

MATCHING EXACTLY OR SEMANTICALLY? AN EXAMINATION OF THE EFFECTIVENESS OF SYNONYM-BASED MATCHING STRATEGY IN CHINESE PAID SEARCH MARKET

Shuai Yang

Glorious Sun School of Business and Management, Donghua University
1882 West Yan'an Road, 200051 Shanghai, China
shuai.yang@dhu.edu.cn

Yiping Song*

Department of Marketing, Fudan University
670 Guoshun Road, 200433 Shanghai, China
ypsong@fudan.edu.cn

Joseph Pancras

Department of Marketing, University of Connecticut
2100 Hillside Road, Unit 1041, Storrs, CT 06269-1041, US
jpancras@business.uconn.edu

ABSTRACT

Paid search advertising has been a widely used marketing tool in both Chinese and English countries. Matching strategy greatly influences the effectiveness of paid search advertising. Extant studies have examined the matching strategy between keywords and ad content in paid search advertising using the English language. However, the rapidly growing Chinese paid search advertising market has been largely ignored. Different from the English market, the Chinese paid search advertising market has a comparatively greater use of synonyms. Considering the high semantic dependence of words and characters in Chinese, we develop a method to classify Chinese keywords according to the information complexity of the keywords. Based on the keyword classification, we use synonym-based matching, defined as the semantic similarity of ad content and the keyword, to study the bidding behavior of Chinese paid search advertisers. Our results indicate that synonym-based matching increases click-through rate, especially for complex keywords that have multiple search attributes. Both the empirical analysis using secondary data from the Chinese paid search market and a subsequent controlled experiment show the robustness of the results. Our results point to the need for understanding the local characteristics (especially language) when studying online paid search advertising in the Chinese market.

Keywords: Paid Search Advertising; Click-through Rate; Exact Matching; Synonym-based Matching; Chinese Paid Search Market.

1. Introduction

Paid search advertising, as a predominant online marketing technique, has been used widely in both English and Chinese language markets. The greater relevance of the ad content to what consumers search for attracts consumers to click on the ads and make subsequent purchases. Extant literature has introduced ‘exact matching’, where the keyword is included in the ad content. However, few studies have examined other matching strategies, especially in the non-English linguistic environment, such as the Chinese paid search market.

Specifically, the Chinese paid search market, as one of the most attractive emerging markets, is considerably different from the English markets. “Synonym is ... a common phenomenon in Chinese [Li 2007, p. 20].” Chinese is regarded as a “language of compound words” [Arcodia 2007; Chen and Chen 2006], as each Chinese word or character is composed of sub-words or sub-characters, each of which has its own meaning [Williams and Bever 2010]. Most Chinese morphemes have a lexical nature and are termed “bound roots” [Packard 2000, p. 77]. Unlike the English words, Chinese words usually are formed through combining bound roots. Since there is a tight semantic relationship between a Chinese word and its sub-words or sub-characters, the meaning of a Chinese word is determined by the

* Corresponding author

meanings of its parts [Guan et al. 2002]. The strong semantic dependence of the Chinese language drives Chinese consumers to use a large set of semantically similar keywords (where there is often an overlap of characters or sub-words) in searching¹.

Consequently, using synonymous keywords in paid search ads has gained increasing attention from Chinese companies. For example, search engines of the Chinese e-commerce websites (including *Taobao*², the leading e-commerce platform in China, and *Baidu*³, the dominant search engine in China) updated the matching strategies of their searching functions from the exact matching to others (mostly synonyms-based matching). Such upgrading in matching makes the search engines more applicable to users, as actually users have little idea on what should be the exact words to reach their desired information. Thus, in the Chinese paid search market, understanding the use of synonyms in search ads is crucial to improving the performance of paid search advertising (e.g., introducing more traffic to advertisers' stores, enhancing the sales).

In this paper, we take a multi-method approach to answer the following research questions: (1) Do paid search ads with synonymous keywords perform differently from ads with exactly same keywords in the Chinese search market? (2) What is the psychological mechanism of consumers' using synonymous keywords in their information search? And (3) what would be the practical implications of using synonymous keywords on paid search advertising in the Chinese market?

Besides the exact matching and non-matching strategies suggested by Rutz and Trusov [2011], we introduce a third type of matching strategy, *synonym-based matching*, defined as the textual characteristic of the ad whereby it uses semantically similar words to the search information or part of search information contained in the keyword. Consistent with extant studies [i.e., Rutz and Trusov, 2011], we find a positive relationship between matching and click-through rates. Besides, we find that the effect of exact matching on the click-through rate of paid search ad is stronger when consumers search for single-attribute keywords, while the effect of synonym-based matching on the click-through rate is stronger when consumers search for multi-attribute keywords. The results are robust to both an empirical analysis using secondary data from the Chinese paid search market, and a subsequent controlled experiment.

The theoretical contributes of our study are threefold. First, we systematically investigate the effectiveness of using synonyms in paid search advertising, which has been largely ignored by previous literature. Our comprehensive findings on synonym-based matching contribute to the better understanding of consumers' search behavior on paid search advertisements and can help the optimization of matching strategies in the Chinese markets. Second, we introduce an information-based approach to classify search keywords. The classification method is a powerful tool to study paid search advertising in Chinese, as it well considers the information complexity of Chinese words and the widely used synonym-based matching in China markets. Third, we underline the psychological mechanism of different matching strategies by employing a laboratory experiment. The finding from the experimental study is a substantial complement to relevant literature on both consumer decision making and online advertising.

Our study also has important implications. It offers strong supports for using synonyms of keywords and synonym-based matching in China paid search markets. Besides, we clearly list the conditions for achieving greater efficiency in paid search advertising. The suggestions on ad designs for different types of keywords can potentially help the advertisers enhance the advertising performance. Furthermore, the semantic classification method is also a potent tool for designing high-performance paid search ads.

The rest of our paper is structured as follows. Section 2 introduces the relevant literature and the theoretical framework of the study. Section 3 describes data source, models, and results of the empirical study (Study 1). Section 4 is the further experimental study (Study 2) for understanding the underlying mechanism from consumer perspectives. Finally, Section 5 discusses the findings, demonstrates the theoretical contributions, economic implications, and computational implications of our results, and describes the limitations and possible future research.

2. Literature Review

2.1 Performance of Paid Search Advertising

¹ For example, the word '打折' (discount, rebate) has its sub-words '打' (hit, strike, break, smash) and '折' (break, discount, down). '打折' has a tight semantic relationship with its sub-words '打' and '折'. Thus, on one hand, words derived from the same character could be synonyms. For example, words '打折' and '折扣' (discount, rebate) are synonyms, both of which contain the same character '折'. On the other hand, words containing synonymous characters could also be synonyms. For example, words '打折' and '降价' (price reduction) are synonyms, which are derived from synonymous characters '折' and '降', respectively.

² *Taobao* developed a new matching approach to allows sellers to bid for semantically similar keywords with the product descriptions on *Taobao Through Train* on June 9, 2011. Before, *Taobao Through Train* only allowed sellers to bid for keywords that exactly match their product descriptions. See details at <http://www.taobao.li/jiedu/20151203/21197.html>

³ *Baidu* updated its phrase match from only showing the results exactly matching the phrase to also showing results matching the synonyms of the phrase. See details at <http://www.admin5.com/article/20120726/447512.shtml>

In paid search, the advertisers bid price for a click (CPC) on a textual advertisement served after a consumer searches for a keyword [Rutz and Bucklin 2011]. Click-through rate (CTR) is defined as the ratio of the number of clicks to the number of times that the advertisement is shown (i.e., $CTR = \text{clicks} / \text{impressions}$). One impression of an ad may or may not lead to a click, so CTR plays a crucial role in measuring the performance of paid search advertising.

Most theoretical research on paid search advertising focuses on the auctions in paid search advertising. The theoretical foundations for analyzing auctions in paid search advertising are described in Edelman et al. [2007] and Varian [2007], which have studied the generalized second price (GSP) auction mechanisms used by Google and Yahoo!. Even-Dal et al. [2007] show how to integrate characteristics of keywords into the advertising auction. Feng and Xie [2007] study how Cost-Per-Click (CPC) advertising auctions can affect the quality signaling function of paid search advertising. Recent theoretical research has examined such topics as identifying pricing strategies for paid-placements on search engines [Sen et al. 2008], how to distribute advertising ranks for bidding advertisers [Chen et al. 2009], how the numbers of advertisers and ad links influence research price selection [Xiao et al. 2009] how direct-response advertisers and brand advertisers benefit differently from the advertising auction [Zhu and Wilbur 2011], how consumer search strategies affect position auctions [Athey and Ellison 2011], and how first-page bid can increase search engine revenues and advertisers' welfare [Amaldoss et al. 2015].

Most empirical studies on paid search advertising focus on the effects of the rank on advertising performance. These studies indicate that the rank of paid search advertising has negative effects on the click-through rate of the advertising. In other words, the lower (top) rank of the paid search ad, the higher is the click-through rate of the ad. Relevant studies show that the effect of advertisement rank would be affected by the characteristics of keywords [Ghose and Yang 2009; Yang and Ghose 2010; Rutz et al. 2011], advertisers' positioning strategies [Animesh et al. 2011], and firm qualities perceived by consumers [Jerath and Ma 2010].

Click-throughs are one of the fundamental measures of online ad performance. Formulating rules to maximize click-throughs requires novel empirical contributions to the online paid search literature, including a new system of coding (Chinese) keywords based on their information content and complexity, appropriate coding of ad positioning strategies based on the empirical context, and methods to account for semantic or synonym-based matching between ad content and keywords in addition to exact matching.

2.2 Matching Strategies in Paid Search Advertising

The Unique Selling Proposition (USP) of a brand in an ad is critical in differentiating the ad from its rivals and attracting consumers. For example, advertisers use primarily two USPs (i.e., price USP and quality USP) to attract two types of consumers – consumers who search primarily for lower prices and consumers who seek higher quality with higher willingness to pay [Animesh et al. 2011].

The matching strategies between ad USPs and keywords are critical to attract and hold consumers' attention. Rutz and Trusov [2011] suggest two types of matching strategies, exact matching and non-matching strategies. Exact matching is defined as the textual characteristic of the ad where the keyword is included exactly in the ad content, and non-matching refers to the ad characteristic where the ad content does not mention the keyword information. The extent of the match between the keyword and the content of the search ad will increase the relevance of the ad to the information searched for and thus increases click-through. Rutz and Trusov [2011] show the positive effect of exact matching on click-through; for example, if a consumer wants to buy a “discounted” airline ticket, he is more likely to click the advertisements exactly containing “discounted” airline ticket after searching this keyword.

Broad match is a matching process adopted by search engines, under which search engines run a paid search advertisement when consumers search for relevant variations of the keyword, such as synonyms, singular and plural forms, possible misspellings, and phrases containing the keyword. In other words, synonym-based matching is a type of broad matching. According to Singh and Roychowdhury's [2008] study, in the scenario that advertisers have the full information about broad match and have the control of budget splitting, broad match could lead to either an improvement or a loss in advertisers' revenue. The finding leaves a dilemma in search engine marketing. Narayanan and Kalyanam [2015] show that broad match has a lower click-through rate than exact match where a paid search advertisement is placed when consumers search for the keyword. Amaldoss et al. [2016] examine the strategic implications of broad match and find that search engines use broad match to reduce keyword management costs and increase their profits.

In this paper, we argue that the effectiveness of using synonyms differs from using other relevant variations of the keyword, such as singular and plural forms. Many searches in Chinese consist of complex keywords with conjunctive information (multi-attribute keywords) and Chinese language has a large number of semantically similar words. To our best knowledge, academic researchers have ignored the effects of using synonyms of the keywords in paid search advertising. To study the effectiveness of using synonyms in paid search advertising, we propose a new matching concept, synonym-based matching. Table 1 shows the differences between our classification of matching strategies and match options on Google.

Table 1: Google Match Option and Our Matching Strategies for Keywords

Classification	Matching	Definition	Example
Google Match Option	Exact Match	Showing ads to customers who are searching for the exact keyword	Keyword: discounted ticket; Ads may show on searches for: discounted ticket
	Phrase Match	Showing ad to customers who are searching for the exact keyword with additional words before or after	Keyword: discounted ticket Ads may show on searches for: buy discounted ticket
	Broad Match	Showing ads on relevant variations of the keyword, including synonyms, singular and plural forms, possible misspellings, and phrases containing the keyword	Keyword: discounted ticket Ads may show on searches for: discounted airline ticket; cheap ticket; ticket with discounted price
Our classification	Exact Matching	When the ad title contains exactly the keyword	Keyword: discounted ticket Ad title: Shanghai Yongle discounted ticket
	Synonym-based Matching	When the ad title contains synonyms of the keyword rather than the keyword itself	Keyword: discounted ticket Ad title: Shanghai Yongle cheap ticket
	Non-Matching	When the ad title does not contain keyword search information	Keyword: discounted ticket Ad title: Shanghai Yongle ticket

2.3 Consumer Search Behavior and Keyword Classification

In previous studies on English paid search advertising, researchers have used different methods to classify keywords. In Table 2, we summarize the existing keyword classifications, along with the comparison to our keyword classification method.

Table 2: Keyword Characteristics Used in Previous Studies and This Paper

Research	Keyword Characteristics Used in the Study	Summary of Keyword Characteristics		
		Generality	Detailed Attributes	Number of Attributes
This paper	Generic keywords, Single- attribute Keywords and Multi- attribute Keywords	Yes	Yes	Yes
Dou et al. [2001]	Broad, Moderate, and Narrow	Yes	No	Yes
Ghose and Yang [2009]	Brand, Retailer and Length	No	Yes	No
Yang and Ghose [2010]	Brand, Retailer and Length	No	Yes	No
Rutz and Bucklin [2011]	Generic and Branded	Yes	No	No
Jansen et al. [2011]	Brand focused keyphrases and Nonbrand focused keyphrases	No	Yes	No
Rutz et al. [2011]	Auto, Buying, Car, Comparison, Image, Grammar, Information, Make, Model, Sale, Search, Company, Web, Truck, Category, Channel, Condition, Inventory, Feature, Mileage, Price, Financial, Selling, Vehicle, Year, New, Used, Word Count, New By Year, Old By Year	No	Yes	No

Among the extant classification methods, the one used by Rutz et al. [2011] is especially relevant to our study. They extracted the attributes of semantic keywords by using WordNet 2.1⁴. Since languages other than English so far have no comparable and reliable software to analyze, despite some efforts in the computer science literature [Guan et al. 2002], we have to take the route of manual keyword categorization. Based on the level of information searched for by consumers, we classify keywords into three categories:

- *Generic* keywords. Consumers could search for generic information using *generic* keywords (e.g., *airfares*).
- *Single-attribute* keywords. Consumers could search for specific, simple keywords such as price-specific keywords (e.g., *discounted airfares*) or place-specific keywords (e.g., *Beijing airfares*).
- *Multi-attribute* keywords. Consumers could search for complex keywords with conjunctive search information, such as price-and-place-specific keywords (e.g., *Shanghai discounted airfares*).

This approach is similar to Rutz et al. [2011], as we also employ a two-stage approach – first decomposing keywords into attributes, and then coding each keyword with the identified attribute(s). Our extension of coding complex keywords with conjunctive search information makes it possible to deal with potentially huge numbers of

⁴ WordNet 2.1 is a semantic classification software for the English language [Miller 1995].

semantically similar keywords used in Chinese consumers' searches. Furthermore, our keyword coding approach could reflect different types of consumer engagement in online search.

2.4 Consumer Search Behavior and Associative Relevance

In paid search, a consumer starts with a keyword, which reflects his search goal. Previous literature on paid search advertising points out that the process of consumer search usually begins with generic keywords and then becomes specific in subsequent steps [Search Engine Watch 2006]. In the initial generic keyword searches, consumers are in the early stage of decision-making process [Rutz and Bucklin 2011]. At this stage, consumers' preferences are construed at a more abstract level and only have a broad idea on what they want [Lambrecht and Tucker 2013]. At the later stage of the search process, consumers know better about what they want and pursue specific ends at the product feature level [Rutz and Bucklin 2011]. When consumers search for multi-attribute keywords, they have developed narrowly construed preferences, which are usually on a concrete and specific level [Lambrecht and Tucker 2013].

Customized ads using exact matching appear attractive to consumers compared to other presented options, because the customized offers are more likely to be perceived as superior fit to customers' preferences [Huber et al. 1982]. The effect of this fit would be stronger when consumers have less developed preferences [Simonson 2005]. On the contrary, the effect will be weaker when consumers have developed narrowly construed preferences (i.e., searching for multi-attribute keywords), because they know their preferences clearly and can transfer this knowledge easily to other options. In other words, consumers are less loyal when they have narrowly construed preferences than consumers with broad preferences [Hoeffler and Ariely 1999; Simonson 2005].

Associative relevance can well explain the effects of synonym-based matching. As a critical concept in understanding creativity, associative relevance refers to the evolutionary and cohesive notion underpinning human thoughts of analogy [Ahmad 2015]. Creativity represents a chaining process of contexts on the basis of analogy and similarity where the same relations or likeness hold [Haskell 2000; Hofstadter and Sander 2013; Ahmad 2015]. In synonym-based matching, the ad content and the keyword are associatively relevant in terms of intent, which can be utilized by consumers in the process of understanding the ad content. When consumers have developed narrowly construed preferences and have good knowledge about their preferences, they are more likely to be engaged in associative relevance when encountering synonym-based matching. Thus, the effect of associative relevance will be stronger when consumers search for multi-attribute keywords.

Hereby, we summarize our research framework as Figure 1. We take a multi-method approach in explaining the effect of synonym-based matching. First, we conduct a secondary data analysis to examine the effect of synonym-based matching on performance of paid search ads. Second, we resort to consumer behavior theories to explain why synonym-based matching has a positive effect on consumers' click behavior through an experimental examination.

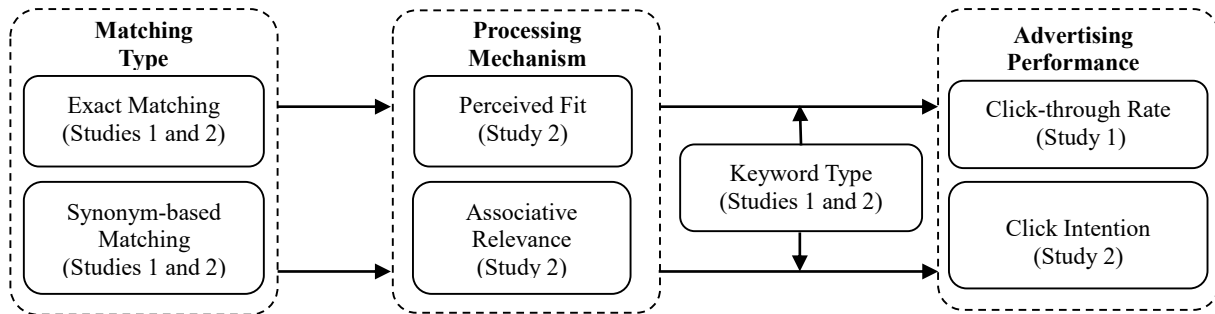


Figure 1: Conceptual Framework

3. Study 1

3.1 Data Source

We collected paid search advertising data from five online airline ticket advertisers in Shanghai. The five advertisers dominate the top seven positions across all keywords in the market, representing the typical SMB dominated search market in China. The process of generating paid search ad data starts from a keyword. Once an advertiser gets a rank for a keyword, one of its paid search advertisements will be displayed on the top left or right part of the search page. If a consumer clicks on the advertisement, s/he will be led to the landing page of the advertiser's website. Our data record consumers' clicks on a paid search advertisement for a given keyword on a daily base. A paid search advertisement for a given keyword consists of a title and a two-line description. The data record both the titles and the descriptions of the advertisements as well as the daily performance of the advertisement (i.e., number of impressions, number of clicks, the average position, and the quality score of the advertisement). The detailed titles and descriptions of search ads allow us to code ad USPs and identify different types of matching between ad content

and keywords. One unique feature of this dataset is that it spans all keyword advertisements of the five airline ticket advertisers for three years. Appendix A1 provides a sample of the raw data.

As described before, we develop a two-stage approach to classify the keywords into information ‘units’ that consumers search for. At the first stage of the approach, we define the classes of information that consumers usually search for. Specifically, we decompose all keywords we collected from online advertisements of airline ticket sellers into attributes and sort these attributes into categories of the basis of “look-alike” characteristics [Lincoln and Guba 1985]. Then, we identify segment groups of these attributes and define these segments as the classes of keyword information. As a result, only two attributes are extracted from the keywords, which are price attribute and place attribute. At the second stage of the approach, we code each keyword with the identified group. In this way, all keywords we have collected are categorized into three types: generic keywords, simple single-attribute (i.e., price-specific or place-specific) keywords, and complex multi-attribute (i.e., price-and-place-specific) keywords.

Table 3: Description of Four Types of Keywords

Keyword Type	Keyword	Total Impressions	Total %imp	Total Clicks	Total %clicks	Average CTR
Generic	Airline tickets	455,683	0.251%	9,617	1.015%	2.110%
	Airfares	88,279,694	48.638%	256,354	27.050%	0.290%
Single-Attribute (Price-specific)	Airfare discount	208,537	0.115%	8,868	0.936%	4.252%
	Discounted airfares	1,505,573	0.830%	49,490	5.222%	3.287%
	Special price airline tickets	12,012,834	6.619%	113,555	11.982%	0.945%
Single-Attribute (Place-specific)	Beijing airfares	716,272	0.395%	10,517	1.110%	1.468%
	Shanghai airline tickets	2,329,897	1.284%	65,335	6.894%	2.804%
	Shanghai airfares	67,972,550	37.450%	271,192	28.616%	0.399%
	Shanghai airlines	5,401,029	2.976%	124,799	13.169%	2.311%
Multi-Attribute	Shanghai discounted airfares	198,232	0.109%	12,514	1.320%	6.313%
	Shanghai special price airfares	2,421,632	1.334%	25,451	2.686%	1.051%
Total		181,501,933	100%	947,692	100%	0.522%

Table 3 summarizes the descriptive statistics of the four type keywords in our data. The statistics show that clicks, conversions, and click-through rates vary markedly across keyword types. Besides, specific keywords (including single-attribute keywords and multi-attribute keywords) have fewer impressions but higher CTRs than generic ones.

3.2 Matching Strategy Identification

Following the previous study [i.e., Rutz and Trusov 2011], we identify the matching strategies according to the relationship between the key USPs in the ad content and the keywords. Besides price and quality USPs, we add several important relevant UPSs that are described in the titles and two-line descriptions of the search ads, such as superior consulting service, speedy booking, convenient payment method, and express delivery.

We measure the matching degree of the ‘bid-pair’ consisting of each keyword-advertisement by comparing its semantic meaning and wordings. A paid search ad would be exact matching (ExactMatch) if it contains the keyword exactly. If it contains only semantically similar words, it would be synonym-based matching (SynonymMatch). If the ad USP does not include any keyword information, the advertisement is non-matching. Consider the example of the keyword ‘discounted airline tickets’ and three ads: Ad 1 ‘Shanghai airline tickets 021-51870253’, Ad 2 ‘discounted airline tickets’, and Ad 3 ‘low-priced airline tickets are here’. On Step 1, this keyword would be classified as a price-oriented search due to the presence of the word ‘discounted’. On Step 2, we identify USPs in the three ads, which are place in Ad 1 (i.e., ‘Shanghai’) and price in both Ad 2 (i.e., ‘discounted’) and Ad 3 (i.e., ‘low-priced’). Hence the matching strategy of Ad 1 could be classified as non-matching on Step 3. On Step 4, we identify “discounted” in the keyword ‘discounted airline ticket’ as the word that describes the search attribute of the keyword. Then, in Step 5, we identify ‘discounted’ in Ad 2 and ‘low-priced’ in Ad 3 as the words that describe the USPs of the two ads respectively. Thus, on Step 6, we classify the matching strategy of Ad 2 as exact matching and that of Ad 3 as synonym-based matching. We depict the process of identifying the matching strategy in Figure 2.

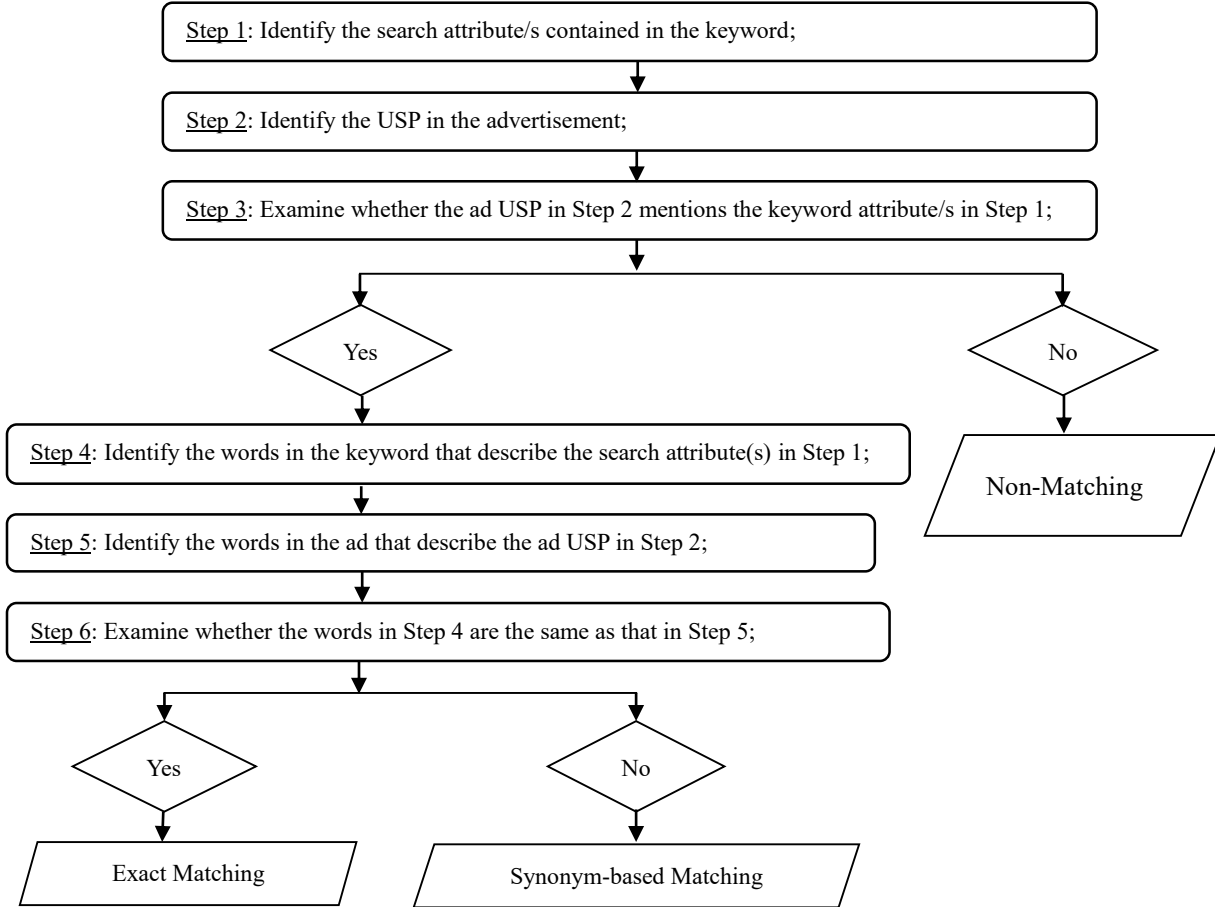


Figure 2: Six Steps to Identify Matching Strategy

3.3 Identification of Other Attributes

We use content analysis and factor analysis to define other attributes in ads that are not captured by the matching measures. First, we use content analysis to extract the semantic attribute(s) in both titles and two-line descriptions. Each word or phrase in the ad title is encoded into an attribute. For example, the title “Shanghai airfares 021-51870253” includes the place and the contact information of the core product. The title “super low-priced airfares for holidays” has price attribute (i.e., “super low-priced”) and occasion attribute (i.e., “holiday”). As the price attributes and the place attributes have been considered in the match between keyword search information, here we exclude these two to avoid redundant repetitions. Ad descriptions are encoded in the same way. Overall, the extracted attributes in ad titles include four types of attributes (i.e., retailer attribute, contact information attribute, “call to action” attribute, holiday attribute), and the attributes in ad descriptions include ten types of attributes (e.g., “popular ticket” attribute, delivery attribute, excellent attribute, payment method attribute). Then, we use factor analysis to unearth the underlying factors, without constraining the number of factors a priori. The results of factor extractions of the titles and the descriptions are shown in Table 4 and Table 5 respectively. Table 6 reports the summary statistics of matching variables and other factors in advertisements.

Table 4: Factor Analysis of Title Attributes of Search Ads

	Title Retailer Factor	Title Holiday Factor
Retailer attribute	.679	-.013
Contact information attribute	.570	-.016
“Call to action” attribute	.854	-.028
Holiday attribute	-.005	1.000

Table 5: Factor Analysis of Description Attributes of Search Ads

	Description Core Product Factor	Description Augmented Product Factor	Description Retailer Factor	Description Payment Method Factor
Destination attribute	.841	-.428	.106	-.198
'Speedy booking' attribute	.834	-.325	-.125	-.098
"Popular ticket" attribute	.795	.427	.314	.187
'Excellent' attribute	.778	.319	-.150	.030
Departure attribute	.761	.324	.265	.026
Delivery attribute	.116	.957	.045	.090
Consulting service attribute	-.128	.948	-.046	.189
Retailer attribute	.029	.040	.999	-.007
Contact information attribute	-.029	-.040	.999	.007
Payment method attribute	-.034	-.060	-.012	.993

Table 6: Summary Statistics of the Paid Search Data (N=34,688)

Variable	Mean	Std. Dev.	Min	Max
Exact Matching	0.406	0.490	0	1
Synonym-based Matching	0.408	0.492	0	1
Title Retailer factor	0	1	-1.259	1.986
Title Holiday factor	0	1	-0.056	0.017
Description Core product factor	0	1	-1.458	1.841
Description Augmented product factor	0	1	-2.158	0.774
Description Retailer factor	0	1	-0.346	5.041
Description Payment method factor	0	1	-0.421	9.397

3.4 Models and Estimation Techniques

The utility of a consumer's choice on clicking on a paid search advertisement follows an Independent and Identically Distributed (i.i.d.) extreme value distribution [Agarwal et al. 2011]. Thus, an Ordinary Least Squares (OLS) approach employed by linear regression models which assume a normal data distribution ($Y \sim N(\mu, \phi)$) is inappropriate for the model estimation of our study [Ghose and Yang 2009; Agarwal et al. 2011]. A generalized linear mixed model (GLMM) allows the response variable following different distributions, including the i.i.d. extreme value distribution. Previous studies on paid search advertising widely use GLMM in model estimation [e.g., Yang et al. 2016]. We therefore employ GLMM to capture the click-through probability of a paid search advertisement for a search keyword in this study.

Since paid search advertisements (level 1) are nested in keywords (level 2), we cast our generalized linear mixed model with both fixed effects and random effects. Pseudo-likelihood techniques are used for parameter estimation. At the first level, a logit model is used to capture the click-through probability of an advertisement j for a keyword i at time t as follows:

$$Pr_{ijt}^{CTR} = \frac{\exp(U_{ijt}^{CTR})}{1 + \exp(U_{ijt}^{CTR})} \quad (1),$$

where U_{ijt}^{CTR} is the latent utility of clicking, which depends on both advertisement characteristics and externalities caused by other competing advertisements.

We have controls for the effects of the position of the advertisement, the landing page quality score, keywords, individual advertisers, time and holiday. For an advertisement j under a keyword i at time t , the latent utility can be expressed as follows:

$$\begin{aligned}
 U_{ijt}^{CTR} = & \alpha_{0i} + \alpha_{1i}ExactMatch_{ijt} + \alpha_{2i}SynonymMatch_{ijt} + \beta_1Position_{ijt} + \beta_2QualityScore_{ijt} \\
 & + \sum_{m=1}^2 (\beta_{3m}TitleFactor_j^m) + \sum_{n=1}^4 (\beta_{4n}DescFactor_j^n) + \sum_{f=1}^4 (\beta_{5f}Firm_j^f) + \beta_6Time_t \\
 & + \beta_7Holiday_t + \varepsilon_{ijt}, \\
 \varepsilon_{ijt} \sim & N(0, \sigma^2)
 \end{aligned} \quad (2),$$

where *ExactMatch* and *SynonymMatch* are dummies representing the matching types between keyword search information and advertisement value proposition. If the matching strategy is exact matching, *ExactMatch* is 1; otherwise, *ExactMatch* is 0. If the matching strategy is synonym-based matching, *SynonymMatch* is 1; otherwise, *SynonymMatch* is 0. When non-matching strategy is used, both *ExactMatch* and *SynonymMatch* are 0.

Position is the position (rank) of the advertisement in the search result. QualityScore is the quality score of the landing page, which measures how relevant the ad, keyword, and the advertiser's landing page are to a consumer seeing the ad [Narayanan and Kalyanam 2015].

TitleFactor and DescFactor are factors in the advertisement title and descriptions respectively. We use content analysis and factor analysis to define other attributes in ad titles and ad descriptions that are not captured by the matching measures. The results show that there are two factors in the ad title, including title retailer factor and title holiday factor (please see details in Table 5). Thus, in Equation 2, m in TitleFactor ranges from 1 to 2. The results of factor extractions of the descriptions show that there are four factors in the ad description, including description core product factor, description augmented product factor, description retailer factor, and description payment method factor (please see details in Table 6). Thus, in Equation 2, n in DescFactor ranges from 1 to 4.

Firm is a dummy variable, which represents the fixed effects that control for differences in click-through probability across advertisers. Since there are five advertisers in our data, we use four advertiser dummy variables in Equation 2 and use one advertiser as the reference. Thus, in Equation 2, f in Firm ranges from 1 to 4.

Time and Holiday are time controls; and ε is the error term.

In keywords used by airfare agencies, four types of keywords can be identified in terms of search information. Therefore three dummy variables enter into the second level of the model to explain unobserved heterogeneity with random coefficients on the intercept, exact matching, synonym-based matching and externalities. Thus, let

$$\begin{aligned}\alpha_i &= [\alpha_{0i}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \alpha_{4i}, \alpha_{5i}] \\ \alpha_{ki} &= \gamma_{k0} + \gamma_{k1} PriceKey_i + \gamma_{k2} PlaceKey_i + \gamma_{k3} MultiKey_i + \delta_{ki}, \\ k &= 1, \dots, 5, \delta_{ki} \sim N(0, \sigma_k^2)\end{aligned}\tag{3}$$

where *PriceKey*, *PlaceKey* and *MultiKey* represent single-attribute (price-specific) keywords, single-attribute (place-specific) keywords, multi-attribute keywords, and δ is the error term. We use the SAS macro GLIMMIX to estimate the model.

3.5 Estimation Results

Table 7 reports the results of our generalized linear mixed model with control functions as well as standard errors provided by the bootstrap method. The results show that click-through rates of specific keywords are higher than those of generic keywords. The coefficients of both exact matching and synonym-based matching are positive and significant, indicating that paid search advertisements that contain the keyword exactly, or semantically similar words in the title can significantly increase the click-through rate. Specifically, exact matching corresponds to a greater increase in the click-through rate than synonym-based matching.

The effects of exact matching and synonym-based matching vary across different types of keywords. We find that compared with generic keywords, single-attribute keywords (i.e., price-specific or place-specific keywords) positively interact with exact matching in affecting click-through rate, while multi-attribute keywords (i.e. price-and-place-specific keywords) have a negative interaction with exact matching. In contrast, single-attribute keywords have a *negative* interaction with synonym-based matching while complex keywords have a positive interaction with synonym-based matching. These results suggest that for single-attribute keywords, advertisements that contain the exact keyword can result in even higher click-through rates; but for multi-attribute keywords, advertisements using synonym-based matching can have higher click-through rates.

Table 7: Generalized Linear Mixed Model Results with Control Functions

	Intercept for Generic Keywords	Single-Attribute Keywords (Price- specific)	Single-Attribute Keywords (Place- specific)	Multi-Attribute Keywords
Constant	4.645*** (0.930)	0.152** (0.039)	-1.177*** (0.016)	8.715*** (0.073)
ExactMatch	0.880*** (0.001)	0.171*** (0.007)	0.806*** (0.013)	-0.636*** (0.002)
SynonymMatch	0.672*** (0.005)	-0.599*** (0.004)	-0.308*** (0.003)	0.305*** (0.001)
Position	-1.789*** (0.158)			
QualityScore	0.353*** (0.004)			
TitleRetailer	-0.222*** (0.008)			
TitleHoliday	0.015*** (0.0002)			
DescCore	0.144*** (0.031)			
DescAugmented	1.003*** (0.002)			
DescRetailer	0.050*** (0.001)			
DescPayment	0.222*** (0.004)			
Fixed Effects of Advertisers	Yes			
Time	-0.018*** (0.001)			
Holiday	0.190*** (0.001)			
Residual from Position	1.818*** (0.158)			
Residual from QualityScore	0.026*** (0.004)			
-2 Res Log Pseudo LL		644770.5		
Pseudo-BIC		645215.5		

Notes: ^ap<0.1; *p<0.05; **p<0.01; ***p<0.001

From a managerial perspective, we calculate Table 8 by using non-matching as the baseline. For all the four types of keywords, both exact matching and synonym-based matching can improve advertising performance, in terms of CTRs and revenues. Assuming that the average conversion rate is 5.9% as in Rutz and Trusov [2011] and the average price of airline tickets is 500 CYN, we find that both exact matching and synonym-based matching can considerably increase five firms' total revenues. Synonym-based matching is especially effective for multi-attribute keywords.

Table 8: Revenue Changes for Different Types of Keywords Using Different Matching Strategies (in CYN)

	Generic Keywords	Single-Attribute Keywords (Price-specific)	Single-Attribute Keywords (Place-specific)	Multi-Attribute Keywords
Exact Matching	1,097,947 (22.35%)	2,117,874 (30.73%)	3,496,399 (57.90%)	988,007 (16.75%)
Synonym-based Matching	541,685 (11.03%)	475,841 (6.90%)	476,049 (7.88%)	1,315,124 (22.29%)

Note: The advertising performance of non-matching is treated as the reference.

Previous literature [i.e., Rutz and Trusov 2011] shows a positive relationship between exact matching and click-through rate. Our results are consistent with their conclusion -- exact matching could improve click-through rate significantly for all four types of keywords. Besides, we find that, for multi-attribute keywords, synonym-based matching is more effective than exact matching. This is interesting since exact matching is always more effective than

synonym-based matching for the other keywords. Therefore, advertisers should use synonym-based matching for multi-attribute keywords and use exact matching for other types of keywords.

This interesting finding on the effectiveness of synonym-based matching raises one question: should advertisers use as many popular words as possible to boost the chances of exact matching and synonym-based matching? To explore possible answers to this question, we select ads that use five USPs or more, a strategy that is facilitated by many popular USPs to attract consumers (e.g., ‘We provide low-priced Shanghai airline tickets with express delivery, superior services, and convenient payment!’). By calculating the average CTR of ads with five USPs or more across four types of keywords, we get 0.258% for multi-attribute keywords, 0.165% for price-specific keywords, 0.126% for place-specific keywords, and 0.006% for generic keywords. All the four CTRs are much smaller than the average CTR of our overall dataset (i.e., 0.522%). This is an evidence showing that using as many USPs as possible in ads appears to be an ineffective strategy.

We also check the effect of synonym repetition, where advertisers use two or more distinct synonyms in an ad title. Using synonym repetition is more likely to induce consumers to take an action and click on the ad. But the action incidence is different from “call to action” words (e.g. ‘call us today’, ‘book airline tickets from us’, ‘buy the ticket now’), which could increase click-through rate [Geddes 2010]. Synonym repetition could connect with a consumer cerebrally by delivering multiple straightforward messages to answer the consumer’s question. Moreover, we find that the number of repetition times influences the CTR in a reversed U-shape pattern. In our data, ads using synonym repetition twice or three times have a high CTR (8.166%), but ads using synonym repetition four times or more produce a very low CTR (0.345%). This finding could be very suggestive to Chinese search ad design.

The ineffectiveness of heavy use of USPs or synonym repetition in ads could be explained by information overload. Consumers would face an ‘information-overload’ problem when paid search advertisements carry on too much information (i.e., many USPs or many synonyms). Given that consumers have limited capabilities to process information [Miller 1956], much information could lead to choice deferrals [Dhar 1997; Iyengar and Lepper 2000] and low effectiveness in product evaluations [Messner and Wanke 2011].

Table 7 also reports the effects of factors in ad titles and two-line descriptions. Holiday information in ad titles has a significant and positive effect on CTR, suggesting that consumers are more likely to click on the ads that indicate holiday information. However, the retailer factor extracted from title attributes has a significantly negative effect on CTR, suggesting that retailer information (e.g., retailer name and contact information) in ad titles would decrease CTR. The four factors extracted from description attributes (i.e., core product factor, augmented product factor, retailer factor, and payment method factor) can increase CTR significantly. It is notable that influence of description augmented product factor is much larger than the other description factors, suggesting that providing valued-added services is effective in increasing CTR on the Chinese paid search market.

Some additional results are reported in Table 7. As expected, we replicate the finding of Animesh et al. [2011] that position has a negative and significant impact on CTR. This suggests that the lower position (rank) of an ad in the search list, the higher is the ad’s CTR. The conclusion is consistent with previous empirical studies of paid search advertising [e.g., Ghose and Yang 2009; Agarwal et al. 2011]. Table 7 also shows a positive and significant relationship between quality score and CTR, which indicates that an ad with a higher quality score will attract more consumer attention, thus leading to a higher CTR. The conclusion is also congruent with previous empirical results [e.g., Ghose and Yang 2009].

4. Study 2

Surprisingly, the results of model estimation show that synonym-based matching performs better than exact matching for multi-attribute keywords. Our findings suggest that the effectiveness of exact matching or synonym-based matching depends on the extent to which consumers have developed their preferences at the point where they conduct the online search. To further understand consumers’ search behavior, we conducted a follow-up experimental study to examine the underlying mechanism.

4.1 The Pilot Focus Group

We first conducted a focus group to comprehensively capture consumers’ understanding of matching strategies on paid search advertising. Six participants including three females and three males expressed their views in the focus group session, which lasted for about one hour. In general, the participants expressed that they perceived fit between the information that they were searching for and the presented ads if ads contained the exact keywords. Because of the perceived fit instigated by the exact keyword in the ad, people are more likely to believe that the ad is worthy of clicking. At this point, the factor driving people to click on a paid search advertisement with an exact keyword is clear. However, the factors that make participants click on an ad with a synonym of the keyword are still ambiguous. Most of the participants expressed that the presence of synonyms of the keyword in an ad made them perceive associative relevance, which could raise their interests to click on the ad. If they felt that the synonyms were associated with their

needs, they would use the synonyms to search for more information. Besides, some participants also offered an alternative plausible explanation -- variety-seeking tendency. To further assess the alternative explanations, we conducted a follow-up laboratory experimental survey.

4.2 Experimental Design and Procedure

The objective of the laboratory experiment is three-fold. First, we aim to check our findings from the secondary data analysis. Second, we intend to test whether the effects of consumers' perceived fit and associative relevance indeed influence consumers' reactions toward exact matching and synonym-based matching. Third, we would rule out some alternative explanations on the effect of synonym-based matching.

The experiment was a 2 (keyword type: single-attribute keyword vs. multi-attribute keyword) \times 3 (matching strategy: non-matching vs. exact matching vs. synonym-based matching) between-subject design. Two hundred and fifty-four undergraduate students from a large university in Shanghai participated in the experiment. Students were randomly assigned to one of the six conditions. In the condition of single-attribute keywords, participants were told to imagine that they were going to buy "a laptop with superior performance". The keyword that we used in the experiment was "*superior performance*" ("性能优越" in Chinese). In the condition of multi-attribute keywords, participants were told to imagine that they were going to buy "a laptop with *superior performance and a large-volume battery*", where the keyword that we used was "*superior performance and a large-volume battery*" ("性能优越且电池容量大" in Chinese). Participants were randomly assigned to either of the two conditions and told to search more information on *Baidu*, the largest Chinese search engine. They were asked to indicate a search term that they would use for the *Baidu* search. The indicated search terms were used for manipulation check. We compared the search terms that indicated by the participants with the keywords that we manipulated in each condition. Only two out of two hundred fifty-four participants mentioned different key information in their search terms from our manipulated keywords. We removed these two samples from our analysis.

After giving his/her search term(s), each respondent was randomly presented an ad with one of the three matching strategies. Specifically, in the condition of single-attribute keywords, the ads with non-matching strategy did not include any search information; the ad with exact matching showed the exact manipulated keyword (i.e., "*superior performance*", "性能优越" in Chinese); and the ad with synonym-based matching contained a synonym (i.e., "*high performance*", "性能高的" in Chinese) of the manipulated keyword. In the condition of multi-attribute keywords, we designed two versions of ads with synonym-based matching (i.e., "*high performance and a large-volume battery*", "性能高的且电池容量大" in Chinese, vs. "*high performance and a long-lived battery*", "性能高的且续航时间长" in Chinese) so as to make the ads under synonym-based matching parallel to the ads under exact matching. The two versions were randomly distributed to respondents.

Table 9: Construct Measurements

Construct	Measurement Items	Source
Perceived Fit	1. There is a good fit between the ad and my search needs	Keller and Aaker [1992]
	2. It is logical for me to click the ad	
	3. It is appropriate for me to click the ad	
Associative relevance	1. I found that the search ad appearing on the page were relevant to my task	Tam and Ho [2006] and Jiang et al. [2009]
	2. I could get some indications from the search ad when doing the shopping task	
	3. I could get some associative search terms I can use from the search ad	
	4. I found that he searches ad appearing on the page were relevant to my interest	
	5. The search ad on the page was consistent with my target product	
Click Intention	1. I feel like clicking the ad now	Taylor and Todd [1995]
	2. I would like to click on the ad as soon as possible	
	3. I would like to click on the ad right away	
Variety-seeking tendency	1. I like a job that offers change, variety, and travel, even if it involves some danger	Steenkamp and Baumgartner [1995]
	2. I am continually seeking new ideas and experiences	
	3. I like continually changing activities	
	4. When things get boring, I like to find some new and unfamiliar experience	

Note: All constructs were measured based on a 7-point Likert scale from "Strongly disagree" to "Strongly agree".

For the dependent variables, we adapted three items from Taylor and Todd [1995] to assess respondents' click intention (Cronbach's Alpha = 0.94), corresponding to the behavioral outcome in the real ad data (i.e., ad clicks). For the independent variables, we derived the followings variables based on relevant theories as well as findings from the focus group session: perceived fit, associative relevance, and variety-seeking tendency. Perceived fit (Cronbach's

Alpha = 0.92) is measured by asking the respondents whether they feel fit between the ad and their search needs. We used three items from Keller and Aaker [1992] to measure perceived fit. Associative relevance (Cronbach's Alpha = 0.91) is measured by five items drawn from Tam and Ho [2006] and Jiang et al. [2009]. Variety-seeking tendency (Cronbach's Alpha = 0.90), representing the extent to which respondents like changes, variety, and new experiences, is measured by four items from Steenkamp and Baumgartner [1995]. Table 9 presents these constructs and their measurements.

4.3 Analysis of Findings

ANOVAs were conducted on perceived fit and associative relevance separately. The results reveal significant effects of the type of matching strategies on both perceived fit and associative relevance ($p < .001$). In specific, the ads with exact matching were perceived significantly fitter than the ads with synonym-based matching ($M_{\text{exact}} = 5.15$ vs. $M_{\text{synonym}} = 3.26$; $p < 0.001$); and the ads with synonym-based matching were perceived significantly fitter than the ads with non-matching ($M_{\text{synonym}} = 3.26$ vs. $M_{\text{non}} = 2.19$; $p < 0.001$). The ads with synonym-based matching were perceived more relevant than the ads with exact matching ($M_{\text{synonym}} = 5.01$ vs. $M_{\text{exact}} = 3.84$; $p < 0.001$); and the same as the ads with exact matching to the ads with non-matching ($M_{\text{exact}} = 3.84$ vs. $M_{\text{non}} = 2.41$; $p < 0.001$).

Next, we ran an ANCOVA of click intention with matching type and keyword type as independent variables and variety-seeking tendency as the covariate. Table 10 shows the results. We found a significant interactive effect between matching type and keyword type ($p < 0.001$). The main effects of matching type ($p < 0.001$) and keyword type ($p = 0.002$) were significant. Planned contrasts revealed that when consumers searched for single-attribute keywords, consumers were more likely to click on the ads with exact matching than the ads with either synonym-based matching ($M_{\text{exact}} = 4.60$ vs. $M_{\text{synonym}} = 4.01$; $p < 0.001$) or non-matching ($M_{\text{exact}} = 4.60$ vs. $M_{\text{non}} = 2.24$; $p < 0.001$). When searching for multi-attribute keywords, consumers were more likely to click on the ads with synonym-based matching than those with exact matching ($M_{\text{synonym}} = 5.45$ vs. $M_{\text{exact}} = 4.35$; $p < 0.001$) or those with non-matching ($M_{\text{synonym}} = 5.45$ vs. $M_{\text{non}} = 2.02$; $p < 0.001$). Figure 3 is the corresponding bar graph.

Table 10: ANCOVA Results on Purchase Intention

Sources of Variation	F Value
Matching Type	185.62***
Keyword Type	6.39**
Variety-Seeking Tendency	6.67*
Matching Type \times Keyword Type	20.83***

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

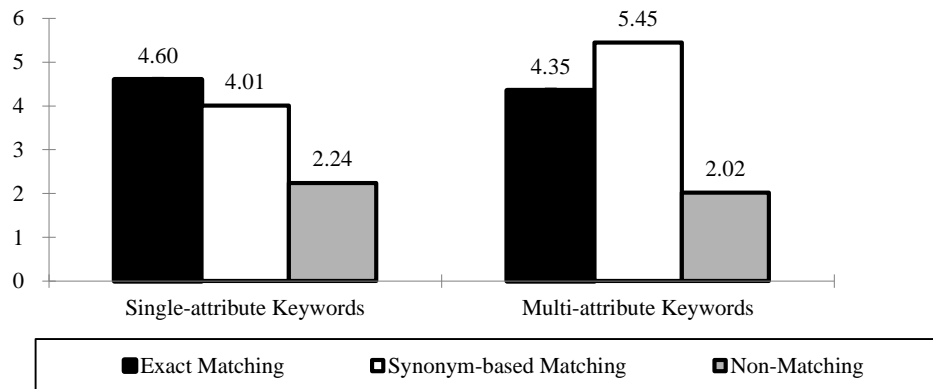


Figure 3: Click Intentions cross Matching Strategies and Keyword Types

Lastly, we examined the mediation roles of perceived fit and associative relevance. This test can help us to better understand the underlying psychological mechanism of customers' responses to different matching strategies. We ran two models to check the mediation roles of perceived fit and associative relevance respectively. In the first model, the independent variable is matching type (exact matching vs. other two matching strategies), and the dependent variable is click intention. Following Zhao, Lynch, and Chen [2010]'s procedure, we computed the bootstrap tests of the indirect effect through by 5,000 samples of 252 data points and a confidence of 95%. Our results show that the indirect effect through perceived fit is significant ($B = 0.55$, $p < 0.001$), but the direct effect from matching type on click

intention is not significant ($B = -0.03, p = 0.88$). The results suggest that perceived fit fully mediates the effect of exact matching on click intention.

In the second model, matching type (synonym-based matching vs. other two matching strategies) still serves as the independent variable. The mediator is associative relevance, and the dependent variable is click intention. Following the same procedure as that used for the first mediation model above, we got the results showing the significant indirect effect through associative relevance ($B = 0.80, p < 0.001$). We also found that the direct effect from the independent variables on the dependent variable was not significant ($B = -0.09, p = 0.62$). The results suggest that associative relevance fully mediates the effect of synonym-based matching on click intention.

5. Conclusion and Discussion

5.1 Conclusion

We study the advertising strategies in the Chinese paid search market, which is characterized by the use of many synonyms. Primarily, our contribution is the development of optimal matching strategies, which can considerably improve advertisers' search advertising performance in the Chinese paid search market. We find that both exact matching and synonym-based matching increase the click-through rate, and their effects vary across the types of keywords. Specifically, exact matching performs better for generic and single-attribute keywords than for multi-attribute keywords. But the effect of synonym-based matching is stronger for multi-attribute keywords than for generic and single-attribute keywords.

Ad design using exact matching could be a formidable task in markets that use a great number of synonyms. The splendid richness of synonyms is not rare in global markets. The Chinese market is typical. Therefore, we suggest that to emphasize single attribute, advertisers should target popular single-attribute keywords and use *exact matching strategy*. If more than one attribute need to be outstanding, advertisers should use *synonym-based matching strategy* and bid for more synonyms of the keywords.

The follow-up experimental study further explains the underlying mechanism of the different effects of matching strategies. The results show that perceived fit fully mediates the effect of exact matching on click intention, and the effect of perceived fit would be stronger when consumers search for single-attribute keywords. The results of the experimental study also show that associative relevance fully mediates the effect of synonym-based matching on click intention, and the effect of associative relevance is stronger when consumers are searching for multi-attribute keywords.

5.2 Theoretical Contributions

Our study contributes to the literature from four aspects. First, we systematically investigate the effectiveness of using matching strategies between keyword information and ad content in paid search advertising. Previously, academic studies [e.g., Rutz and Trosov 2011] have paid most attention to the effects of exact matching, but ignored the effects of using synonyms of the keywords. In this study, we introduce a new matching strategy, synonym-based matching, which provides better and more comprehensive understanding of marketing strategies on the Chinese paid search market. Our study underlines the importance of incorporating local marketing characteristics into frameworks for studying paid search marketing, especially for the markets using language that are very different from English.

Second, we contribute to the literature on semantic classification in paid search advertising by developing a new classification method for the Chinese paid search market. Previous studies on English paid search advertising have classified keywords using different dimensions [e.g., Dou et al. 2001; Ghose and Yang 2009; Rutz et al. 2011; Rutz and Bucklin 2011; Jansen et al. 2011]. But there is no comparable or reliable tool to classify Chinese keywords. Our classification method considers the information complexity of Chinese keywords and the synonym-based matching that is widely used in the Chinese market. This method is a powerful tool to study Chinese paid search advertising, and could also be useful for relevant studies on other paid search markets that are close to the Chinese one, such as many other markets in Asia.

Third, we contribute to the literature on consumer search theories. An increasing body of paid search advertising literature has attempted to understand consumers' searching behavior from psychological perspective. Rutz and Bucklin [2011] show that consumers would search generic keywords when they are at the early stage of decision-making process and then pursue specific ends at the product feature level. Similarly, Lambrecht et al. [2011] show that online advertising data can be used to understand interconnections between different stages of consumers' purchase process. Lambrecht and Tucker [2013] also agree that in the initial generic keyword searches, consumers' preferences are construed at a more abstract and broader level, and then are developed onto a concrete and specific level. Building on prior work, our study further demonstrates that even when consumers have developed a specific preference level, searching multi-attribute keywords differs from searching single-attribute keywords. Using synonym-based matching on multi-attribute keywords can increase consumers' click intention.

Additionally, we reveal the underlying psychological mechanism of consumers' search behavior in Chinese paid search advertising. Though Rutz and Trosov [2011] have depicted the positive effect of using the exact keywords in ad content, they have not explored the psychological mechanism of the effect. Using a controlled laboratory experiment, we find that perceived fit fully mediates the effect of using the exact keywords on click intention, and associative relevance fully mediates the effect of using synonyms of the keywords on click intention. This knowledge is a substantial complement to literature on both consumer decision making and online advertising.

5.3 Practical Implications

Our study also sheds light on several aspects of the practical implications. Using results of the generalized linear mixed model, we estimate the average CTRs, denoted by 'estimated CTR' in Figure 4. Then we compute the average CTRs for these keywords when we apply exact matching strategy for generic and single-attribute keywords and synonym-based matching strategy for multi-attribute keywords. Figure 4 shows that if advertisers adopt our recommendations, the average CTRs would be increased by 0.532%, 2.486%, 1.358%, and 1.626% for generic keywords, price-specific keywords, place-specific keywords, and multi-attribute keywords, respectively. The lift is especially higher for price-specific and multi-attribute keywords since most advertisements for generic and place-specific keywords have been already using exact matching strategy.

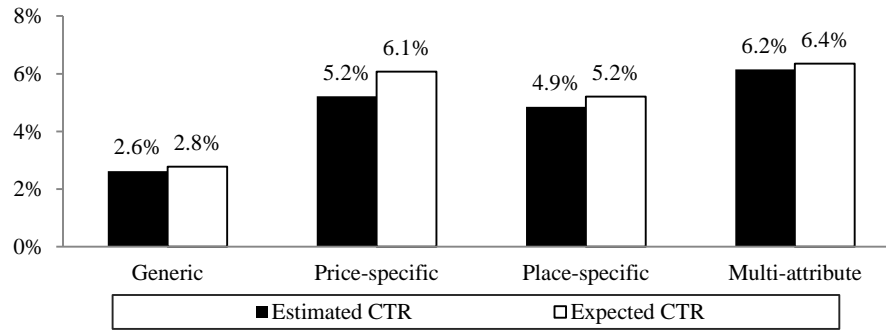


Figure 4: Average Click-through Rates across Keywords

From a managerial perspective, we investigate the effects of exact matching and synonym-based matching on the CTRs of different types of keywords. Figure 5 compares the CTRs of five advertisers' current matching strategies with the CTRs gained from the recommended matching strategies. We find that the CTRs of ads using *non-matching strategy* can be increased by 0.264%, 4.871%, 1.534% and 1.190% for price-specific keywords, place-specific keywords, and multi-attribute keywords, respectively. If advertisers change the matching strategy from *synonym-based matching* to *exact matching* for generic keywords and price-specific keywords, CTRs will be increased by 0.895% and 2.199%, respectively. If advertisers change from *exact-matching* to *synonym-based matching* for multi-attribute keywords, CTR will be increased by 1.975%.

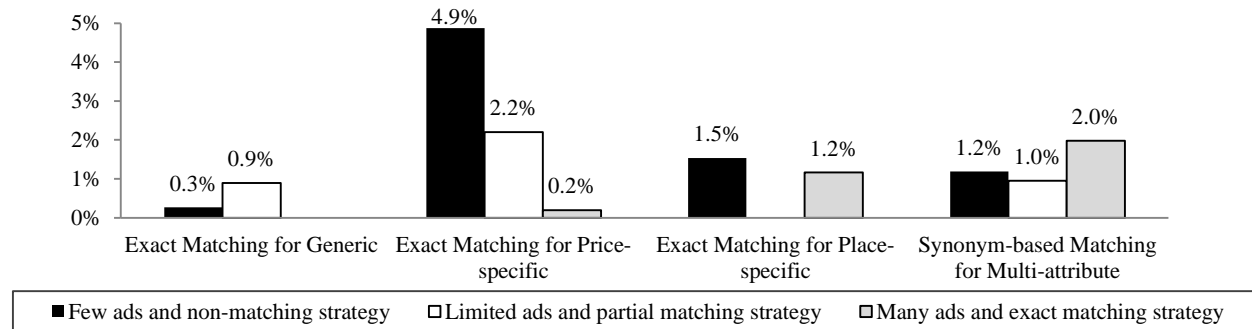


Figure 5: Increased CTRs for three strategies across Keywords

Given the fact that the number of generic keywords is limited, we suggest advertisers use different ads for all generic keywords through exact matching. Since consumers would search for a large number of specific keywords, it is impossible to design customized exact matching ads for all specific keywords, especially when advertisers have very limited advertising budgets. For single-attribute keywords, advertisers should select the popular keywords and

customize ads for each of these keywords, so as to maximize CTR. But for multi-attribute keywords, advertisers can increase CTR by designing a limited number of ads with synonym-based matching strategy.

Additionally, we discuss the practical implications on search engines. In paid search, advertisers bid on keywords based on cost-per-click (CPC, advertisers only pay when their advertisement is clicked) [Zhu and Wilbur 2011]. Search engines such as Google ranks advertisers on a score called Ad Rank, which is a product of CPC and Quality Score assigned by Google: $AdRank = CPC \times QualityScore$ [Zhu and Wilbur 2011]. The positions of paid search ads are in descending order of Ad Ranks. Google does not reveal the exact procedure of assigning Quality Score. It only tells that Quality Score is a function of (1) the relevance of the ad and the keyword, (2) the ad's past CTRs, (3) the quality of the advertiser's landing page, and (4) other relevance factors: $QualityScore = f(relevance, CTR, quality, other)$.

To provide users with relevant paid search ads to their queries, Google uses relevance of the ads and keywords as one important factor in calculating Quality Score. The more relevant the keywords are, the higher the Quality Score of the ad. Google considers ads using synonyms as less targeted than ads using exact matching. As a result, paid search ads using synonyms will have lower Quality Scores because of their lower level of relevance.

Our results indicate that exact matching is not always the most effective in increasing CTR. In specific, synonym-based matching is more effective than exact matching for complex (multi-attribute) keywords. Therefore, the calculation of the relevance of the ads and keywords should depend on the keyword type rather than on the matching type. For multi-attribute keywords, paid search ads that use synonyms of the keyword might have a higher level of relevance than or at least the same level of relevance as those ads that use exact keywords. This indicates that paid search ads using keyword synonyms should be granted a higher Quality Score.

5.4 Limitations and Future Research

Our paper also has a number of limitations that could be promising directions for future research. Using our analysis framework to examine the paid search advertising in English markets could help to check whether synonym-based matching has the same importance. It is possible that synonym-based matching is more relevant to a high-context culture such as China [Hall 1976], where people pay more attention to the contextual information, and where they are more likely to be influenced by other people's opinions [Teng and Laroche 2006], and where the logographic language enables accurate visual judgments [Tavassoli and Lee 2003]. These factors may lead to more searches and clicks on ads using semantically related words by Chinese consumers. Future studies can explore the impact of these factors on the effectiveness of synonym-based matching so as to offer further and deeper understanding on the paid search advertising in the high-context culture.

Acknowledgement

We acknowledge the financial supports from the National Science Foundation in China (No. 71172036, 71531006, and 71602026), China Scholarship Council (No. 201506105044), and the Fundamental Research Funds for the Central Universities in China (No. 15D210808 and 16D110811).

REFERENCES

- Agarwal, A., K. Hosanagar, and M. D. Smith, "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Consumer Research*, Vol. 48, No. 6: 1057-1073, 2011.
- Ahmad, G., "Instructional Strategy Based on Associative Relevance Improves e-Learning," *International Journal of e-Education, e-Business, e-Management and e-Learning*, Vol. 5, No. 3: 165-172, 2015.
- Amaldoss, W., K. Jerath, and A. Sayedi, "Keyword Management Costs and 'Broad Match' in Sponsored Search Advertising," *Marketing Science*, Vol. 35, No. 2: 259-274, 2016.
- Amaldoss, W., P. S. Desai, and W. Shin, "Keyword Search Advertising and First-Page Bid Estimates: A Strategic Analysis," *Management Science*, Vol. 61, No. 3: 507-519, 2015.
- Animesh, A., S. Viswanathan, and R. Agarwal, "Competing 'Creatively' in Sponsored Search Markets: The Effect of Rank, Differentiation Strategy, and Competition on Performance," *Information Systems Research*, Vol. 22, No. 1: 153-169, 2011.
- Arcodia, G. F., "Chinese: A Language of Compound Words?" *Proceedings of the 5th Décembrettes: Morphology in Toulouse*, ed. Fabio Montermini, Gilles Boyé, and Nabil Hathout, 79-90. Somerville, MA: Cascadilla Proceedings Project, 2007.
- Athey, S., and G. Ellison, "Position Auctions with Consumer Search," *Quarterly Journal of Economics*, Vol. 126, No. 3: 123-1270, 2011.
- Chen, J., D. Liu, and A. B. Whinston, "Auctioning Keywords in Online Search," *Journal of Marketing*, Vol. 73, No. 4: 125-141, 2009.
- Chen, T. M. and J. Y. Chen, "Morphological Encoding in the Production of Compound Words in Mandarin Chinese," *Journal of Memory & Language*, Vol. 54, No. 4: 491-514, 2006.

- Dhar, R., "Consumer Preference for a No-Choice Option," *Journal of Consumer Research*, Vol. 24, September: 215-231, 1997.
- Dou, W., R. Linn, and S. Yang, "How Smart Are 'Smart Banners'?" *Journal of Advertising Research*, Vol. 41, No. 4: 31-44, 2001.
- Edelman, B., M. Ostrovsky, and M. Schwarz, "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, Vol. 97, No. 1: 242-259, 2007.
- Even-Dal, E., M. Kearns, and J. Wortman, "Sponsored Search with Contexts," In *Internet and Network Economics*, X. Deng and F. C. Graham (eds.), Vol. 4858: 312-317, Springer, Berlin, 2007.
- Feng, J. and J. Xie, "Performance-Based Advertising: Price and Advertising as Signals of Product Quality," Working paper, University of Florida, Gainesville, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1027296, 2007, retrieved on March 31, 2016.
- Geddes, B., *Advanced Google AdWords*, Wiley Publishing Inc., Indianapolis, Indiana, 2010.
- Ghose, A. and S. Yang, "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, Vol. 55, No. 10: 1605-1622, 2009.
- Guan, Y., X. Wang, X. Kong, and J. Zhao, "Quantifying Semantic Similarity of Chinese Words from Hownet," In *IEEE Proceedings of ICMLC'02*, 234-239, 2012.
- Hall, E. T., *Beyond Culture*, Garden City, NY: Anchor Press, 1976.
- Haskell, R. E., *Transfer of Learning: Cognition, Instruction, and Reasoning*, London: Academic Press, 2000.
- Hausman, J. A., "Specification Tests in Econometrics," *Econometrica*, Vol. 46, No. 6: 1251-1271, 1978.
- Hoeffler, S. and D. Ariely, "Constructing Stable Preferences: A Look into Dimensions of Experience and Their Impact on Preference Stability," *Journal of Consumer Psychology*, Vol. 8, No. 2: 113-39, 1999.
- Hofstadter, D. and E. Sander, *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*, New York: Basic Books, 2013.
- Huber, J., J. W. Payne, and C. Puto, "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, Vol. 9, No. 1: 90-98, 1982.
- Iyengar, S. S. and M. R. Lepper, "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, Vol. 79, December: 995-1006, 2000.
- Jansen, B. J., K. Sobel, and M. Zhang, "The Brand Effect of Queries and Advertisements in Keyword Advertising," *International Journal of Electronic Commerce*, Vol. 16, No. 1: 77-106, 2011.
- Jerath, K. and L. Ma, "A 'Position Paradox' in Sponsored Search Auctions," *Marketing Science*, Vol. 30, No. 4: 612-627, 2010.
- Jiang, C., K. H. Lim, Y. Sun, and J. Z. Peng, "Does Content Relevance Lead to Positive Attitude toward Websites? Exploring the Role of Flow and Goal Specificity," In *AMCIS 2009 Proceedings*, San Francisco, California, USA, August, 2009.
- Keller, L. K. and D. A. Aaker, "The Effects of Sequential Introduction of Brand Extensions," *Journal of Marketing Research*, Vol. 29, No. 1: 35-50, 1992.
- Lambrecht, A. and C. Tucker, "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, Vol. 50, No. 5: 561-576, 2013.
- Lambrecht, A., K. Seim, and C. Tucker, "Stuck in the Adoption Funnel: The Effect of Interruptions in the Adoption Process on Usage," *Marketing Science*, Vol. 30, No. 2: 355-367, 2011.
- Li, S. H. *A Systemic Functional Grammar of Chinese: A Text-Based Analysis*. Continuum: HB, 2007.
- Lincoln, Y. S. and E. G. Guba, "Naturalistic inquiry." Beverly Hills, CA: Sage Publications, 1985.
- Messner, C. and M. Wanke, "Unconscious Information Processing Reduces Information Overload and Increases Product Satisfaction," *Journal of Consumer Psychology*, Vol. 21, January: 9-13, 2011.
- Miller, G. A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity For Processing Information," *Psychological Review*, Vol. 63, March: 81-97, 1956.
- Miller, G. A., "WordNet: A Lexical Database for English," *Communications of the ACM*, Vol. 38, No. 11: 39-41, 1995.
- Narayanan, S. and K. Kalyanam, "Position Effects in Search Advertising and Their Moderators: A Regression Discontinuity Approach," *Marketing Science*, Vol. 34, No. 3: 388-407, 2015.
- Packard, J. L., "The Morphology of Chinese: A Linguistic and Cognitive Approach," Cambridge, UK: Cambridge University Press, 2000.
- Petrin, A. and K. Train, "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, Vol. 47, No. 1: 3-13, 2010.
- Rutz, O. J. and M. Trusov, "Zooming In on Paid Search Ads – A Consumer-Level Model Calibrated on Aggregated Data," *Marketing Science*, Vol. 30, No. 5: 789-800, 2011.

- Rutz, O. J. and R. E. Bucklin, "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, Vol. 48, No. 1: 87-102, 2011.
- Rutz, O. J., M. Trusov, and R. E. Bucklin, "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?" *Marketing Science*, Vol. 30, No. 4: 646-665, 2011.
- Search Engine Watch, "Delving Deep inside the Searcher's Mind," Special report presented at *Search Engine Strategies Conference*, San Jose, CA, 2006.
- Sen, R., J. D. Hess, S. Bandyopadhyay, and J. Jaisingh, "Pricing paid placements on search engines," *Journal of Electronic Commerce Research*, Vol. 9, No. 1: 33-50, 2008.
- Simonson, I., "Determinants of Customers' Responses to Customized Offers: Conceptual Framework and Research Propositions," *Journal of Marketing*, Vol. 69, No. 1: 32-45, 2005.
- Singh, S. K. and V. P. Roychowdhury, "To Broad-Match or Not To Broad-Match: An Auctioneer's Dilemma," *Proceedings of the Fourth Workshop on Ad Auctions*, Chicago, Illinois, USA, 2008.
- Steenkamp, J. and H. Baumgartner, "Development and Cross-Cultural Validation of A Short Form of CSI as A Measurement of Optimum Stimulation Level," *International Journal of Research in Marketing*, Vol. 12, No. 2: 97-104, 1995.
- Sudhir, K., "Structural Analysis of Manufacturer Pricing in the Presence of a Strategic Retailer," *Marketing Science*, Vol. 20, No. 3: 468-481, 2001.
- Tam, K. Y. and S. Y. Ho, "Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes," *MIS Quarterly*, Vol. 30, No. 4: 865-890, 2006.
- Tavassoli, N. T. and Y. H. Lee, "The Differential Interaction of Auditory and Visual Advertising Elements with Chinese and English," *Journal of Marketing Research*, Vol. 40, No. 4: 468-481, 2003.
- Taylor, S. and P. Todd, "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research*, Vol. 6, No. 2: 144-176, 1995.
- Teng, L. and M. Laroche, "Interactive Effects of Appeals, Arguments, and Competition across North American and Chinese Cultures," *Journal of International Marketing*, Vol. 14, No. 4: 110-128, 2006.
- Varian, H., "Position auctions," *International Journal of Industrial Organization*, Vol. 25, No. 6: 1163-1178, 2007.
- Villas-Boas, J. M. and R. S. Winer, "Endogeneity in Brand Choice Models," *Management Science*, Vol. 45, No. 10: 1324-1338, 1999.
- Williams, C. and T. Bever, "Chinese Character Decoding: A Semantic Bias?" *Reading & Writing*, Vol. 23, No. 5: 589-605, 2010.
- Xiao, B., W. Yang, and J. Li, "Optimal Reserve Price for The Generalized Second-Price Auction in Sponsored Search Advertising," *Journal of Electronic Commerce Research*, Vol. 10, No. 3: 114-129, 2009.
- Yang, S., and Ghose, A., "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science*, Vol. No. 4: 602-623, 2010.
- Yang, S., S. Lin, J. R. Carlson, and W. T. Ross, "Brand Engagement on Social Media: Will Firms' Social Media Efforts Influence Search Engine Advertising Effectiveness?" *Journal of Marketing Management*, Vol. 32, No. 5-6: 526-557, 2016.
- Zhao, X., J. G. Lynch, and Q. Chen, "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis," *Journal of Consumer Research*, Vol. 37, No. 2: 197-206, 2010.
- Zhu, Y. and K. C. Wilbur, "Hybrid Advertising Auctions," *Marketing Science*, Vol. 30, No. 2: 249-273, 2011.

Appendix A1. A Sample of the Raw Data

Date	Company	Keyword	Title	Line1	Line2	Impression	Click	Position	Quality
8/27/09	A	discounted airfares	Shanghai discounted airfares 021-51870253	80% off special price air tickets speed booking	special price international lines free delivery	276	2	6.5	7
8/26/09	A	airfares	Shanghai discounted airfares 021-51870253	80% off special price air tickets speed booking	special price international lines free delivery	253	10	6.7	4
8/26/09	A	airfares	Shanghai cheap airline ticket Shanghai discounted airline ticket	021-51870253 international domestic airfares from Shanghai	low price, excellent services, free delivery in Shanghai	6	2	1	4
8/27/09	A	Shanghai airfares	cheap Shanghai airfares surprisingly low holiday tickets	provide domestic and international airline tickets	booking consulting hotline: 021-51870253	406	5	9.4	4
8/26/09	A	Shanghai discounted airfares	Shanghai international airfares, did you buy an expensive one?	provide domestic and international airline tickets	booking consulting hotline: 021-51870253	2	1	3	6
8/27/09	A	airline tickets	Shanghai discounted airfares 021-51870253	80% off special price air tickets speed booking	special price international lines free delivery	238	4	8.7	6
8/28/09	A	airline tickets	Shanghai cheap airline ticket Shanghai discounted airline ticket	021-51870253 international domestic airfares from Shanghai	low price, excellent services, free delivery in Shanghai	510	17	6.5	6
8/27/09	A	special price air tickets	Shanghai cheap airline ticket Shanghai discounted airline ticket	021-51870253 international domestic airfares from Shanghai	low price, excellent services, free delivery in Shanghai	848	15	6.3	6
8/26/09	A	special price air tickets	Shanghai discounted airfares 021-51870253	80% off special price air tickets speed booking	special price international lines free delivery	158	6	5.1	6

Note: The advertisement was originally in Chinese.

Appendix A2. Endogeneity Issues

As the position decided by search engines for a paid search ad is influenced by current CPC and prior CTR of the keyword [Ghose and Yang 2009], position is likely to be endogenous. We conduct the Hausman [1978] test using CPC and lagged CTR as instruments as in Ghose and Yang [2009], since Google decides the keyword position by considering both the current CPC and the quality score which is determined by that keyword's prior CTR [Varian 2007; Athey and Ellison 2011]. According to our results, the null hypothesis that position is exogenous should be rejected at the 1% level of significance ($p < 0.001$).

Because matching strategies are strategic decisions made by both the advertiser and the competitors, they could also be endogenous. Following Sudhir [2001], we test for endogeneity of matching and quality score in two ways. First, we examine the correlation between estimated residuals from the model where only position is treated as endogenous. The correlations are not significant ($p > 0.1$), which indicates that there should be no endogeneity bias if we treat matching as exogenous. Second, we conduct the Hausman's [1978] test using lagged matching as instruments with position as endogenous. The results indicate that we cannot reject the null hypothesis that matching strategies are exogenous ($p = 0.158$). We use an easily available set of instruments, lagged matching, for matching, using an argument analogous to that used by Villas-Boas and Winer [1999]. The idea is that while CTR in any given period is likely to be affected by matching strategies of the same period, it is less likely to be influenced by matching strategies of the previous period. Thus lagged matching is an appropriate instrument for the matching variable in a given period.

Similarly, we use lagged quality score as instrument of quality score. However the correlation between residuals and quality score is significant ($p < 0.001$), indicating that there should be endogeneity bias if quality score are treated as exogenous. The Hausman's (1978) test using lagged quality score as instruments for quality score also indicates that quality score are not exogenous ($p < 0.001$). To correct for the endogeneity of position and quality score, we use the control function approach [Petrin and Train 2010] that can handle endogeneity in choice models⁵, and obtain bootstrapped standard errors.

⁵ Petrin and Train [2010] describe the use of control functions to correct for potential endogeneity of a single variable. We have confirmed in an email exchange with Professor Kenneth Train that the method can be applied to correct for multiple potential endogenous variables, where the multiple control functions and the error term in the utility function are assumed to be jointly normal.