal of Marketing*, 00222429221129200.

Greene, WH William H . (2012a), *Econometric analysis 7th Ed*, *Prentice Hall*.

Holbrook, Morris B. and Donald R. Lehmann (1980), "Form versus content in predicting Starch scores.," *Journal of Advertising Research*.

Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra (2009), "Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods," *Journal of Marketing*, 73 (2).

Jerath, Kinshuk, Liye Ma, and Young Hoon Park (2014), "Consumer click behavior at a search engine: The role of keyword popularity," *Journal of Marketing Research*, 51 (4).

Jerath, Kinshuk, Liye Ma, Young Hoon Park, and Kannan Srinivasan (2011), "A 'position paradox' in sponsored search auctions," *Marketing Science*, 30 (4).

Jeziorski, Przemyslaw and Sridhar Moorthy (2018), "Advertiser prominence effects in search advertising," *Management Science*, 64 (3).

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online demand under limited consumer search," *Marketing Science*, 29 (6).

Laurent, Gilles and Jean-NOËL Kapferer (1985), "Measuring Consumer Involvement Profiles," *Journal of Marketing Research*, 22 (1).

Li, Jingjing, Ahmed Abbasi, Amar Cheema, and Linda B. Abraham (2020), "Path to Purpose? How Online Customer Journeys Differ

for Hedonic Versus Utilitarian Purchases," *Journal of Marketing*, 84 (4).
Lordan, Oriol (2014), "Study of the full-service and low-cost carriers network configuration," *Journal of Industrial Engineering and Management*.
De los Santos, Babur and Sergei Koulayev (2017), "Optimizing click-through in online rankings with endogenous search refinement," *Marketing Science*, 36 (4).
Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta (2004), "Response modeling with nonrandom marketing-mix variables," *Journal of Marketing Research*.
Morozov, Ilya, Stephan Seiler, Xiaojing Dong, and Liwen Hou (2021a), "Estimation of preference heterogeneity in markets with costly search," *Marketing Science*, 40 (5).
Narayanan, Sridhar, and Kirthi Kalyanam (2015), "Position effects in search advertising and their moderators: A regression discontinuity approach," *Marketing Science*, 34 (3).
Nelson, Phillip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2).
Nelson, Phillip (1974), "Advertising as Information," *Journal of Political Economy*, 82 (4).
Petty, Richard E., John T. Cacioppo, and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *Journal of Consumer Research*, 10 (2).
Rossi, Peter E. and Greg M. Allenby (2003a), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3).
Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2006a), *Bayesian Statistics and Marketing*, *Bayesian Statistics and Marketing*.
Rutz, Oliver J. and Michael Trusov (2011), "Zooming in on paid search ads-A consumer-level model calibrated on aggregated data," *Marketing Science*, 30 (5).
Rutz, Oliver J., Michael Trusov, and Randolph E. Bucklin (2011), "Modeling indirect effects of paid search advertising: Which keywords lead to more future visits?," *Marketing Science*, 30 (4).
Traylor, Mark B. and W. Benoy Joseph (1984), "Measuring consumer involvement in products: Developing a general scale," *Psychology & Marketing*, 1 (2).
Ursu, Raluca M. (2018), "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 37 (4).

Wooldridge, Jeffrey M (2002), "Econometric Analysis of Cross Section and Panel Data," *MIT Press,* 108 (2).

Abstract

본 논문은 온라인 커머스에서 순위 효과의 이질성을 연구하여 마케팅 관리자가 최적의 디지털 마케팅 전략을 설정하는데 유용한 시사점을 제안하였다. 온라인 커머스는 급격하게 크고 있다는 특성 때문에, 위치 효과 혹은 순위 효과는 마케팅 연구에서 중요한 주제이다. 검색 광고가 아닌 자연 검색 결과 (Organic Result)가 더 중요하다는 것이 보고되었지만, 최근의 연구는 대부분 검색 광고 (Sponsored Search)에 초점을 맞추고 있다. 이 연구는 온라인 커머스에서 유기적 결과의 효과에 초점을 맞추었다는 점에서 새로운 연구이다. 연구에서 상품 수준에서의 분석을 위해 키워드 검색량을 통한 가중평균을 활용하여 새로운 개념의 순위지수를 제안하였다. 모형에서 순위 효과는 제품 수준, 제품군(Category) 수준과 순위지수 간 교호작용으로 식별되었다. 한편, 시간에 따라 변화하는 제품 수준의 변수와 이원고정효과 (Two-way Fixed Effect)가 공변량으로 활용되었다. 차원의 저주 (Curse of Dimensionality) 문제를 해결하기 위해 제품의 일부를 무작위로 선정하여 분석을 진행했다. 모형 추정 결과는 제품 간 가격의 분산과 이질성이 높은 제품군이 낮은 순위 효과를 갖고 있음을 시사한다. 또한, 정보 비대칭이 있는 제품군은 낮은 순위 효과를 갖는 경향을 보였다. 한편, 평균 가격, 제품 속성 및 경쟁 강도와 같은 제품군 내 상품들의 전반적인 특성은 순위 효과에 영향을 미친다는 증거를 발견할 수 없었다. 결론적으로, 마케팅 관리자들은 만약 제품이 높은 수준의 순위 효과를 갖는 제품군에 속한다면 온라인 커머스에서의 검색엔진최적화(Search Engine Optimization)를 통한 마케팅

전략을 고려할 수 있다. 이 논문은 광범위한 데이터셋을 활용한 온라인 시장의 전반적인 묘사를 통해 마케팅 문헌을 새로운 영역으로 확장했다는 점에서 의의를 가진다. 계층적 모델링과 내생성을 고려한다면 향후 연구는 보다 강력하고, 엄격한 인과관계를 규명할 수 있을 것으로 기대된다.

Master's Thesis of Mingi Goo

Heterogeneous Rank Effects in Online Marketplace

온라인 커머스에서 랭킹 효과의 이질성 분석

February 2023

College of Business Administration
Seoul National University
Marketing Major

Heterogeneous Rank Effects in Online Marketplace

Mingi Goo

Submitting a master’s thesis of
Business

October 2022

College of Business Administration
Seoul National University
Marketing Major

Mingi Goo

Confirming the master's thesis written by
Mingi Goo
December 2022

Chair	주 우 진	(Seal)
Vice Chair	김 병 도	(Seal)
Examiner	김 준 범	(Seal)

Abstract

This paper studies the rank effect heterogeneity in the online marketplace and suggests a practical implication for marketing managers to set the optimal digital marketing strategies. Because of the increasing economy of online marketplaces, the position or rank effect is a crucial issue in the marketing literature. The latest literature has focused on the effects of sponsored search results on search engine advertising, though it is known that organic results are more critical than search ads. This research is novel to focus on the effect of organic results in the online marketplace. For analysis on the unit of product level, this paper constructs the rank index through weighted average by keyword search volumes. In the model, the rank effect was specified by the interaction of product-level and category-level averaged variables with the rank index, with the covariates of product-level time-variant variables and two-way fixed effects. Some products were selected randomly to escape the curse of dimensionality. The estimation result suggests that product sales increased in rank and the number of Q&A and reviews. Meanwhile, categories with high price dispersion experienced a lower rank effect, and categories with information asymmetry experienced a lower rank effect. The overall characteristics of the category, such as average price, product attributes, and competition intensity, do not have a significant rank effect. In conclusion, I suggest that marketing managers implement search engine optimization in online marketplaces if their products are in the category with a higher rank effect. This paper finally took a snapshot of the online marketplace by exploiting a vast dataset and extending the marketing literature to the new area. Future research considering hierarchical modeling and endogeneity can investigate more robust and rigorous causality.

Keyword: Rank effect, Online Marketplace, Product category, Digital Marketing, Search Engine Optimization, Marketing Strategy

Student Number: 2021-29714

Table of Contents

Chapter 1. Introduction

Online commerce is rapidly growing. Especially during the COVID-19 pandemic, online commerce has grown fast and massive. In 2021, the two online commerce giants of South Korea, Naver, and Coupang, announced that their sales increased by 35% and 54% compared to 2020 and recorded 1.47 trillion Korean Won and 18.4 billion dollars, respectively.[1]

Consumers can access the assessment, information, and recommendation of others in online marketplaces much more quickly than in traditional marketplaces as references for their purchases. If a product is positioned at the top of a website, a consumer might consider it a good product, so she is more likely to purchase it. In this sense, it is significant for sellers to expose their products at the top of a website. This kind of marketing is already a popular concept in digital marketing; if the product is located at the top position of the search engine results page via paid advertising, it is called search engine marketing (SEM). If a seller or a company tries to win the top position in an organic search result, it is called search engine optimization (SEO).

So far in marketing literature, numerous works have analyzed the effect of SEM on search engines (Chan and Park 2015; Dou et al. 2010; Ghose, Ipeirotis, and Li 2014; Jerath, Ma, and Park 2014). Some

[1] https://www.fnnews.com/news/202203120850523475

of them have also studied rank effects in the online marketplace (Morozov et al. 2021a; Ursu 2018). Although they successfully figured out the heterogeneous position effect, they did not deal with the heterogeneity of the position effect across product categories. This paper generates new knowledge from this point of view.

It is an essential issue for marketing managers to consider a new marketing campaign for their products in an online marketplace. The managers would wonder whether the campaign should aim to improve the brand value or improve its position at the top. To answer this question, this paper estimated the position effect in the most purchased categories and identified how the characteristics of categories affect the position effect. If the product for which a marketing manager wants to implement a campaign has a sensitive position effect, she would be advised to execute the SEO. If not, she would make her brand valuable first in the long-term perspective.

This research took the empirics-first (EF) approach rather than the theory-first approach (Golder et al., 2022). Although the EF approach does not seem rigorous from the theory view, the EF approach is more suitable here because of the nature of this research, which studies highly empirical and managerial questions.

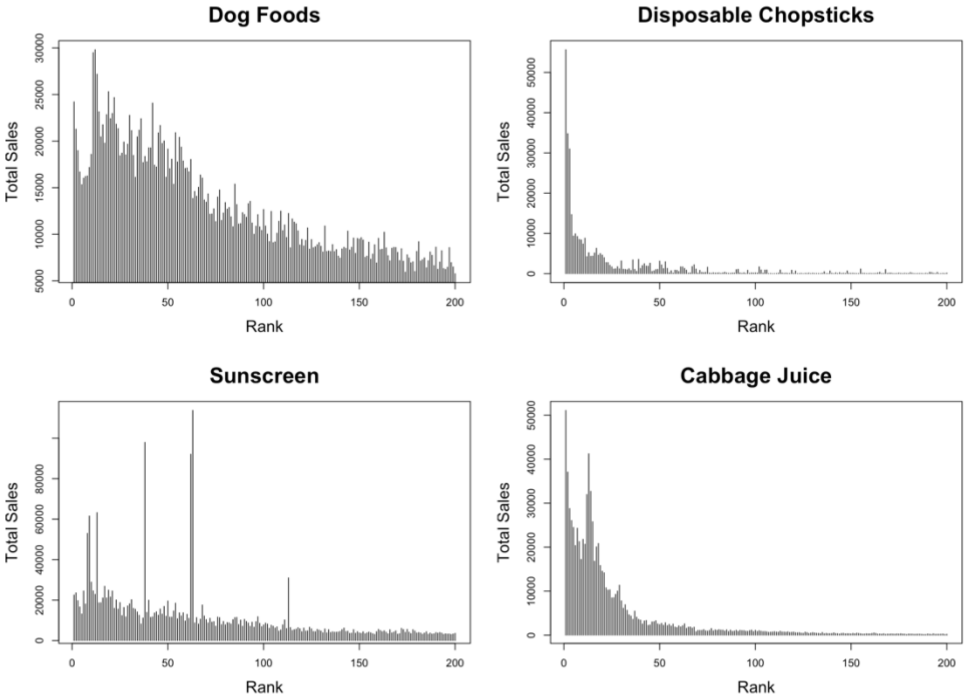

Figure 1. Casual observations of relationships between total sales and the rank of a few keywords.

Figure 1 shows relationships between rank and sales volume for a few keywords. For example, the graph in the upper left of figure 1 stands for the entire sales when the data was collected and summed in the same ranks from 1 to 200 of the keyword of dog foods. As the graphs show, the relationships between total sales and the ranks are remarkably different in the keywords. In the dog foods keyword, the total sales decrease in rank almost linearly. In the disposable chopsticks keyword, the total sales decrease significantly and exponentially in rank, while in the sunscreen keyword, it decreases slowly. In the cabbage juice keyword, it reduces sharper than sunscreen but more slowly compared to disposable chopsticks.

The primary research question of this paper starts from this graph: what is in the relationship between product or product category characteristics and rank effects? Which product category has a more sensitive rank effect? If this question could be answered, the marketing managers of online commerce can make decisions on their marketing strategy, not just from their intuition and experiences but from the data. To this end, this research used all available data in the online marketplace and connected them to the pieces of marketing literature.

The remaining part of this paper is organized as follows. The next section reviews the precedent literature, primarily on the position effect on websites and consumer behavior theories related to categorizing products. Section 3 introduces the data. Section 4 shows the models specifying heterogeneity across the position effect of categories. Section 5 shows the results of the model estimation and interpretation. Section 6 concludes.

Chapter 2. Literature Review

2.1. Position Effect

Numerous pieces of literature in the marketing field have already studied the position effect online from various views. Especially most of them investigate how the product characteristics affect the position effect. The most relevant literature to this research is how effects differ in position by product categories on mobile. Bart, Stephen, and Sarvary (2014) predominantly investigated what product-related conditions of mobile display advertising (MDA) are beneficial in influencing consumer attitudes and purchase intention. When it comes to product-related conditions, scholars adopted utilitarian consumption and hedonic consumption as one axis and involvement as the other axis. They finally concluded that in utilitarian and high-involvement settings, the effect of MDA is relatively significant, so the setting is more worthwhile to execute than the other settings.

Other relevant work to this research investigates the endogeneity of rank effect and the rank effect on click-through rates (CTR). Earlier, Ghose and Yang (2009) already found the rank effect on the search engine in the context of sponsored marketing. Ghose, Ipeirotis, and Li (2014) expanded this topic to the product search engine, an online travel agency (OTA) where consumers search for hotels. Ursu (2018) further investigated the effect of rankings on the OTA platform. There are some implications for these works. One is that ranking itself is an endogenous variable. Those position effect literature constructed models to deal with the endogeneity or implemented randomized experiments to tackle this issue. The other significant problem is that they dealt with click-through rates (CTRs), not with sales. As Ursu (2018) found out, the rank effect does influence CTRs but not sales.

Although many works have already investigated the heterogeneity of the position effect, they have been in the context of sponsored search on search engines. One mainstream is to identify the heterogeneity of rank effects across keywords, as the work of Rutz, Trusov, and Bucklin (2011). They proposed different strategies to bid or select a keyword to advertise for marketing managers in other purposes and situations. Ghose and Yang (2009) and Agarwal, Hosanagar, and Smith (2011) found that a lower position makes fewer click-through rates, but the conversion rates increase. Especially this effect gets more substantial for more specific keywords. Conclusively, they showed that the topmost position is not always profitable. Blake, Nosko, and Tadelis (2015), in sponsored results on eBay, one of the most dominant e-commerce of the United States, found a specific heterogeneity of the position effect across keywords. The brand keywords had no short-term benefit, while the non-brand keywords significantly affected consumers who rarely purchase. However, the frequent purchasers are not influenced by those ads, resulting in the total returns not being profitable.

Another stream is heterogeneity across advertisers. Jerath et al. (2011) investigated the position paradox in sponsored search that the superior firms with lower positions gain more clicks than the inferior ones with higher ranks. Given that the higher position needs more bids to win, the result may imply that the brand power exceeds the position effect in some contexts. Dou et al. (2010) also proposed a brand positioning strategy in information system literature: they found that the unknown brand can get a favorable reputation when its position is higher than the already-renowned brands in their research. Narayan and Kalyanam (2015) found that the position effect is more substantial when the advertiser is small, and consumers have rare experiences with the keyword. The position effect is weaker when the keyword hints at a specific brand or product information. Baye, De Los Santos, and Wildenbeest (2016) and Jeziorski and Moorthy (2018) further found in organic product search and search advertising that the prominence of advertisers could reinforce the position effects, so the prominent advertisers do not necessarily need the top position.

Even if these works successfully studied the position effect, their setting was in sponsored search, and only a few were in organic search. Also, they focused on the search engine and the CTR of the sites. This paper is different from previous research since it has a different context, and it studies product sales in organic search. The reason why organic search is essential is already found. Jerath, Ma, and Park (2014) found that most consumers' clicks are for organic search results rather than sponsored ones, although the sponsored ones are located on the upper side than the organic ones in SERPs (Search Engine Result Pages). Furthermore, they found that consumers who search less popular keywords tend to click more per search. Finally, they connected the consumer’s search behavior with involvement: the involvement of a consumer is correlated inversely to the popularity of a keyword.

The nature of the position effect is highly connected to the search cost and product uncertainty. Since consumers' search starts from the top-ranked items (Granka et al. 2004), searching for lower-ranked items requires more search costs. Earlier marketing literature already showed that the assumption that the search attractiveness decreases as the search cost increases and product uncertainty increases is rational and modeled the concept quantitatively (Kim, Albuquerque, and Bronnenberg 2010). In this context, it is reasonable to speculate that consumers would be likely to search for a higher position due to the search cost increasing as they search for a lower position, and they would be possible to search for a lower position to decrease product uncertainty if the product they are searching has a high level of uncertainty.

2.2. Purchasing Behavior

An early approach in economics to categorize products is whether consumers can determine the product quality before purchasing. If so, the product is called search goods, or otherwise, it is called experience goods. (Nelson 1970, Nelson 1974) Internet, however, significantly increased the capability of consumers to search for a product and has changed their search behaviors. The scholars discovered that consumers who search for experience goods tend to spend more time on a page and be more affected by other consumers' feedback, whereas consumers who search for search goods tend to explore more pages (Huang, Lurie, and Mitra 2009).

Early marketing literature first considered the concepts of involvement and differences between brands (Assael 1987; Kotler and Armstrong 1988), which cause different product purchasing behavior of consumers. Involvement has been considered as an axis of characteristics of products or consumer purchasing behavior. Petty, Cacioppo, and Schumann (1983), Holbrook and Lehmann (1983), and Celsi and Olson (1988) investigated that involvement moderates the effect of advertisement. Consumers pay more attention to an ad in a high-involvement setting than a low one. Bloch (1983) made a connection between involvement and risk. Taylor and Joseph (1984) claimed that high- or medium-priced and durable goods have high involvement, while low-priced and non-durable goods have low involvement. Laurent and Kapferer (1985) broadened the nature of involvement by identifying its profile; it is related not solely to risk but also other constructs, including hedonic value.

Meanwhile, modern consumer behavior marketing literature considers hedonic and utilitarian characteristics or consumptions for categorizing products or consumers' behavior. Dhar and Wertenbroch (2000) investigated the different choice behavior between hedonic and utilitarian goods. Childers et al. (2001) also found that purchase motivation differs in online settings. Li et al. (2020) discovered that consumers employ different search paths in purchase characteristics; for utilitarian purchases, they use social media and product pages while searching for third-party reviews for hedonic purchases.

Chapter 3. Data

3.1. Data Source and Shopping Environment

Storelink provided the data for this paper. Storelink is a South Korean startup company to offer marketing solutions in online commerce. The provided data is collected from Naver Shopping, operated by a leading Korean search engine, Naver.com, similar to Amazon.com in the United States and Alibaba in China. The online service aggregates different seller items by the same categories or keywords. Once a consumer requests a query on the shopping site, it provides corresponding search results, in which sponsored products appear at the top of the list while organic results follow. Figure 2 shows an example of search results after requesting the keyword “sunscreen” in Korean, where the blue box indicates a sponsored product. Storelink collects data on the position of each product in the organic list, from the top to the 200^{th}, depending on each searched keyword, which is selected thanks to its dominance, popularity, or high demand from the marketing consulting clients. The data contains numerous product information, such as prices and product characteristics.

Sellers upload their products’ descriptions on Naver Shopping using a standardized format for different products so that consumers can consistently view essential information specific to each product type. As depicted in Figure 2, product characteristics such as

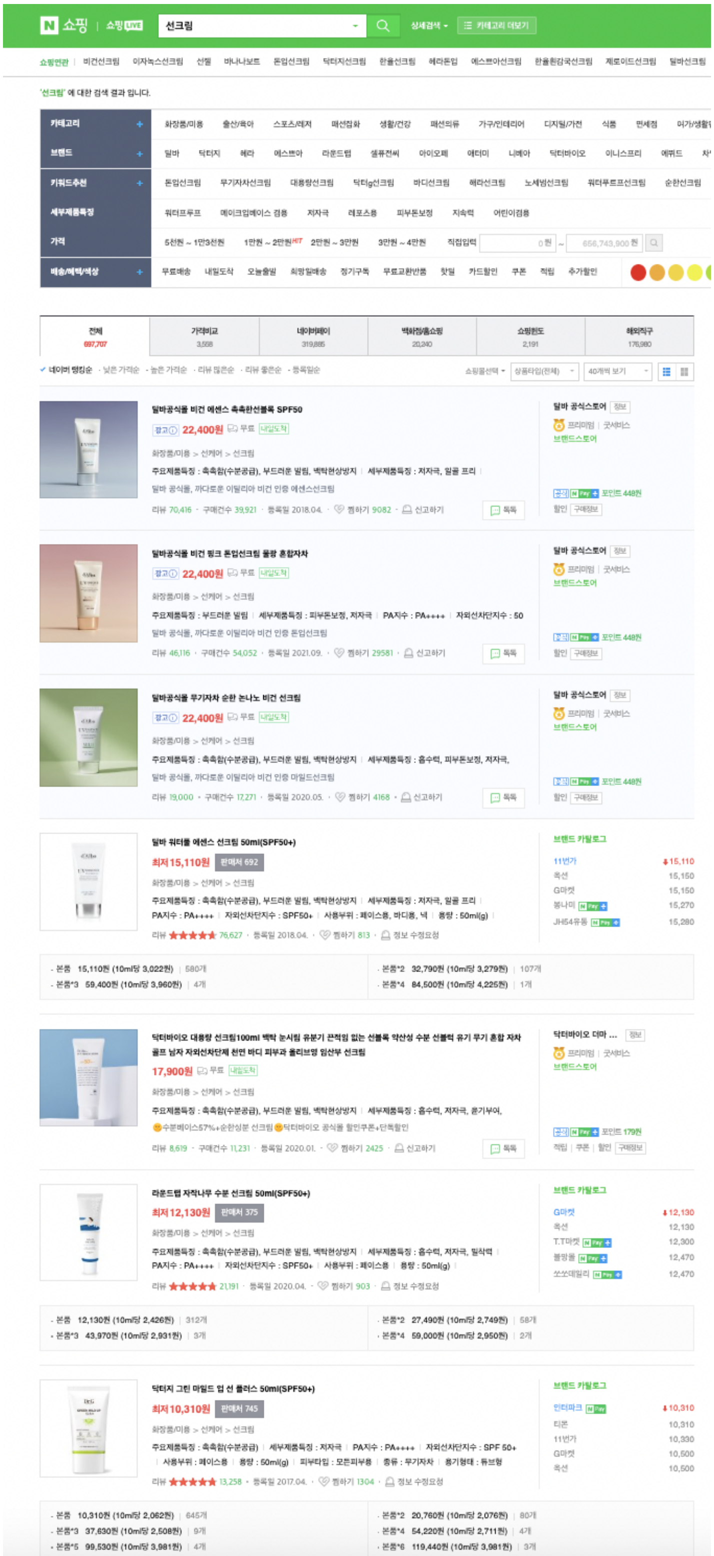

Figure 2. An example of SERPs, which resulted from a query "sunscreen."

moisturizing can be found. This research exploited this information by counting how many product characteristics are provided on the websites, and this variable will be called the number of attributes hereafter. Meanwhile, the shopping portal categorizes products into four depths. For example, a product from Figure 2 can be categorized into the sunscreen category (3rd depth) of the sun-related product category (2nd depth) of cosmetics (1st depth). Nevertheless, most of the products are classified into third depth. This research targets the heterogeneous rank effects across these third-depth categories.

Additionally, the cumulative number of product reviews has also been collected. These are kinds of electronic word of mouth in which consumers who have bought the product leave messages about the product. Although the review data used in this paper do not imply any further information, such as the positiveness or negativeness of the reviews, it still includes the number of reviews for a product cumulatively. The data also includes the number of questions and answers (Q&A). If consumers have questions about the product they are willing to buy, they can directly ask the sellers about the product to resolve the uncertainty about the product. The reviews are notes consumers leave after purchasing, whereas the questions and answers are notes consumers leave before purchasing. In addition, the promotion information is also gathered. This research considered the simple specification: if any price discount promotion is being executed for a product at the time, this will be 1; otherwise, 0. The sales are also collected. It indicates the total sales of the previous three days. Although the data is the total sales, not daily sales, this variable is used directly. Plus, missing values in sales data are recorded as "0," which makes it impossible to distinguish whether the product was not sold or the data was missed. The data with zero-value sales are filtered out to avoid misinterpretation.

Storelink also collects how many keywords are searched. Although this is not an accurate number, they calculated it using their algorithm; Naver.com provides the proportion of the number of keywords searched each day to the maximum number across a given period. Also, the shopping portal provides an exact value of the number of keywords searched in a day. Collecting these two pieces of information, the marketing company conducts reasonable computation about the keyword search volumes. However, it only collects some of the keywords containing some popular or marketing-targeting keywords. Also, the data for unpopular keywords with a volume below zero is not provided if the number is below 10. Therefore, it is assumed that the uncollected keywords are not frequently searched, and their volumes are substituted with 5, the average number of 0 to 10.

3.2. Data Selection and Description

I prepared a sample dataset for the empirical analysis to reduce the computation burden since the data contains observations of more than 200 million. First, the daily data is aggregated into a weekly basis. If an item is observed at least a day in a week, it can be included in the week's observation. After the aggregation, the data have 4.7 million observations of 420 thousand items from the 37^{th} week of 2021 to the 38^{th} week of 2022. Categories in which the average number of products across the observed period is less than 200 are dropped. Since the company collects 200 items on a keyword daily, it is rational that if a category contains pertinent information, one category will include at least 200 items since it can contain different results from different keywords. Namely, a category can cover products more than a keyword. The data of products observed less than three times were dropped for the panel analysis. This process selected 177,491 products from 189 categories, totaling 2,233,692 observations (unbalanced panel).

The rank variable increases when the position is lower. For ease of interpretation, minus rank $\hat{r}_{jkt,w}$ is used.

$$\hat{r}_{jkt,w} = -Rank_{jkt,w} \qquad -(1)$$

where $Rank_{jkt,w}$ stands for the rank value of product j from the category k and keyword w at time t. The rank variable is derived from various keywords. Table 1 shows an example of a part of the data.

Although all the data in table 1 indicates the same product of the same date, they are gathered from different keywords. As a result, a product's data can contain various rank values while others are equal. Direct use of this data might cause bias since duplicated data is used. The rank index r_{jkt} is considered in order to avoid the potential issue, defined by the weighted average of the searched volume of the keyword.

$$r_{jkt} = \frac{\sum_{w}^{W} \widehat{r_{jkt,w}} \times n_{j,w}}{\sum_{w}^{W} n_{j,w}} \qquad -(2)$$

where $n_{j,w}$ stands for the searched volume of the keyword w where the product j is shown.

Since the original data were highly skewed, all the variables, except for the promotion binary variable and the variable indicating time lag from the first product registration on the marketplace, were log-transformed. Table 2 shows the descriptive statistics of the raw data. The value of one is added for some variables when the minimum is zero for log transformation or to keep the values from getting negative.

Three-level variables were considered in this research, apart

Keyword	Product Name	Rank	Date	Price	Keyword Search Volume	...
Dog Summer Clothes	Cute Dog Clothes	1	2022-10-01	7400	12	...
Dog	Cute Dog Clothes	74	2022-10-01	7400	5698	...
Dog Clothes	Cute Dog Clothes	35	2022-10-01	7400	4513	...
...	...	...	...	...	...	...
Dog Hoody	Cute Dog Clothes	5	2022-10-01	7400	48	...
Pet Supplies	Cute Dog Clothes	42	2022-10-01	7400	39	...

Table 1. An example of raw data which shows a data of same product from different keywords.

from the rank index. First, the time-variant product-level variable vector X_{jkt} was used. The variables include (1) the product price, (2) the cumulative number of reviews, (3) the cumulative number of Q&A, (4) the indicator variable indicating whether the promotion is executed, and (5) the lagged time from the first registration of the product on the marketplace (called as time difference hereafter). Product-level average variable vector $\bar{Y}_{jk}$ is also considered, (1) the average price.

of the product across time, (2) the average number of reviews across time, (3) the average number of Q&A across time, and (4) the average time difference across time. That is, for a time-variant variable x_{jkt}, an element of X_{jkt}, gives the time-variant variable $\bar{y}_{jk}$, an element of $\bar{Y}_{jk}$, by $\bar{y}_{jk} = \frac{1}{T_{jk}} \sum_t^{T_{jk}} x_{jkt}$, where T_{jk} refers to the observed number of times of the product j from category k. The average variables $\bar{Y}_{jk}$ contain the time-invariant information of a product by doing so.

The category-level variable vector $\bar{Z}_k$ is considered as well. This vector includes (1) the average characteristics across products in category k, (2) the standard deviation of the prices in the category, (3) the number of brands in the category, (4) the average number of Q&As across products, (5) the average number of the reviews across

	mean	std	min	25%	50%	75%	max
$sales_{jkt}$	56.72	642.18	1.00	2.33	6.71	23.17	128398.00
$price_{jkt}$	30789.40	69225.44	10.00	8000.00	15900.00	29900.00	7421455.0
r_{jkt}	94.55	56.79	1.00	45.00	91.43	143.00	200.00
$promotion_{jkt}$	0.48	0.50	0.00	0.00	0.00	1.00	1.00
$Q\&A_{jkt}$	41.12	298.29	0.00	0.00	0.00	11.00	24616.50
$review_{jkt}$	1931.43	8110.98	0.00	81.00	347.00	1271.00	688306.00
$timediff_{jkt}$	591.17	634.70	0.00	166.00	405.00	786.00	6702.17

Table 2. Descriptive statistics of data.

products, (6) the average number of the attributes of the products in the category, and (7) the standard deviation of the number of the attributes of the products in the category. Please note that these variables are not directly derived from X_{jkt} or $\bar{Y}_{jk}$, which only have information about the selected products of the categories. Instead, $\bar{Z}_k$ includes the information of all products in category k. Note that for any variable x, $\mu_{x,k} = \frac{1}{\sum_{j=1}^{J} T_{jk}} \sum_{j=1}^{J} \sum_{t}^{T_{jk}} x_{jkt}$, namely, the average value across products in a category, and $\sigma_{x,k} = \sqrt{\frac{\sum_{j,k}^{J,k} \sum_{t}^{T_{jk}} \left(x_{jkt} - \mu_{x,kt}\right)^2}{\sum_{j,k}^{J,k} T_j}}$, the standard deviation across products in a category, and $brand_k$ refers to the number of brands being sold in category k.

Chapter 4. Model

4.1. Model Development

The primary model is considered with product- and time-fixed effects. (Wooldridge 2002a).

$$log\, s_{jkt} = \beta_0 + \beta_j + \beta_t + \alpha_{jk}\, log\, r_{jkt} + \gamma X_{jkt} + e_{jkt} \qquad -(3)$$

where β_0 stands for the intercept, β_j stands for the product fixed effect, β_t stands for the time fixed effect, e_{jkt} is an error term following the iid distribution with mean zero. Following the specification of Wooldridge (2002a), each fixed effect contains $J-1$ product-fixed effect for a category k and $T-1$ time-fixed effect for a product j of category k for the identification. Remark that the rank effect α_{jk} is specified with heterogeneity across products. Like Ghose and Yang (2009), α_{jk} is specified with the mean value of products and categories:

$$\begin{aligned} log\, s_{jkt} = \beta_0 + \beta_{jk} + \beta_t + \left(\alpha_0 + \alpha_1 \bar{Y}_{jk} + \alpha_2 \bar{Z}_k\right) log\, r_{jkt} + \gamma X_{jkt} \\ + e_{jkt} \end{aligned} \qquad -(4)$$

in which the coefficients can be estimated with the ordinary least squares. Here, α_0 indicates the main effect of the rank, and α_1 and α_2 stand for the impact of products' characteristics on the rank effect and the effect of categories' characteristics on the rank effect, respectively. By identifying $\{\alpha\} = \{\alpha_0, \alpha_1, \alpha_2\}$, the model describes the market-average effect of the product-level and category-level on the rank effect.

Chapter 5. Results

5.1. Estimation

In this analysis, the large number of products causes the curse of dimensionality: to estimate all the fixed effects, the model has to estimate $m + \sum_{k=1}^{K} J_k + \mathrm{T}$ coefficients, where m means the number of independent variables. In this case, the total dimension is 18+ 177,491+ 52=177,561. Instead, fifty products from 150 categories were randomly selected to reduce further computation burden. Some products have been registered as categories of more than one. It can happen if a seller registers a product in different categories on different days. The final number of the products analyzed is 7,447 products, not 7,500.

		Model (1)	Model (2)	Model (3)	Model (4)
β_0	Intercept	6.3314***	6.6229***	6.5216***	6.619***
$\gamma\ (X_{jkt})$		(0.327)	(0.335)	(0.331)	(0.339)
	$\log price_{jkt}$	-0.3768***	-0.378***	-0.376***	-0.3797***
		(0.018)	(0.018)	(0.018)	(0.018)
	$promotion_{jkt}$	0.0751***	0.0734***	0.0678***	0.0665***
		(0.021)	(0.021)	(0.021)	(0.021)
	$\log Q\&A_{jkt}$	0.0304***	0.0311***	0.0301***	0.0306***
		(0.002)	(0.002)	(0.002)	(0.002)
	$\log review_{jkt}$	0.0332***	0.0316***	0.0339***	0.0321***
		(0.006)	(0.006)	(0.006)	(0.006)
	$timediff_{jkt}$	0.0000	0.0000	-0.0000	-0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
α_0	$\log r_{jkt}$	0.3058***	0.3967***	0.9417***	0.8267***
		(0.005)	(0.048)	(0.109)	(0.111)
$\{\alpha_1\}\ (\bar{Y}_{jk})$	$\mu_{price,jk}$		-0.0158***		0.001
			(0.004)		(0.005)
	$\mu_{Q\&A,jk}$		-0.041***		-0.0367***
			(0.004)		(0.004)
	$\mu_{review,jk}$		0.0274***		0.0269***
			(0.003)		(0.003)
	$\mu_{timediff,jk}$		-0.0001***		-0.0001***
			(0.0000)		(0.0000)
$\{\alpha_2\}\ (\bar{Z}_k)$	$\mu_{price,k}$			-0.0155	-0.008
				(0.016)	(0.018)
	$\sigma_{price,k}$			-0.0261**	-0.0248*
				(0.013)	(0.013)
	$brandcnt_k$			0.004	0.0034
				(0.006)	(0.006)
	$\mu_{Q\&A,k}$			-0.0105***	-0.0089***
				(0.003)	(0.003)
	$\mu_{review,k}$			-0.0158**	-0.0217***
				(0.008)	(0.008)
	$\mu_{attributes,k}$			-0.017	-0.0237
				(0.015)	(0.015)
	$\sigma_{attributes,k}$			-0.0392**	-0.0315*
				(0.016)	(0.016)
	Fixed Effects	Yes	Yes	Yes	Yes
	R^2	0.8390	0.8400	0.8400	0.8400
	$Adjusted R^2$	0.8240	0.8250	0.8240	0.8250
	F-statistics	55.1	55.22	55.17	55.26

Table 3. The estimation results of models.
Notes. Standard errors are in parenthesis.
$*** < .01$; $** \ p < .05$; $* \ p < .1$

5.2. Results

Table 3 shows the estimation results of the primary model, equation 4. Model (4) refers to the primary model, while model (1), model (2), and model (3) stand for the model without $\bar{Y}_{jk}$, $\bar{Z}_k$, and both $\bar{Y}_{jk}$ and $\bar{Z}_k$, respectively. Although all models show similar R^2, adjusted R^2, model (4) presents slightly higher than any other models do. Thus, I will mainly discuss and interpret the result of model (4).

The result shows that the category characteristics $\bar{Z}_k$ covers essential information on the rank effect. The coefficients of X_{jkt} were similar between fixed effect variables, but the main effect α_0 of rank r_{jkt} was underestimated in the models without $\bar{Z}_k$. In model (4) and model (3), α_0 was estimated as .83 and .94, respectively, whereas that of model (1) and (2) was .31 and .40. Additionally, $\mu_{price,jk}$ was insignificant in model (4) ($p>.6$). At the same time, it was significant in model (2) ($p<.01$). This may imply that the nature of $\mu_{price,jk}$ is from $\bar{Z}_k$ than itself.

The estimated γ, the coefficients of X_{jkt}, suggests that the estimates have face validity. The estimated price elasticity γ_{price} is −.38. Note that the coefficient estimated from the log-log model can be directly interpreted as the elasticity. The elasticity is in the reasonable interval analyzed in the meta-analysis (Bijmolt, Van Heerde, and Pieters, 2005). The number of reviews, a kind of word of mouth (WOM), was also positively related to product sales as the literature ($\gamma_{review} = .0269$; Chevalier and Mayzlin, 2006). The number of Q&A also positively affected product sales ($\gamma_{Q\&A} = .0306$).

There were some interesting findings in the category-level estimates. First, a category with significant price variance experienced a weak rank effect (Coefficient of $\sigma_{price,k} = -.0248$). Similarly, a category with a significant variance in the number of attributes experienced a weak rank effect (Coefficient of $\sigma_{attributes,k} = -.0315$). On the other hand, the average price and the average number of attributes do not impact the rank effect (p-value of coefficient of $\mu_{price,k} = .651$, p-value of coefficient of $\mu_{attributes,k} = .116$). This result means that the characteristics influencing the rank effect are not the average level of price and product attributes but the price dispersion and heterogeneity across products. It may be because consumers expect to find a more satisfying product or price by searching to lower ranks if the category has a considerable price variance and heterogeneity across products. One significant issue is that the p-value of the coefficient of $\mu_{attributes,k}$ is marginally higher than the 10% significance level. It implies that the variable is significant if the sample, model, or estimating method is changed. If this is the case, the potential result will imply that overall category characteristics — such as utilitarian-consumed vs. hedonic-consumed, information goods vs. search goods, and high-involvement goods vs. low-involvement goods — can also affect the rank effect.

On top of that, categories with more Q&As and reviews tended to have smaller rank effects (coefficient of $\mu_{Q\&A,k} = -.0089$, coefficient of $\mu_{review,k} = -.0217$). The category where consumers have more significant uncertainty about their purchasing products may have a weaker rank effect since they search for low-ranked products. This interpretation is plausible since the two variables can be interpreted as measuring the information asymmetry in a category. They resolve product uncertainty by providing information as word of mouth (Berger, 2014).

Additionally, the coefficients of an averaged number of reviews have different signs by its level (coefficient of $\mu_{review,jk} = .0269$, coefficient of $\mu_{review,k} = -.0217$). It can mean that categories in which consumers tend to make WOM have a weaker rank effect, but products with more reviews tend to have a more substantial rank effect in a given category. The lagged time was not significant in the level of product (p-value of coefficient of $timediff_{jkt} > .8$), whereas it was significant in the level of average (coefficient of $\mu_{timediff,jk} = -.0001$). Even if the effect seems extremely marginal, since the coefficient is not normalized, it cannot be concluded that it has no impact. It may hint that as the period a product was registered is older, the rank effect becomes weaker. Competition intensity across brands was not a significant predictor for the rank effect (p-value of coefficient of $brandcnt_k > .6$).

Chapter 6. Discussion

6.1. Conclusion and Managerial Implication

In summary, the model estimates suggested that the rank effect was significantly related to product and category characteristics. While the category characteristics may influence the rank, the rank effect decreased in the heterogeneity and price dispersion. In contrast, overall characteristics and average price did not have sufficient evidence to affect the rank. Furthermore, a category with information asymmetry tended to attenuate the rank effect. This result proposes the possibility that different rank effects across product categories exist, and the heterogeneity, price dispersion, and information asymmetry make the rank effect different across product categories.

Marketing managers, who consider the profitability of performance marketing, such as SEM or SEO in online commerce, can be recommended to implement the SEO on the online marketplace if their products belong to a homogeneous, low price-dispersion category with sufficient information provided to consumers online. When it is ambiguous to identify the marketing strategy that is more profitable than other strategies, they can directly calculate the expected profit from the marketing campaign and speculate which one fits their purpose and circumstances more. If they conclude that SEO is not sufficiently profitable, they can instead consider price-discount promotions or establish long-term strategies such as brand marketing.

This research contributed to the marketing literature in three points. Firstly, this study exploited a new and unique dataset and produced a different result. Meanwhile, by taking the empirics-first approach, this research focused on investigating data patterns and connecting them to the existing theories. Second, this research took a snapshot of the online marketplace by studying numerous product categories simultaneously rather than focusing on a category. Although existing studies in the online marketplace have already proceeded with their research, they focused on a category such as a hotel industry (Ursu, 2018). In contrast, this paper analyzes various market categories and analyzes them. In this sense, this paper highlights studies on the general online marketplace.

The most important contribution of this research is to identify the heterogeneous rank effect in the online marketplace. The latest literature spotlighted the search, especially on the sponsored search in search engines. Blake, Nosko, and Tadelis (2015) studied the heterogeneous keyword effect in online commerce (eBay), but the rank effect was not focused on. This paper extended the latest topic of marketing literature to a new area.

6.2. Limitations and Directions for the Future Research

As discussed, this research only employed some of the data because of the curse of dimensionality. Although the sample is randomly selected, it would be more robust research after using all the data, and its generalizability will also be reinforced.

Furthermore, the model in this paper estimated the rank effect by interaction for simplicity. For rigor, the hierarchical method can estimate the same model (Rossi, Allenby, and McCulloch, 2003). If so, the rank effect will be a function of product and category characteristics.

One limitation of this research is that the endogeneity in the rank variable needs to be dealt with appropriately. Recent marketing literature has already found that the rank variable coefficient estimates suffer from endogeneity because the former variables significantly affect the present rank value. This nature of the rank variable triggers the simultaneity bias (Ghose and Yang, 2009). In this sense, the rank variable and the error term may be positively correlated. That is, $cor(r_{jkt}, e_{jkt}) > 0$. Thus, the rank effects in this study might be overestimated than actual values.

The past literature suggested various methodologies to manage the endogeneity for causal inferences. Ghose and Yang (2009) and De los Santos and Koulayev (2017) solved the problem by estimating simultaneous equations. Similarly, Baye, De los Santos, and Wildenbeest (2016) tackled the issue using the ranks of the same products or keywords from the other search engine as the instrumental variable. However, this research exploited the rank index, the weighted average of rank values on keyword search volume, so it is difficult to find appropriate instrumental variables. Instead, some researchers proposed different approaches. Rutz and Trusov (2011) dealt with the endogeneity issue with the latent instrumental variable approach proposed by Ebbes et al. (2005). Meantime, Narayanan and Kalyanam (2015) carried out their research on regression discontinuity design. These two approaches can be appropriate for the rank index approach. Other than that, randomizing experiment by Ursu (2018) can also be one of the possible choices. By doing these directions, future research is expected to provide more rigorous causality in rank effect and purpose empirical solutions for the contemporary marketing issues in the online marketplace.

Bibliography

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, location, location: An analysis of the profitability of position in online advertising markets," *Journal of Marketing Research*, 48 (6).

Bajari, Patrick, Jeremy T. Fox, and Stephen P. Ryan (2008), "Evaluating wireless carrier consolidation using semiparametric demand estimation," *Quantitative Marketing and Economics*, 6 (4).

Bart, Yakov, Andrew T. Stephen, and Miklos Sarvary (2014), "Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions," *Journal of Marketing Research*, 51 (3).

Baye, Michael R., Babur de los Santos, and Matthijs R. Wildenbeest (2016), "What's in a name? Measuring prominence and its impact on organic traffic from search engines," *Information Economics and Policy*, 34.

Berger, Jonah (2014), "Word of mouth and interpersonal communication: A review and directions for future research," *Journal of Consumer Psychology*.

Bijmolt, Tammo H.A., Harald J. van Heerde, and Rik G.M. Pieters (2005), "New empirical generalizations on the determinants of price elasticity," *Journal of Marketing Research*, 42 (2).

Blake, Thomas, Chris Nosko, and Steven Tadelis (2015), "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment," *Econometrica*, 83 (1).

Bloch, Peter H (1981), "An Exploration Into The Scaling Of Consumers' Involvement With A Product Class," *Advances in Consumer Research*, 8 (1).

Celsi, Richard L. and Jerry C. Olson (1988), "The Role of Involvement in Attention and Comprehension Processes," *Journal of Consumer Research*, 15 (2).

Chan, Tat Y. and Young Hoon Park (2015), "Consumer search activities and the value of ad positions in sponsored search advertising," *Marketing Science*, 34 (4).

Chevalier, Judith A. and Dina Mayzlin (2006), "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*.

Childers, Terry L., Christopher L. Carr, Joann Peck, and Stephen Carson (2001), "Hedonic and utilitarian motivations for online retail shopping behavior," *Journal of Retailing*, 77 (4).

Dhar, Ravi and Klaus Wertenbroch (2000), "Consumer choice between hedonic and utilitarian goods," *Journal of Marketing Research*, 37 (1).
Dou, Wenyu, Kai H. Lim, Chenting Su, Nan Zhou, and Nan Cui (2010), "Brand positioning strategy using search engine marketing," *MIS Quarterly: Management Information Systems*, 34 (SPEC. ISSUE 2).
Ghose, Anindya and Sha Yang (2009), "An empirical analysis of search engine advertising: Sponsored search in electronic markets," *Management Science*, 55 (10).
Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li (2014), "Examining the impact of ranking on consumer behavior and search engine revenue," *Management Science*, 60 (7).
Golder, P. N., Dekimpe, M. G., An, J. T., van Heerde, H. J., Kim, D. S., & Alba, J. W. (2022), "Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation," *Journal of Marketing*, 00222429221129200.
Greene, WH William H . (2012a), *Econometric analysis 7th Ed*, *Prentice Hall*.
Holbrook, Morris B. and Donald R. Lehmann (1980), "Form versus content in predicting Starch scores.," *Journal of Advertising Research*.
Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra (2009), "Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods," *Journal of Marketing*, 73 (2).
Jerath, Kinshuk, Liye Ma, and Young Hoon Park (2014), "Consumer click behavior at a search engine: The role of keyword popularity," *Journal of Marketing Research*, 51 (4).
Jerath, Kinshuk, Liye Ma, Young Hoon Park, and Kannan Srinivasan (2011), "A 'position paradox' in sponsored search auctions," *Marketing Science*, 30 (4).
Jeziorski, Przemyslaw and Sridhar Moorthy (2018), "Advertiser prominence effects in search advertising," *Management Science*, 64 (3).
Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online demand under limited consumer search," *Marketing Science*, 29 (6).
Laurent, Gilles and Jean-NOËL Kapferer (1985), "Measuring Consumer Involvement Profiles," *Journal of Marketing Research*, 22 (1).
Li, Jingjing, Ahmed Abbasi, Amar Cheema, and Linda B. Abraham (2020), "Path to Purpose? How Online Customer Journeys Differ

for Hedonic Versus Utilitarian Purchases," *Journal of Marketing*, 84 (4).

Lordan, Oriol (2014), "Study of the full-service and low-cost carriers network configuration," *Journal of Industrial Engineering and Management*.

De los Santos, Babur and Sergei Koulayev (2017), "Optimizing click-through in online rankings with endogenous search refinement," *Marketing Science*, 36 (4).

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta (2004), "Response modeling with nonrandom marketing-mix variables," *Journal of Marketing Research*.

Morozov, Ilya, Stephan Seiler, Xiaojing Dong, and Liwen Hou (2021a), "Estimation of preference heterogeneity in markets with costly search," *Marketing Science*, 40 (5).

Narayanan, Sridhar, and Kirthi Kalyanam (2015), "Position effects in search advertising and their moderators: A regression discontinuity approach," *Marketing Science*, 34 (3).

Nelson, Phillip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2).

Nelson, Phillip (1974), "Advertising as Information," *Journal of Political Economy*, 82 (4).

Petty, Richard E., John T. Cacioppo, and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *Journal of Consumer Research*, 10 (2).

Rossi, Peter E. and Greg M. Allenby (2003a), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3).

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2006a), *Bayesian Statistics and Marketing*, *Bayesian Statistics and Marketing*.

Rutz, Oliver J. and Michael Trusov (2011), "Zooming in on paid search ads-A consumer-level model calibrated on aggregated data," *Marketing Science*, 30 (5).

Rutz, Oliver J., Michael Trusov, and Randolph E. Bucklin (2011), "Modeling indirect effects of paid search advertising: Which keywords lead to more future visits?," *Marketing Science*, 30 (4).

Traylor, Mark B. and W. Benoy Joseph (1984), "Measuring consumer involvement in products: Developing a general scale," *Psychology & Marketing*, 1 (2).

Ursu, Raluca M. (2018), "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 37 (4).

Wooldridge, Jeffrey M (2002), "Econometric Analysis of Cross Section and Panel Data," *MIT Press,* 108 (2).

Abstract

본 논문은 온라인 커머스에서 순위 효과의 이질성을 연구하여 마케팅 관리자가 최적의 디지털 마케팅 전략을 설정하는데 유용한 시사점을 제안하였다. 온라인 커머스는 급격하게 크고 있다는 특성 때문에, 위치 효과 혹은 순위 효과는 마케팅 연구에서 중요한 주제이다. 검색 광고가 아닌 자연 검색 결과 (Organic Result)가 더 중요하다는 것이 보고되었지만, 최근의 연구는 대부분 검색 광고 (Sponsored Search)에 초점을 맞추고 있다. 이 연구는 온라인 커머스에서 유기적 결과의 효과에 초점을 맞추었다는 점에서 새로운 연구이다. 연구에서 상품 수준에서의 분석을 위해 키워드 검색량을 통한 가중평균을 활용하여 새로운 개념의 순위지수를 제안하였다. 모형에서 순위 효과는 제품 수준, 제품군(Category) 수준과 순위지수 간 교호작용으로 식별되었다. 한편, 시간에 따라 변화하는 제품 수준의 변수와 이원고정효과 (Two-way Fixed Effect)가 공변량으로 활용되었다. 차원의 저주 (Curse of Dimensionality) 문제를 해결하기 위해 제품의 일부를 무작위로 선정하여 분석을 진행했다. 모형 추정 결과는 제품 간 가격의 분산과 이질성이 높은 제품군이 낮은 순위 효과를 갖고 있음을 시사한다. 또한, 정보 비대칭이 있는 제품군은 낮은 순위 효과를 갖는 경향을 보였다. 한편, 평균 가격, 제품 속성 및 경쟁 강도와 같은 제품군 내 상품들의 전반적인 특성은 순위 효과에 영향을 미친다는 증거를 발견할 수 없었다. 결론적으로, 마케팅 관리자들은 만약 제품이 높은 수준의 순위 효과를 갖는 제품군에 속한다면 온라인 커머스에서의 검색엔진최적화(Search Engine Optimization)를 통한 마케팅

전략을 고려할 수 있다. 이 논문은 광범위한 데이터셋을 활용한 온라인 시장의 전반적인 묘사를 통해 마케팅 문헌을 새로운 영역으로 확장했다는 점에서 의의를 가진다. 계층적 모델링과 내생성을 고려한다면 향후 연구는 보다 강력하고, 엄격한 인과관계를 규명할 수 있을 것으로 기대된다.