A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising

Oliver J. Rutz

Randolph E. Bucklin

Garrett P. Sonnier

August 2010

* Oliver J. Rutz is Assistant Professor of Marketing, Yale School of Management, 135 Prospect Street, New Haven, CT 06520, oliver.rutz@yale.edu. Randolph E. Bucklin is the Peter W. Mullin Professor, UCLA Anderson School, 110 Westwood Plaza, Los Angeles, CA 90095, rbucklin@anderson.ucla.edu. Garrett P. Sonnier is Assistant Professor of Marketing at the University of Texas at Austin, 1 University Station, Austin, TX 78712. The authors wish to thank a collaborating firm for providing the data used in this study.

Abstract

We present a modeling approach to assess the purchase conversion performance of individual keywords in paid search advertising. An integrated framework models both conversion and clickthrough rates, while also accounting for the potentially endogenous position of the text advertisement. Because it is in part determined by the paid search auction, the position of the ad in the listing served in response to a search is expected to be endogenous. Standard paid search data available to advertisers lacks the competitive information that might be used to instrument for position, posing a problem for structural or reduced form models of keyword performance. The proposed model handles position endogeneity using a latent instrumental variable approach. The model is applied to keyword-level paid search data containing daily information on impressions, clicks and reservations for a major lodging chain. Results show that addressing position endogeneity as well as the correlation between click-through and conversion is warranted. A comparison in a holdout sample suggests that campaign management using the proposed model outperforms methods used in practice.

Keywords: Internet, Advertising, Paid Search, Bayesian Methods

INTRODUCTION

Paid search advertising allows companies to address consumers directly during their electronic search for products or services. When a consumer searches the Web with the help of an Internet search engine, the terms he enters to initiate a search are known as keywords. In practice, a keyword can consist of multiple words, such as “Hotels Los Angeles.” A company in the lodging business, for example, can address this consumer directly by bidding for specific keywords and creating a text ad that will be shown when a consumer searches on those keywords. In paid search, advertisers bid their maximum willingness to pay for a click on a paid search ad. An automated auction-type algorithm then determines position (e.g., 1st or 3rd) of the ad in the sponsored listings section of the results page.

Paid search differs from traditional advertising as companies typically do not pay for exposures (as for most types of banner ads or offline advertising), but for actual clicks on their paid search ads. Also, paid search campaigns require the management of an extensive list of keywords often numbering in the tens of thousands. Because some of the keywords are widely searched while many, if not most, generate very little traffic, an advertiser’s keyword list typically has a long tail. As the activity data become sparse in the tail of the list, evaluating performance on these keywords becomes difficult. This is frustrating in practice as sparsely searched keywords are potentially good advertising investments but it is difficult to gauge their performance.

Search engines routinely provide daily information on paid search advertising to firms and advertisers typically manage their paid search campaigns based upon such data. Assuming a direct-marketing approach to performance evaluation, we could calculate the marginal benefit of spending for each keyword, comparing advertising-related profit-per-sale with advertising-related cost-per-sale[1]. If that difference is positive, a keyword generates a positive return for the company from a direct-marketing perspective. Using standard paid search data, we can compute cost per sale as the ratio of cost per click to conversion rate. However, a problem occurs when the observed conversion rate (number of sales divided by number of clicks) for a given keyword is zero. Even when non-zero, the cost per sale measure may be based on very few observations, and hence subject to substantial error. Since average conversion rates for keywords in paid search are very low, this measurement problem occurs quite often. For example, the average conversion rate from paid search click to purchase in the travel industry was 2.1% in the first quarter of 2004 (ClickZ 2008). Thus, even on a monthly basis most keywords simply do not generate any sales, thereby precluding the calculation of cost per sale at the keyword level. Even when the true long-term conversion rate for a keyword is positive, it could take some time for a sale to occur and the resulting conversion ratio still remains subject to large error.

Given the challenges of evaluating cost per sale at the keyword level, managers often resort to ad hoc model-free strategies. For example, evaluating paid search at the campaign level by aggregating across all keywords is a crude but straightforward strategy. The manager can compare spend on the campaign versus sales attributable to the campaign. However, this does not aid in allocating advertising spending across individual keywords, especially those in the long tail. The manager may refine this strategy by aggregating keywords into groups. Alternatively, the manager may deploy metrics other than conversion rate, such as impressions or click-through-rate (CTR). These approaches almost always require the manager to arbitrarily select a decision rule (e.g., keep all keywords with CTR ≥ 1%). Furthermore, these approaches do not allow the manager to assess the conversion performance of individual keywords.

[1]For ease of exposition, we will use cost-per-sale instead of advertising-related cost-per-sale.

Model-based strategies for assessing keyword conversion rates hold the promise of helping managers make better decisions. However, standard paid search data are fraught with issues. The major search engines (Google, Yahoo! and MSN) provide advertisers with data on their own campaign performance (i.e., impressions, clicks, position and cost). These data are aggregated on a daily keyword level. None of the major search engines provides competitive data or allows advertisers to infer which competitors were listed together with their own ad. Also, companies cannot infer the exact position of an ad on which a consumer has clicked or the consumer's searches prior to clicking.

A key question to address in any model of paid search advertising is the effect of the position of the text ad on the consumer's click and conversion decisions. A key challenge in doing so is that position is very likely to be endogenous, which may bias estimates of its effect. Position is determined by the outcome of the auction, which is a function of past consumer clicking behavior and competitive bids. The lack of competitive bid information raises concerns over omitted variables. In addition, typical paid search data reports position as the daily average position at the keyword level, raising concerns over measurement error. It is well known that omitted variables and measurement error can induce regressor-error dependencies.

Existing models of paid search advertising have explicitly modeled the auction via reduced form simultaneous equations (Ghose and Yang 2009). In this approach, equations for click-through and conversion are specified along with equations for the advertiser's decision (i.e., the firm's bid) and the search engine's decision (i.e., position). In lieu of actual bid information, Ghose and Yang (2009) utilize cost-per-click as a proxy for bid. Reduced form simultaneous equations model typically need exclusion restrictions to ensure identification. For example, Ghose and Yang (2009) exclude contemporaneous position from their cost-per-click equation, in essence treating cost-per-click as exogenously determined. This assumption may be questioned as the search engine's decision where to rank a text ad affects the cost borne by advertisers. While Ghose and Yang (2009) find no contemporaneous correlation between position and cost-per-click, their data are aggregated into weekly observations, which may mask the correlation. It is unclear whether this condition would generally hold in daily paid search data. Furthermore, using firm cost as a proxy for bid, as well as the lack of competitive bid information, may induce an errors-in-variables problem which can complicate identification and estimation of simultaneous equations systems (Hsaio 1976; Hausman 1977). Given these issues, borne largely of data limitations, it is of interest to consider an alternative model of keyword conversion performance.

The objective of this paper is to develop an alternative, model-based approach to assess the performance of individual keywords on a daily level that directly addresses the endogeneity of position in the equations for both the consumer's click and conversion decisions. We conceptualize conversion as a binary choice decision conditional on a click (i.e., a user has clicked on the paid search ad and has been taken to the landing page on the company's website). We model click-through as a binary choice conditional on search. The conversion and the click-through models are linked by correlated unobserved shocks. Our Bayesian shrinkage estimator infers conversion rates for keywords with very few or no conversions (and improves the estimated conversion rates for other keywords) by exploiting the similarity among keywords to produce a shrinkage-based conversion rate estimate for each one.

Instrumental variable (IV) techniques offer one means of addressing the problem of position endogeneity in the click and conversion equations. However, the absence of full information on the paid search auction results in a scarcity of candidates for observed instruments in standard paid search data. Furthermore, any observed instruments may be weak, or worse, invalid. While IV estimation can correct the bias in model parameter estimates, weak instruments may result in relatively large standard errors (Stock et al. 2002). We therefore propose to account for the endogeneity of position by augmenting observed instruments with latent instrumental variables (LIV) (Ebbes et al. 2005; Zhang et al. 2009). The LIV technique uses a latent variable model to account for regressor-error dependencies and, as such, addresses the issues of instrument availability, weakness, and validity.

Our intended contribution is both methodological and substantive. From a methods perspective, we provide a solution to the measurement problem of individual keyword performance on a daily level that addresses the potential endogeneity of position via instrumental variables. This provides a complementary method to existing simultaneous equation approaches that model the auction (Ghose and Yang 2009). In our case, we find that ignoring the endogeneity problem biases the estimates of the effect of position on click-through and conversion. Substantively, we find that conversion rates differ systematically across keywords and that keyword characteristics and position are significant predictors of conversion rates. Using the estimated model parameters, we demonstrate in a holdout sample that the keyword list generated by our model-based approach yields superior profits compared with the status quo.

The remainder of the paper is structured as follows. First, we provide a brief overview of paid search advertising and discuss some relevant literature. We then present our modeling approach, dataset, and results. Next we discuss the implications of our findings and illustrate how to improve the performance of a paid search campaign by individual keyword management. We finish with a conclusion, the limitations of our approach, and discuss future research in the realm of search engines and marketing.

We briefly describe the paid search process from the perspective of both the advertiser and the consumer. Consider the scenario in which a consumer searches using the keyword “Hotels Los Angeles”. An advertiser has selected this keyword and created a text advertisement for their offerings. Advertisers bid the maximum dollar amount willing to pay for a click on a text ad served in response to a search for a keyword. The actual cost-per-click (CPC) and position of the text ads are determined by a proprietary, auction-style algorithm. In general, CPC and position are a function of the own bid, the bids of the competing firms (unobserved by the focal advertiser) and other metrics that focus on past ad performance (e.g., past click-through rate). The search results page will display non-sponsored results (organic search results) and the text ads from the focal advertiser next to text ads from other advertisers (paid search results). From the advertiser’s perspective, one impression has been generated which is attributed to the keyword “Hotels Los Angeles”. The consumer views the advertiser’s text ad and chooses to click on link provided in the text ad. He is transferred to the landing page on the advertiser’s website. The keyword “Hotels Los Angeles” has now generated a click. The consumer then decides whether to reserve a hotel room. If he does so, the keyword “Hotels Los Angeles” is now associated with a reservation.

Despite the impressive growth and scale of paid search, there has been little academic study of this advertising service in the literature, especially in marketing. Recent theoretical papers investigate paid search as a pure second price auction (Edelman and Ostrovsky 2007, Edelman et al. 2007) and paid search advertising as a product differentiation/signaling game (Chen and He 2009). In related work, Wilbur and Zhu (2009) investigate click fraud in paid search auctions from a game theoretic perspective. Empirical models of paid search advertising are the subject of two recent working papers in marketing. Goldfarb and Tucker (2009) investigate how regulation affects paid search ad pricing and show that search engines profit when regulation limits the advertisers' other advertising options. Song and Mela (2009) develop a dynamic structural model of paid search advertising for a small business-to-business search engine specializing in industrial software. Their model makes use of competitive bid data, which is not available in the standard paid search data provided by the major Internet search engines.

Using a paid search dataset for a retail chain that advertises on Google, Ghose and Yang (2009) propose a simultaneous equation approach to analyzing search engine advertising that models the consumer's click-through and conversion decisions along with the search engine's decision on position and the advertiser's decision on cost-per-click (CPC). CPC is used as a proxy for the advertiser's bid, which is unobserved. Ghose and Yang (2009) note that cost and bid are highly correlated. Nonetheless, using a proxy for the firm's own bid, along with missing competitive bid data, may induce an errors-in-variables problem, complicating identification and estimation of the system.[2] A related issue is the exclusion restrictions necessary for identification of systems of simultaneous equations. These restrictions are sometimes difficult to justify and generalize to other modeling settings.[3]

MODELING APPROACH

As discussed previously, a major problem with measuring conversion in paid search is that it is sparse, particularly in the tail of an advertiser's keyword list. Thus, for many keywords,

[2] Firm bidding data could ostensibly be obtained from a cooperative firm. However, firms cannot collect data on competitive bids without the search engine releasing this information. Firms also cannot collect data on which competitor occupies which position in the auction without the help of the search engine. It is unclear at this point in time whether Google or any of the other major search engines will ever release this information, either to companies or academic researchers.

[3] As noted in the introduction, Ghose and Yang (2009) exclude contemporaneous position from their cost-per-click equation. While Ghose and Yang (2009) find no contemporaneous correlation between position and cost-per-click, their data are aggregated into weekly observations, which may mask the correlation. In our daily data, we find significant contemporaneous correlation between position and cost-per-click.

advertisers cannot simply calculate conversion rates based on observed data, making it impossible to evaluate which keywords are profitable (i.e., generate margins on purchases which exceed advertising costs). The ultimate goal of our modeling approach is to address the sparseness problem and improve the measurement of conversion rates at the individual keyword-level. To this end, we build an integrated model of click-through and conversion that accounts for the endogeneity of keyword position. Our model is suited to the paid search data generally available to firms. Typical paid search data do not include information at the visitor level (i.e., clickstream data) which precludes us from modeling the purchasing decision of individual users.[4] Google, as well as the other major search engines, only provide keyword-level aggregate data on a daily basis. Thus, firms are not able to tie their site visitors to individual paid search ads in terms of specific position and cost per click. Even in the event that some firms could accomplish this, cookie-based clickstream data suffer from their own limitations and do not allow companies to "identify" visitors prior to purchase (i.e., leverage demographic information).[5]

In our model, we investigate whether the conversion probability given a click-through to the advertiser's web site can be measured based on available keyword-centric information alone. We acknowledge that this is a second-best option, but it is the only viable one given the data available to managers. Our approach hinges on the notion that conversion rates differ systematically across keywords. In our data (and other data sets we have examined), this is indeed the case. In April 2004, for example, daily conversion rates at the keyword level ranged

[4] Clickstream data assigns each visitor to the site an individual ID (cookie). Clickstream data would allow researchers to connect a reservation to a specific click and that click, in turn, to a specific impressions and keyword. Google has not provided impression and keyword data on a user specific level in the past (smallest aggregation level currently is hourly) and experts in the field believe that Google has no intention of doing so in the future.

[5] Alternatively, one could imagine having panel data for Internet users, collected and compiled by a syndicated supplier such as comScore or Nielsen. Unfortunately, the panel approach is likely to break down in the evaluation of paid search keywords. For example, it is not clear whether even a large panel of consumers would make enough searches in the product or service category to get around the sparseness problem; indeed, such a problem could very well be far more severe with panel data versus the data provided by the search engine.

between 0% and 50% with an average of 0.9% and a standard deviation of 0.07. This variation in conversion could indicate that different keywords “attract” different types of consumers who end up purchasing at different rates. (Similar patterns are also present in multiple datasets of more recent vintage.) In other words, consumers reveal information about themselves through their choice of search terms. Here, we focus on whether using information on the keywords alone allows us to measure keyword conversion rate. An implicit assumption of our approach is that consumers who use a certain keyword have similar objectives and behavior. Our model is built with keywords as the unit of analysis so that we can improve the estimation of conversion rates at the keyword level. It allows us to explore whether conversion rates differ systematically across individual keywords and if so, whether we can explain those differences based on observable keyword covariates.

One approach to modeling conversion would be to condition on a click-through, akin to brand choice models that condition on purchase incidence. However, if the click and conversion decisions are correlated, as seems likely, this can lead to selection bias. While aggregate data do not allow the construction of a nested model as in the purchase incidence-brand choice literature, we are able to link a keyword-level click-through model with a keyword-level conversion model via correlated error shocks (Berry et al. 1995). As discussed previously, the position of the text ad, an important covariate in the click and conversion equations, is most likely endogenous. The problem of position endogeneity in the click and conversion equations is akin to the problem of endogenous schooling in wage regressions (Ebbes et al. 2005). This problem has been addressed by instrumental variables and motivates our approach to account for position endogeneity of position in the click and conversion equations. In paid search, as in many other applications, valid instruments are difficult to obtain. To overcome this problem, we incorporate the latent-instrumental variable (LIV) approach (Ebbes et al. 2005; Zhang et al. 2009) into our model of click and conversion.

Model Specification: The Conversion Model

We employ a binary logit model to investigate the probability of conversion conditional on a visitor reaching the company landing page via a click. The daily clicks for each individual keyword are used as choice occasions, whereas the daily conversions for each individual keyword represent the "successful" choices. Based on the binary logit model, the conversion probability, P_{wt}^{con}, for keyword w at time t is given by

$$(1) \qquad P_{wt}^{con} = \left(\frac{\exp(pos_{wt}\alpha_w^{con} + x_{wt}^{con\prime}\beta_w^{con} + \xi_{wt}^{con})}{1 + \exp(pos_{wt}\alpha_w^{con} + x_{wt}^{con\prime}\beta_w^{con} + \xi_{wt}^{con})} \right)$$

Where pos_{wt} is keyword position, x_{wt}^{con} is a vector of keyword level covariates, $\theta_w^{con} = \left[\alpha_w^{con} \ \ \beta_w^{con\prime} \right]'$ is a vector of are keyword level parameters with $\theta_w^{con} \sim N\left(\bar{\theta}^{con}, \Sigma_\theta^{con}\right)$, and ξ_{wt}^{con} is an error shock.

In a standard application of the logit model, the data contain one choice outcome (observation) per time period. Since clickstream data are not available, we cannot link a specific click (choice occasion) to a specific reservation (successful choice). For each individual keyword we observe the numbers of clicks and reservations on a daily basis. Thus, our data typically have more than one "choice" occasion per time period. In light of the above, we therefore use the following likelihood function to estimate the parameters of the logit model:

$$(2) \qquad Likelihood^{con} = \prod_t \prod_w (P_{wt}^{con})^{conversions_{wt}} (1 - P_{wt}^{con})^{(clicks_{wt} - conversions_{wt})}$$

where t is time and w is keyword.

Model Specification: Linking Click-through and Conversion.

Before a consumer can arrive at the company's landing page via paid search, he must decide whether or not to click on the paid text ad of the company that is displayed in response to his search. While we cannot link a specific click to a specific conversion, we can connect the conversion and click-through decisions in our proposed aggregate framework. Similar to conversion, we model the click-through decision using a binary logit. Note that the click-through decision is modeled conditional on a visitor searching for a keyword that has been bid on by the advertiser.[6] Based on the binary logit model, the click-through probability, P_{wt}^{cl}, for keyword w at time t is given by

$$(3) \qquad P_{wt}^{cl} = \left(\frac{\exp(pos_{wt}\alpha_w^{cl} + x_{wt}^{cl\prime}\beta_w^{cl} + \xi_{wt}^{cl})}{1 + \exp(pos_{wt}\alpha_w^{cl} + x_{wt}^{cl\prime}\beta_w^{cl} + \xi_{wt}^{cl})} \right)$$

where pos_{wt} is keyword position, x_{wt}^{cl} is a vector of keyword level covariates, $\theta_w^{cl} = \left[\alpha_w^{cl} \;\; \beta_w^{cl\prime}\right]'$ is a vector of are keyword level parameters with $\theta_w^{cl} \sim N\left(\bar{\theta}^{cl}, \Sigma_\theta^{cl}\right)$ and ξ_{wt}^{cl} is an error shock. The likelihood for the click-through model is similar to the likelihood for the conversion model given in (2).

We link the click-through and the conversion model given by (1) and (3) via the error shocks, $\xi_{wt}^{cl}, \xi_{wt}^{con}$

$$(4) \qquad \begin{pmatrix} \xi_{w,t}^{cl} \\ \xi_{w,t}^{con} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{cl,cl} & \sigma_{cl,con} \\ \sigma_{cl,con} & \sigma_{con,con} \end{pmatrix}\right],$$

where the parameters of the covariance matrix are to be estimated.

[6] We refrain from modeling the consumer's choice of keywords. This would entail modeling the choice among the very large number of possible search terms a consumer could use. Also, we only observe searches on the keywords for which the company has bid.

Model Specification: Accounting for the Auction

In paid search, a modified second price auction determines a keyword's position as well as the cost-per-click paid by the advertiser. To date, the major search engines have not revealed the inner workings of their auction mechanisms. It is known that in addition to the actual bids past performance of the ad in terms of click-through rate, as well as measures of landing page quality, are taken into account. Often, additional features such as the performance of the ad group (companies can group similar keywords together) or the performance the whole campaign are taken into account. From an advertiser's perspective, the paid search auction can be best described as a black box.[7] In addition to not knowing the precise workings of the auction, advertisers also do not have access to basic competitive information such as who else was listed and in which position competitors were listed. The major U.S. search engines will, in all likelihood, never make competitive bids available to advertisers.

Position is seen as a key criterion for the success of a paid search campaign and managers focus on "getting it right." From an information processing perspective, research has found that information displayed in list format is generally investigated from top to bottom. Assuming that consumers inspect the sponsored listings until they find an ad that meets the threshold for clicking, a paid search ad in a higher position will, most likely, be viewed by more consumers than an ad in a lower position. Another argument for the importance of position is derived from the search engines' auction mechanism. Supposedly, a "better fitting" ad (i.e., an ad that has a

[7] No competitive information is given by the search engines. In addition the proprietary auction algorithms are described in vague terms. Google, for example, describes it auction mechanism in 2004 as ad being determined by maximum cost-per-click (CPC) bid and a keywords past performance, which is determined by a "measure of past CTR". Note that in 2004 Google had not yet introduced the notion of the Quality Score, which it now defines as follows: the Quality Score is determined by: i) the historical click-through rate (CTR) of the keyword, ii) the account history, which is measured by the CTR of all the ads and keywords in the account, iii) the historical CTR of the display URLs in the ad group, iv) the relevance of the keyword to the ads in its ad group, v) the relevance of the keyword and the matched ad to the search query, vi) the account's performance in the geographical region where the ad will be shown, and vii) other relevance factors."

higher CTR based on past performance) will be ranked higher by the search engine given the same bid. If consumers are aware of this, the position can be understood as a signal of "fit" and as such becomes a valuable input to the click-through decision as well as the conversion decision.

Although position is strategically important, treating position as an exogenous covariate in the click and conversion equations is questionable. First, a company's bid strategy and past click-through performance enter the auction. Second, competitive actions (bids) influence the company's position through the auction mechanism. As competitive bid information is unavailable, omitted variable concerns loom. Lastly, typical paid search data report position as the daily average position at the keyword level. Thus, observed position contains measurement error. One way to alleviate the endogeneity concern would be to explicitly model the underlying auction. Some researchers have been successful in addressing the problem by leveraging bidding history information using data from a non-major search engine for specific software products (Yao and Mela, 2009). However, given the current information-sharing policies of the major search engines, it is highly unlikely that competitive bidding data will available any time soon. At present, even less "sensitive" information such as the number of competitive firms bidding for the same keyword is not offered. Ghose and Yang (2009) propose to address this problem by estimating a reduced form model of keyword position. However, given a lack of competitive bid information along with the exclusion restrictions necessary for identification, it is useful to consider an alternative model of keyword conversion performance. We propose a model that can address the position endogeneity problem without resorting to a structural or reduced form model of the auction and the necessary – but potentially untenable – assumptions required.

Assuming the availability of valid instruments, one alternative to address the position endogeneity problem is to use instrumental variables (IV) estimation. In this case, we correlate the unobserved demand shocks ξ_{wt}^{cl} and ξ_{wt}^{con} with the IV equation as described below. We express position as a linear function of observed instruments:

(5) $$pos_{wt} = z_{wt}^{IV\prime}\phi + \xi_{wt}^{IV}$$

where z_{wt}^{IV} is a vector of observed instruments, ϕ is a vector of parameters to be estimated and ξ_{wt}^{IV} is an error term.

To complete the IV specification, we define the correlation structure between the error term ξ_{wt}^{IV} and the click-through and conversion demand shock, ξ_{wt}^{cl} and ξ_{wt}^{con}, as follows:

(6) $$\begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \\ \xi_{wt}^{IV} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{cl,cl} & \sigma_{cl,con} & \sigma_{cl,IV} \\ \sigma_{cl,con} & \sigma_{con,con} & \sigma_{con,IV} \\ \sigma_{cl,IV} & \sigma_{con,IV} & \sigma_{IV,IV} \end{pmatrix}\right],$$

where the elements of the covariance matrix are to be estimated.

The IV approach requires valid, observed instruments to be available. Instruments that are correlated with the error term in the model are invalid. Ideally, instruments are highly correlated with the endogenous covariate but the stronger the correlation between the instrument and endogenous covariate, the more likely the instrument is to be invalid. The corollary, of course, is that valid instruments are often weak. This fundamental difficulty in finding good instruments leads us to avoid relying solely upon observed instruments in the estimation of our model. Rather, we extend the latent instrumental variable (LIV) framework (Ebbes et al. 2005; Ebbes et al. 2009; Zhang et al. 2009) to account for position endogeneity in our click-through and conversion models.

The LIV estimator belongs to the family of frugal IV estimators that do not require observed instruments. These include the higher moments (HM) estimator, the identification through heteroskedasticity (IH) estimator, and the LIV estimator (Ebbes et al. 2005; Ebbes et al. 2009). An advantage of the LIV estimator in our setting is that it is a likelihood based approach while the HM and IH are method-of-moments estimators. Thus, the LIV approach is amenable to Markov chain Monte Carlo (MCMC) estimation.[8] Ebbes et al. (2005) discuss the properties of the LIV estimator, demonstrate its performance in simulation, and apply it to the problem of estimating the effect of education on income. In a recent marketing application, Zhang et al. (2009) apply latent and observed instrumental variables to study the effect of visual attention to feature advertisements in a sales regression framework.

In the LIV approach, a latent variable model is 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 estimate of the effect of an endogenous covariate (e.g., position) on the desired actions (e.g., click-through and conversion). This framework was originally developed in a linear regression setting. We extend the LIV approach to our choice model framework for analyzing paid search advertising. Via data augmentation, the LIV model introduces a latent categorical variable with C categories. For identification, we require $C \geq 2$. We define the LIV equation for position as a function of the latent categorical instrument, γ_{wt}, and ω, the category means. It is given by

(7) $$pos_{wt} = \gamma_{wt}'\omega + z_{wt}^{IV\prime}\phi + \xi_{w,t}^{LIV} .$$

The latent instrument γ_{wt} follows a C-dimensional multinomial distribution with probabilities $\{\pi_1, \pi_2, ..., \pi_C\}$, where π_c is the probability that the c^{th} latent instrument is one; a value of one

[8] For a more detailed discussion of frugal IV estimators see Ebbes et al. (2009).

indicates that keyword w belongs to category c at time t. We define the link between the LIV error term, ξ_{wt}^{LIV}, and the click and conversion shocks as

$$(8)\qquad \begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \\ \xi_{wt}^{LIV} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{cl,cl} & \sigma_{cl,con} & \sigma_{cl,LIV} \\ \sigma_{cl,con} & \sigma_{con,con} & \sigma_{con,LIV} \\ \sigma_{cl,LIV} & \sigma_{con,LIV} & \sigma_{LIV,LIV} \end{pmatrix}\right],$$

where the elements of the covariance matrix are to be estimated. For details on the sampling procedure, please see the Appendix.

EMPIRICAL APPLICATION

Data

Our data encompass one calendar quarter of the paid search campaign for a major lodging chain on the Google search engine. The daily data span April to June, 2004. The company used 301 keywords in its campaign. The data consist of the standard information advertisers receive from Google and complementary, additional information purchased from a third party data provider. The standard information supplied by Google is daily data on an individual keyword level. For each keyword (e.g., Hotels Los Angeles) we have information on cost (in $), average position served (ranking, e.g., 2.3), and number of impressions and clicks. The additional third party data provides daily information on the number of reservations for each keyword.

We enhance the data by introducing semantic keyword characteristics. The keywords used have certain common characteristics that are specific to the lodging industry (e.g., a city or a holiday destination is included). We "decompose" each of the 301 keywords along the following set of characteristics:

- *Branded*: Is the company brand name included? 99 keywords are branded.
- *US*: Is the keyword for a US location? 223 keywords are for a US location.
- *State*: Does the keyword include a state name? 52 keywords include a state name.

- *City*: Does the keyword include a city name? 210 keywords include a city name.
- *Hotel*: Does the keyword include the word hotel or other lodging related phrases such as accommodation, motel, or room? 222 keywords include hotel or other related phrases.
- *#Words*: How many words are used? The mean number of words is 2.65 with a variance of 0.55.

For both the click-through and the conversion decision we use position and the semantic keyword characteristics as covariates, along with a keyword-specific intercept. For the IV model, we require some observed variables to instrument for position. In demand models with endogenous prices, input prices are often used as instruments (Kuksov and Villas Boas 2008). Lagged prices, lagged shares, cost, and prices in other markets are also used as instruments for endogenous prices (Yang et al. 2003). As noted in the introduction, standard paid search data are scarce with respect to candidate instruments. Similar to using lagged prices and cost information as instruments for price, we use lagged position, current cost-per-click, and lagged click-through rate as instruments for position in the IV model.[9] For the LIV model, we add the latent instruments as defined in (7), with $C = 2$, to these observed instruments. While one may consider a larger C, Ebbes et al. (2005) demonstrate that the LIV estimator is robust to under specifying the number of categories.

We use the data for April 2004 as an estimation sample. In April 2004 the campaign generated 2,281,023 impressions, 14,302 clicks and 518 reservations. The average position was 6.0 and the company spent $5,106.74 on the campaign. The average click-through rate (percentage of impressions that led to a click) was 0.6% and the average conversion rate

[9] Note that in 2004 Google had not yet introduced the notion of Quality Score, which could also serve as an observed instrument. Also note during the observation period our firm did not use multiple landing pages. As such, there is no effect of landing page on conversion as landing pages do not differ across keywords.

(percentage of clicks that led to a reservation) was 3.6%. The average cost-per-click was $0.36 and the average cost-per-reservation was $9.86. We use the data from May and June 2004 as a hold-out sample (2,983,085 impressions, 38,878 clicks, 1,348 reservations, $12,548 cost and 6.3 average position). The performance of the paid search campaign in May and June is very similar to April. In terms of conversion rate (3.5%) and cost-per-reservation ($9.31), there is little difference when compared to the estimation sample.

We believe that reviewing a paid search campaign on a monthly basis is a reasonable policy. Even niche keywords can accumulate multiple clicks over that time period. Shorter time periods result in a significant number of keywords that do not generate any clicks. Without at least one click we cannot estimate a conversion rate and are unable to evaluate the performance of keywords on an individual basis. Longer estimation periods (e.g., two months or more) are not attractive from a management perspective. Failure to identify "underperforming" keywords can quickly lead to significant losses. For example, one paid search manager we spoke with indicated that his company used to wait for 10 purchases before estimating a conversion rate. However, for keywords with a low number of clicks and a low conversion rate, that meant waiting almost a year to evaluate a keyword. In a fast moving advertising "market" like paid search, the dynamics of keywords change rapidly due to the competitive aspects of the underlying auction. The longer a company is not measuring the performance of its paid search strategy, the more money it is potentially losing.

Estimation Results

We estimate three models on the April paid search data. Model M1 is a click-through and conversion model linked by correlated unobserved error shocks. Model M2 is a click-through and conversion model linked by correlated unobserved error shocks using IV estimation. Model

M3 is a click-through and conversion model linked by correlated unobserved demand shocks using LIV estimation. We have 8,497 observations (daily) for 301 keywords, resulting in, on average, 28 observations per keyword (some keywords had zero clicks on certain days). The 8,497 observations represent 14,302 clicks. We observe 518 reservations (or successful choices), for an average conversion rate of 3.6%. For each model, we compute the log marginal density (LMD) as described in Newton and Raftery (1994). We find that model M3, which is estimated via the LIV method, has the best in-sample fit.

---Insert Table 1 About Here---

The estimates of the position coefficients in the click-through and conversion equations are presented in Table 2. Relative to the model that accounts for position endogeneity with observed instruments (i.e., model M2), the coefficient estimates from the model that does not account for endogeneity (i.e., model M1) are biased upwards. Consistent with the weak instrument argument, the coverage intervals for the point estimates in M2 are much wider, which indicates the standard errors of the estimates are larger. In the LIV approach (i.e., model M3), we find that the effect of position is smaller than in either model M1 or M2. For linear models, IV estimates with weak instruments have been shown to exhibit bias in the same direction as ordinary least squares estimates (Ebbes et al. 2005; Bound et al. 1995). Compared to the IV case, we also find that the precision of the estimates in model M3 has increased versus model M2. In sum, the empirical results suggest that the observed instruments are weak and that incorporating latent instruments provides an attractive alternative approach to handling position endogeneity.

---Insert Table 2 About Here---

Table 3 reports the estimated LIV category means and probabilities as well as the coefficient estimates for the observed instruments. If the LIV categories are not well separated (i.e., the category means and probabilities are approximately equal), the LIV estimator behaves in a similar fashion to the IV estimator with weak instruments. In our case, the LIV categories are well separated, helping to overcome the apparent weak instrument problem with the IV model. Table 4 presents the covariance matrix for the click-through, conversion, and LIV errors. We find significant covariance between the click-through and conversion error shocks, indicating the click-through and conversion decisions should be jointly modeled. We also find significant covariance between both the click-through and conversion error shocks and the LIV error.

---Insert Tables 3 and 4 About Here---

Table 5 presents the coefficient estimates for the LIV model. As expected, the intercept terms are strongly negative, reflecting the low probabilities of click-through and conversion. We find that position affects click-through; ads in higher positions are more likely to be clicked.[10] Interestingly, we also find that keywords with a higher position have a higher conversion rate, all else equal. Ghose and Yang (2009) report a similar empirical finding based on their weekly data. Chen and He (2009) argue that "position is a credible signal to the consumer" and recommend that firms should use position as such. Our empirical finding that an ad in a higher position attracts consumers with a higher propensity to convert lends support to this idea.

---Insert Table 5 About Here---

We now turn to the effect of (semantic) keyword characteristics on keyword performance. These are reported in Table 5. We find both similarities and differences across the click-through and conversion parameters. "Generic" has a negative effect on both click-through and conversion. A branded keyword generally performs better on all aspects of paid search (e.g., Rutz and Bucklin 2010) due to the likelihood that the user's decision process is more advanced at the time and that there is less intense competition for the keyword in the auction. "State" also has

[10] As position is measured from 1 (high) to 10 (low), the coefficient estimates for position are negative.

a negative effect on both click-through and conversion. The remaining keyword characteristics differ in their effects on click-through and conversion. While “#Words” has a positive effect in click-through, it has no effect on conversion. “US” has no effect on click-through, but a negative effect on conversion. “City” and “Hotel” have a negative effect in click-through, but no effect on conversion.

The results for the (semantic) keyword characteristics can be thought of as a toolbox for building keywords. According to our data and model, the best keyword for the company, is the *broadest branded* keyword, without any other characteristics (i.e., “BRAND NAME”). Indeed, the keyword “BRAND NAME” has the highest conversion rate for the company. Adding a word characteristic such as “Los Angeles” to the “BRAND NAME” narrows the resulting keyword (i.e., “BRAND NAME Los Angeles”) and lowers the conversion rate – again according to our model and data). Keyword characteristics also help the model better discriminate performance across keywords. They enable us to explain more of the observed variance in conversion rates and to better estimate conversion rates at the keyword level, especially in the long tail. In sum, incorporating keyword characteristics improves the measurement of individual keyword performance.

IMPLICATIONS FOR KEYWORD LIST MANAGEMENT

We now explore whether our proposed model would allow managers to improve the performance of a paid search campaign, at an individual keyword level, going forward. Specifically, we introduce and test an approach that uses the model-based conversion rates to manage the keyword list so as to improve the paid search campaign in future periods. We note that we are not proposing a method for optimizing a paid search campaign. Our approach focuses on individual keyword performance measurement, per se, as an essential building block for a more comprehensive optimization approach.

Based on the data from April 2004, we use our three estimated models to determine which keywords are "attractive" and should be retained in the campaign versus which keywords are "unattractive" and should be dropped. For each model a potentially different list of attractive keywords is generated due to differences in estimated conversion rates. Using the resulting keyword lists, we evaluate holdout performances in the May-June 2004 period and compare them to a status quo strategy that retains all 301 original keywords.

As the basis for keyword selection, we will use a cost-per-reservation threshold $\left(CPR_{threshold}\right)$ to discriminate between attractive and unattractive keywords. We use the estimated keyword-level conversion rates to calculate the average monthly cost-per-reservation $\left(CPR_{w}^{monthly}\right)$ for each keyword *w*. (This is computed by dividing a keyword's average monthly cost per click by its estimated conversion rate.) Note that our approach permits us to obtain this figure for all of the keywords on the campaign list. Without a model for conversion rate, the data only allow us to calculate monthly CPR for 84 keywords – the remaining 217 keywords in the long tail would be assigned an infinite monthly CPR. For each model we then rank order the keywords by estimated $CPR_{w}^{monthly}$ and retain those keywords for which estimated $CPR_{w}^{monthly} \geq CPR_{threshold}$. This produces different keyword lists corresponding to each model.

Evaluating Performance in a Hold-Out Period

The company did not change the keyword list that it used in April for the remainder of the quarter. This allows us to make an assessment of the performance of the "attractive" or retained keyword lists from each model in the May-June holdout period. In practice, the performance of the different keyword lists could be evaluated on profitability. Because we could not obtain confidential profit margin information for these data, we base our assessment on the comparative performance of the status quo strategy (keep all keywords) versus the lists generated by the model-based strategies. We do know that the average price range for a room is between \$75 and \$100 per night. Assuming an average of 1.5 nights per trip and a profit margin of 30%, the actual cost-per-reservation threshold should be in the range of \$30 - \$50 for the firm to avoid losses in paid search advertising. We investigated CPR thresholds ranging from \$20 - \$60 and found that the comparative performance was independent of the CPR threshold. Thus, we set \$30 as the $CPR_{threshold}$ and discuss findings based on that assumption.

Comparing the different model-based keyword lists with the status quo, we find that all three model-based strategies outperform it in terms of implied profits (returns to paid search in excess of the assumed threshold CPR of \$30). The results are reported in Table 6. The keyword list based on the proposed LIV model (M3) selects 151 keywords and increases profits over the status quo by 7% for May-June. This represents an improvement of \$12/keyword over two months. The comparisons in Table 6 also show that the LIV model performs better than the other two model specifications, M1 and M2.[11] Thus, not only does accounting for position endogeneity help to gauge the effect of position correctly, it also may provide practitioners with the potential for superior keyword selection and future campaign performance.

---Insert Table 6 About Here---

CONCLUSION

Paid search campaigns have become a crucial part of the marketing budget for most firms. Our objective in this paper has been to develop a modeling approach which we hope will

[11] We have also tested a model free click-through-rate (CTR) based approach to generating a keyword list. It entails choosing a CTR threshold x and keeping all keywords with CTR $\geq x$. Managing by CTR is popular in practice, in part because CTR is available for most keywords. However, the CTR threshold x is chosen arbitrarily. Managing by CTR allows for individual word performance evaluation, but it is not clear whether this strategy is consistent with profit maximization. Indeed, no CTR threshold was able to increase profits to the level of our proposed model.

aid companies in managing these campaigns at the keyword level. The performance of a paid search campaign can be evaluated by cost-per-sale, or in the case of a lodging chain, cost-per-reservation (CPR). However, most keywords do not lead to reservations on a regular basis. In our sample, only 84 keywords out of 301 led to reservations, making it impossible to calculate a meaningful CPR for the remaining 217 keywords based on the data alone. Should the company immediately drop the keywords in this long tail? Probably not, but it does need a valid method for estimating the future conversion probabilities for those keywords so that the list of words can be managed. This raises the question how best to produce such estimates and whether or not they could be used to improve campaign management going forward.

In developing a model to address these issues, we conceptualized conversion as a binary choice conditional on click. We then integrate a conversion model with a click-through model to account for the possible correlation across both decisions. Both models are designed to be estimated on the standard, aggregated paid search data available from search engines. An important strategic variable in these data and in the click-through and conversion models is the position of the text ad served in response to a search. Position is likely to be endogenous as it is determined by the paid search auction. Furthermore, typical paid search data sets report position as the average daily position, raising concerns over measurement error. An issue with modeling the auction either via a structural or reduced form approach is the fact that competitive bid information is not part of standard paid search data. In addition, firm bidding data is also often difficult to obtain. To address the issue of position endogeneity, we propose a latent instrumental variable approach. Our results show that our proposed model provides the best fit to the data when compared to models that treat position as exogenous or use only observed instruments. We find that endogeneity concerns about text ad position are valid and that failure to account for it is likely to lead to biased parameter estimates. Further, a standard instrumental variables approach indicates that available instruments in paid search data seem to be weak. Our proposed LIV model augments the observed instruments, thereby correcting the bias and without adversely impacting the precision of the estimates.

We use the model to estimate daily conversion probabilities for each keyword using a Bayesian shrinkage approach and the semantic similarities across keywords. Differences in conversion probabilities can be explained by the keyword itself (keyword heterogeneity), keyword performance measures (position) and keyword characteristics. We find that all of the above are predictive of conversion rate. Specifically, knowing which keyword was used to initiate the search helps to predict the probability of conversion. We use the estimated daily conversion probabilities to calculate the monthly cost per reservation for each keyword. Using the estimated conversion rates, we create keyword subsets by keeping the attractive keywords and dropping the remainder. Using holdout data we evaluate the implied profit performance of the model-based keyword lists against the status quo of maintaining the entire list. The LIV model-based list outperforms the status quo, as well as the alternative models. This result underscores the importance of accounting for position endogeneity in modeling keyword conversion performance.

We base our model on data that are readily available to paid search advertisers from the major US-based search engines. Our strategy can be implemented – and the performance of a campaign measured – based upon such a data set. The ability to measure allows the manager to test different position/cost combinations and decide, based on the measured outcomes, which ones are best. These data, however, also have notable limitations. First, there is no information on competition which precludes us from modeling the actual auction. Without modeling the auction we cannot determine bidding strategies. Unfortunately, it does not appear that such a dataset will be available to advertisers anytime soon. Companies do not have access to their competitors bidding strategies and, from the perspective of the search engines, it seems preferable to keep this confidential. Lastly, we do not have clickstream-type data and cannot model the consumer’s choice process. Overcoming these limitations are promising areas for future research. A visitor-centric panel dataset, if available, could be used to investigate whether and how different keywords attract distinct visitor segments. A potential project could study whether consumers’ characteristics and observed search behavior allow us to determine the extent to which factors like position can be explained by consumer heterogeneity. Also, research that investigates why position is a credible signal to buyers in terms of both click through and conversion would be valuable.

REFERENCES

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Bound, J., D.A. Jaeger, and R.M. Baker. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450.

Business Week (2007). So many Ads, so Few Clicks. November 12th, 38.

Chen, Y., & C. He (2009), "Paid Placement: Advertising and Search on the Internet", *Working Paper,* University of Colorado, CO.

ClickZ (2008). Benchmarking an Average Conversion Rate. http://www.clickz.com/showPage.html?page=3362641

Ebbes, P., M. Wedel, and U. Boeckenholt (2009), "Frugal IV Alternatives to Identify the Parameter for an Endogenous Regressor," *Journal of Applied Econometrics*, 24, 446-468.

Ebbes, P., M. Wedel, T. Steerneman, and U. Boeckenholt (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," *Quantitative Marketing and Economics*, 3, 365-392.

Edelman, B., & M. Ostrovsky (2007), "Strategic Bidder Behavior in Sponsored Search Auctions," *Decision Support Systems,* 43(1), 192-198.

Edelman, B., & M. Ostrovsky, M. Schwarz (2007), "Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review,* 97(1), 242-259.

Ghose, A. and S. Yang (2009), "An Empirical Analysis of Sponsored Search in Online Advertising," *Management Science*, 55(10), 1605-1622.

Goldfarb, A., & C. Tucker (2009), "Search Engine Advertising: Pricing Ads to Context," *Working Paper*, University of Toronto, ON.

Hausman, J.A. (1977), "Errors in Variables in Simultaneous Equations Models", *Journal of Econometrics*, 5, 389-401.

Hsiao, C. (1976), "Identification and Estimation of Simultaneous Equations Models with Measurement Error", *International Economic Review*, 17(2), 319-339.

Kuksov, D. and J. M. Villas-Boas (2008), "Endogeneity and Individual Consumer Choice," *Journal of Marketing Research*, 10(6), 702-714.

PricewaterhouseCoopers (2007). IAB Internet Advertising Revenue Report. http://www.iab.net/media/file/resources_adrevenue_pdf_IAB_PwC_2007_Final.pdf

Rutz, O.J., & R.E. Bucklin (2010), "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, forthcoming.

Willbur, K. C. and Y. Zhu (2009), "Click Fraud," *Marketing Science*, 28 (2), 293-308.

Yao, S. and C. F. Mela (2009), "A Dynamic Model of Sponsored Search Advertising," *Working Paper*, Northwestern University, Il.

Yang, S., Y. Chen, and G. Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," Quantitative Marketing and Economics, 1, 251-275.

Zhang, J., M. Wedel, and R. Pieters (2009), "Sales Effects of Visual Attention to Feature Ads: A Bayesian Mediation Analysis," *Journal of Marketing Research*, 46(3), 669-681.

Table 1: Model Fit Statistics

Model	Endogeneity[1]	Log-Marginal Density
M1	Not Applicable	-60,268
M2	Instrumental Variables (IV)	-60,237
M3	Latent Instrumental Variables (LIV)	-60,213

[1] Indicates how position endogeneity is addressed.

Table 2: Coefficient Estimates for Position

		Coefficient Estimates[1]	
Model	**Endogeneity**	**Click-Through**	**Conversion**
M1	Not Applicable	-0.43 (-0.49, -0.37)	-1.38 (-1.53, -1.04)
M2	Instrumental Variables (IV)	-0.31 (-0.61, -0.08)	-1.19 (-1.66, -0.69)
M3	Latent Instrumental Variables (LIV)	-0.25 (-0.31, -0.19)	-1.03 (-1.33, -0.78)

[1] We report the posterior mean and 95% coverage interval. Posterior mean estimates of the unobserved heterogeneity are omitted for brevity.

Table 3: Estimation Results – LIV Parameters

		Coefficient Estimates[1]
LIV Category Means	ω_1	6.27 (6.02, 6.54)
	ω_2	0.94 (0.88, 1.01)
LIV Category Probabilities	π_1	0.79 (0.53, 0.96)
	π_2	0.21 (0.07, 0.43)
Observed Instruments	$Position_{t-1}$	0.81 (0.80, 0.82)
	CPC_{t-1}	-0.07 (-0.15, -0.02)
	CTR_{t-1}	-1.28 (-1.54, -1.03)

[1] We report the posterior mean and 95% coverage interval.

Table 4: Covariance Matrix for Error Shocks-LIV Model

	ξ^{cl}	ξ^{con}	ξ^{LIV}
ξ^{cl}	0.18 (0.01)	0.16 (0.02)	0.12 (0.01)
ξ^{con}	-	2.27 (0.08)	0.34 (0.06)
ξ^{LIV}	-	-	1.46 (0.02)

Table 5: Estimation Results – Choice Parameters

		Coefficient Estimates[1]	
		Click-Through	**Conversion**
Covariates	*Intercept*	-2.55 (-2.83, -2.32)	-3.61 (-4.12, -3.17)
	Position	-0.25 (-0.31, -0.19)	-1.03 (-1.33, -0.78)
Semantic Word Characteristics	*Generic*	-0.98 (-1.24, -0.73)	-1.27 (-2.13, -0.17)
	# words	0.29 (0.19, 0.39)	-0.27 (-0.66, 0.08)[2]
	US	0.20 (-0.01, 0.48)[2]	-1.32 (-2.06, -0.56)
	State	-1.05 (-1.27, -0.81)	-1.09 (-1.73, -0.31)
	City	-0.54 (-0.86, -0.22)	0.21 (-0.70, 0.90)[2]
	Hotel	-0.39 (-0.60, -0.16)	-0.88 (-2.22, 0.25)[2]

[1] We report the posterior mean of the mean and 95% coverage interval. Posterior mean estimates of the unobserved heterogeneity or omitted for brevity
[2] Not significant.

Table 6: Hold-out Performance: Comparison of Model Strategies

	# Key-words	# Reser-vations	Cost	Cost per Reservation	Implied Profit
Do Nothing (Status Quo)					**$ 27,891.36**
All Keywords selected	301	1,348	$ 12,548.64	$ 9.31	
M1					**$ 28,288.75**
Keywords selected	129	1,214	$ 7,470.39	$ 6.70	
Keywords not selected	172	134	$ 3,694.88	$ 32.97	
M2					**$ 28,827.36**
Keywords selected	164	1,265	$ 8643.84	$ 7.21	
Keywords not selected	137	83	$ 2,521.43	$ 41.28	
M3					**$ 29,705.06**
Keywords selected	151	1,256	$ 7,470.39	$ 6.35	
Keywords not selected	150	92	$ 3,694.88	$ 40.71	

APPENDIX

Sampler

1) Generate β_w^{cl} and β_w^{con} using a random walk Metropolis-Hastings (MH) step based on the logit likelihood L^d given in equation (2) combined with a multivariate normal prior:

$$\beta_w^d \mid \ldots \propto L^d \cdot \left(\left(\beta_w^d\right) \sim MVN\left(\mu_\beta^d, \Sigma_\beta^d\right)\right), \text{ where } d \in \{cl, con\}.$$

2) Generate μ_β^d and Σ_β^d:

$$\mu_\beta^d \sim N(\tilde{\mu}_\beta^d, \tilde{\Sigma}_\beta^d), \text{ where } \tilde{\Sigma}_\beta^d = \left[n\left(\Sigma_\beta^d\right)^{-1} + \Sigma_b^{-1}\right]^{-1} \text{ and } \tilde{\mu}_\beta = \tilde{\Sigma}_\beta\left[n\left(\Sigma_\beta^d\right)^{-1}\overline{\beta} + \Sigma_b b\right]$$

$$\Sigma_\beta^d \sim IW\left(\nu_1 + n, \nu_2 + \sum_{w=1}^{n}(\beta_w^d - \mu_b^d)(\beta_w^d - \mu_b^d)'\right),$$

where $d \in \{cl, con\}$, $b = 0_L$, $\Sigma_b = 10^6 \times I_L, \nu_1 = 2$, and $\nu_2 = \nu_1 I_L$.

3) Generate ξ_{wt}^{cl} and ξ_{wt}^{con} using a random-walk MH step based on the likelihood as given by the equation (2) combined with a multivariate normal prior:

$$\begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \end{pmatrix} \mid \ldots \propto L^{cl} \cdot L^{con} \cdot \left(\begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \\ \xi_{wt}^{LIV} \end{pmatrix} \sim MVN\left(0, \Lambda\right)\right),$$

where $\xi_{wt}^{LIV} = pos_{wt} - \omega\,\gamma_{wt} - \phi x_{wt}^{IV}$ and $\Lambda = \begin{pmatrix} \sigma_{cl,cl} & \sigma_{cl,con} & \sigma_{cl,LIV} \\ \sigma_{cl,con} & \sigma_{con,con} & \sigma_{con,LIV} \\ \sigma_{cl,LIV} & \sigma_{con,LIV} & \sigma_{LIV,LIV} \end{pmatrix}$.

4) Generate ω :

$$\omega \mid \ldots \sim MVN\left(\mu_w, \Sigma\right),$$

where $\Sigma = \left(\Sigma_0^{-1} + \frac{1}{\sigma_{LIV,LIV} - s}\sum_{t=1}^{T}\gamma_{wt}\gamma_{wt}'\right)^{-1}$, $\mu_w = \Sigma\left(\Sigma_0\mu_0 + \frac{1}{\sigma_{LIV,LIV} - s}\sum_{t-1}^{T}\gamma_{wt}\,pos_{wt}^*\right)$, where $s=\sigma_{LIV,LIV}-\left(\sigma_{cl,LIV}\quad \sigma_{con,LIV}\right)\begin{pmatrix}\sigma_{cl,cl} & \sigma_{cl,con}\\ \sigma_{cl,con} & \sigma_{con,con}\end{pmatrix}\begin{pmatrix}\sigma_{cl,LIV}\\ \sigma_{con,LIV}\end{pmatrix}$,

$pos^{*}_{wt}=pos_{wt}-\phi x_{wt}$, $\mu_0=\underline{0}$ and $\Sigma_0=10^6 I_C$.

5) Generate Λ:

$$\Lambda\,|\,\ldots\sim IW\left(\begin{array}{l}\nu+WT,\\ \left(S+\sum_{w=1}^{W}\sum_{t=1}^{T}\left[\begin{pmatrix}\xi^{cl}_{w,t}\\ \xi^{con}_{w,t}\\ pos_{wt}\end{pmatrix}-\begin{pmatrix}0\\ 0\\ \omega\gamma_{wt}-\phi x_{wt}\end{pmatrix}\right]\left[\begin{pmatrix}\xi^{cl}_{wt}\\ \xi^{con}_{wt}\\ pos_{wt}\end{pmatrix}-\begin{pmatrix}0\\ 0\\ \omega\gamma_{wt}-\phi x_{wt}\end{pmatrix}\right]'\right)\end{array}\right),$$

where $\nu=3+3$ and $S=\nu I_3$.

6) Generate γ_{wt} as a categorical variable with a posterior probability given by

$\Pr\left(\gamma_{wt}=c\right)=\dfrac{L\left(\omega,\xi,\gamma^{(c)}_{wt},\phi,\Lambda\right)\times\pi_c}{\sum_{i=1}^{C}L\left(\omega,\xi,\gamma^{(i)}_{wt},\phi,\Lambda\right)\times\pi_i}$, where $\gamma^{(c)}_{wt}$ denotes that keyword w at time t is

assigned to class c, $L\left(\omega,\xi,\gamma^{(c)}_{wt},\phi,\Lambda\right)$ is the likelihood function of the LIV model

evaluated at $\gamma^{(c)}_{wt}$, and π_c is the prior probability of class c membership.

7) Generate π:

$\pi\,|\,\gamma\sim Dirichlet(1+K_1,\ldots,1+K_C)$, where the vector K_c denotes the sum of $\gamma^{(c)}_{wt}$ over all c,

i.e., $K_c=\sum_{w=1}^{W}\sum_{t=1}^{T}\gamma_{w,t}(c=k)$.