Price Information Patterns in Web Search Advertising: An Empirical Case Study on Accommodation Industry

Guanting Tang
School of Computing Science
Simon Fraser University
Burnaby, Canada
gta9@cs.sfu.ca

Yupin Yang
Beedie School of Business
Simon Fraser University
Burnaby, Canada
yya29@sfu.ca

Jian Pei
School of Computing Science
Simon Fraser University
Burnaby, Canada
jpei@cs.sfu.ca

Abstract—Unlike advertising in traditional media, web search advertising content can be easily customized with little cost. In this paper, we apply content analysis and regression models on $11,818$ unique ads related to the accommodation industry to empirically investigate how advertisers customize price information in their web search advertising content. To the best of our knowledge, our study is the first of this kind. We find that advertiser characteristics, such as website traffic, product quality, and position in the distribution chain, affect both the amount and forms of price information in its search advertising content. Moreover, the use of price information by an advertiser depends on query characteristics, such as search volume, cost per click ("CPC"), and specific words (e.g., trademark, location, price cue) in queries. Our empirical findings shed new light on how to effectively manage price information in search advertising, and suggest new research opportunities on web search advertising.

Keywords-search advertising; content analysis; regression

I. INTRODUCTION

Web search advertising (hereafter "search advertising") has become increasingly important, and has been the dominant Internet advertising sector [1]. In 2009, search engine advertising accounted for about 47% of the $22 billion spent on Internet advertising in the United States [2]. The growth in search advertising expenditure in recent years has spurred a significant increase in research on search advertising. Most of the research focused on auction mechanism design used in search advertising [3], [13] and effectiveness of position and query characteristics on click-through rates or conversion rates [15], [38]. Surprisingly, although Ad content is an important component of search advertising, it has been almost neglected in search advertising research, with the only exception of two recent studies on how queries [32] and trademarks [9] in ad content affect consumers' actions. Different from traditional advertising media, search

advertisements are triggered when consumers search a query in search engines. Search advertising tools (e.g., Google AdWords) offered by search engines enable advertisers to track their ads' performance. Thus, it is easier for advertisers to customize and optimize their search advertising content according to queries and past performance of their search advertisements. In addition, multiple ads are usually displayed at the same time to compete for consumers' attention.

More and more companies tend to allocate more and more resources to search advertising. It is increasingly important for advertisers to decide what information should be included and how the information should be expressed in their ad content in order to attract consumers' attention and increase revenues. Given the great flexibility of customization and the high importance of search advertising content, in this paper, we conduct an empirical investigation into how advertisers design their search advertising content, price information in particular, according to advertiser characteristics (e.g., website traffic, product quality, position in the distribution chain) and query characteristics (e.g., query length, search volume, CPC).

For consumers, price is a critical factor in making purchase decisions. Consumers perceive price information not only as the cost of obtaining a product, but also as a cue to infer product quality [22], [26]. Because of its importance in consumers' purchase decisions, price information has been an essential component in advertising content. Compared with other product attributes that may vary across industries, price information is easier to identify, classify, and generalize to other industries. Thus, in this paper we focus on the price information in search advertising content.

In particular, we choose the accommodation industry to conduct a case study for the following reasons. First, online travel and leisure accounts for a significant percentage (about 6%) of Internet advertising [2]. Second, the high degree of competition in the accommodation industry makes price information an important component of search advertising content. Third, search advertising is heavily used by different channel members in the accommodation industry, such as service providers (e.g., hilton.com), travel agencies

(e.g., expedia.com), price search websites (e.g., kayak.com), and third-party information providers (e.g., tripadvisor.com). Furthermore, the provision of accommodation service varies from very basic (e.g., hostels) to luxury (e.g., five-star hotels and resorts). These variations allow us to examine the impacts of advertiser characteristics on the use of price information. Last but not least, many companies in the accommodation industry are early adopters of search advertising. Thus, they have already passed the learning stage in designing effective advertising content, which makes our results reliable and generalizable.

In this paper, we first apply content analysis to extract price information in search advertisements. Then, we apply regression models to study the relationships between the amount and form of price information used in search advertising, and the characteristics of queries and advertisers.

To the best of our knowledge, we are the first to empirically examine the relationships between search advertising content and characteristics of queries and advertisers. Our analysis discovers a handful of interesting patterns and possible explanations, which shed new light on how to effectively manage price information in search advertising, and suggests new research opportunities on search advertising.

The rest of the paper is organized as follows. In Section II, we review the related work. We then describe our data sets in Section III, and develop our empirical models in Section IV. Section V reports the results and findings. We conclude the paper in Section VI.

II. RELATED WORK

Our study builds on and extends three streams of the existing literature: search advertising, price advertising, and content analysis. We briefly review the three streams and discuss how this paper contributes to the literature.

Although textual content has a great impact on consumer behaviors, little attention was paid to the textual content of search advertisements, either empirically or theoretically. Recently, some researchers, such as Chiou and Tucker [9] and Rutz and Trusov [31], studied the impacts of textual content (e.g., use of queries or trademarks) of search advertising on consumer behaviors (e.g., click-through). In one of its tips for creating successful ads, Google suggests including prices, promotions, and exclusives in the text, so that consumers can use this information to make decisions.

Despite the importance of price information in consumer behaviors, the previous research did not examine the use of price information in search advertising. We investigate not only the amount and forms of price information used in search advertising, but also the relationships between these variables and the characteristics of queries and advertisers. Given that utilization of search advertising content is based on past performance, our findings can shed light on the impacts of price information on consumer behaviors, which is an interesting direction for future work.

Price is an important marketing mix variable that has been intensively studied in literature. Tellis [34] proposed a unified taxonomy and pointed out that consumer characteristics and company objectives are two basic dimensions underlying price principles, strategies, and tactics. One line of price research studies how the processing of price information is influenced by consumer characteristics, such as price consciousness [14], product/brand familiarity, and prior knowledge and experiences regarding prices [23], and how price information impacts on consumers' evaluations and choices [21]. The other line of research studies the relationship between price and other marketing variables, such as advertising [20] and product quality [26].

Since price is important for consumers who are making purchase decisions, it is critical for firms to convey price information in their advertising. Previous research found that an increase in price advertising leads to higher price sensitivity among consumers, lower prices [20], and lower product quality perception [37]. However, previous research on price advertising often ignores how price information is expressed in ad content, even though expression forms have a large impact on consumers' evaluations and choices [36]. Our paper studies the relationships between expression forms of price information and characteristics of advertisers and queries, which reflect the characteristics of consumers.

Since Kassarjian [19] introduced the content analysis method into consumer research, content analyses have been applied to study message content in a variety of advertising contexts such as TV advertising [28], environmental advertising [6], teleshopping [5], use of print advertising visuals [24], and advertising of services [35]. Because search advertising is a new advertising medium, its content has not been explored in previous research. However, the unique characteristics of search advertising, such as query-triggered ads, and the great flexibility of search advertising in content customization allow advertisers to better target their customers. Thus, it is worthwhile to make an investigation in search advertising content. Our paper is the first to explore search advertising content and its relationship with characteristics of queries and advertisers.

III. DATA SETS

To analyze how the use of price information in search advertising is impacted by the characteristics of queries and advertisers, we use three data sets: a query data set, a search advertisement data set, and an advertiser data set. The data sets are downloadable at http://www.cs.sfu.ca/%7Ejpei/Software/AdsDataSets.zip.

A. The Query Data Set

A *query* contains one or multiple words. We collected queries and their characteristics, including global monthly searches and approximate cost per click (CPC for short), from the accommodation category in the Travel & Tourism

section of Google AdWords [16]. The measure of *global monthly searches* is the average number of searches for each query on Google over the past 12 months in all locations, languages, devices, and query match types [17]. *Approximate CPC* is the cost of a query that a customer might have to pay if she bid on it. It is calculated by the average costs of ads on all positions of that query [17].

The query data set was crawled from Google AdWords in April 2011. There are 3,151 unique queries in total. 2,124 of them have ads. In our study, we consider all 1,542 queries that return at least an ad related to accommodation. The 582 deleted queries include those do not yield any ads related to the accommodation types investigated in this study.

We constructed the query length variable, which is the number of words in a query, and three dummy variables indicating whether a query contains a trademark (e.g., "Hilton"), a location name (e.g., "New York"), and at least a price cue (e.g., "cheap", "deal", "discount"), respectively. A *price cue* here is defined as a word, which is used to persuade customers that the current price offers good value compared to competitors' prices, past prices or future prices [4]. A query may contain more than one price cue.

B. The Search Advertisement Data Set

The content of search advertisements is obtained by submitting the queries from the query data set to search engine Google. Advertisers often randomize several advertisements for a query. Thus, we submitted each query 30 times to make sure that our data includes most of the advertisements. The same advertisement may be recorded more than once. We regard two advertisements identical only if they satisfy three conditions: (1) their entire content, including the headline and description, is the same; (2) they are triggered by the same query; and (3) they come from the same advertiser. The replicated advertisements were deleted from the data set. In addition, there are 1,027 queries that do not yield any paid ad, which were excluded from the query data set.

After obtaining the content of the ads, we constructed the ad length variable, which is the number of meaningful words in an advertisement, including the headline and description. We excluded from the meaningful words non-price related symbols, such as "&", "-", "+", and "—", and stop words. The set of stop words are "*a, about, above, across, along, among, amongst, an, and, around, as, at, before, behind, below, beside, besides, but, by, during, either, etc., for, from, in, into, near, not, nor, of, on, onto, or, over, per, than, the, then, to, too, under, unto, upon, with, within, without, yes*".

We categorized the price information into three forms: (1) actual price number, (2) percentage discount, and (3) relative price information [35]. An *actual price number* was identified by a currency word (e.g., dollar), abbreviations (e.g., USD), or a symbol (e.g., $, £, €) with a number. We constructed a dummy variable to indicate whether an advertisement contains at least an actual price number.

Group 1	Group 2	
affordable	amazing	hot (hottest)
bargain (bargains)	better	huge
cheap (cheaper, cheapest)	best	irresistible
competitive	big	incredible
deal (deals)	budget	less
discount (discounts, discounted)	check	low (lower, lowest)
last minute	compare	more
money	exclusive	no (cost/fee)
match	fantastic	offer (offers)
price (prices)	find	offering (offerings)
sale (sales)	free	pay
save	get	promotion (promotions)
saving (savings)	good	receive
rate (rates)	great	special
reduced	guarantee	super
	guaranteed	unbeatable

Table I: List of Relative Price Words

Meanwhile, we recorded the nearby words, such as "just," "only," and "per-night", that are used to describe the actual price numbers. These words were used later to calculate the total number of words for price information. Following a similar process, the *percentage discount* was identified by the percentage sign (%) and price cues like "save" and "off". A dummy variable was constructed to indicate whether at least a percentage discount is included in an advertisement. We also recorded nearby words like "up" and "off" that are used to describe percentage discounts.

Calculating the number of words that convey relative price information is a more time-consuming and complicated process. First, we went through all ads under investigation and compiled a list of words that are used as price information. Second, we divided those words into two groups. The words in the first group are those used to describe only price information (e.g., "cheap", "savings", "deal", "discount"). The words in the second group are those used to describe either price or other attributes. For example, "guarantee" can refer to a price guarantee or a service quality guarantee. Therefore, the words in the second group need further information from nearby words. A word in the second group is counted as price information only if the three words immediately preceding or following it contain a price cue from the words in the first group. Table I gives the complete list of the two groups of relative price words. We counted the number of words contained in both groups to get the number of words for relative price information. Last, we used the ratio between the number of relative price words and the number of meaningful words in an advertisement to construct the percentage of relative price words variable.

After obtaining the price information in these three forms, we aggregate it together to calculate the amount of price information in an advertisement, which is measured by the ratio of the number of all price-related words against the total number of meaningful words.

In addition to the ad content, the position of each ad was also recorded in the data set. On a search result page, a paid ad appears either underneath the search bar or on the

right-hand side. Since a query is crawled multiple times, it is possible that an ad appears on multiple places. We recorded all positions for an ad, and constructed two variables to indicate the position of the ad. First, we used a dummy variable to indicate whether an ad appears on the right-hand side. In the cases where it appears in both locations, the dummy variable indicates whether it appears on the right-hand side more than 50% of times. Very few ads (40 out of 11,818) appear in both places. Since a maximum of three ads can appear underneath the search bar, we did not distinguish the relative position (1st, 2nd, 3rd) of these ads in this paper. However, the relative positions of the ads on the right-hand side were recorded in a variable. When an ad appears in multiple positions on the right-hand side, the average of the positions was recorded in the variable.

C. The Advertiser Data Set

To identify the advertisers we used the domain part of the URLs appearing in the advertisements. For example, we treated www.expedia.ca and www.expedia.com the same advertiser instead of two. The advertisers that are not in the accommodation business (e.g., sears.com, ask.com) were excluded from our data set.

Most of the advertiser characteristics were constructed based on the advertisers' website information (especially on the "About Us" pages). First, we categorized each advertiser in one of four types based on its role in the distribution chain of the accommodation industry: (1) service provider (e.g., Hilton), (2) travel agency/intermediary (e.g., Expedia), (3) price search website (e.g., Kayak), and (4) third-party information provider (e.g., Tripadvisor). Second, we examined the category and characteristics of the accommodation that an advertiser offers, sells, or provides information for, and considered three types of accommodation: hostels, hotels, and other accommodation options (e.g., cottages, bed and breakfast, farmhouses, and vacation homes). Some advertisers offer only one type of accommodation, while others may offer multiple types of accommodation. Therefore, we constructed three dummy variables to indicate whether an advertiser offers hostels only, hotels only, or multiple types of accommodation. Third, a dummy variable was constructed to indicate whether an advertiser offers other products such as flights and car rental as well. Fourth, a dummy variable was constructed to indicate whether an advertiser is part of a hotel chain (e.g., Hilton). Last, we used the traffic estimator website Alexa to gather the global ranks of the advertisers' website traffic. Even though the global rank may fluctuate over time, the relative position of the advertisers are relatively stable. Therefore, we constructed the advertiser ranking variable according to the rank of the advertisers indicated by the global ranks from Alexa. Note that 106 of the 963 advertisers cannot be found in Alexa. As these missing advertisers usually have very low website traffic and therefore a very low rank, we assigned

the maximum ranking number ($866 = 963 - 106$) to them.

After all, our data sets comprise 11,818 advertisements, $1,542$ queries, and 963 advertisers. The basic statistics of the variables are summarized in Table II.

IV. REGRESSION MODELS

In this section, we apply regression models to examine the relationships between the use of price information in search advertising content and the characteristics of queries and advertisers.

We first study how the amount of price information used in search advertising content is related to the query characteristics and advertiser characteristics using a linear regression model as follows:

$$Y_{1ijk} = \alpha_1 + \beta_1 * X_i + \gamma_1 * Z_j + \delta_1 * H_k + \epsilon_{1ijk} \quad (1)$$

where i, j, k are the index numbers for queries, advertisers, and ads. The dependent variable Y_{1ijk}, which represents the amount of price information in an advertisement, is measured by the ratio of the number of price-related words against the total number words in the advertisement. X_i, Z_j, and H_k, respectively, represent the independent variables for query i, advertiser j, and ad k. The independent variables X_i on the query side include the variables in Table III, the independent variables Z_j on the advertiser side include the variables in Table IV, and the independent variables H_k include two variables: $Right_{Side_k}$ is a dummy variable indicating whether ad k is on the right-hand side, and $Position_k$ is the average position of ad k if it is on the right-hand side. Parameters β_1, γ_1, δ_1 are the coefficients for the independent variables, α_1 is a constant term, and ϵ_{1ijk} is the error term and assumed to be independent and identically distributed following normal distributions.

Not only the amount of price information in search advertising content influences consumers' perception and behaviors, but also the expression forms of price information is important in shaping consumers' perception and impacting their behaviors. Therefore, we study the use of three price forms: actual number, percentage discount, and relative price information in search advertising. The following two logistic regression models and one linear regression model are applied to investigate the relationships of these three price forms with the characteristics of queries and advertisers.

$$Logit(P(Y_{2ijk}|X_i, Z_j, H_{ijk})) = \alpha_2 + \beta_2 * X_i + \gamma_2 * Z_j \\ + \delta_2 * H_k + \epsilon_{2ijk} \quad (2)$$

$$Logit(P(Y_{3ijk}|X_i, Z_j, H_{ijk})) = \alpha_3 + \beta_3 * X_i + \gamma_3 * Z_j \\ + \delta_3 * H_k + \epsilon_{3ijk} \quad (3)$$

$$Y_{4ijk} = \alpha_4 + \beta_4 * X_i + \gamma_4 * Z_j + \delta_4 * H_k + \epsilon_{4ijk} \quad (4)$$

where independent variables X_i , Z_j , H_k are the same as the ones used Equation 1 for comparison. Dependent variable Y_{2ijk} is a dummy variable used to indicate whether

Variable	Mean	Std.dev.	Min.	Max.	No. of Obj.
Query Variables					
Global Monthly Searches	125,943.40	433,427.10	480	6,120,000	1,542
Approximate CPC	1.42	0.82	0.1	5.77	1,542
Query Length	3.03	1.06	1	8	1,542
Trademark Dummy	0.21	0.41	0	1	1,542
Location Dummy	0.87	0.34	0	1	1,542
Price Cue Dummy	0.05	0.21	0	1	1,542
Ads Content Variables					
Advertisement Length	13.11	3.27	6	29	11,818
Actual Price Number Dummy	0.24	0.43	0	1	11,818
Percentage Discount Dummy	0.25	0.43	0	1	11,818
Percentage of Relative Price Words	0.18	0.15	0	0.75	11,818
Percentage of All Price-related Words	0.25	0.19	0	0.85	11,818
Right-hand-side Dummy	0.82	0.38	0	1	11,818
Average Position on the Right-hand-side	3.00	2.07	1	8	9.737
Advertisers Variables					
Service Provider Dummy	0.62	0.49	0	1	963
Travel Agency/Intermediary Dummy	0.30	0.46	0	1	963
Price Search Website Dummy	0.02	0.13	0	1	963
Third-party Information Providers Dummy	0.07	0.25	0	1	963
Hostel-Only Dummy	0.05	0.22	0	1	963
Hotel-Only Dummy	0.47	0.50	0	1	963
Other Accommodation-Only Dummy	0.42	0.49	0	1	963
Multiple-Accommodation Dummy	0.05	0.22	0	1	963
Other-products Offering Dummy	0.10	0.29	0	1	963
Chain Dummy	0.22	0.42	0	1	963
Website Global Rank	6,280,754	8,820,014	224	27,590,488	963
Advertiser Website Traffic Ranking	477.06	270.59	1	866	963

Table II: Summary of the Basic Statistics on the Data Sets.

an advertisement contains an actual price number. Similarly, dependent variable Y_{3ijk} is also a dummy variable used to indicate whether an advertisement contains a percentage discount. Given that Y_{2ijk} and Y_{3ijk} are dummy variables, logistic regression models are applied to study the use of price forms of actual price number and percentage price discount. Different from the forms of actual price number and percentage discount, the form of relative price information is measured by the ratio of relative price-related words to the total number of words contained in an advertisement. Given that the dependent variable Y_{4ijk} is continuous, a linear regression model is used to study the use of relative price information. ϵ_{2ijk}, ϵ_{3ijk}, and ϵ_{4ijk} are the error terms for the models and assumed to be independent and identically distributed following normal distributions. α_2, α_3, α_4 are constant terms. β_2, γ_2, δ_2, β_3, γ_3, δ_3, β_4, γ_4, δ_4 are the coefficients for the independent variables for these models.

Stata [33] is used to estimate the parameters β_1, γ_1, δ_1, β_2, γ_2, δ_2, β_3, γ_3, δ_3, β_4, γ_4, δ_4 in the four regression models, and the results are presented in the next section.

V. RESULTS AND FINDINGS

Table V shows the estimated results for the four regression models. We discuss the amount of price information and the uses of the three price forms in relationships with query characteristics, advertiser characteristics, and the locations and positions of advertisements.

Variable	Meaning
$Log(Q_{GMS_i})$	mean-adjusted variable of the log transformation of the global monthly searches variable for query i
$(Log(Q_{GMS_i}))^2$	quadratic transformation of $Log(Q_{GMS_i})$
Q_{CPC_i}	mean-adjusted variable of approximate CPC for query i
$(Q_{CPC_i})^2$	quadratic transformation of Q_{CPC_i}
Q_{Length_i}	number of words in query i
$Q_{Trademark_i}$	dummy variable indicating whether query i contains a trademark
$Q_{Location_i}$	dummy variable indicating whether query i contains a location name
$Q_{Pricecue_i}$	dummy variable indicating whether query i contains at least a price cue (e.g., cheap, deal, discount, or last minute)

Table III: The variables included in X_i on the query side.

A. Query Characteristics

The query characteristics often indicate the characteristics of consumer segments, the stage in the decision-making process, and/or the degree of complexity of information processing tasks. Consumer decision-making processes have been extensively studied in the literature of information processing and consumer behavior [25], [7], [27]. The key factors influencing consumer decision processes are task characteristics (e.g., number of alternatives available, numerical vs. verbal information); individual differences (e.g., personal values, prior knowledge and experience); and the interactions between them [25]. The query characteristics including search volume, query length, trademark, CPC, and location in some degree reflect these factors that impact consumer decision process. Next, we summarize how these query characteristics are related with the use of price infor-

	Dependent Variable							
	Model (1)		Model (2)		Model (3)		Model (4)	
	Percentage of all Price-related Words		Actual Price Number		Percentage Discount		Percentage of Relative Price Words	
Independent Variable	Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
Constant	0.171***	0.010	-0.314	0.201	-3.987***	0.264	0.097***	0.008
Query Variable								
$Log(Q_{GMS_i})$	-0.003	0.002	0.275***	0.051	0.081***	0.040	-0.002	0.002
$(Log(Q_{GMS_i}))^2$	0.006***	0.002	-0.173***	0.040	0.213***	0.030	0.003*	0.002
Q_{CPC_i}	0.006**	0.002	0.240***	0.049	-0.027	0.040	0.006***	0.002
$(Q_{CPC_i})^2$	-0.002***	0.001	-0.201***	0.031	0.060***	0.021	-0.002***	0.001
Q_{Length_i}	-0.011***	0.002	-0.153***	0.034	-0.041	0.027	-0.005***	0.001
$Q_{Trademark_i}$	-0.021***	0.006	-0.386***	0.144	-0.565***	0.101	0.005	0.005
$Q_{Location_i}$	0.016***	0.006	0.929***	0.130	0.110	0.094	0.004	0.005
$Q_{Pricecue_i}$	0.084***	0.005	0.990***	0.102	0.257***	0.082	0.058***	0.004
Advertiser Variable								
A_{Rank_j}	0.000***	0.000	-0.003***	0.000	0.002***	0.000	0.000	0.000
$(A_{Rank_j})^2$	-4.21e-07***	0.000	1.34e-06**	0.000	-7.73e-06***	0.000	-1.70e-07***	0.000
$A_{Service_{Provider_j}}$	-0.016*	0.008	0.232	0.141	0.493**	0.241	-0.010	0.007
$A_{Tavel_{Agent_j}}$	0.056***	0.007	-0.055	0.118	1.538***	0.187	0.048***	0.006
$A_{Search_{Website_j}}$	0.151***	0.009	-3.687***	0.281	2.225***	0.197	0.115***	0.007
A_{Hostel_j}	0.124***	0.007	1.060***	0.118	-4.268***	1.018	0.091***	0.006
A_{Hotel_j}	0.041***	0.006	-2.246***	0.118	2.389***	0.187	0.048***	0.005
$A_{M_{Accom_j}}$	0.061***	0.007	1.066***	0.122	2.474***	0.193	0.027***	0.006
$A_{Other_{Product_j}}$	0.027***	0.005	-1.214***	0.114	0.005	0.065	0.027***	0.004
A_{Chain_j}	-0.024***	0.008	-0.737***	0.154	-0.636***	0.187	0.003	0.006
Ads Variable								
$Right_{Side_k}$	0.019***	0.005	0.145*	0.088	-0.322***	0.077	0.009***	0.004
$Position_k$	-0.001*	0.001	-0.104***	0.017	-0.016	0.014	-0.001*	0.001
Goodness of fit R-square	0.29		0.38		0.25		0.22	

Note: Significance at 0.01, 0.05 and 0.1 levels are denoted by ***, **, and * respectively.

Table V: Summary of Results on the Estimates of the Parameters for All the Models.

Variable	Meaning
A_{Rank_j}	website traffic ranking of advertiser j, which is mean-adjusted
$A_{Rank_j}^2$	square of the variable A_{Rank_j}
$A_{Service_{Provider_j}}$	dummy variable indicating whether advertiser j is a service provider
$A_{Travel_{Agent_j}}$	dummy variable indicating whether advertiser j is a travel agent
$A_{Search_{Website_j}}$	dummy variable indicating whether advertiser j is a price search website
A_{Hostel_j}	dummy variable indicating whether advertiser j offers/sells only hostel accommodation
A_{Hotel_j}	dummy variable indicating whether advertiser j offers/sells only hotel accommodation
$A_{M_{Accom_j}}$	dummy variable indicating whether advertiser j offers/sells multiple types of accommodation
$A_{Other_{Product_j}}$	dummy variable indicating whether advertiser j offers other products
A_{Chain_j}	dummy variable indicating whether advertiser j is part of an accommodation chain

Table IV: The variables included in Z_j on the advertiser side.

mation in search advertising.

1) Query Search Volume: Our models include the mean-adjusted and log-transformed variables of query search volume, measured by global monthly searches, and its quadratic term. Therefore, the coefficients for these two terms show the relationships between query search volume and the uses of price information including the amount and three price forms in search advertising. To visualize the relationships, we plot Fig. 1 using the estimated coefficients in Table V.

As shown in Fig. 1(a), the amount of price information has a U-shape pattern with query search volume. This means that more price information is used for advertisement with both low and high search volume. In the information processing literature, Bettman and Park [7] found that consumers who are in the early stage of decision process use more attribute-based processing than brand-based processing, engage in more comparison, and use the elimination-by-attribute strategy. They also found that consumers with less prior knowledge or less experience engage in more attribute-based processes. Given that price information is an important attribute for accommodation services and relatively easy to compare, price information is more likely to be presented in the search ad content for less knowledgeable and less experienced consumers in the early stage of decision process. Also, the queries with a higher number of global monthly searches are usually used by consumers who have less information and/or are in the early stage of the decision process [11]. Therefore, price information is more likely to be presented in these high search volume queries. However, why more price information is used for the advertisement with low search volume queries? One possible reason may be that queries with very low search volume are used for niche markets or unknown markets, and low pricing strategy is the more effective way for these advertisers to attract more consumers.

As we look further into the three price forms in relationships with query search volume. They are also non-linear and vary across price forms. The uses of percentage discount (Fig. 1(c)) and relative price words (Fig. 1(d)) follow a
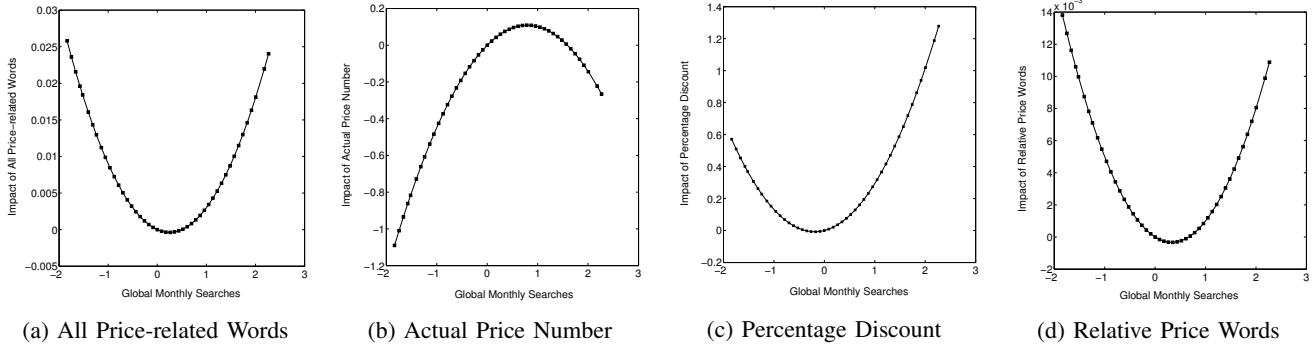
(a) All Price-related Words   (b) Actual Price Number   (c) Percentage Discount   (d) Relative Price Words

Figure 1: Relationships between the use of price information (y axis) and query search volume (x axis). "Global Monthly Searches" is represented by the mean-adjusted and log-transformed variable of the actual global monthly searches numbers ($\log(Q_{GMS_i})$).

U-shape pattern as the total amount of price information. Surprisingly, the use of actual price number (Fig. 1(b)) follows an inverted U-shape pattern. The reasons behind the less use of actual price number for very high search volume queries could be that consumers searching for these very high search volume queries are in the earliest stage of decision making. The actual price numbers probably have not entered their consideration yet. Another reason could be that these consumers have little or no prior knowledge, and could not process the actual price numbers in an unbiased way. Rao and Sieben [27] found that the acceptable price range is lower for people with very low prior knowledge. Thus, consumers with little or no prior knowledge might reject advertisements with actual price numbers because they may consider the price too high. However, percentage discount or relative price words provide price comparison information easier for consumers with low knowledge to interpret and accept. Therefore, more price information in percentage discounts and relative price words but not actual price number are presented for the very high search volume queries.

2) Query Length and Trademarks: As Table V indicates, the coefficients for the query length and trademark in query are significantly negative for the amount of price information used in an advertisement. These negative coefficients mean that less price information is shown for consumers searching for longer queries or queries containing a trademark. When consumers search using a query of more words or a query containing a trademark, they are highly likely to have some prior knowledge about prices, and also likely to be in the late stage of their decision-making process researching specific information or engage in brand-based processing of information. Thus, less price information is used by the advertisers for these queries.

In terms of the price forms, the coefficients of actual price number and relative price words are significantly negative while the coefficient of percentage discount is not significant for the query length variable. Consumers who search using longer queries are more likely to have already done some

research or have more knowledge, and need further information on other attributes. Thus, price information such as actual price number and relative price words are used less in a longer query.

The coefficient for the trademark variable is significantly negative for actual price number and percentage discount. However, there is no significant difference between use of relative price words in queries containing a trademark and those without a trademark. Consumers who search using queries with a trademark usually have prior experience and some degree of preference toward the trademark. They are probably looking for further detailed information in some specific attributes to confirm their preferences. Therefore, price information, such as actual price number or price discount, is not attractive to them, and price information is used less for queries with a trademark. However, the use of relative price words is not clearly dependent on the trademark variable. A possible explanation is that advertisers who have a price advantage use relative price words to confirm or reassure consumers about their prior beliefs and preferences. However, the percentage of all price-related words is lower in queries with a trademark than in those without a trademark.

3) CPC: Similar to query search volume, our models include both mean-adjusted query approximate CPC and its quadratic term. According to the estimates of the coefficients for these two variables (Table V), we plot Fig. 3.

The amount of price information used in an advertisement has an inverted-U shape with query CPC. Our data sets show that, on average, queries with CPCs above \$2.50 generate 1.6 more advertisements than queries with CPCs below \$2.50. With more ads competing for consumers' attention, price becomes a differentiating factor in advertisers' ability to attract traffic. As the number of available alternatives increases, consumers in the early stage of their search engage more in information search and multistage decision processes, and conduct more elimination-by-attribute [7], [10], [12]. When more competitors' ads are available, price information becomes an important attribute for getting past the first stage of the consumer choice process, which could lead to more
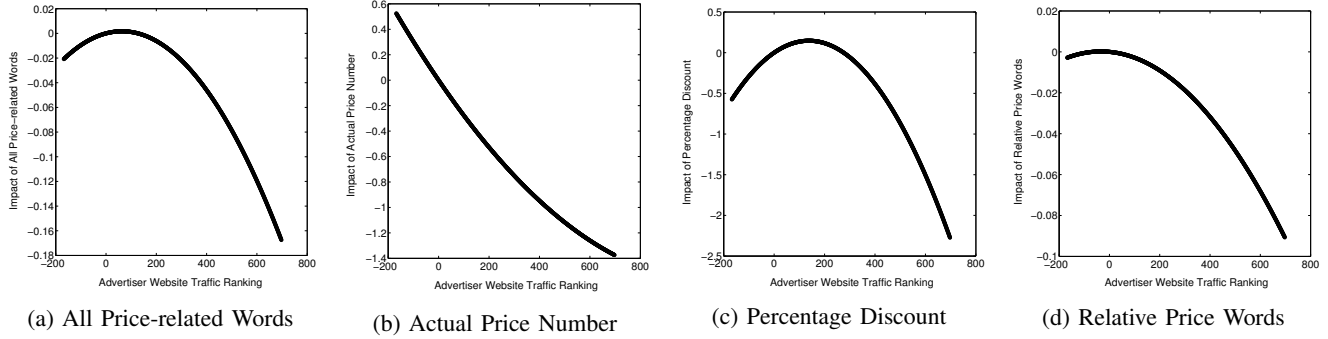
| (a) All Price-related Words | (b) Actual Price Number | (c) Percentage Discount | (d) Relative Price Words |

Figure 2: Relationships between the use of price information (y axis) and advertiser website traffic ranking (x axis). "Advertiser Website Traffic Ranking" is represented by mean-adjusted popularity ranking (A_{Rank_j}).
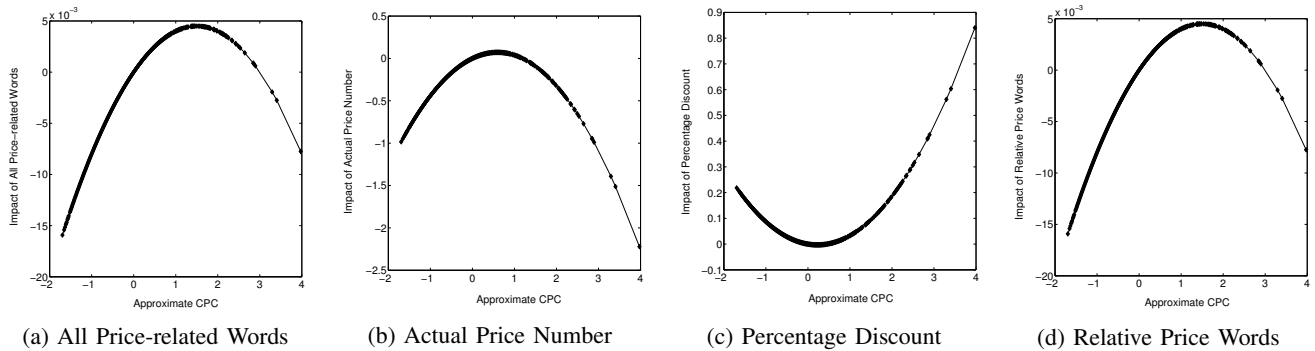

| (a) All Price-related Words | (b) Actual Price Number | (c) Percentage Discount | (d) Relative Price Words |

Figure 3: Relationships between the uses of price information (y axis) and query CPC (x axis). "Approximate CPC" is represented by the mean-adjusted query approximate CPC (Q_{CPC_i}).

clicks. Thus, more price information may be used in ads for queries of higher CPCs. However, an advertiser is willing to pay more per click if a query can generate higher revenue, which is usually driven by higher prices. The advertisers offering high-price accommodations are targeting less price-sensitive consumers. Thus, price information is used less for queries with very high CPCs.

Fig. 3 shows that the uses of actual price numbers and relative price words follow the similar pattern as the total amount of price information. However, the use of the percentage discount form starts to increase when the approximate CPC reaches a certain high level. As mentioned earlier, the advertisers offering high-price accommodations are more likely to bid on the queries with very high CPCs. Thus, percentage discount is more likely to be used than actual price numbers by these advertisers because their actual price numbers are already very high compared with those of low-price products.

4) Miscellanea: As Table V shows, the coefficient for the location variable is significant positive for the amount of price information used in an ad. However, this result is driven by the more use of the form of actual price numbers. There is no significantly more use of percentage discount or relative price words. A recent consumer study conducted by travel industry research authority PhoCusWright Inc. [29] found nearly half of travelers had a particular place in mind,

but are flexible about travel dates. Therefore, more precise price information, such as actual price numbers, can attract the attention of consumers who look for lower prices in order to decide their travel dates and locations.

The coefficients for the price cue variable are significantly positive for all three price forms. Consumers who search for queries with price- and promotion- related words, such as "cheap", "deal", "last minute", and "discount", are usually price sensitive, and price information is likely to be used to attract these consumers. These results are consistent with Becker and Connor [8], who showed that personal values influence consumers' attitudes towards price as an attribute, and that these attitudes, in turn, influence their choices.

B. Advertiser Characteristics

1) Customer Loyalty and Accommodation Chains: Similar to query search volume and approximate CPC, we also include both the mean-adjusted popularity ranking variable and its quadratic terms for advertisers in the models. Using the coefficients calculated, we plot Fig. 2 to visualize the relationships between the uses of price information and advertiser website traffic rankings.

The amount of price information used in an ad is the highest for the advertisers with medium-high website traffic ranks. In previous literature, Raju *et al.* [30] showed that brands with larger brand loyalty use price promotion less

frequently. Similarly, Iyer and Pazgal [18] found that a store with a greater number of loyal customers is less likely to use price comparison sites. Advertisers with more website traffic have a greater number of loyal customers. Thus, price information is expected to be used less by advertisers with high website traffic. However, advertisers with low traffic may be engaging in a focused marketing strategy. Therefore, it may also be true that advertisers with very low traffic use more special feature information instead of price information. As a result, we observe an inverted U-shape effect of the advertiser website traffic ranking.

The price information in the forms of percentage discount and relative price words shows the similar patterns as the total amount of price information. However, the use of actual price numbers decreases as advertisers' website traffic decreases. It might seem at first that the result for actual price numbers contradicts previous theories and findings [30], [18]. However, if we consider the differences among the three price forms, there is an explanation. Percentage discount and relative price words are used for price comparison or price promotions, while actual price numbers are used to indicate the exact price information or to signal product quality. Therefore, instead of a price promotion, actual price numbers are used more often in popular websites to signal product quality or provide price information.

Table V also shows that the percentage of all price-related words is 2.4% less for advertisers that are part of a product chain. Advertisers that are part of a large accommodation chain have a larger customer base. Thus, less amount of price information is observed in their search advertising. Meanwhile, these chain advertisers are also unlikely to use actual price numbers and price discount in their search advertising. However, the use of relative price words is not significantly different no matter advertisers that are part of a chain or not. It may be possible that a chain with an established brand name would prefer to use relative price words instead of actual price numbers and percentage discount to reinforce the consumers' perception.

2) Types of Advertisers: The coefficients for different types of advertisers indicate that the percentage of all price-related words is the highest for the price search websites, about 9.6% higher than travel agents, the next highest type. Service providers use the least amount of price information in their search advertising. These results are consistent with in the findings of Kaul and Wittink's meta-analysis of price advertising in 22 product categories [20]. They found that resellers/dealers engaged in more price advertising than manufacturers.

Preferences for price forms also vary across types of advertisers. For example, price search websites and travel agents prefer to use percentage discount and relative price words instead of actual price numbers. Given that price search websites and travel agents often offer products/services from multiple service providers and/or other product categories such as flights and car rental, it is less effective for them to advertise actual price numbers. However, service providers like to use actual price numbers or a percentage discount.

3) Quality: As shown in Table V, the amount of price information is 8.3% higher for advertisers offering low quality accommodation (e.g., hostels) than for those offering high quality accommodation (e.g., hotels). Given that low quality firms have a price advantage while high quality firms have an advantage on product features, it is unsurprising that low quality firms advertise more on price than high quality firms.

However, these advertisers do have different preferences on the price forms. Since the actual prices of the advertisers offering low quality accommodation are already very low, advertising actual price numbers will attract consumer's attention. However, the advertisers offering high quality accommodation prefer to use a percentage discount, because their prices are higher than those of low quality firms, even after discount.

4) Advertisers of Cross-Category Selling: As shown in Table V, advertisers that also sell other products on average use 2.7% more price information than those selling only accommodation. Firms selling products across multiple categories have the cost advantage and gain more value from an additional customer acquisition. Therefore, they are more likely to use price information. Meanwhile, their price form is mainly dominated by the use of relative price words rather than actual price numbers or percentage discount. Price levels and discount percentages are usually different across product categories. For example, consumers may perceive 20% off as a good deal for car rental but not for accommodation. Therefore, in a short search advertisement, advertisers who sell multiple products cannot convey their price advantages through numbers. Thus, relative price words are a more appropriate form for advertisers to be consistent across multiple product categories.

C. Advertisement Location and Position

As shown in Table V, ads on the right-hand side have 2% more price information in their content than ads underneath the search bar do, mainly dominated by the uses of actual price numbers and relative price words. Even though the total amount of price information is lower for ads underneath the search bar, these ads are more likely to contain a percentage discount in their content. One possible reason is that ads underneath the search bar are usually from advertisers with high website traffic. Thus, more consumers already have prior knowledge and experience with them, and may be motivated to purchase from them by a price discount. Among the ads on the right-hand side, the ones at the top positions contain more price information than those at the bottom positions. The advertisers of the top position ads usually pay higher prices for their positions.

Given that price information is the information considered by most consumers in their decision-making process, it will be a safer strategy for these advertisers to put a little bit more price information in their content.

VI. CONCLUSIONS

Search engines have been changing consumers' shopping behaviors and decision-making processes. It becomes very important for advertisers to create effective content in their search advertisements to attract more attention from consumers. This paper empirically investigates the relationships between ad content and characteristics of queries and advertisers. By examining these relationships, we shed light on what ad content works best for different queries and advertisers. Our results suggest that advertisers should customize their ad content, particularly the price information, in accordance with their own characteristics and the characteristics of the queries they are bidding on.

Our findings can help search engine practitioners to develop tools for novice advertisers to customize their ad content based on their own characteristics, and also help advertisers to analyze industry competition structure by investigating similarity and difference among their ad content.

As future work, we plan to investigate some predictive models, such as SVM classifiers and naive Bayes classifiers, for search ads composition, based on the query, advertising content and advertiser characteristics we discussed in this paper.

REFERENCES

[1] Iab internet advertising revenue report, 2004 full year results. *Interactive Advertising Bureau*, 2004.
[2] Iab internet advertising revenue report, 2010 full year results. *Interactive Advertising Bureau*, 2010.
[3] G. Aggarwal, *et al*. Bidding to the top: Vcg and equilibria of position-based auctions. In *WAOA*, 2006.
[4] E. T. Anderson and D. I. Simester. *Handbook of pricing research in marketing*, 2009.
[5] P. J. Auter and R. L. Moore. Buying from a friend: A content analysis of two teleshopping programs. *Journalism Mass Comm. Q.*, 1993.
[6] S. Banerjee, *et al*. Shades of green: A multidimensional analysis of environmental advertising. *J. Advertising*, 1995.
[7] J. R. Bettman and C. W. Park. Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *J. Cons. Res.*, 1980.
[8] B. B.W. and C. P.E. The influence of personal values on attitude and store choice behavior. In *EMAC*, 1982.
[9] L. Chiou and C. Tucker. How does the use of trademarks by third-party sellers affect online search? *Market. Sci.*, 2012.
[10] L. E. Crow, *et al*. Industrial buyers' choice strategies: A protocol analysis. *J. Marketing Res.*, 1980.
[11] DoubleClick. Search before the purchase. http://goo.gl/l73dI, 2005.
[12] H. J. Einhorn. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organ. Behav. Hum. Perf.*, 1971.
[13] J. Feng, *et al*. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS J. on Computing*, 2007.
[14] A. Gabor and C. W. J. Granger. On the price consciousness of consumers. *J. R. Stat. Soc.*, 1961.
[15] A. Ghose and S. Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Manage. Sci.*, 2009.
[16] Google. https://adwords.google.com, 2012.
[17] Google AdWords Help. Understanding keyword tool columns. http://goo.gl/rchmu, 2012.
[18] G. Iyer and A. Pazgal. Internet shopping agents: Virtual co-location and competition. *Market. Sci.*, 2003.
[19] H. H. Kassarjian. Content analysis in consumer research. *J. Cons. Res.*, 1977.
[20] A. Kaul and D. R. Wittink. Empirical generalizations about the impact of advertising on price sensitivity and price. *Market. Sci.*, 1995.
[21] T. Mazumdar and K. B. Monroe. The effects of buyers' intentions to learn price information on price encoding. *J. Retailing*, 1990.
[22] K. B. Monroe. Buyers' subjective perceptions of price. *J. Marketing Res.*, 1973.
[23] K. B. Monroe. The influence of price differences and brand familiarity on brand preferences. *J. Cons. Res.*, 1976.
[24] S. E. Moriarty. A content analysis of visuals used in print media advertising. *Journalism Q.*, 1987.
[25] G. N. Punj and D. W. Stewart. An interaction framework of consumer decision making. *J. Cons. Res.*, 1983.
[26] A. R. Rao and K. B. Monroe. The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *J. Marketing Res.*, 1989.
[27] A. R. Rao and W. A. Sieben. The effect of prior knowledge on price acceptability and the type of information examined. *J. Cons. Res.*, 1992.
[28] A. Resnik and B. L. Stern. An analysis of information content in television advertising. *J. Marketing*, 1977.
[29] C. Rheem. Empowering inspiration: The future of travel research. PhoCusWright, Sherman, CT, 2012.
[30] J. S. Rju, *et al*. The effects of brand loyalty on competitive price promotional strategies. *Manage. Sci.*, 1990.
[31] O. J. Rutz and M. Trusov. Zooming in on paid search ads—a consumer-level model calibrated on aggregated data. *Market. Sci.*, 2011.
[32] O. J. Rutz, *et al*. Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Market. Sci.*, 2011.
[33] Stata. http://www.stata.com/.
[34] G. J. Tellis. Beyond the many faces of price: An integration of pricing strategies. *J. Marketing*, 1986.
[35] L. W. Turley and S. W. Kelley. A comparison of advertising content: Business to business versus consumer services. *J. Advertising*, 1997.
[36] M. Viswanathan and T. L. Childers. '5' calories or 'low' calories? what do we know about using numbers or words to describe products and where do we go from here? *Adv. Consum. Res.*, 1997.
[37] L. Xu, *et al*. Price advertising by multiple channel members. 2011.
[38] S. Yao and C. F. Mela. A dynamic model of sponsored search advertising. *Market. Sci.*, 2011.