Contemporary Management Research
Pages 121-142, Vol. 14, No. 2, June 2018
doi:10.7903/cmr.18285

The Chinese E-Commerce Search Advertising Business: A Case Study of Taobao

Joseph Richards
California State University Sacramento
E-Mail: richardsj@csus.edu

Min Li
California State University Sacramento
E-Mail: limin@csus.edu

ABSTRACT

Given the rapid growth of the e-commerce market in China, there is a growing need to understand the characteristics of the search advertising strategy employed by dominant e-commerce firms in China. This paper presents a case study to understand the search advertising strategy employed by Taobao based on two sample data sets obtained from the online shopping site taobao.com operated under the Chinese e-commerce giant Alibaba. The first data set contains features about the advertisements displayed after the customer's search using keyword(s). The second data set contains features about the advertisements clicked by customers. These features and the features specific to customers such as their location, time of search, and time of clicking an advertisement in the data sets are studied to reveal insights into the online advertising strategy employed by Taobao in response to customer searches. The results of data analysis show that there is often a relationship among the variables such as the number of advertisements displayed, prices of advertised products, whether a keyword was used in the search, the time elapsed between the start of search and when an advertisement was clicked, the number of keywords used in search, the likelihood of clicking on an advertisement, and price discounts of advertised products. Specific relationships between these variables are formulated as hypothesis and then tested. The insights gained from the test results can be used to develop better business models for sponsored search advertising.

Keywords: Online Search, Sponsored Search, Online Advertising, Taobao

INTRODUCTION

Alibaba and its affiliates, such as the consumer-to-consumer shopping websites Taobao and Tmall, form one of the largest e-commerce firms in the world. China's e-commerce market will soon be bigger than the markets in America, Britain, France, Germany, and Japan combined ("The Alibaba Phenomenon," 2013). The Chinese e-commerce market, with many home-grown cutting-edge innovations, is very different from that of other developed economies. As just a few examples, Alipay, Alibaba's online-payment system, uses an escrow system releasing funds to sellers after buyers receive purchased goods. Alifianance serves as a microlender to consumers and small firms. It is of great value to study the specific features of these innovations and then learn from them through other e-commerce firms like Amazon and eBay as such firms venture into large emerging markets. The search advertising business model carried out by Taobao in response to user search and browsing provides a rich source of data for analysis and important insights. This paper analyzes the search advertising business carried out by Taobao using a sample of its customer search and browsing data.

Understanding the search phase of the consumer choice process has been deemed important for good reason, albeit in a traditional shopping context (Kim, Albuquerque, & Bronnenberg, 2011). Pre-choice consumer search can reveal aspects of consumer consideration sets as well as the organization of the search (Roberts & Lattin, 1997). Much attention has been devoted to understanding the effectiveness of advertising through traditional media such as magazines, newspapers, and TV advertisements. The growing importance of non-traditional media like online advertising platforms requires marketers to understand the effectiveness of online advertising (Laroche, Kiani, Economakis, & Richard, 2013). In the U.S., 91% of Internet users report finding information using a search engine, and it is reasonable to assume that many of these online searches are for finding information about products (Kulkarni, Kannan, & Moe, 2012).

Today China is at the forefront of e-commerce development, which is growing at a rapid pace. There were 731 million Internet users in China by the end of 2016, and 95% of them access the Internet from mobile devices[1]. The number of online shoppers exceeded 466 million in 2016[2]. There were over 400 million active users in 2016 on Taobao.[3] Using Taobao's search engine, consumers obtain product

[1] https://www.techinasia.com/china-731-million-internet-users-end-2016

[2] https://www.statista.com/statistics/277391/number-of-online-buyers-in-china/

[3] http://expandedramblings.com/index.php/taobao-statistics/

information including prices, user reviews, quality, and the seller’s credit evaluation. Information including any click on advertisements by the customer, any search links displayed to the customers, the user query process, and click-stream is also recorded in the search engine (Wei, Geng, Ying, & Shuaipeng, 2014). Though the e-commerce market in China is rapidly developing and evolving, there has been growing evidence that it is very different from other markets in terms of customer behavior and the behavior of sellers. Many global firms do not appear to understand all aspects of drivers of customer satisfaction online in China (Stanworth, Warden, & Shuwei Hsu, 2015). Many well-established companies from the West such as eBay and Google have surprisingly stumbled upon their e-commerce efforts in China. Chinese customers evaluate online service experiences in subtly but substantially different ways from their western counterparts (Ching Yick Tse & Ho, 2009). Therefore, it is important to have a more thorough understanding of the nature of the business model adopted by major Chinese firms. Through a case study, this research offers some important insights into search advertising and how the data might be utilized by Taobao.

Characteristics of Online Search and Advertising: Relevant Literature This paper focuses only on the sponsored search advertising on the Taobao platform. A customer searches a term in a search engine and receives search results and links. Some of these links are sponsored by advertisers and clicking these generates payments to search engines. Therefore, search engines want to maximize the click-through rate (CTR), the ratio of the number of clicks an advertisement received, and the number of times the advertisement was shown. Researchers at Google and Microsoft have produced a lot of research to predict CTR. Richardson, Dominowska, and Ragno (2007) estimate CTR using Microsoft's Bing search engine data and consider common features such as landing page, bid word (keyword), title, body, display URL, clicks, and views. After clicking on an advertisement, a customer is redirected to the landing page. The keyword or bid word is the query corresponding to the displayed ad. The title of the advertisement is shown to the customer. The body includes a description of the advertisement. At the bottom of the advertisement, the customer sees the display URL chosen by the search engine. The clicks value records how many times the advertisement has been clicked. An advertisement displayed to a customer in a page view is called an advertisement impression by Graepel, Candela, Borchert, and Herbrich (2010). They categorize the features as advertisement features, query features, and context features. Advertisement features are bid words or keywords, the title and text of the advertisement, the URL of the landing page, and advertiser-related information. Query features are the search keywords and related expansion, cleaning, and stemming. Context features include time, locations (display and geographic), search history, and other data related to the customer. Ramirez,McMahan et al. (2013) demonstrate the challenges of predicting CTR in a large-scale machine learning problem due to the sparse nature of the data with many missing values and very few clicked advertisements. Ciaramita, Murdock, and Plachouras (2008) investigate the method and features for sponsored search data from Yahoo! and find that it is more likely for viewers to click items at the top of ranked search results. Many of these data features appear in the two sample data sets from Taobao used in this case study. This research presents a case study to understand the search advertising strategy employed by Taobao based on Taobao's customer search and advertising data. Some insights valuable for merchants sponsoring online advertisements as well as for Taobao's competitors are also uncovered. The next section describes the data and its main features. A case study using these data is then presented in detail, including a discussion on the implications of data analysis and potential avenues for further research.

Data and FeaturesThe following information is displayed after a customer searches for a product by entering a keyword on taobao.com: four columns of pictures and texts about the product in the center and two additional columns of product advertisements with searched keyword bid by various merchants on the right and at the bottom. In these two additional columns, the original higher prices of the advertised products are usually crossed out and replaced by the discounted price. The keywords entered by customers are matched to the keywords (called bidwords by Taobao) in the search engine on which merchants have placed bids. The position (right side of the page or at the bottom) and order of advertisements are determined by Taobao's search engine. The advertisement with the highest bid is usually placed at the top. Every time a potential customer clicks the advertisement, Taobao is paid the bid price by the merchant, just as in Google's AdWords. In addition to sponsored search advertising, Taobao also provides other services including display advertising and behavioral targeting. In display advertising, visual items such as banners are displayed on a website to drive traffic to advertisers' own websites. In behavioral targeting, customers' interests are predicted based on their past online behavior and other features (Svensén, Xu, Stern, Hanks, & Bishop, 2011). The two data sets for this case study were sampled from Taobao's paid search ranking (pay-for-performance) system. To make the size of the data files manageable, sampling was restricted to a 5-minute interval. Chinese characters in these two data sets take up several gigabytes of storage space. We describe below the variables captured in the two data sets.

Page View Log file

Table 1 describes some important variables in the page view log file. Information about the displayed advertisements triggered by a customer’s search using keyword(s) is recorded in this file. There are 997,407 observations and 65 variables. Many are ID-type variables identifying cities, provinces, keywords that merchants placed bids on, customer queries, advertising customers, advertisement groups, etc. Table 1 describes variables such as adhighestprice, adprice, adrelativeposition, adscoretag, and adsnumperpage. These variables belong to the advertisement features described by (Graepel et al., 2010). The time and location of the customer when viewing this page are stored in three variables: pvtime, city, and province, belonging to the context features described by (Graepel et al., 2010). The variable acookie tracks advertising customers, and the variable sessionid tracks specific advertisements. The variable rawquery_keyword contains the search keywords and its related expansion. The content of the variable bidword is usually part of what is stored in the variable rawquery_keyword. Both variables belong to the query features in (Graepel et al., 2010). Additional explanation of these variables is given in Table 1.

Every time a customer enters some keyword(s) as identified by the variable rawquery_keyword to search for a product, Taobao’s search engine displays organic results about the searched product in the central area of the page with four columns of pictures and texts. In addition, advertisements sponsored by merchants who have placed bids on a part or whole (identified as bidword) of searched keyword(s) are displayed on the right and at the bottom of the search webpage. For each search session, the variable sessionid assigns one value to a display of sponsored advertisements on the right side of the page and another value to a display of other sponsored advertisements at the bottom. There are often eight advertisements on the right and five advertisements at the bottom of the page. Thus, there are thirteen sponsored advertisements in total on this page, and the value 13 is recorded by the variable adsnumperpage. The eight-five combination is the most common in this data set. So for one search session, one customer, as identified by acookie in Table 1, could correspond to two sessions, as identified by sessionid, and thirteen observations in the data file. There are 26,919 unique values of acookie or customers. Often one customer searched for several products using different keywords and generated thirteen (or other values for combinations other than eight-five) observations each time. In five minutes, these 26,919 customers generated 997,407 observations with 18,824 unique keywords (rawquery_keyword in Table 1). Table 2 lists frequencies of the variable adsnumperpage. Over one-third of such searches triggered 13 advertisements (8 advertisements on the right and 5 advertisements at the bottom).

These advertisements contain common search keywords like high-heeled shoes, and backpacks. Over 18% of these searches had only 3 advertisements displayed by the search engine. These searches involve long or unusual search keywords such as “long-sleeved summer dress for middle-aged women” or “12B37D.” 11% of the searches triggered 16 advertisements. However, 4,005 of these 4,099 searches contain no value for the variable rawquery_keyword. Clearly these customers did not perform a search with a keyword but clicked some links that triggered 16 advertisements. Over 8% of these searches activated 5 advertisements, but over half of these correspond to missing values for rawquery_keyword. The data displayed in Table 2 omitted 220 instances of missing values for the variable adsnumperpage.

Table 1 Description of variables in the Page View Log file

Variable Name	Description of Variable
acookie	A variable identifying the customer on the website.
adhighestprice	The highest price of the advertised product from which the price is marked down.
adprice	The price of the advertised product (the marked down price from adhighestprice).
adrelativepostion	The position of the ad on the page ranging from 1 to 75 with 1 indicating the best position.
adscoretag	The number of times an ad has been clicked.
adsnumperpage	The number of advertisements per page.
bidword	The keyword a merchant placed the highest bid on.
city	The city in which the customer is located.
province	The province in which the customer is located.
pvtime	Time when page view took place; used to calculate time until click.
rawquery_keyword	The keyword a customer searches for on Taobao.com. The terms contained in this variable include the *bidword*. A search for these keywords activates advertisements for display. If the customer clicks on any of these advertisements, Taobao gets paid.
sessionid	The ID of each session of page view. Sponsored advertisements displayed on the right after a search are considered as part of one distinct session ID while sponsored advertisements displayed at the bottom from this search are considered part of another distinct session ID. It can be used to link to the ID in the click log files variable, *pvid*. This variable was generated randomly with 32 alphanumeric characters in length.

Table 2 Frequency of The Number of Advertisements per Page (adsnumperpage), Ordered from The Most Frequent to The Least Frequent

adsnumperpage	Frequency	Percent	Cumulative Frequency	Cumulative Percent
13	13827	37.06	13827	37.06
3	6834	18.32	20661	55.38
16	4099	10.99	24760	66.36
5	3130	8.39	27890	74.75
6	1902	5.10	29792	79.85
24	1771	4.75	31563	84.60
32	1320	3.54	32883	88.13
20	1045	2.80	33928	90.94
27	727	1.95	34655	92.88
1	685	1.84	35340	94.72
12	645	1.73	35985	96.45
7	451	1.21	36436	97.66
4	179	0.48	36615	98.14
15	148	0.40	36763	98.53
8	143	0.38	36906	98.92
40	137	0.37	37043	99.28
10	91	0.24	37134	99.53
30	78	0.21	37212	99.74
2	35	0.09	37247	99.83
50	23	0.06	37270	99.89
60	18	0.05	37288	99.94
9	11	0.03	37299	99.97
11	11	0.03	37310	100.00
13	13827	37.06	13827	37.06

The variable *adprice*, representing the prices of products displayed in the advertisements, ranges from 5 to 999 Chinese Yuan4 with an average of 150.45 and median of 117 Yuan. This is the price discounted from another variable

[4] 1 Chinese Yuan equals around 0.16 US dollars as of March 2018.

adhighestprice, the highest list price or regular price of the advertised product. For almost all advertisements, *adhighestprice* is crossed out and *adprice* is displayed next to it. The higher *adhighestprice* has 164 as the average and 122 Yuan as the median, ranging from 5 to 9,900 Chinese Yuan. The number of times an advertisement has been clicked is recorded in the variable *adscoretag.* It ranges from 0 to 999, with a mean of 373 and median of 311.

Click Log File

The second sample data file is a log of advertisement clicks between 2 pm and 11:41 pm on the same afternoon the page view log file was sampled. It contains 312,660 observations and 90 variables. Table 3 below describes the variables used in this case study.

Table 3. Description of Variables in The Click Log File

Variable Name	Description of Variable
pvid	An identification (ID) variable similar to *sessionid* in the page view log file. It can be used to link to the ID in the click file's variable, *sessionid* (see Table 1). This variable was generated randomly with 32 alphanumeric characters in length. An exact match indicates the customer clicked the ad after viewing the ad on the page.
clicktime	Time when ad was clicked. It can be used to calculate time until click.
clickprice	The price of the clicked product.
clickpriceorigin	The original price of the clicked product.
highestprice	The highest price of the advertised product from which the price is marked down.
fromdomainname	Indicate the part of taobao.com the click originated from. There are a number of sites of taobao.com, e.g., taobao's search engine or tmall (online store fronts at taobao.com).
clickcity	The city in which the clicking customer is located.
clickprovince	The province in which the clicking customer is located.
keyword	The keywords a customer searches for on the site.
customerid	ID assigned to merchandise advertisers
clickcookie	A variable identifying customers who clicked on the advertisements.
pvtime	Time when page view took place; used to calculate time until click.

The variable *pvtime* records the time visiting the webpage, ranging from 2 pm to 2:05 pm. The variable *clicktime* records the time of the click on the advertisment after the page view, ranging from 2:00:01 pm to 11:41:32 pm on the same day. There are 312,660 observations from customers viewing pages and then clicking on advertisements. There are 211,575 unique customers who clicked on the advertisements based on the number of unique values of the variable *clickcookie*. 97,961 advertisers or merchants sponsored advertisements based on the number of unique values of *customerid*. From the page viewing timestamp (*pvtime*) and the timestamp of clicking an advertisement (*clicktime*), half of the customers clicked an advertisement within 51 seconds after viewing the advertisement, and 75% of the customers clicked an advertisement within 169 seconds.

The origin or URL which the advertisement came from is recorded in the variable *fromdomainname*. Over 43% of the clicks originated from s.taobao.com, Taobao's main search engine. A vertical shopping search engine called etao.com also sent over some advertisements. This search engine was once powered by Microsoft's Bing. Another domain worth mentioning is list.taobao.com, a market list by category of products. Other specialized shopping websites run by Taobao occupy the rest of the domains. Close to half (44.31%) of the clicked advertisements are part of 8 & 5 advertisement combination per page (*adsperpage*) described in the page view file. 24 advertisements per page occurred in over 12% of the clicked advertisements. Other values of *adsperpage* are negligible.

The frequencies of the relative position (the variable *relativeposition*) of the clicked advertisements on a page are summarized in Table 4. This variable contains integer values ranging from 1 to 75. These values indicate the relative positions of the advertisement compared to other advertisements on the webpage. The frequencies in Table 4 indicate that the relative position of 1 corresponds to the most desirable position on the webpage, as it received the greatest number of clicks. The number of clicks in Table 4 decreases as the value of *relativeposition* increases, indicating less desirable positions on the webpage. It is in a merchant's interest to have the advertisement placed in a position with a smaller value to increase the likelihood that the advertisement will be clicked.

Table 4 Frequencies of The Relative Positions (*relativeposition*) of The Clicked Advertisements on A Page, Top 16 positions

relativeposition	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	42904	13.72	42904	13.72
2	38422	12.29	81326	26.01
3	34769	11.12	116095	37.13
4	22854	7.31	138949	44.44
5	19331	6.18	158280	50.62
6	17033	5.45	175313	56.07
7	15403	4.93	190716	61
10	15187	4.86	205903	65.86
11	15048	4.81	220951	70.67
9	14413	4.61	235364	75.28
8	13715	4.39	249079	79.66
12	13294	4.25	262373	83.92
13	11620	3.72	273993	87.63
14	4235	1.35	278228	88.99
15	3911	1.25	282139	90.24
16	3663	1.17	285802	91.41

CASE STUDY

In this section, we explore a number of research hypotheses using these two sample data sets. The rationale for these hypotheses and implications of the results from the point of view of Taobao’s strategy are presented and then discussed. The overarching goal of focusing on specific hypothesis is to deduce the strategic thinking undertaken by Taobao in its search advertising strategy as evidenced by the response the user receives from the Taobao search platform.

Hypotheses and Analysis

Substantial price dispersion occurs online (Bodur, Klein, & Arora, 2015). If the advertisements shown in response to user search are for similar products, we expect the prices of products to be similar, and the price variation is expected to be lower compared to the case otherwise. This assumption leads to the following hypothesis.

Hypothesis 1: There is a relationship between the number of product advertisements shown and the distribution of prices of products advertised.

From the mean and the standard deviation of the prices of advertised products shown, we calculate the coefficient of the variation of product prices, which is the ratio of the standard deviation to the mean. The coefficient of variation shows the amount of variation relative to the typical price shown on the advertisements. It is observed that the coefficients of variation of prices of advertised products differ widely as the number of advertised products changes. The coefficient of variation is highest when 60 or 2 advertisements are shown, according to the Tukey multiple comparison procedure. Another relationship of interest is whether the variation changes in a certain direction when the number of advertisements increases or decreases. A higher variation in a larger number of advertisements may signal that Taobao is displaying quite dissimilar products and hence with wide ranging prices. However, this will also be the result when Taobao is assuming in certain search contexts that the search customer may be focused on browsing a variety of product categories rather than focused on a product category. Similarly, if the variation is lower when the number of advertisements is fewer, it could be the result of Taobao showing advertisements of a narrow range of competing products in a similar price range. From a strategy perspective, Taobao is likely judging that the search customer has narrowed down the products of interest, and hence it is best to show advertisements of products within similar quality and hence similar prices. Using regression analysis, we test the relationship between the coefficient of variation of prices and the number of advertisements. The result shows that these two variables are not significantly related.

The above analysis did not show any relationship between the number of product advertisements and the distribution of prices in the advertisements. However, we should not rule out the existence of any moderating variable in this relationship. Price dispersion in a market is typically attributed to imperfect information and consumer search costs (Brynjolfsson, Dick, & Smith, 2010). When Taobao has less information about the consumer or the consumer’s intentions, it implies that this imperfect information situation leads to Taobao showing more varied products with correspondingly different prices. For example, if a customer initiates the search with a search keyword(s), the presence of a keyword can convey significant information about the intention of the customer. Taobao can use this information to alter the mix of product advertisements. In other words, the search keyword(s) can moderate the relationship observed between the number of advertisements and price variation of advertised products. This intuition leads to the next hypothesis.

Hypothesis 2: The coefficient of variation of prices shown in advertisements has a relationship with the number of advertisements when the search data is divided into two groups based on whether a search keyword(s) is present or not.

A two-way Analysis of Variance (ANOVA) analysis with interaction is performed with the factors: the number of advertisements (grouped into two levels, "high" and "low," based on a median split) and whether the search keyword is present or not. The high and low split is based on the median of the number of advertisements, which is 13. The response is the coefficient of variation of advertised prices. Two-way ANOVA shows that the effect of the number of advertisements is significant, F (1,87938) = 1288.3, p < .0001. The dummy yes or no variable for search keyword presence is also significant, F (1,87938) = 86.94, p<.0001. The interaction between the number of advertisements and the dummy variable keyword is also significant: F (1, 87938) = 434.08, p<.0001. The significant interaction demonstrates that the relationship between the number of advertisements and the coefficient of variation is moderated by the search keyword dummy variable. When the search keyword(s) is present, the mean coefficient of variation for the high-level advertisement group is 0.456, and the mean for the low-level advertisement group is 0.420. In the absence of search keyword(s), the mean coefficient of variation for the high-level advertisement group is 0.483, and the mean for the low-level advertisement group is 0.349. Therefore, we can reasonably infer Taobao's strategy of using the presence of search keyword in determining the price range of products displayed and the number of advertisements shown. From Hypothesis 2, we also infer that the number of advertisements and the presence or absence of keyword(s) in a search could be related. When the user takes the effort to initiate a keyword-based search, we can fairly assume that the user is reasonably focused and involved in the search results. Therefore, the presence of the keyword can be used as a signal of user involvement and attention, which could be capitalized by Taobao to improve advertisement effectiveness. The degree of involvement represents the attention and concern given by a person to a specific event. When more attention or concern is given to a specific message, the individual processes the content of the message more in depth (Greenwald & Leavitt, 1984; Paivio, 1986), which ultimately affects advertising effectiveness. Involvement is the prototypical moderator of the relationship between search and choice (Hofacker, Malthouse, & Sultan, 2016). Given these considerations, the number of advertisements displayed, depending whether the customer inputs a search keyword or not, may reveal the type of advertising strategy employed by Taobao. If a keyword is entered, the advertisements shown are likely to be specific to the keyword-related products. Alternatively, when the customer browses the Taobao search engine without entering any keyword, Taobao does not have any cue regarding the intention of customer search, and as a result could be showing advertisements based on other cues such as prior search history, if any, of the customer. These explanations lead us to the next hypothesis.

Hypothesis 3: The number of advertisements shown depend on whether one initiates the search with a keyword.

The distributions of the number of advertisements displayed in searches are compared with and without the keyword(s). The mean number of advertisements with and without the search keyword(s) is calculated and tested for difference in significance. Since the two data groups have unequal variances (using SAS Folded F method: p-value<0.0001), the Satterthwaite method for unequal variance (using Satterthwaite approximation for degrees of freedom) is employed for testing means. It is found that there is significant difference in the means for the number of advertisements between the two groups (p-value<0.0001). Fewer advertisements are displayed when a search is initiated with keyword(s) than without keyword(s).

In the click log data, the search keyword(s) is absent for 61,399 clicks, while the search keyword(s) is present for 251,261 clicks. The amount of time spent by the customer before initiating the click is also known. Consumers often substitute purchase through one channel for another because of variations in search time and effort (Wang, Lin, Tai, & Fan, 2016). Since time is precious for the user, higher involvement and attention to the search task should lead to less time available for other distractions. This would then result in faster progression to clicking an advertisement when the search is done with a keyword. The presence of a keyword in a search implies more involvement and attention from the user, as in the case assumed in Hypothesis 3. The following hypothesis ensues according to this reasoning.

Hypothesis 4: The amount of time spent by the customer before clicking any advertisement, if that happens, depends on whether a search keyword is present or not.

For the two groups based on the presence or absence of keyword(s), a *t*-test is performed to compare the amount of time spent before initiating the click, and the difference is found to be significant. On average, customers who did not search with any keyword(s) spent 263 seconds more before clicking an advertisement than those who searched with specific keyword(s). It appears that customers initiating a search without search keyword(s) spent more time browsing and deciding which advertisements to click. This finding about customer search behavior has potential strategic value for search websites, as they can develop strategies to capture customer "eyeball time" when no keyword is present.

The keywords in the data file are Chinese characters. The number of Chinese characters ranges from one character to twenty-one characters, with an average of 8.11 and a median of 8. The number of Chinese characters in the query can be used to represent the number of search keywords as an approximation. We reasoned in Hypotheses 3 and 4 that the presence of keyword(s) showed the users' involvement and attention to the search task. If this logic is correct, the more keywords used in a search, the higher the degree of attention and involvement by the user. This higher degree of involvement and attention will lead to less time elapsed between the start of a search and when an advertisement is clicked. The reasoning is captured by the following hypothesis.

Hypothesis 5: The amount of time elapsed between the start of search and when an advertisement is clicked depends on the number of keywords searched.

We regress the amount of time elapsed between the start of a search and the time when an advertisement is clicked to the number of Chinese characters. The regression is significant (p-value<0.0001). From the regression coefficient, we observe that, on average, the amount of time elapsed decreased by 10.66 seconds for each additional character in the search query. Price perceptions seem to constitute an important determinant of consumer behavior (Ferreira & Coelho, 2015). Higher prices may convey positive cues about an offer, such as its higher quality (Lichtenstein, Bloch, & Black, 1988). Product price indicates the level of financial uncertainty associated with a transaction, with risk associated with higher-priced products being greater than with lower-priced products (Sweeney, Soutar, & Johnson, 1999). Therefore, the prices displayed on product advertisements would have some effect on the user's response to seeing these advertisements. For instance, for expensive products, consumers would maximize effort and minimize risk by insisting on additional detailed information such as checking available product information (Walia, Srite, & Huddleston, 2016). Furthermore, consumers check on average a small number of attributes per product (Dörnyei, Krystallis, & Chrysochou, 2017), which makes sense for a user who is constrained by time and information processing cost. Since time is precious for the user, it is natural to assume that the user's response time before clicking an advertisement is related to the prices of products shown in advertisements. This rationale leads to the next hypothesis.

Hypothesis 6: The amount of time taken before a customer clicks on an advertisement is related to the prices of products in the displayed advertisements.

Simple regression is performed to relate the amount of time taken before the customer clicked an advertisement to the price of the clicked product (clickprice), and the regression coefficient is significant. The result indicates that the higher the price of the clicked product, the more time is taken before the customer clicked on an advertisement. On average, an increase of one Chinese yuan led to an increased advertisement viewing time of 0.08 seconds. These results show that customers were clearly spending more time reviewing the advertisements of more expensive products.

As noted earlier, use of keywords by the user in a search indicates higher involvement by the user compared to when no keywords are used. As consumers become more involved, they tend to become less price sensitive about the product category of interest, and involved consumers are willing to pay more for brands to which they are loyal (Ramirez & Goldsmith, 2009). This theoretical understanding leads us to the next hypothesis.

Hypothesis 7: The price of products clicked depends on whether any search keyword(s) is present or not.

Using the t-test, we evaluate the effect of missing keywords on the price of the clicked product (clickprice), and the result is significant. On average, the price of the clicked product is 2.44 yuan lower when the keyword is missing than otherwise. It appears that Taobao takes the cue from keyword(s) to show advertisements for pricier products than the case when keyword(s) are not present. The finding is also consistent with the prior theoretical finding that involvement leads users to become willing to pay higher prices.

Since the degree of involvement is also related to the number of keywords, it is natural to ask directly after Hypothesis 7 whether the number of keywords searched is related to the prices of clicked products. The next hypothesis captures this rationale.

Hypothesis 8: The number of keywords used in a search session is related to the prices of clicked products.

Simple regression is performed, relating the price of the clicked product (clickprice) to the number of Chinese characters (keywords). The significant regression result indicates that the more Chinese characters (keywords) are included in an advertisement, the lower the price of the clicked product. On average, the price of the clicked product decreases by 2.19 yuan with each additional Chinese character. This result suggests that fewer searched Chinese characters indicate a greater level of customer certainty in knowing what product or brand she/he is searching for, and the price of the clicked product tends to be higher, assuming well-known brands have higher prices. There is extensive marketing literature on consumer information search and its relation to consumer involvement. Specifically, involvement affects information search behavior in web-based user search situations (Hodkinson & Kiel, 2003). Clicking on advertisements is part of the information search undertaken by the user in purchase decisions. Therefore, it is natural to hypothesize that there is a relationship between whether a search keyword is present (which indicates user involvement) and the likelihood of clicking an advertisement, which is codified below.

Hypothesis 9: The likelihood of clicking an advertisement depends on whether the query keyword is present or not.

For search sessions that resulted in advertisement clicks, the query keyword(s) is not present in around 15% of these sessions. On the other hand, for the sessions without an advertisement click, the query keyword(s) is not present in almost 30% of these sessions. Clearly, missing keyword(s) corresponds to smaller likelihood for an advertisement being clicked by a customer. The t-test is also performed to compare the number of keywords (*rawquery_keyword*) between the sessions with and without clicking on an advertisement and found a significant difference (p-value<0.0001). A natural corollary of Hypothesis 9 is to find out whether the number of keywords involved in a search, which shows the degree of involvement, is also related to the likelihood of clicking an advertisement. This relationship is stated in the next hypothesis.

Hypothesis 10: The likelihood of clicking an advertisement depends on the number of keywords present.

The average number of keywords searched by customers who clicked on an advertisement is 9.45 Chinese characters, while the average number of keywords searched by customers who did not click on an advertisement is 7.69 Chinese characters. If more search keywords indicate a level of greater uncertainty as we assumed in Hypothesis 8, this uncertainty leads to customers browsing advertisements through clicks in greater proportion than when they are more certain. In the rationale leading to Hypothesis 7, we provided literature support for the relationship between involvement and price sensitivity. Involvement is also related to information search, as discussed in Hypotheses 8 and 9. Therefore, the likelihood of clicking on an advertisement, which is associated with information search, and the price discount shown on the advertisement, should show a relationship. This rationale leads to the final hypothesis.

Hypothesis 11: There is a relationship between the likelihood of clicking an advertisement (captured by the variable adscoretag) and the price discount shown on the advertisement.

The discount rate is calculated based on *adprice* and *adhighestprice* in the page view log file data. The variable *adscoretag* captures the popularity of an advertisement reflecting the number of clicks the advertisement received in the past. Regressing *adscoretag* on the discount rate does not show a significant relationship. Thus, advertisements with higher *adscoretag* are not associated with Taobao undertaking a strategy to heavily discount products. Also, it is possible that a discounting strategy does not apparently lead to customers clicking on those advertisements for which such discounting is offered.

DISCUSSIONThe first hypothesis explored whether there is a relationship between the number of advertisements and the coefficient of variation in the prices of products shown in the advertisements. The result showed no significant relationship. However, this relationship could be moderated when a keyword(s) is present in the search query. This relationship is further explored in Hypothesis 2. Once the search keyword(s) as a moderating variable is accounted for, the relationship between the number of advertisements displayed and the coefficient of variation in the prices of products shows a significant relationship. The significant interaction between the number of advertisements and the presence or absence of the keyword variable on the coefficient of variation demonstrates that the presence or absence of a keyword is an important component of the search advertising strategy by Taobao. The coefficient of variation is significantly larger when the search keyword(s) is present, which proves that the presence of the search keyword(s) apparently cues Taobao to display a wider range of prices in advertisements. Research Hypothesis 3 showed that fewer advertisements are displayed when a search is initiated with a keyword(s) than without a keyword(s). What this implies is that with a keyword(s) present, Taobao is taking cues from the keyword(s) and displays a smaller number of focused product advertisements that matches the keyword(s). Without a search keyword(s), Taobao is less sure about what the customer is looking for, and therefore to elicit the customer’s attention, Taobao casts a wider net with more varied product advertisements. Analysis for Hypothesis 4 provided an important insight. Customers who did not search with any keyword (absent query keyword) spent 263 seconds more before clicking on an advertisement than those who searched with a specific keyword. In other words, when no keyword search is involved, the customer, on average, spends more time browsing before clicking on an advertisement, which presents itself as an opportunity for Taobao to monetize this customer time more effectively. It is also likely that some customers who initiate a search with no keyword(s) are just casually browsing the search webpage leading to more time spent before clicking on an advertisement. Such customers present this as a good opportunity for Taobao to capture their “eyeball time.”Research Hypothesis 5 also revealed an important insight. The amount of time elapsed between the start of a search and when an advertisement is clicked decreased when more words (or characters in the Chinese language context) are searched. When more characters (keywords) are included in the search, it is likely that the customer is more uncertain about which product she/he is looking for, and therefore, the customer is eager to get to know one of advertised products displayed. It is also likely that, when a customer uses more search keywords, Taobao is able to produce a selection of “interesting” advertisements based on the cues provided by the keyword(s) and hence elicit faster clicks.

Analysis for Hypothesis 6 showed that search customers waited for a longer amount of time when clicking on advertisements of higher-priced products. This additional wait time implies that the customer is processing additional information when seeing such advertisements, and the customer is likely engaging in complex decision making, because higher prices entail more risk. From Taobao’s strategic point of view, it can optimize the number of advertisements or redesign advertisements to reduce any complexity of information processing. Obviously, other factors such as customer knowledge, involvement, level of trust, and a host of other demographic and psychographic characteristics of the user are additional considerations that affect the wait time before the user clicks on an advertisement.In analysis for Hypothesis 7, we observed that the prices of products clicked were lower when no search keyword is present. Apparently, when customers are searching with specific keywords, they are more seriously looking for a product and are not clicking on an advertisement purely on impulse, and they are also clicking well-known products that are priced higher. It is also the case that Taobao displays higher-priced product advertisements in general when a keyword is present. Hypothesis 8 revealed that, when the number of keywords increases, the prices of clicked advertised products tend to decrease. If a customer is searching with more keywords (Chinese characters), it presumably indicates less certainty of the customer as to what is desired or what is known. In such cases, Taobao casts a wider net by showing more varied products both in kind and quality, resulting in lower prices on average. When the number of keywords (Chinese characters) is fewer, the search results in the user clicking on advertisements with higher product prices, presumably because Taobao is displaying branded or well-known products based on cues from the search keyword(s). This important insight suggests the optimization potential in the decisions involved in displaying advertisements based on the number of keywords searched. Hypotheses 9 and 10 focused on the role of keywords in affecting the likelihood of clicking advertisements. When a customer is searching with keyword(s), either Taobao is taking cues from the keyword(s) and displaying more relevant advertisements, or such customers are more likely to know what they are looking for and can focus on relevant advertisements faster. In Hypothesis 10, we observe that more keywords result in a higher likelihood of clicking on an advertisement. It is apparent that more keywords imply less certainty by the customer regarding what is searched. In such situations, the customer is likely to initiate a click to know more about an advertised product if the customer is seeking to reduce the risk of making a bad choice. It is also possible that, with more keyword(s), Taobao is able to cue more about customer's intentions and able to display more relevant advertisements.Research Hypothesis 11 explored whether there is a significant relationship between a price discount in the advertisement and the popularity of that advertisement. The result showed no significant relationship, implying that Taobao price discounting strategy does not affect the number of clicks of advertisements when these discounts are advertised. In general, one would expect that, when prices are more discounted, the search engine would engage in more advertising to broadcast these discounts, resulting in such advertisements receiving more clicks. This result shows that there is room for Taobao to develop more effective advertising approaches to capitalize on the higher discounts of advertised products.The above research hypotheses presented interesting conclusions about the operational impact and associated strategies of the Taobao's search advertising platform. The analysis presented in this paper is one of the first of its kind to reveal the search-based advertising strategy using actual transactional data from Taobao. To obtain more conclusive findings, future research should focus more on specific hypotheses with more controls added to account for other confounding factors. We hope these analyses will lead to avenues for further research to test the generality of customer behavior in online search advertising contexts. Furthermore, the observed behavior of customers is complicated by the continuing tweaks or strategizing employed by Taobao to optimize customer response. This creates specific challenges for researchers who do not control or moderate the strategic and tactical responses from Taobao. Controlled experiments are needed for more conclusive findings. Overall, as China emerges as the world's largest e-commerce market, the analytical sophistication employed in monetizing customer search in the e-commerce realm will only increase. This research contributes in some measure to our understanding of this important market.

REFERENCES

The Alibaba Phenomenon. (2013, March 23). *The Economist, 406 (5)*.

Bodur, H. O., Klein, N. M., & Arora, N. (2015). Online Price Search: Impact of Price Comparison Sites on Offline Price Evaluations. *Journal of Retailing, 91*(1), 125-139. http://dx.doi.org/10.1016/j.jretai.2014.09.003

Brynjolfsson, E., Dick, A. A., & Smith, M. D. (2010). A nearly perfect market? *Quantitative Marketing and Economics, 8*(1), 1-33. http://dx.doi.org/10.1007/s11129-009-9079-7

Ching Yick Tse, E., & Ho, S.-C. (2009). Service Quality in the Hotel Industry. *Cornell Hospitality Quarterly, 50*(4), 460-474. http://dx.doi.org/10.1177/1938965509338453

Ciaramita, M., Murdock, V., & Plachouras, V. (2008). *Online learning from click data for sponsored search*.

Dörnyei, K. R., Krystallis, A., & Chrysochou, P. (2017). The impact of product assortment size and attribute quantity on information searches. *The Journal of Consumer Marketing, 34*(3), 191-201.

Ferreira, A. G., & Coelho, F. J. (2015). Product involvement, price perceptions, and brand loyalty. *The Journal of Product and Brand Management, 24*(4), 349-364.

Graepel, T., Candela, J. Q., Borchert, T., & Herbrich, R. (2010). *Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine*.

Greenwald, A. G., & Leavitt, C. (1984). Audience involvement in advertising: Four levels. *Journal of Consumer Research, 11*(1), 581-592. http://dx.doi.org/10.1086/208994

Hodkinson, C., & Kiel, G. (2003). Understanding web information search behavior: An exploratory model. *Journal of End User Computing, 15*(4), 27-48.

Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big Data and consumer behavior: imminent opportunities. *The Journal of Consumer Marketing, 33*(2), 89-97.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2011). Mapping Online Consumer Search. *Journal of Marketing Research (JMR), 48*(1), 13-27. http://dx.doi.org/10.1509/jmkr.48.1.13

Kulkarni, G., Kannan, P. K., & Moe, W. (2012). Using online search data to forecast new product sales. *Decision Support Systems, 52*(3), 604-611. doi:10.1016/j.dss.2011.10.017

Laroche, M., Kiani, I., Economakis, N., & Richard, M.-O. (2013). Effects of Multi-Channel Marketing on Consumers' Online Search Behavior: The Power of Multiple Points of Connection. *Journal of Advertising Research, 53*(4), 431-443. http://dx.doi.org/10.2501/JAR-53-4-431-443

Lichtenstein, D. R., Bloch, P. H., & Black, W. C. (1988). Correlates of price acceptability. *Journal of Consumer Research, 15*(2), 243-252.

McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., . . . Golovin, D. (2013). *Ad click prediction: a view from the trenches.* Paper presented at the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).

Paivio, A. (1986). *Mental Representation: A Dual-Coding Approach.* New York: Oxford University Press.

Ramirez, E., & Goldsmith, R. E. (2009). Some Antecedents of Price Sensitivity. *Journal of Marketing Theory and Practice, 17*(3), 199-213.

Richardson, M., Dominowska, E., & Ragno, R. (2007, May 8-12). *Predicting clicks: estimating the click-through rate for new ads.* Paper presented at the 16th international conference on World Wide Web, Banff, Alberta, Canada.

Roberts, J. H., & Lattin, J. M. (1997). Consideration: Review of Research and Prospects for Future Insights. *Journal of Marketing Research (JMR), 34*(3), 406-410.

Stanworth, J. O., Warden, C. A., & Shuwei Hsu, R. (2015). The voice of the Chinese customer. *International Journal of Market Research, 57*(3), 459-481. http://dx.doi.org/10.2501/IJMR-2015-037

Svensén, M., Xu, Q., Stern, D., Hanks, S., & Bishop, C. M. (2011). *Broad vs Narrow: Modelling Strategies for Online Behavioural Targeting.* Paper presented at the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising San Diego.

Sweeney, J. C., Soutar, G. N., & Johnson, L. W. (1999). The role of perceived risk in the quality-value relationship: a study in a retail environment. *Journal of Retailing, 75*(1), 77-105.

Walia, N., Srite, M., & Huddleston, W. (2016). Eyeing the web interface: the influence of price, product, and personal involvement. *Electronic Commerce Research, 16*(3), 297-333. http://dx.doi.org/10.1007/s10660-015-9200-9

Wang, Y.-m., Lin, H.-h., Tai, W.-c., & Fan, Y.-l. (2016). Understanding multi-channel research shoppers: an analysis of Internet and physical channels. *Information Systems and eBusiness Management, 14*(2), 389-413. http://dx.doi.org/10.1007/s10257-015-0288-1

Wei, D., Geng, P., Ying, L., & Shuaipeng, L. (2014, May 31 2014-June 2 2014). *A prediction study on e-commerce sales based on structure time series model and web search data.* Paper presented at the The 26th Chinese Control and Decision Conference (2014 CCDC).