MARKETING SCIENCE
CELEBRATING 30 YEARS
Vol. 30, No. 5, September–October 2011, pp. 789–800
ISSN 0732-2399 | EISSN 1526-548X | 11 | 3005 | 0789

informs®
http://dx.doi.org/10.1287/mksc.1110.0647

Zooming In on Paid Search Ads—A Consumer-Level Model Calibrated on Aggregated Data

Oliver J. Rutz
Foster School of Business, University of Washington, Seattle, Washington 98195, rutz@uw.edu

Michael Trusov
Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742, mtrusov@umd.edu

We develop a two-stage consumer-level model of paid search advertising response based on standard aggregated data provided to advertisers by major search engines such as Google or Bing. The proposed model uses behavioral primitives in accord with utility maximization and allows recovering parameters of the heterogeneity distribution in consumer preferences. The model is estimated on a novel paid search data set that includes information on the ad copy. To that end, we develop an original framework to analyze composition and design attributes of paid search ads. Our results allow us to correctly evaluate the effects of specific ad properties on ad performance, taking consumer heterogeneity into account. Another benefit of our approach is allowing recovery of preference correlation across the click-through and conversion stage. Based on the estimated correlation between price- and position-sensitivity, we propose a novel contextual targeting scheme in which a coupon is offered to a consumer depending on the position in which the paid search ad was displayed. Our analysis shows that total revenues from conversion can be increased using this targeting scheme while keeping cost constant.

Key words: Internet; paid search advertising; aggregate data; choice modeling; Bayesian methods
History: Received: January 25, 2010; accepted: March 1, 2011; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Pradeep Chintagunta served as associate editor for this article. Published online in *Articles in Advance* June 20, 2011.

1. Introduction

Paid search advertising—or simply, paid search—is the leading customer acquisition tool of Internet marketers. In 2009, paid search accounted for roughly 47% of the $22.7 billion spent on Internet advertising, about double that of Internet display advertising (PricewaterhouseCoopers 2010). The basic idea of paid search is as simple as it is intriguing: consumers can be directly addressed during their electronic search for products or services. In paid search, companies select specific keywords and create text ads that the search engine serves when a consumer searches for these keywords. A typical paid search ad[1] is composed of three elements: a headline, the main body text, and a display URL.[2] After designing ads and pairing them with keywords, companies bid their maximum willingness to pay for clicks on these ads. An automated auction-type algorithm then determines the position of the ad in the sponsored listings section of the results page. If consumers click on the ad, they are taken to the company's website (landing page), where an array of traditional marketing instruments (e.g., price or promotion) can be used to lure consumers into purchase. Consequently, multiple interrelated decision variables are used to optimize the performance of paid search campaigns,[3] including choice of keywords, target position on the results page, maximum bid amount, textual content, and layout of the ad and landing-page design.

The purpose of this paper is to develop an empirical model that will assist paid search practitioners in making some of these decisions. In particular, we are interested in evaluating the effects of *ad position* within the search results page and *textual properties* of the ad on consumers' actions (i.e., click-through and purchase). Whereas the latter is a research question that, to the best of our knowledge, has not been addressed in the marketing literature, the former has

[1] For the convenience of the reader, we will refer to "ad" instead of "paid search ad" for the remainder of the paper.

[2] For some examples, please see the appendix "Designing Effective Ringtone Ads" in the electronic companion. (An electronic companion to this paper is available as part of the online version that can be found at http://mktsci.pubs.informs.org/.)

[3] In paid search, sets of keywords are generally referred to as campaigns.

received a notable amount of attention. Nevertheless, we argue that the existing research has some important shortcomings, and hence, additional inquiry into this question is warranted.

One of the key challenges faced by empirical studies of paid search advertising lies in the nature of widely available paid search data. The data Google and other major search engines provide to their advertisers are aggregated on a keyword/ad level. For each keyword/ad pair, the search engine provides advertisers with summary information on the number of impressions and clicks as well as average position and average cost per click (CPC) calculated over some period of time. In the marketing literature, a number of approaches were proposed to model the standard Google-type aggregate paid search data (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Rutz et al. 2011). These models explore the differences and similarities across keywords and, from a practical perspective, are useful in forecasting the performance of individual keywords. We will refer to these models as *keyword-centric*.

One notable limitation of these models is that they assume that online shoppers who use a certain keyword are *homogeneous* in their preferences and responses to marketing instruments of paid search. From our perspective, this is an ad hoc assumption akin to segmenting consumers a priori. Although segmentation is one approach to account for unobserved consumer heterogeneity, segmenting consumers should be an integral part of the model (e.g., Kamakura and Russell 1989), which is not the case in current keyword-centric approaches. As a result, the interpretation of paid search covariates, e.g., ad position, implies that all consumers using a certain keyword have the same response to the covariate in question. Whether this is a valid assumption is an empirical question and cannot be answered using a keyword-centric approach. It is not clear whether findings from keyword-centric models—for example, the effect of position—are a true representation of consumers' response to position or an artifact of the homogeneity assumption. In our view the online shopping process—e.g., product search, response to advertising, and purchase—is naturally described as a sequence of *consumer* decisions, and hence it is sensible to model it accordingly. In a consumer-level approach (we refer to this as *consumer-centric*), the decision process can be captured using intuitively appealing economic primitives (e.g., based on consumer utility maximization). Moreover, the consumer-level model may offer richer insights on the distribution of preferences among online shoppers not restricted by an a priori keyword-based segmentation. To uncover heterogeneous preferences from aggregated data, we adopt the Bayesian framework developed by Musalem et al. (2008, 2009) and extend it to a two-stage consumer-level model in which the outcome of the second-stage decision (conversion) is conditional on the outcome of the first-stage decision (click). The proposed approach allows us to correctly specify heterogeneity on a consumer level and alleviates concerns stemming from the treatment of unobserved heterogeneity in keyword-centric models.

One of the points of difficulty in paid search research is the treatment of the text ad's position stemming from endogeneity concerns (e.g., Ghose and Yang 2009, Yang and Ghose 2010). One way to alleviate endogeneity concerns would be to explicitly model the underlying auction. Indeed, some researchers have been successful in addressing the problem by leveraging bidding history information in a unique data set from a specialized search engine in the software space (Yao and Mela 2011). Yet given the current information-sharing policies of the major U.S. search engines (i.e., Google and Bing), it is highly unlikely that competitive bid data will be seen any time soon. Alternatively, an instrumental variables (IV) approach can be used to account for position endogeneity. However, an IV approach requires the availability of suitable instruments—a contentious issue in any IV application. We circumvent the need to find suitable instruments by extending the latent instrumental variable framework proposed by Ebbes et al. (2005).

This paper contributes to marketing research in several ways. Whereas most industry practitioners would acknowledge that the textual properties of the paid search ad play an important role in driving consumers' responses to the ad in the click-through decision, previous paid search studies have neglected to explore this aspect of paid search and focused solely on keyword properties. From our perspective, quantification of these effects has direct implications for ad design. Our first contribution is to fill this gap in the literature. Moreover, drawing from trade publications and academic literature on classified advertising,[4] our study is the first of its kind, to the best of our knowledge, to propose a theoretical justification for the effects of different design attributes on paid search ad performance. Our third contribution is the extension of the Musalem et al. (2008, 2009) framework to a two-stage consumer-level model of click-through and conversion that is based on the economic primitives to explicitly account for consumer preferences while also accounting for differences across keywords. As another methodological contribution, we extend the LIV framework to choice models to account for position endogeneity. Finally, on the substantive end, we find that consumers' preferences with regard to

[4] We thank an anonymous reviewer for this helpful suggestion.

response to ad position and price are correlated. In the case of the collaborating firm, consumers who are more likely to click on the ad when the ad appears in one of the top positions tend to be more price sensitive. This empirical finding is an interesting one because it presents a new opportunity for contextual targeting. We leverage it by proposing a novel (price) promotion tied to the position in which the ad was shown.

The rest of this paper is structured as follows. First, we offer a brief overview of the current state of research on paid search ad design. We then present our model, data set, and results. We conclude with a discussion of the implications of our findings for paid search practitioners.

2. Ad Design in Paid Search

Designing an effective ad is perhaps one of the hottest topics among paid search practitioners and is extensively discussed in numerous online forums. The underlying theme of almost all of these discussions is that the performance (i.e., traffic generation) of the individual ad is largely determined by how it is designed. Numerous design recommendations are offered by both industry gurus and search engines. For example, Google AdWords recommends "keep[ing] ad content simple" and focusing on unique features, including information on prices and promotions, using a "strong call-to-action" and including keywords in the ad text. Although there are a number of trade publications and online sources that offer advice on how to design effective paid search ads, academic research on this topic is quite scarce. First, most of the empirical studies in the academic literature focus on the characteristics of key phrases but not the ads (e.g., Ghose and Yang 2009). Second, the paid search ad format is still a very recent invention for the well-developed field of advertising design, and perhaps, academics focusing on linguistics and advertising simply have not caught up with it yet. Conceivably, the closest (but still quite distinct) type of advertising with some academic research is classified advertising (e.g., Bruthiaux 1996).[5] Even for classified advertising, however, there are very few published empirical studies, which is probably due to the challenges associated with collecting ad performance data (Bruthiaux 2000).

As a result, paid search advertising opens up new opportunities for empirical analysis by offering almost unlimited samples of ad designs paired with instant field performance data. However, there is little foundation on theoretical grounds for deriving hypotheses with regard to ad performance as yet. To fill this gap in the academic literature, we develop a basic framework on paid search ad design, largely borrowing from the sociolinguistics literature, trade publications, and empirical studies across different academic disciplines. Our approach can be summarized as follows. We are interested in identifying design elements that have an effect on consumers' response to the ad. Our first step is to create a list of features that can be used to characterize a paid search ad. This list must be complete to the extent that any randomly selected ad targeted to a specific consumer segment can be described by the features in this list. Our second step is to form a set of theoretically sound hypotheses on how these features may influence the consumers' response to the ad. Finally, using the proposed model, we test whether the hypothesized effects are supported by our empirical data. Because of space constraints, we offer a thorough discussion of the first two steps in the appendix "Designing Effective Ringtone Ads" in the electronic companion. We report results pertaining to our hypotheses in the results section and in the electronic companion.

3. Model

3.1. Motivation

The key premise of our modeling approach is that response to paid search advertising is inherently a consumer-level decision and, hence, can vary across consumers. The process can be viewed as a sequence of two choices. First, a consumer decides to click on the ad depending on whether the ad seems appealing or not. Second, conditional on the first decision and depending on the offer attractiveness, the consumer makes a purchasing (conversion) decision. Possible correlation between consumers' preferences across click and conversion needs to be taken into consideration to control for self-selection bias (e.g., if position-sensitive consumers are also more price sensitive, then ignoring the selection will lead to an attenuated estimate of price elasticity). Also, these correlations (if found) may be leveraged in managerial applications.

In this section we introduce a consumer-level two-stage model meant to capture the above-mentioned process. Aggregation of consumer-level decisions is observed in the form of daily click and conversion summary statistics provided by search engines to paid search advertisers. We show how the distribution of consumer preferences can be inferred from these data.

3.2. Click-Through Model

We model the utility of clicking for person i using keyword w at time t based on observable covariates such as ad position, characteristics of the keyword, ad content characteristics, and search environment characteristics. The utility of clicking u_{iwt}^{cl} is given by

$$u_{iwt}^{\mathrm{cl}} = \beta_i^{\mathrm{cl}} x_{wt}^{\mathrm{cl}} + \xi_{wt}^{\mathrm{cl}} + \varepsilon_{iwt}^{\mathrm{cl}}, \tag{1}$$

[5] We thank an anonymous reviewer for this suggestion.

where β_i^{cl} are parameters to be estimated; x_{wt}^{cl} are observable keyword-specific covariates, including an intercept; and $\varepsilon_{iwt}^{\mathrm{cl}}$ is distributed extreme value.[6] Although some rudimentary information on competition is provided by Google through search environment characteristics, details on the dynamic competitive landscape are not available. Given that these time-varying factors may affect consumer utility, we include a zero-centered and normally distributed time-varying keyword-specific demand shock, $\xi_{wt}^{\mathrm{cl}} \sim \mathrm{N}(0, v_{\mathrm{cl}}^2)$, in the model.

3.3. Conversion Model

We model the conversion decision similarly to the clicking decision (for details on predictors x_{wt}^{con}, see empirical application in §4), and the utility of conversion is given by

$$u_{iwt}^{\mathrm{con}} = \beta_i^{\mathrm{con}} x_{wt}^{\mathrm{con}} + \xi_{wt}^{\mathrm{con}} + \varepsilon_{iwt}^{\mathrm{con}}, \tag{2}$$

where β_i^{con} are the parameters to be estimated; x_{wt}^{con} are observable keyword-specific covariates, including an intercept; $\xi_{wt}^{\mathrm{con}} \sim \mathrm{N}(0, v_{\mathrm{con}}^2)$ is a time-varying demand shock; and $\varepsilon_{iwt}^{\mathrm{con}}$ is the distributed extreme value.

3.4. An Integrated Model of Click-Through and Conversion

To integrate the click-through and conversion decisions, we model a full covariance structure for parameters β_i^{cl} and β_i^{con}. We should note that for the conversion stage we only observe consumers who clicked on the ad. To be able to correctly estimate the covariance, we augment the parameter vectors for the nonclickers assuming their behavior is governed by the same correlation structure as the remainder of consumers. This addresses selection as well as the econometric problems created by different sample sizes across choices (please see the electronic companion for details). With the augmented values, we set up our model as follows:

$$\begin{pmatrix} u_{iwt}^{\mathrm{cl}} \\ u_{iwt}^{\mathrm{con}} \end{pmatrix} = \begin{pmatrix} x_{wt}^{\mathrm{cl}} & 0 \\ 0 & x_{wt}^{\mathrm{con}} \end{pmatrix} \begin{pmatrix} \beta_i^{\mathrm{cl}} \\ \beta_i^{\mathrm{con}} \end{pmatrix} + \begin{pmatrix} \xi_{wt}^{\mathrm{cl}} \\ \xi_{wt}^{\mathrm{con}} \end{pmatrix} + \begin{pmatrix} \varepsilon_{iwt}^{\mathrm{cl}} \\ \varepsilon_{iwt}^{\mathrm{con}} \end{pmatrix}, \tag{3}$$

where

$$\begin{pmatrix} \beta_i^{\mathrm{cl}} \\ \beta_i^{\mathrm{con}} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} b^{\mathrm{cl}} \\ b^{\mathrm{con}} \end{pmatrix}, \begin{pmatrix} \Omega_{\mathrm{cl,\,cl}} & \Omega_{\mathrm{cl,\,con}} \\ \Omega_{\mathrm{cl,\,con}} & \Omega_{\mathrm{con,\,con}} \end{pmatrix} \right] \quad \text{and}$$

$$\begin{pmatrix} \xi_{wt}^{\mathrm{cl}} \\ \xi_{wt}^{\mathrm{con}} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{\mathrm{cl}}^2 & 0 \\ 0 & v_{\mathrm{con}}^2 \end{pmatrix} \right].$$

For model identification reasons, we assume independently distributed errors $\varepsilon_{iwt}^{\mathrm{cl}}$ and $\varepsilon_{iwt}^{\mathrm{con}}$ and orthogonality in demand shocks.[7]

3.5. Likelihood Function

The probability $P(\cdot)$ of clicking (converting) based on the assumption of extreme value errors $\varepsilon_{iwt}^{\mathrm{cl}}(\varepsilon_{iwt}^{\mathrm{con}})$ is given by the following equation:

$$P(u_{iwt}^{\cdots}) = \frac{\exp(\beta_i^{\cdots} x_{wt}^{\cdots} + \xi_{wt}^{\cdots})}{1 + \exp(\beta_i^{\cdots} x_{wt}^{\cdots} + \xi_{wt}^{\cdots})}. \tag{4}$$

We define the latent indicator z_{iwt}^{cl} that is equal to 1 if "augmented" consumer i clicks after searching with keyword w at time t, and 0 otherwise (z_{iwt}^{con} is defined correspondingly). We next define $N_{wt}^{\mathrm{cl}}(N_{wt}^{\mathrm{con}})$ as the observed number of clicks (conversions) for keyword w at time t, and N_{wt}^{imp} represents the number of impressions (searches). Note that the assignment of indices to consumers is arbitrary in this case, and, without loss of generality, we assign the first $N_{wt}^{\mathrm{cl}}(N_{wt}^{\mathrm{con}})$ indices out of all $N_{wt}^{\mathrm{imp}}(N_{wt}^{\mathrm{cl}})$ indices to consumers who click (convert after clicking) on keyword w at time t.[8] We treat the unobserved individual choices z_{iwt}^{cl} and z_{iwt}^{con} as parameters to be simulated from their posterior distributions (Musalem et al. 2009). The augmented likelihood of our integrated click-through and conversion model is given by the following equation:

$$L = \left[\begin{array}{l} \left[\displaystyle\prod_{t=1}^{T} \prod_{w=1}^{W} \prod_{i=1}^{N_{wt}^{\mathrm{con}}} P(u_{iwt}^{\mathrm{cl}})^{z_{it}^{\mathrm{cl}}} P(u_{iwt}^{\mathrm{con}})^{z_{it}^{\mathrm{con}}} \right] \\ \quad \times \left[\displaystyle\prod_{t=1}^{T} \prod_{w=1}^{W} \prod_{i=N_{wt}^{\mathrm{con}}+1}^{N_{wt}^{\mathrm{cl}}} P(u_{iwt}^{\mathrm{cl}})^{z_{it}^{\mathrm{cl}}} (1 - P(u_{iwt}^{\mathrm{con}}))^{(1-z_{it}^{\mathrm{con}})} \right] \\ \quad \times \left[\displaystyle\prod_{t=1}^{T} \prod_{w=1}^{W} \prod_{i=N_{wt}^{cl}+1}^{N_{wt}^{\mathrm{imp}}} (1 - P(u_{iwt}^{cl}))^{(1-z_{it}^{\mathrm{cl}})} \right] \end{array} \right] \cdot I_{[(Z^{\mathrm{cl}},\, Z^{\mathrm{con}}) \in S]}, \tag{5}$$

where

$$S = \left[(Z^{\mathrm{cl}}, Z^{\mathrm{con}})\colon \sum_{i=1}^{N_{wt}^{\mathrm{imp}}} z_{iwt}^{\mathrm{cl}} = N_{wt}^{\mathrm{cl}}, \sum_{i=1}^{N_{wt}^{\mathrm{cl}}} z_{iwt}^{\mathrm{con}} = N_{wt}^{\mathrm{con}} \right]. \tag{6}$$

For information on priors and details on estimation, please see the appendix "Estimation" in the electronic companion.

[6] The proposed setup models click-through and conversion *across* a set of keywords. As was pointed out by the associate editor, residuals could be correlated across keywords. We tested for this in our data and found no evidence of correlation (details on the tests are available from the authors on request). However, in case correlation is a valid concern, a full covariance structure across keywords may need to be modeled. For a relatively small number of keywords, this can be done directly. However, as the data dimension increases, one may want to consider restricting the correlation structure, e.g., using copulas (Danaher and Smith 2011).

[7] Model identification discussion and simulation analysis are provided in the appendix titled "On Identification" in the electronic companion.

[8] Note that these indices remain fixed at all iterations of the Gibbs sampler. This alleviates concerns with regard to label switching (Musalem et al. 2009).

3.6. Addressing Position Endogeneity—A Latent Instrumental Variable Approach

In paid search, endogeneity concerns with respect to position loom large (e.g., Ghose and Yang 2009). First, position is a firm decision variable similar to price and therefore can be set based on the expected performance of the ad (simultaneity). Second, position is the outcome of an auction and is thus influenced by competition, which is not observed by the focal firm (omitted variables). Third, position is only reported as a daily average (errors-in-variables). Without the ability to model the auction as a result of nonavailability of competitive data, an IV approach can be used to account for endogeneity, assuming *suitable* instruments are available. We define the IV equation as follows:

$$\mathrm{pos}_{wt} = \phi_{wt} x^{\mathrm{IV}}_{wt} + \zeta^{\mathrm{IV}}_{wt}, \tag{7}$$

where pos_{wt} is the position of keyword w at time t, x^{IV}_{wt} are keyword-specific instruments, ϕ_{wt} are the parameters to be estimated, and ζ^{IV}_{wt} is an error term.

Following Yang et al. (2003), we allow for correlation between the error term ζ^{IV}_{wt} and the click-through demand shock ξ^{cl}_{wt}:

$$\begin{pmatrix} \xi^{\mathrm{cl}}_{wt} \\ \zeta^{\mathrm{IV}}_{wt} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \upsilon^2_{\mathrm{cl}} & \upsilon_{cl,\,\mathrm{IV}} \\ \upsilon_{cl,\,\mathrm{IV}} & \upsilon^2_{\mathrm{IV}} \end{pmatrix} \right]. \tag{8}$$

A caveat of the IV approach is the availability of suitable instruments, i.e., to find variables that are correlated with the endogenous covariate as well as with the dependent variable but not with the error term in the model. The most obvious candidate—lagged position—reflects the same strategic decision/unobserved competitive landscape that creates endogeneity concerns to begin with making it a potentially invalid instrument.

A recently developed method, the so-called latent instrumental variables (LIV) (Ebbes et al. 2005, 2009), alleviates the need to find suitable instruments. In the LIV approach, a latent variable model is used to account for dependencies between the endogenous covariate and the error by introducing unobserved discrete binary variables. These latent variables are used to decompose the endogenous covariate into a systematic part that is uncorrelated with the error and one that is possibly correlated with the error. This allows for an unbiased estimation of the effect of an endogenous covariate, such as position, on the desired action, such as click-through.[9] Originally developed in a regression setting, the LIV framework is extended to a choice model in which the endogenous covariate is observed only on an aggregate level at time t. We define the LIV equation for position based on a given number of C binary latent variables as follows:

$$\mathrm{pos}^{\mathrm{cl}}_{wt} = \omega \gamma^{c}_{wt} + \zeta^{\mathrm{LIV}}_{wt}, \tag{9}$$

where γ^c_{wt} is a $(C \times 1)$ binary vector of $C-1$ zeros with a nonzero element indicating that keyword w belongs to category c at time t, ω is a $(1 \times C)$ vector of category weights to be estimated, and $\zeta^{\mathrm{LIV}}_{wt}$ is the LIV error term.

Following the treatment of instruments as specified in (7) and (8), we link the LIV error $\zeta^{\mathrm{LIV}}_{wt}$ with the error of the demand shock ξ^{cl}_{wt} as follows:

$$\begin{pmatrix} \xi^{\mathrm{cl}}_{wt} \\ \xi^{\mathrm{con}}_{wt} \\ \zeta^{\mathrm{LIV}}_{wt} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \upsilon^2_{\mathrm{cl}} & 0 & \rho_{\mathrm{cl,\,LIV}} \\ 0 & \upsilon^2_{\mathrm{con}} & 0 \\ \rho_{\mathrm{cl,\,LIV}} & 0 & \upsilon^2_{\mathrm{LIV}} \end{pmatrix} \right], \tag{10}$$

where $\rho_{\mathrm{cl,\,LIV}}$ is the covariance of ξ^{cl}_{wt} and $\zeta^{\mathrm{LIV}}_{wt}$, and $\upsilon^2_{\mathrm{LIV}}$ is the variance of $\zeta^{\mathrm{LIV}}_{wt}$.

For details on estimation, please see the appendix "Estimation" in the electronic companion.

4. Empirical Example

4.1. Data

We use a novel data set from the ringtone industry to test our model. Our data contain Google AdWords information on the top 80 keywords over 20 days in 2007. These keywords represent the company's top keywords for the time period, alleviating concerns with regard to sparse data as encountered by Rutz et al. (2011). We focus on the top keywords for business reasons as well as data reasons. Based on our own experience working with multiple companies across industries and categories on their paid search campaigns, we find that most firms spend most of their budget on a relatively small number of keywords. Understanding how these top keywords perform is of critical importance in managing a successful paid search campaign. From a data perspective, non-top keywords generally display very low search volume coupled with low click-through and conversion rates, resulting in very sparse data. Our model is set up to use shrinkage across keywords and can deal with small click-through and conversion rates. However, model performance in terms of ability to recover the heterogeneity structure is deteriorating when using very sparse data, i.e., for keywords not

[9] A shortcoming of the LIV framework is that what-if analyses are not straightforward as they would be in the case of observed instruments. Potentially, a what-if analysis could be based on draws of the latent IV from its (empirical) posterior distribution, which is similar to drawing an augmented variable such as goodwill based on the empirical distribution in forecasting. However, such an approach is likely to be inferior to observed instruments, which, unfortunately, are hard to find (if at all) in paid search data.

at the top (please see the electronic companion for details).

The data include the typical information on the keyword, daily number of impressions, clicks, conversions, average ad positions, and average cost. For all keywords, the collaborating firm used an "exact match" option to match a search query with a keyword. Over the observation period, on average, we observe 3,555 impressions (searches), 487 clicks, and 29 conversions per day. The average click-through rate (CTR) is 13.7%, the average conversion rate is 5.9%, and the average CPC is \$0.67. Based on these performance and cost metrics, the average cost per conversion is \$11.40 for our data. In our case, conversion represents consumers signing up for a three-month contract with the ringtone provider, allowing for the download of 10 ringtones per months. The data also contain the actual subscription price on a daily basis. The company was running frequent promotions, so we observe variations in subscription price across time (variance, \$0.71). Although the company wishes that we do not reveal the exact price, we are at liberty to mention that the monthly subscription cost is around \$10. The level of aggregation with regard to price is similar to the situation in which we observe aggregate store data: for example, average prices instead of individual-level prices (Yang et al. 2003). Additionally, the data set includes new information not previously used in marketing research—namely, measures for the level of competition and the level of search volume, which were collected using Google AdWords API services.[10] The level of competition is an ordinal variable from 0 to 5, where 0 is the lowest level. The level of search volume represents the number of search queries on Google.com matching each keyword; again, 0 is the lowest level. The information on the search volume can be used to account for the relative popularity of the keyword among consumers. This metric allows us to include comparative information and investigate whether the response differs across keywords with different levels of popularity. In our data set, both measures are static over the period of the data and thus might capture important static aspects of competition with respect to other paid search ads and organic results, but they do not allow accounting for dynamics (which we do by using time-varying keyword-specific demand shocks).

Additionally, we have information on the whole ad; namely, we have the *headline*, for example, "Stealth Mosquito Ringtone," and the *body* of the ad, "Yeah, the one the adults can't hear. Tones in 30 seconds." We have developed a general framework to generate a set of predictor variables that capture different aspects of ad content and design, such as creating interest or making an appealing offer (see Table 1 for an example and the appendix "Designing Effective Ringtone Ads" in the electronic companion for details on the procedure). Over our observation period, there is no variation in the URL contained in the ad, and all ads feature the same URL. The same is true for the landing page. Thus, we do not use the URL and landing page in our model as possible predictors.

Table 1 Ad Predictor Variables—Examples

	Example
Keyword	Supersonic ringtone
Headline	Stealth Supersonic Ringtone
Line 1	Tones to your phone—Get it now!
Line 2	Yeah, the one the adults can't hear
Keyword in headline	Yes—code as "1"
Keyword in body	No—code as "0"
Call to action[a]	Yes: "Get it now"—code as "1"
Keyword word count	2
Headline word count	3
Body word count	14
Flesch reading ease score	0.941

[a]We use two independent coders.

The ringtone industry provides us with an excellent opportunity to estimate the effects of position and other search-related covariates on consumer response. In essence, all consumers buy the same product—access to a subscription service that allows for a certain number of downloads per month. Compared with most other product categories, the ringtone industry does not provide much differentiation in terms of products or pricing. This allows us to get a "cleaner" estimate of the effects of covariates on the click and conversion decisions of consumers compared with previous studies. For example, Rutz et al. (2011; hotel room reservations in multiple geographic locations), Ghose and Yang (2009) and Yang and Ghose (2010) (wide range of products from an online retailer) face the problem of having a number of attributes (unobservable to the researcher) that could influence the purchasing decision and, hence, could be misattributed as keyword effects.

4.2. Model Selection

We compare our model to a set of alternative models that we designed along the lines of the key issues that we deem important when it comes to modeling paid search data. First, we propose that although it is important to account for differences across keywords, unobserved consumer heterogeneity should

[10] It is our understanding that Google recalibrates these measures using a sliding window mechanism. However, when we repeated our API queries over some extended period of time, we did not find any changes in the data. Hence, competitive measures enter our model as time-invariant covariates and provide a baseline for differences in competition and volume across keywords.

Table 2 Model Comparison

	Individual level	Integrated	Model fit: Log marginal density	Model fit: Log Bayes factor[a]
Full model	✓	✓	−8,090	—
Model 1[b]	×	✓	−8,912	822
Model 2[c]	✓	×	−8,256	166

[a]In relation to the best model, i.e., the full model.
[b]Model 1 is a keyword-centric model in the spirit of Ghose and Yang (2009).
[c]Model 2 treats two decisions as not connected and does not account for selection.

go beyond an a priori segmentation imposed by keyword use. Thus, we compare our consumer-centric model with a keyword-centric aggregate model. In this model, keywords are used as proxies for unobserved consumer heterogeneity. We find that our proposed model strongly outperforms the keyword-centric model (Bayes factor of 822; see Table 2). In addition to the superior fit, our proposed model allows for correct interpretation of the estimates. Take, for example, position: our model captures response to position from a consumer standpoint. In contrast, in a keyword-centric model, different keywords "react" differently to positions, which intuitively is not very appealing.

Second, click-through and conversion should be modeled in an integrated framework that controls for selection bias in the conversion stage. An alternative view could be independence between click-through and conversion decisions. In this case, we model whether a consumer clicks or not and, conditional on click-through, use an independent model of conversion. In this setup, there is no correlation in consumer preferences across the two decisions and no correction for selection bias. We find that our integrated model of click-through and conversion fits better (Bayes factor of 166; see Table 2) compared with independent models.

4.3. Model Estimates

We start this section with a short description of our covariates. As we argue above, we use two different sets of predictors for each decision stage (click-through and conversion).

4.3.1. Click-Through Stage Predictors. At the time of the click-through decision, the consumer observes the ad's text as well as the ranking provided by the search engine that can be seen as a proxy for how well the ad and, more importantly, the firm behind the ad, matches the consumer's search query. The ads are ranked between 1 ("top of the page") and 8 ("bottom of the page"). The *Competition* and the *Volume* metrics are observed to be between $[0, 5]$ and $[0, 2]$, respectively.[11] A lower number is indicative of less competition (volume). Note that the consumer does not observe these metrics. Instead, he or she actually sees competitive ads. For highly competitive keywords, it is likely that all ad slots are filled, whereas for some less popular keywords, this may not be the case. In terms of volume, high-volume keywords are typically more attractive in search engine optimization (SEO)[12] and therefore present stronger competition to paid ads in the form of "organic" search results. We use these two metrics as static proxies for the competitive landscape a consumer is exposed to (as discussed before, competitive information is not available from search engines). We also include two keyword-specific covariates. First, we generate a covariate measuring the breadth of the search—similar to Rutz and Bucklin's (2011) distinction between generic versus branded keywords. In our case, brands do not play a role, and we define a 0–1 covariate called *Broad* if the keyword includes more broad (non-Broad) information, e.g., "blues ringtone" ("AC/DC ringtone"). Second, in our data, keywords that include specific ringtones are either for songs or TV shows, so we define a 0–1 covariate called *TV show* if the keyword includes a TV show.

Next, we turn our attention to the ad copy. In our data, each keyword is linked with a unique ad copy. Although the actual keyword used in a search query may reveal a search objective, it is the information contained in the ad that is being evaluated by the consumer and ultimately drives choice. Therefore, we argue that ads are an important component of the decision. Leveraging information the ad provides, we propose a new set of measures to differentiate keyword/ad combinations.[13] We start with ad features that are designed to *catch the consumer's attention and create interest*. First, we consider low-level stimuli such as visual characteristics, which may play a role in attracting the consumer's initial attention to certain areas of the screen. In the domain of paid search text

[11] According to Google, *advertiser competition* measures the number of advertisers bidding on each keyword relative to all keywords across Google. This represents a general low-to-high quantitative guide to help determine how competitive ad placement is for a particular keyword. *Search volume* measures the approximate average monthly number of search queries matching each keyword. These statistics apply to searches performed on Google and the search network over a recent 12-month period.

[12] The basic idea is that high-volume keywords are very attractive, and hence, the major players' SEO strategy focuses on these keywords. Long-tail keywords, on the other hand, represent a feasible SEO strategy for smaller players. Thus, long-tail keywords compete less strongly with sponsored results than high-volume keywords (see, for example, http://www.seobook.com/why-it-makes-sense-target-longtail-keywords-first).

[13] Please see the appendix "Designing Effective Ringtone Ads" in the electronic companion for more details.

ads, we define visual characteristics as brightness of the ad as well as density of the text. If the keyword appears in the headline or the body text, it will appear in bold font—making the ad "brighter." We define *Keyword in headline* and *Keyword in body* as covariates to measure the visual impact of the ad in terms of brightness. We implement these measures as indicator variables and code the case in which the keyword appears as "1." The headline appears in a larger font; we use the log of *Headline word count* as an additional measure of brightness. The density of the ad is measure by the log of the *Body word count*. Next, we turn to high-level attributes that will generate attention conditional or whether low-level attributes have attracted the consumer's gaze. Contextual characteristics such as *Keyword in headline* defined above can also capture whether attention will be generated by providing a match between the search (i.e., keyword) and the search results (i.e., text ad).

The advertising literature suggests that although the main purpose of a headline is to capture attention and generate interest, the main function of the ad body is to *stimulate the consumer's desire* for the product and *to create real conviction in a product's superiority to competitors* (Vestergaard and Schroder 1985). By investigating thousands of ad copies, we find virtually no evidence of superiority claims in the ringtone space. It seems that beyond discount and promotion offerings, most of the advertisers follow Google's advice to stay very specific and list product/service/phone/media format features in hopes that their ad will be seen by a consumer with matching interests. As with the header, we expect that features matching a consumer's interest(s) revealed through a search query translate into a higher likelihood of perceiving the ad as relevant. Hence, we use *Keyword in body* as a proxy for the match. Additionally, we have calculated the Flesch reading ease score to represent the readability of the body of the ad. Note that a higher score indicates an ad that is easier to read. Finally, after having attracted the gaze, generated attention, and made a convincing offer, *getting action* is the final step in customer acquisition. In our data a "call to action" is often included in the ad; for example, "Get it now!"

We also explore whether consumers' expectations with regard to the product information affects the clicking decision. As a natural candidate for product information, we have selected product price, which in our data varies over time. This choice is typical for many existing studies with forward-looking consumers (e.g., Erdem et al. 2003). From a practical perspective, the assumption that consumers are shopping for ringtone plans over an extended period of time—and can learn price variation—cannot be ruled out until tested (it is not very likely, given a relatively low involvement product and contract duration which prevents frequent repeat purchases). To account for this, we have incorporated average price as well as price trend dummies calculated over a moving window in our click model. We did not find any empirical support for these effects in our data.

4.3.2. Conversion Stage Predictors. After click-through, the consumer is on the company website and can decide to purchase based on the product and its price. In addition, the keyword itself might be informative in predicting a purchase event. For example, a keyword could be a proxy for the stage of the consumer's search process. For example, consumers early on often use broad search terms (Hotchkiss 2006). In this early stage, the main goal is information search but not purchase. As a result, for broad search terms, the conversion rates are often found to be low. At a later stage of the search process, narrower terms are regularly used that include specific information—for example, brands. At this stage, consumers are willing to buy, and conversion rates are higher. We allow for this phenomenon by including keyword characteristics (*Broad* and *TV show* as defined previously) and keyword-specific demand shocks in the conversion model.

4.3.3. Estimation Results. First, we are going to discuss the estimates from the click-through model (see Table 3). Given that the proposed model allows for different effects for different consumers, in the following discussion we are referring to the "mean" effects. As expected, the intercept is negative, with

Table 3 Parameter Estimates for Ringtones

		Estimate	
Decision	Variable	Mean	95% coverage interval
Click-through	*Intercept*	**−4.13**	**(−4.35, −3.97)**
	Broad	**−0.48**	**(−0.71, −0.17)**
	TV show	**1.59**	**(1.31, 1.91)**
	Competition	**−0.49**	**(−0.67, −0.34)**
	Volume	**−1.49**	**(−1.71, −1.09)**
	Keyword in headline	**0.95**	**(0.77, 1.15)**
	Keyword in body	−0.13	(−0.42, 0.19)
	Call to action	**1.83**	**(1.67, 1.98)**
	Keyword word count (log)	−0.18	(−0.38, 0.05)
	Headline word count (log)	**−0.62**	**(−0.98, −0.38)**
	Body word count (log)	**−0.93**	**(−1.11, −0.76)**
	Flesch reading ease	−0.11	(−0.32, 0.15)
	Position (log)	**−2.21**	**(−2.38, −2.02)**
Conversion	*Intercept*	**−1.81**	**(−2.08, −1.67)**
	Broad	**−0.97**	**(−1.23, −0.71)**
	TV show	**0.34**	**(0.09, 0.52)**
	Price	**−1.09**	**(−1.20, −0.94)**

Note. Parameter estimates in boldface are significant.

Table 4 **Posterior Mean (Standard Deviation) of Covariance Matrix for Demand Shocks**

	ξ_{wt}^{cl}	ξ_{wt}^{con}	ξ_{wt}^{LIV}
ξ_{wt}^{cl}	2.08 (0.24)		
ξ_{wt}^{con}	—	2.90 (0.42)	
ξ_{wt}^{LIV}	0.08 (0.03)	—	0.08 (0.005)

a mean of −4.13 because of the low average CTR. Keyword-specific factors (captured by *Broad* and *TV show*) are also important and allow us to link the search (and with it the different stages of search) to CTR—we find that broad keywords have a lower CTR (−0.48) and that keywords for TV show ringtones have a higher CTR then keywords for songs (1.59). This is similar to the findings of Ghose and Yang (2009) and Rutz and Bucklin (2011) with respect to differences between generic and branded keywords. The effect of log of position is negative, with a mean of −2.20. As expected, a higher position index (further down in the rankings) leads to a lower CTR. We find strong evidence of position endogeneity. First, the effect of position is attenuated when treating position as exogenous (mean, −1.55). Second, based on our LIV approach, we find that the correlation between the keyword-specific demand shock and the LIV error is 0.19. The LIV parameters are well separated, providing evidence for endogeneity in position and the need to account for it (see Tables 4 and 5). We have estimated our model with $C = 2$ and $C = 3$. For $C = 3$, two of the latent categories show very little separation (i.e., the parameter means are very similar), which, according to Ebbes et al. (2005, p. 370), is evidence for a smaller number of categories: "If the groups found by the LIV model are not well separated, it resembles a situation in classical IV where the instruments are weak," and a small number of categories is sufficient. Next, the measures of *Competition* and *Volume* capture the effect of competition—higher levels of *Competition* (*Volume*) lead to a lower CTR; the mean is −0.49 (−1.49). This is in line with expectations. First, in a more competitive environment, CTR is lower compared with a less competitive environment, everything else equal. Second, a more searched keyword has to compete more head-on with organic searches; thus, the CTR is lower.

Table 5 **Estimates for LIV**

	Estimate	
Variable	Mean	95% coverage interval
ω_1	**1.01**	**(0.95, 1.06)**
ω_2	**0.31**	**(0.28, 0.34)**
p	**0.25**	**(0.20, 0.30)**

Note. Parameter estimates in boldface are significant.

We now turn our attention to the keyword/ad copy. We find that some of the attention attractors matter: *Keyword in headline* has a positive effect (0.95), whereas *Keyword in body* is not effective. The density of the ad also matters; both proposed measures (log of *Headline word count* and log of *Body word count*) are negative and significant, which suggest that in our data set, consumers seem to favor less dense ads. We find that keyword appearance in the ad body and the Flesch reading score does not affect the click-through behavior. Finally, we find that a *Call to action* indeed affects CTR (1.83), as expected. From a managerial perspective, it is important to understand how ad characteristics can affect ad performance of a keyword/product combination. We find that some ad characteristics affect CTR, and we investigate their effect in terms of lift in CTR in our data. Including a keyword in the headline improves CTR by 7.9%, whereas adding a call to action has an effect of 32.9% on CTR. In terms of ad density, we find that decreasing the density (by one word) with regard to headline (body) increased ad performance measured by CTR by 2.2% (3.7%).

Next we discuss the conversion model results. In our case, we use an intercept, two keyword characteristics, and price in the conversion stage. We believe that ringtone subscriptions are the same product, and we do not have to take product attributes into account in the conversion decision compared with other industries that offer differentiated products. Clearly, if the products are not homogeneous, the differences in product attributes might influence conversion rates and should be included in the model. Again, we find that the intercept is negative (mean, −1.81). Searches based on broad keywords convert worse (−0.97), and searches for TV show ringtones convert better than searches for songs (0.34). The price coefficient has a mean of −1.08. As with position, price could be potentially endogenous in our setting. We have used the LIV-Hausman test (Ebbes et al. 2005) and find no evidence for endogeneity concerns with regard to position.

One of the benefits of the proposed integrated framework is that it allows inferring correlation in consumer preferences across the two decisions. For example, we find that consumers who are more responsive to position are also more price sensitive (correlation, 0.22). In other words, a consumer who is more likely to be influenced by one marketing action—position—is also more attuned to another marketing instrument—price. We use this insight to propose a contextual targeting scheme in the next section.

4.3.4. Managerial Implications—Contextual Targeting. The greatest marketing opportunity created by search engine advertising is the ability to target consumers based on their current interests revealed through the search query. Additionally, Google and other major search engines offer several tools that allow advertisers to further tailor their offerings to specific segments of consumers (e.g., based on geographical location, time of day, language, device platform). Based on the insights generated by our model, we are able to offer a novel targeting opportunity that can be used to improve conversion performance. Specifically, we found that consumers' preferences with regard to response to position and price are positively correlated (0.22). In other words, consumers who are more responsive when it comes to position, i.e., are more prone to click on an ad in position 1 versus 5, all else equal, are also more price sensitive. To illustrate this result, let us consider a simplified example of having just two keyword/ad combinations—Ad 1, which is usually displayed in top positions, and Ad 2, which is typically shown at the bottom. It is important to note that these two ads never appear together on the same page and are associated with distinct keywords; hence it is plausible to assume that the two ads are being seen by two distinct sets of consumers (for the sake of simplicity, we do not consider a possible change in search criteria for most consumers[14]). Our results then suggest that the average "clicker" for Ad 1 is more price sensitive than the average "clicker" for Ad 2.

We propose to exploit this finding in a contextual targeting scheme. Based on the positive correlation between position and price, we recommend to customize price using position information: for ads shown in top positions (e.g., 1 or 2), more aggressive price incentives should be offered to stimulate conversion compared with the ads shown in a lower position (e.g., 6). From a practical standpoint, this price customization can be implemented, for example, by using exit or midsession pop-up coupons that offer steeper discounts to the consumers who responded to ads in top positions. The proposed approach does not require the firm to make adjustments to its current bidding strategy; it simply exploits heterogeneity in price sensitivity, which is inferred from response to ad position. Also, offering price incentives via targeted coupons integrates well into the current ringtone industry practice of frequent promotions. Retail coupon promotions are often targeted based on either demographics or observed (purchase) behavior (e.g., Rossi et al. 1996). In the domain of paid search, neither demographic information nor past purchase behavior are available to inform a coupon decision. Our approach allows for the use of the one piece of behavioral information that is observed: the position of the ad the consumer has clicked on.

As mentioned before, position is only reported as a daily average after the day is over. Thus, we base the targeting of the coupon on the previous-day ad position, which can easily be automated using standard paid search. Although a full-blown optimization is beyond the scope of this paper, we illustrate how coupons can be targeted based on a previous-day ad position. We pick a cutoff point, e.g., position X, and we assume that for all ads above position X, a coupon with a face value of 2.5% (5%, 7.5%) was served as a pop-up. Note that the firm could potentially also exploit the fact that consumers who are clicking on lower-position ads are less price sensitive by increasing price. However, the firm might price itself out of a competitive market doing this. Without a model that accounts for competitors' pricing, a recommendation with regard to raising price would not be advisable. For each cutoff/face-value scenario, we integrate over the estimated parameters to generate conversion shares conditional on observed clicks. Based on these conversion shares, we calculate profit for the scenario.[15] Although we are not at liberty to report exact profit figures, we report the increase in profit in percentage terms. We find that a coupon with a 5% face value and a cutoff position of 3 yields the highest increase in profit of 2.7% (see Table 6 for details). We also find evidence for a typical promotion issue. Decreasing the price for top positions leads to an increase in profit by increasing conversion over loss in profit as a result of lower prices. Decreasing prices for bottom positions, however, does not increase overall profits. This is similar to issues faced with traditional price promotions: whereas reduced prices attract nonloyal consumers and generate incremental sales, loyal consumers simply get the product at a discounted price without necessarily purchasing more, resulting in a net revenue loss for this segment of consumers. A critical question in promotions is how the promotional bump can be decomposed into incremental sales and sales that would have occurred at the regular price anyway. Our contextual targeting scheme allows distinguishing between consumers who might be swayed by the promotion to buy and consumers who are not price sensitive enough leveraging position information.

Again, we do not advocate changing positions based on response to price but merely suggest that position can be used to target coupons in lieu of

[14] What we are proposing is to exploit the current practice of the major search engines, which do not serve multiple ads of the *same* firm on the *same* search results page. Hence, each keyword and the associated ad can be viewed as isolated "markets." We acknowledge that the opportunity for arbitrage may occur if the consumer chooses to run multiple searches and inspects several ads by the firm associated with different discount coupons.

[15] Note that no additional costs are incurred in our targeting scheme as the campaign is not changed. Profits are calculated based on the margins of the firm and the changes in revenues due to changes in subscriptions (conversions).

Table 6 **Change in Revenue Through Contextual Targeting**

	Percentage of change in price (serve coupon for positions < cutoff)		
Cutoff	2.5	5.0	7.5
2	1.1	1.9	0.6
3	2.0	**2.7**	1.0
4	1.4	2.2	0.7
5	1.7	1.9	0.8
6	1.4	1.7	0.5

demographics or past purchase behavior. We also note that the proposed simulation-based analysis does not take into consideration potential consumer learning and/or competitive reaction.

5. Summary

This paper develops an empirical model that allows the effects of ad position and the textual properties of the ad on consumer actions to be evaluated solely based on the limited data that are typically available from search engines. Accurate evaluation of these effects is of high importance to business practitioners, as evident from online discussion forums, trade journals and professional conferences, and our own interactions with several paid search firms. To the best of our knowledge, no empirical study in the academic marketing literature has modeled the textual properties of paid search ads. Moreover, as we argue in this paper, the existing models are not well suited to perform such an analysis because a keyword-centric perspective on paid search advertising is taken—consumers are assumed to be homogeneous within keyword—and unobserved consumer heterogeneity is only captured across keywords. These models are not based on economic primitives that explicitly account for consumer preferences. Therefore, economic interpretation and policy experiments with these models are somewhat problematic. For example, the interpretation of the effect of the key element of paid search advertising—ad position—is unclear in a keyword-centric model. In these models, it is a *keyword* that responds to a specific position, because keywords are used as a proxy for consumer preferences. However, the effect of position should be measured by modeling *consumers* who are more or less likely to inspect the ad and click on it depending on position. In this paper, we argue that a consumer-centric approach to paid search modeling offers a more plausible description of the underlying choice process captured in aggregated outcomes provided by search engines. Furthermore, we argue that, by effectively ignoring possible heterogeneity among consumers, the existing models are missing an important characteristic of the market environment that over the past two decades of marketing research has become a standard for empirical modeling.

Our model enables us to provide a deeper look into the mechanisms of paid search and, most importantly, provides a first account of the effects of *textual properties* and *design attributes* of ads in response to paid search. We argue that the optimal ad design needs to be tailored to specific product/market conditions, and we hope that the proposed model can assist in this process through improved ad feature performance assessment. Based on the insights generated by our model, we propose a novel contextual targeting scheme. So far, targeting has not been possible in paid search beyond the keyword because data necessary for standard targeting are not available. We find our proposed targeting scheme allows for increasing revenue by 2.7% without changing campaign cost.

From a methodological perspective, we expand the state-of-the-art data augmentation approach proposed by Musalem et al. (2008, 2009) by developing a two-stage consumer-level model of click-through and conversion based on aggregate paid search data that takes selection into account. We find a significant correlation across the click-through and conversion decision that needs to be addressed when modeling paid search. We also extend the LIV framework proposed by Ebbes et al. (2005, 2009) to the choice modeling domain.

The limitations of the available search engines data served as a key motivator for this study. Ideally, a comprehensive consumer-centric model of choice in the paid search domain should explicitly incorporate information about all alternatives presented to a consumer on a search results page. However, until this information becomes available, business practitioners are left to choose among the models constructed on aggregated performance statistics provided by search engines. We hope that the proposed approach will help practitioners make better decisions in planning and executing paid search campaigns.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mktsci.pubs.informs.org/.

Acknowledgments

The authors thank the collaborating firm for providing the data used in this study. The authors are grateful to Peter Ebbes, Andrés Musalem, and Michel Wedel for their helpful comments and suggestions on this work.

References

Bruthiaux, P. 1996. *The Discourse of Classified Advertising: Exploring the Nature of Linguistic Simplicity*. Oxford University Press, New York.

Bruthiaux, P. 2000. In a nutshell: Persuasion in the spatially constrained language of advertising. *Language Comm.* **20**(4) 297–310.

Danaher, P. J., M. S. Smith. 2011. Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Sci.* **30**(1) 4–21.

Ebbes, P., M. Wedel, U. Böckenholt. 2009. Frugal IV alternatives to identify the parameter for an endogenous regressor. *J. Appl. Econometrics* **24**(3) 446–468.

Ebbes, P., M. Wedel, U. Böckenholt, T. Steerneman. 2005. Solving and testing for regressor-error (in)dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quant. Marketing Econom.* **3**(4) 365–392.

Erdem, T., S. Imai, M. P. Keane. 2003. Brand and quantity choice dynamics under price uncertainty. *Quant. Marketing Econom.* **1**(1) 5–64.

Ghose, A., S. Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Sci.* **55**(10) 1605–1620.

Hotchkiss, G. 2006. Inside the searcher's mind: It's a jungle in here! White paper, Enquiro. http://www.enquiro.com/.

Kamakura, W. A., G. J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structuring. *J. Marketing Res.* **26**(4) 379–390.

Musalem, A., E. T. Bradlow, J. S. Raju. 2008. Who's got the coupon: Estimating consumer preferences and coupon usage from aggregate information. *J. Marketing Res.* **45**(6) 715–730.

Musalem, A., E. T. Bradlow, J. S. Raju. 2009. Bayesian estimation of random-coefficients choice models using aggregate data. *J. Appl. Econometrics* **24**(3) 490–516.

PricewaterhouseCoopers. 2010. IAB Internet Advertising Revenue Report. Report, IAB and PricewaterhouseCoopers, New York. http://www.iab.net/media/file/IAB-Ad-Revenue-Full-Year-2009.pdf.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Rutz, O. J., R. E. Bucklin. 2011. From generic to branded: A model of spillover in paid search advertising. *J. Marketing Res.* **48**(1) 87–102.

Rutz, O. J., R. E. Bucklin, G. P. Sonnier. 2011. A latent instrumental variables approach to modeling keyword conversion in paid search advertising. Working paper, University of Washington, Seattle.

Vestergaard, T., K. Schroder. 1985. *The Language of Advertising*. Oxford, Basil Blackwell, UK.

Yang, S., A. Ghose. 2010. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* **29**(4) 602–623.

Yang , S., Y. Chen, G. M. Allenby. 2003. Bayesian analysis of simultaneous demand and supply. *Quant. Marketing Econom.* **1**(3) 251–275.

Yao, S., C. F. Mela. 2011. A dynamic model of sponsored search advertising. *Marketing Sci.* **30**(3) 447–468.