

OLIVER J. RUTZ, RANDOLPH E. BUCKLIN, and GARRETT P. SONNIER*

The authors present a modeling approach to assess the purchase conversion performance of individual keywords in paid search advertising. The model facilitates estimation of daily keyword conversion and click-through rates in a sparse data environment while accounting for the endogenous position of the text advertisement served in response to a search. Position endogeneity in paid search data can arise from both omitted variables and measurement error. The authors propose a latent instrumental variable approach to address this problem. They estimate their model on keyword-level paid search data containing daily information on impressions, clicks, and reservations for a major lodging chain. They find that higher positions increase both the click-through and conversion rates. When advertisements are served in higher positions, approximately one-third of new conversions is due to increased click-through while approximately two-thirds are due to increased conversion rates. The authors show that the keyword list generated on the basis of their estimated conversion rates outperforms the status quo list as well as lists generated by observed conversion and click-through rates.

Keywords: Internet, advertising, paid search, Bayesian methods

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

Paid search advertising allows companies to address consumers directly during their electronic search for products or services. When consumers search the web with the help of an Internet search engine, the terms they enter 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 advertisement that will be shown when a consumer searches on those keywords. In paid search, advertisers bid their willingness to pay for a click on a paid search advertisement. An automated auction-type

algorithm then determines position (e.g., first, third) of the advertisement in the sponsored listings section of the results page.

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

Search engines routinely provide daily information to firms on their own paid search advertising, and advertisers typically manage their paid search campaigns on the basis of such data. Evaluating individual keyword performance can be approached in several ways. First, we might consider taking a direct marketing type of approach, in which we calculate the marginal benefit of spending for each keyword, comparing advertising-related profit per sale with advertising-

*Oliver J. Rutz is Assistant Professor of Marketing, Foster School of Business, University of Washington (e-mail: orutz@uw.edu). Randolph E. Bucklin is the Peter W. Mullin Professor, UCLA Anderson School, University of California, Los Angeles (e-mail: rbucklin@anderson.ucla.edu). Garrett P. Sonnier is Assistant Professor of Marketing, Red McCombs School of Business, University of Texas at Austin (e-mail: garrett.sonnier@mcombs.texas.edu). The authors thank a collaborating firm for providing the data used in this study. Michel Wedel served as associate editor for this article.

related cost per sale.¹ If that difference is positive, a keyword generates a positive return for the advertiser from a direct marketing perspective. Using standard paid search data, we can compute cost per sale as the ratio of cost per click (CPC) 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 the observed conversion rate is nonzero, the cost-per-sale measure may be based on very few observations and thus subject to substantial error. Because average conversion rates for keywords in paid search are low, this measurement problem occurs often. For example, the average conversion rate from a paid search click to purchase in the travel industry was 2.1% in the first quarter of 2004 (Eisenberg 2004). Thus, even on a monthly basis, most keywords simply do not generate any sales. This precludes calculating cost per sale at the keyword level, rendering the approach untenable. 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 significant error.

Given the difficulties in evaluating cost per sale at the keyword level, a second approach is to turn to ad hoc model-free strategies. For example, evaluating paid search at the campaign level, aggregating across all keywords, is a crude but straightforward strategy. The manager can compare spending on the campaign versus sales attributable to the campaign. However, this does not help allocate ad spending across individual keywords, especially those in the long tail.² The manager may refine this approach 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 an arbitrary decision rule, (e.g., keep all keywords with $CTR \geq 1\%$). Furthermore, these approaches do not allow the manager to assess the conversion performance of individual keywords.

A third approach is to use a model-based strategy to assess keyword conversion rates. However, standard paid search data present the modeler with numerous challenges. The major search engines (Google, Yahoo!, and Bing) provide advertisers with data on their own campaign performance (i.e., impressions, clicks, position, and cost). These data are typically aggregated to the daily keyword level. However, none of the major search engines provide competitive data or allow advertisers to infer which competitors were listed along with their own advertisements. To model paid search advertising performance, the effect of the position of the text advertisement on the consumer's click and conversion decisions needs to be addressed. A key challenge in doing so is that position is likely endogenous. 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 an omitted variables issue. In addition, typical paid search data report position as the daily average position at the keyword

level; this raises a measurement error issue. Both omitted variables and measurement error can induce regressor-error dependencies and bias the estimates of model coefficients.

In one of the few empirical models of paid search advertising to be published to date in marketing, Ghose and Yang (2009) model keyword performance using simultaneous equations. In their 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 both own and competitive bid information, Ghose and Yang use CPC as a proxy for bid in a reduced-form model of the auction. To identify the model, Ghose and Yang exclude contemporaneous position from their CPC equation, treating CPC as exogenously determined. However, in paid search auction algorithms, CPC and position are codetermined outcomes of the auction. CPC is determined according to the advertiser's bid, the advertiser's past ad performance, and competitive bids and *competitive* ad performance. As noted previously, the competitor information is not observed. In addition, Google reports CPC as an average, subjecting the observed CPC to measurement error. Together, these issues may induce an errors-in-variables problem, which can complicate identification and estimation of simultaneous equations systems (Hausman 1977; Hsiao 1976). Given the potential errors-in-variables problem and the general unavailability of details on Google's auction, the assumption that CPC is exogenous may be strong.

Despite the concerns over the endogeneity of CPC, Ghose and Yang (2009) report no contemporaneous correlation between position and CPC in their data. However, their data are aggregated into weekly observations, which may mask the correlation. Weekly aggregation raises another important point. Search engines now make data available on a daily and even hourly level, which indicates industry demand for finer levels of temporal aggregation. While the aggregation bias becomes less pronounced as the data become finer grained, sparseness becomes a more significant problem. In particular, low click-through and conversion rates lead to data sets dominated by zeros. With daily data, a significant problem arises with the position and CPC equations in the simultaneous equations approach. While each search will be connected with a position, CPC is only recorded if a consumer clicks on the advertisement. This leads to CPC data with a significant mass concentrated at zero while position is always observed. A position equation that uses current CPC as a covariate is also questionable because current CPC is zero if no click occurred. However, position is a function of the actual unobserved bid amount, which was not zero. As major search engines make daily, and even hourly, paid search data available, sparse data with few clicks and conversions are likely to be the rule rather than the exception going forward. For these reasons, the CPC assumptions made in the simultaneous equations approach may be problematic, making it worthwhile to consider alternative models.

The objective of this article is to develop a model-based approach to assessing the conversion performance of individual keywords—on a daily level—that directly addresses the endogeneity of position in both the consumer's click and conversion decision equations. We model conversion as a binary choice decision conditional on a click; that is, a con-

¹For ease of exposition, in this article, we use cost per sale instead of advertising-related cost per sale.

²Our data set contains 301 keywords, 84 of which generate reservations in the one-month estimation period. These 84 keywords account for approximately 85% of all clicks.

sumer has clicked on the paid search advertisement 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. We use a Bayesian shrinkage estimator to infer conversion rates for keywords with very few or no conversions and to improve the estimated conversion rates for other keywords. We do this by exploiting the similarity among keywords to produce a shrinkage-based estimate of conversion rate 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, Wright, and Yogo 2002). We therefore propose to account for the endogeneity of position using a latent instrumental variable (LIV) approach (Ebbes, et al. 2005; Zhang, Wedel, and Pieters 2009). The LIV technique uses a latent variable model to account for regressor-error dependencies and, as such, circumvents the issues of instrument availability, weakness, and validity.

Our intended contribution is both methodological and substantive. From a methods perspective, we provide an alternative to the simultaneous equations approach (Ghose and Yang 2009) to measure individual keyword performance. Specifically, we address the endogeneity of keyword position with latent instrumental variables. We believe that the LIV approach is well suited to address position endogeneity in the sparse data conditions that might hamper the simultaneous equation approach when using daily paid search data. We find that ignoring the endogeneity problem biases the estimates of the effect of position on click-through and conversion. A simultaneous equations approach applied to our daily paid search data, which treats CPC as exogenous, does not alleviate the position endogeneity bias. Although IV estimation alleviates some of the bias in position estimates, it does so at the cost of decreased precision of the estimates. This is consistent with the behavior of estimators utilizing weak instruments. Concerns also linger over the validity of the available observed instruments given the issues of measurement error and the lack of information on the paid search auction. In contrast, LIV estimation ably addresses the endogeneity of position without adversely affecting precision.

There is an ongoing debate in the paid search practitioner community over the effect of position on conversion (Ballard 2011; Brooks 2004; Van Wagner 2010). Google itself has argued that position has little to no effect on conversion (Friedman 2009). In the academic literature, the nascent empirical findings on the effect of position on conversion are mixed. Some researchers have found that conversion rates improve at higher positions (Ghose and Yang 2009), while others have found the opposite effect (Agarwal, Hosanagar, and Smith 2011). Substantively, we find in our data that conversion rates improve as ad position improves. More significant, we find that the elasticity of conversions with respect to position (holding constant the number of

clicks) is actually higher than the elasticity of clicks with respect to position. We use the model to decompose the increase in conversions from position improvements to that which is due to a better conversion rate, holding constant the number of clicks, and that which is due to the higher number of clicks. We find that 35% of the increased conversions are due to the increase in clicks, and 65% are due to the higher conversion rate (holding the number of clicks constant). To the best of our knowledge, we are the first to demonstrate this surprising empirical finding. We also demonstrate the value of model-based keyword list management by evaluating the performance of an existing list of paid search keywords in a holdout data set. Our model yields estimates of conversion rates for all keywords, even those for which the observed data are sparse and performance difficult to assess. We demonstrate in a holdout sample that the keyword list generated by our model-based approach yields superior results compared with alternative model-based and model-free strategies, such as managing by observed conversion or CTR.

We structure the remainder of the article as follows: We begin with a brief overview of paid search advertising and discuss some relevant literature. We then present our modeling approach, data set, 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 conclude with a discussion of the limitations of our approach and suggest further research in the realm of search engines and marketing.

ANALYZING PAID SEARCH ADVERTISING

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." The advertiser has selected this keyword and created a text advertisement for the offering. Advertisers bid the dollar amount they are willing to pay for a click on a text advertisement served in response to a search for this keyword. The actual CPC and position of the text advertisements 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 CTR). The search results page will display nonsponsored results (organic search results) and the text advertisement from the focal advertiser next to text advertisements from other advertisers (paid search results). From the advertiser's perspective, one impression attributed to the keyword "Hotels Los Angeles" has been generated. The consumer views the advertiser's text advertisement and chooses to click on a link provided in it. He or she 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 or she does so, the keyword "Hotels Los Angeles" is now associated with a reservation (i.e., a conversion).

Paid search provides academics with a wide array of questions to investigate, such as keyword choice, campaign management, and paid versus organic search, among others. Dhar and Ghose (2010) provide an excellent overview of the state of paid search research and discuss areas for fur-

ther research on the topic. Despite the impressive growth and scale of paid search advertising, there is scant published work on paid search, especially in marketing. Recent theoretical studies investigate paid search as a pure second price auction (Edelman and Ostrovsky 2007; Edelman, Ostrovsky, and Schwarz 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. From an empirical perspective, Goldfarb and Tucker (2011) investigate how regulation affects paid search ad pricing and show that search engines profit when regulation limits the advertisers' other advertising options. Yao and Mela (2011) develop a dynamic structural model of paid search advertising for a small business-to-business search engine specializing in industrial software.³ Rutz and Bucklin (2011) document spillover effects between generic and branded sets of keywords in an aggregate-level model.

Related to our work, recent studies have focused on evaluating keyword performance in a direct marketing framework. Using a paid search data set for a retail chain that advertises on Google, Ghose and Yang (2009) propose a simultaneous equations 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 CPC. The authors use CPC as a proxy for the unobserved advertiser's bid. As we noted previously, using a proxy for the firm's own bid, along with missing competitive bid data and measurement error, may induce an errors-in-variables problem.⁴ Furthermore, the exclusion restrictions necessary for identification of systems of simultaneous equations are sometimes difficult to generalize to other modeling settings.⁵ Rutz and Trusov (2011) take a different approach to evaluating keyword performance by estimating a consumer-level model on Google AdWords data using a data augmentation framework. Their focus is on the major keywords in a campaign, and they do not address the sparseness issues that are a central motivation of our study.

An ongoing debate among paid search practitioners is whether conversion rates vary by keyword position. In a 2009 blog post, Google chief economist Hal Varian argues that conversion rates do not vary much by position (Friedman 2009). Although some evidence from the practitioner community suggests conversion rates increase at higher positions (Brooks 2004), other evidence has been found to support Varian's contention (Ballard 2011; Van Wagner

2010). The empirical evidence from the academic marketing literature that measures the effect of position on conversion is similarly mixed. Ghose and Yang (2009) find that higher position yields higher conversion rates. The argument for higher position yielding higher conversion rates stems from a potential association between position and quality or trust perceptions (Ghose and Yang 2009). However, Agarwal, Hosanagar, and Smith (2011) argue that Ghose and Yang's (2009) findings are perhaps driven by aggregation across categories as well as keyword positions that range from 1 to 131. Using data from a field experiment with randomized bidding, Agarwal, Hosanagar, and Smith focus on the top seven positions and find conversion rates increase at lower positions.⁶ The argument for higher position yielding lower conversion rates stems from the notion that buyers higher up the purchase funnel often click on advertisements in higher positions without purchasing while buyers lower down the purchase funnel are more likely to visit lower positions.

MODELING APPROACH

As we discussed previously, a major problem with measuring conversion in paid search is that it occurs infrequently, particularly in the tail of an advertiser's keyword list. Thus, for many keywords, advertisers cannot simply calculate conversion rates on the basis of observed data, making it impossible to evaluate which keywords are profitable (i.e., which keywords generate margins on purchases that exceed advertising costs). The goal of our modeling approach is to address this 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 typically 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.⁷ Google, as well as the other major search engines, only provides keyword-level aggregate data on a daily basis. Thus, firms are not able to tie their site visitors to individual paid search advertisements in terms of specific position and CPC. 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 before purchase (i.e., leverage demographic information).⁸

³Their model makes use of competitive bid data, which are not available in the standard paid search data provided by the major Internet search engines.

⁴Firm 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. Whether future competitive pressures will entice search engines to share more of their data with their advertisers remains to be seen. (We thank an anonymous reviewer for raising this point.)

⁵As we noted at the beginning of the article, Ghose and Yang (2009) exclude contemporaneous position from their CPC equation. Although the authors find no contemporaneous correlation between position and CPC, their data are aggregated into weekly observations, which may mask the correlation. In our daily data, we find significant contemporaneous correlation between position and CPC.

⁶Agarwal, Hosanagar, and Smith (2011) also report a similar finding using Ghose and Yang's (2009) simultaneous equations approach for high-specificity keywords from an online retailer's campaign.

⁷Clickstream data assign each visitor to the site an individual ID (cookie). These data would allow researchers to connect a reservation to a specific click and that click, in turn, to a specific impression 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.

⁸Alternatively, it is conceivable to have 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 than with the data provided by the search engine.

In our model, we investigate whether the conversion probability given click-through to the advertiser’s website can be measured using available keyword-level 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 between 0% and 50%, with an average of .9% and a standard deviation of .07. This variation in conversion could indicate that different keywords “attract” different types of consumers who ultimately purchase at different rates. (Similar patterns are also present in multiple data sets 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 enables 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 enables us to explore whether conversion rates differ systematically across individual keywords and, if so, whether we can explain those differences using observable keyword covariates.

One approach to modeling conversion would be to condition on 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. Although 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 using correlated error shocks (Berry, Levinsohn, and Pakes 1995). As we previously noted, the position of the text advertisement, 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 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 employ the LIV approach (Ebbes et al. 2005; Zhang, Wedel, and Pieters 2009) in 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 by 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. According to the binary logit model, the conversion probability, P_{wt}^{con} , for keyword w at time t is given by

$$(1) \quad P_{wt}^{con} = \frac{\exp(\text{pos}_{wt}\alpha_w^{con} + x_{wt}^{con'}\beta_w^{con} + \xi_{wt}^{con})}{1 + \exp(\text{pos}_{wt}\alpha_w^{con} + x_{wt}^{con'}\beta_w^{con} + \xi_{wt}^{con})}$$

where pos_{wt} is keyword position, x_{wt}^{con} is a vector of keyword-level covariates, $\theta_w^{con} = [\alpha_w^{con} \beta_w^{con}]'$ is a vector of keyword level parameters with $\theta_w^{con} \sim N(\bar{\theta}^{con}, \Sigma_{\theta}^{con})$, and ξ_{wt}^{con} is a zero-centered demand shock.

In a standard application of the logit model, the data contain one choice outcome (observation) per period. Because 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 period. Therefore, we use the following likelihood function to estimate the parameters of the logit model:

$$(2) \quad \text{Likelihood}^{con} = \prod_t \prod_w (P_{wt}^{con})^{\text{conversions}_{wt}} (1 - P_{wt}^{con})^{(\text{clicks}_{wt} - \text{conversions}_{wt})}$$

where t is time and w is keyword.

Model Specification: Linking Click-Through and Conversion

Before consumers can arrive at the company’s landing page through paid search, they must decide whether to click on the paid text advertisement of the company displayed in response to their search. Although 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 we model the click-through decision conditional on a visitor searching for a keyword that has been bid on by the advertiser.⁹ Based on the binary logit model, the click-through probability, P_{wt}^{cl} , for keyword w at time t is given by

$$(3) \quad P_{wt}^{cl} = \frac{\exp(\text{pos}_{wt}\alpha_w^{cl} + x_{wt}^{cl'}\beta_w^{cl} + \xi_{wt}^{cl})}{1 + \exp(\text{pos}_{wt}\alpha_w^{cl} + x_{wt}^{cl'}\beta_w^{cl} + \xi_{wt}^{cl})}$$

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

We link the click-through and the conversion model given by Equations 1 and 3 through the demand shocks, ξ_{wt}^{cl} , ξ_{wt}^{con} , as follows:

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

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

Model Specification: Accounting for the Auction

In paid search, a modified second price auction determines a keyword’s ad position as well as the CPC the adver-

⁹We refrain from modeling the consumer’s choice of keywords. This would entail modeling the choice among the large number of possible search terms a consumer could use. In addition, we only observe searches on the keywords for which the company has bid.

tiser paid. 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 advertisement in terms of CTR, 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 of the whole campaign are taken into account. From an advertiser’s perspective, the paid search auction can be best described as a black box.¹⁰ In addition to not knowing the precise workings of the auction, advertisers do not have access to basic competitive information such as who else was listed and in which position competitors were listed. Although changes in the competitive landscape may result in search engines becoming more transparent regarding the inner workings of their auctions, transparency does not seem to be on the near horizon.

Position is considered 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 typically investigated from top to bottom. Assuming that consumers inspect the sponsored listings until they find an advertisement that meets the threshold for clicking, a paid search advertisement in a higher position will, most likely, be viewed by more consumers than an advertisement in a lower position. Another argument for the importance of position is derived from the search engines’ auction mechanism. Supposedly, a “better fitting” advertisement (i.e., an advertisement that has a 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, it becomes a valuable input to the click-through decision as well as the conversion decision.

Although position is strategically important to advertisers, treating it as an exogenous covariate in the click and conversion equations is a questionable modeling approach. 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. Because competitive bid information is unavailable, omitted variable concerns loom. Last, 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

smaller search engine for specific software products (Yao and Mela 2011). As we noted previously, however, it is unlikely that competitive bidding data will be made available by the major search engines in the near future. 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. As we discussed previously, the approach Ghose and Yang (2009) propose has certain limitations with respect to identification and its ability to handle sparse data in the position and CPC equations. 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 IV estimation. In this case, we correlate the unobserved demand shocks ξ_{wt}^{cl} and ξ_{wt}^{con} with the IV equation as described next. We express position as a linear function of observed instruments:

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

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 covariance structure between the error term and the click-through and conversion demand shock, ξ_{wt}^{cl} and ξ_{wt}^{con} , as follows:

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

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

The IV approach requires that valid, observed instruments be available. Instruments that are correlated with the error term in the model are invalid. Ideally, instruments are highly correlated with the endogenous covariate. However, 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. While this fundamental tension is regularly a chief concern, the mere availability of observed instruments is also an issue. A popular candidate for instruments is lagged variables such as lagged position and lagged CPC. However, if these variables are measured with error (as is the case with paid search data sets that report only average position and average CPC) or if the lagged variables are correlated with the contemporaneous position, lagged variables are invalid as an instrument (Angrist and Krueger 2001). Weak or invalid instruments can result in parameter estimates that are more biased than those obtained by simply ignoring the endogeneity problem (Stock, Wright, and Yogo 2002; Zhang, Wedel, and Pieters 2009). Rather than relying on the availability of valid and nonweak instruments, we use the LIV framework (Ebbes et al. 2005; Ebbes, Wedel, and Boeckenholt 2009; Zhang, Wedel, and Pieters 2009) to account for position endogeneity in our click-

¹⁰No competitive information is given by the search engines. In addition, the proprietary auction algorithms are described in vague terms. Google, for example, describes its auction mechanism in 2004 as being determined by maximum CPC bid and a keyword’s 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: (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 (see <http://support.google.com/adwords/bin/answer.py?hl=en&answer=49174>).

through and conversion models. The approach is designed to circumvent the issues of instrument availability, validity, and weakness.

The LIV estimator belongs to the family of frugal IV estimators that do not require observed instruments. This family includes the higher moments estimator, the identification through heteroskedasticity estimator, and the LIV estimator (Ebbes et al. 2005; Ebbes, Wedel, and Boeckenholt 2009). An advantage of the LIV estimator in our setting is that it is a likelihood-based approach amenable to Markov chain Monte Carlo estimation (for a more detailed discussion of frugal IV estimators, see Ebbes, Wedel, and Boeckenholt 2009). To identify the parameter(s) of interest, the LIV model exploits nonnormality of the endogenous regressors. In a recent marketing application, Zhang, Wedel, and Pieters (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, conversion). Although the framework was originally developed in a linear regression setting, we apply it to our choice model framework for analyzing paid search advertising. Using data augmentation, the LIV model introduces a latent categorical variable with C categories. Ebbes et al. (2005) show that the model is identified by the likelihood with the requirement that the number of categories is at least two and that the category means are different. The intuition for this result follows from the observation that with only a single category, the systematic portion is constant, and there is no information with which to estimate the parameter of interest. It also follows that as any of the category probabilities approach 0 or 1, the latent instrument is weakened in its ability to identify the parameter of interest. Ebbes et al. (2005) also show that when the distribution of the endogenous variable tends toward a unimodal symmetric distribution, identification becomes more problematic.

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

$$(7) \quad \text{pos}_{wt}^{\text{LIV}} = \gamma'_{wt}\omega + \xi_{w,t}^{\text{LIV}}$$

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

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

where the elements of the covariance matrix are to be estimated (for details on the sampling procedure, see the Appendix).

EMPIRICAL APPLICATION

Data

Our data encompass one calendar quarter of the paid search campaign for a major regional lodging chain on the Google search engine. The company used a list of 301 keywords in its campaign. The daily data span April 2004–June 2004. The data consist of the standard information advertisers receive from Google and 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 CPC (in U.S. dollars), average position served (ranking, e.g., 2.3), and number of impressions and clicks. The additional third-party data provide daily information on the number of reservations for each keyword.

We describe the keywords on the list by introducing semantic keyword characteristics. The keywords have certain common characteristics specific to the lodging industry (e.g., a city or a holiday destination is included). These characteristics can be used to explain differences in click-through and conversion performance across keywords. Note that the goal is not the selection of a single “optimal” keyword but rather to evaluate the performance of an existing keyword list used by the firm. Although some keywords perform better than others in terms of generating click-through and conversion, the more relevant question is this: Do the profits generated by the keywords on which bids are placed exceed the cost per sale? This will depend on the estimated conversion rates for the different keywords, explained in part by keyword characteristics. Table 1 presents some basic information on the keyword characteristics used in our analysis.

For both the click-through and the conversion decisions, 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, Chen, and Allenby 2003). As we noted previously, standard paid search data lack good candidate instruments. Similar to using lagged prices and cost

Table 1
KEYWORD CHARACTERISTICS

Keyword Characteristics	Description	Number of Keywords
Branded	Keyword includes company brand name	99
U.S.	Keyword is for a U.S. location.	223
State	Keyword includes a state.	52
City	Keyword includes a city.	210
Hotel	Keyword includes lodging related phrase such as hotel, model, or accommodation.	222
Number of words ^a	Number of words used in the search term.	—

^aThe mean number of words is 2.65 with a variance of .55.

information as instruments for price, we use lagged position, lagged CPC, and lagged CTR as instruments for position in the IV model.¹¹ While lagged variables are always available as instruments, the ease of availability is accompanied by concerns about instrument validity. For example, position and CPC are reported as averages. Thus, measurement error concerns cast doubt on the validity of lagged position and CPC as instruments. In addition, as we noted previously, position and CPC are codetermined in the auction. Moreover, past CTR also factors into the auction. In summary, there are several reasons to be concerned about the validity of these instruments. Furthermore, as concern about validity abates, concern about instrument weakness rises. As we discussed previously, the LIV method does not require the availability of observed instruments. For our LIV model, we use latent instruments, as defined in Equation 7, which alleviates concerns with regard to instrument validity and weakness.

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 CTR (percentage of impressions that led to a click) was .6%, and the average conversion rate (percentage of clicks that led to a reservation) was 3.6%. The average CPC was \$.36, and the average cost per reservation (CPR) was \$9.86. We use the data from May and June 2004 as a holdout sample (2,983,085 impressions, 38,878 clicks, 1348 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 CPR (\$9.31), there is little difference when compared with the estimation sample.

We believe that reviewing a paid search campaign on a monthly basis is a reasonable policy. 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 ten 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” such as 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 four models on the April paid search data. Model M0 consists of a click-through and conversion equation that does not account for position endogeneity. Model M1 is a simultaneous equations approach that includes

equations for position and CPC as in Ghose and Yang (2009). Model M2 is a click-through and conversion model linked by correlated unobserved error shocks using IV estimation with the observed instruments. Model M3 is a click-through and conversion model linked by correlated unobserved demand shocks using LIV estimation.¹² We have 8497 observations (daily) for 301 keywords, resulting in an average of 28 observations per keyword. (Some keywords had 0 clicks on certain days.) The 8497 observations represent 14,302 clicks. We observe 518 reservations (or successful choices), for an average conversion rate of 3.6%. Table 2 reports model fit statistics. For each model, we calculate the deviance information criterion (DIC; Spiegelhalter et al. 2004). We find that model M3 has the best in-sample fit according to the DIC.

Table 3 presents the estimates of the position coefficients in the click-through and conversion equations. We find relatively little difference in the estimated position effects across models M0, which treats position as exogenous, and M1, which uses the simultaneous equation approach. The posterior estimates of the mean effect of position are close across these two models, and the 95% coverage intervals have a large degree of overlap. Furthermore, we find that the CPC and position errors are *not* correlated with the click-through and conversion demand shocks. Thus, whatever slight differences between M0 and M1 exist, they are not attributable to M1 expunging endogeneity bias.

Estimates of the position effects from model M2, which uses observed instruments, are smaller in magnitude when

¹²We estimate the model with C = 2, C = 3, and C = 4 categories. We find that a model with C = 3 provides the best fit to the data. Note that Ebbes et al. (2005) show that the LIV model is robust to the misspecification of C.

Table 2
MODEL FIT STATISTICS

Model	Endogeneity ^a	DIC ^b
M0	No endogeneity correction	121,496
M1	Simultaneous equations ^c	121,489
M2	IV	121,473
M3	LIV	121,465

^aIndicates how position endogeneity is addressed.

^bDeviance information criterion (Spiegelhalter et al. 2004).

^cBased on Ghose and Yang (2009).

Table 3
COEFFICIENT ESTIMATES FOR POSITION

Model	Endogeneity ^a	Coefficient Estimates ^b	
		Click-Through	Conversion
M0	No endogeneity correction	-.43	-1.38
		(-.49, -.37)	(-1.53, -1.04)
M1	Simultaneous equations	-.39	-1.37
		(-.48, -.30)	(-1.76, -1.03)
M2	IV	-.31	-1.19
		(-.61, -.08)	(-1.66, -.69)
M3	LIV	-.24	-.98
		(-.31, -.17)	(-1.25, -.71)

^aIndicates how position endogeneity is addressed.

^bWe report the posterior mean and 95% coverage interval. We omit posterior mean estimates of the unobserved heterogeneity for brevity.

¹¹Note that in 2004 Google had not yet introduced the Quality Score, which could also serve as an observed instrument.

compared with M0. However, concerns remain about the validity and weakness of the observed instruments, which are lagged values of position, CPC, and CTR. Although it is difficult to definitively assess instrument validity, we believe skepticism is warranted in light of measurement error and the unknowns in the paid search auction process. We also note that the coverage intervals on the position coefficients for model M2 are extremely wide by comparison, indicating a relative lack of precision. This is consistent with the behavior of a weak instrument, which alleviates bias at the expense of precision.

In model M3, our LIV model, posterior mean estimates of position effects on click-through and conversion are 44% and 29% smaller, respectively, than in model M0. The posterior mean estimates of the position effects in model M0 are not contained in the 95% coverage intervals for model M3. For the click-through models M0 and M3, the 95% coverage intervals do not overlap at all; for the conversion models, there is slight overlap of the intervals. The estimates from model M3, the LIV model, are also smaller than that of model M2, the observed IV model, and the coverage intervals also are narrower. Collectively, these results indicate that the LIV model provides a good solution to the problem of position endogeneity in daily paid search data.

Table 4 reports the estimated LIV category means and probabilities. As we previously discussed, if the LIV categories are not well separated (i.e., the category means and probabilities are approximately equal), identification of the parameter on the endogenous covariate is hampered. In our case, the LIV categories are well separated, aiding identification. Table 5 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 that 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.

Table 6 presents the coefficient estimates for the LIV model. As we expected, the intercept terms are strongly negative, reflecting the low probabilities of click-through and conversion. We find that position affects click-through: Advertisements in higher positions are more likely to be clicked.¹³ Notably, we also find that keywords with a higher position have a higher conversion rate, all else being 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 advertisement in a higher position attracts consumers with a higher propensity to convert lends support to this idea. Using the model results, we simulate the effect of a 10% improvement in average positions on the expected number of clicks and the expected number of conversions. We find that a 10% position improvement generates a 7% improvement in the expected number of clicks (i.e., a click elasticity of .7). Holding the number of clicks constant, a 10% position improvement generates a 12% improvement in expected conversions (i.e., a conversion elasticity of 1.2). Decomposing the total increase in expected conversions in our data, we find that 35% of the conversion improvement is due to the improvement in clicks and 65% is due to the improved conversion rate (holding constant the number of clicks). Our finding that the majority of the improvement in the overall number of conversions is due to the improved

¹³Because position is measured from 1 (high) to 10 (low), the coefficient estimates for position are negative.

Table 4
ESTIMATION RESULTS: LIV PARAMETERS

		<i>Coefficient Estimates^a</i>
LIV category means	ω_1	12.63 (12.47, 12.79)
	ω_2	6.31 (6.28, 6.36)
	ω_3	2.59 (2.52, 2.64)
LIV category probabilities	π_1	.04 (.03, .05)
	π_2	.61 (.59, .62)
	π_3	.35 (.34, .36)

^aWe report the posterior mean and 95% coverage interval.

Table 5
COVARIANCE MATRIX FOR ERROR SHOCKS: LIV MODEL

	ξ^{cl}	ξ^{com}	ξ^{LIV}
ξ^{cl}	.15 (.01)	.08 (.01)	.06 (.02)
ξ^{com}	—	.87 (.04)	.25 (.05)
ξ^{LIV}	—	—	1.52 (.03)

Table 6
ESTIMATION RESULTS: CHOICE PARAMETERS

	<i>Coefficient Estimates^a</i>	
	<i>Click-Through</i>	<i>Conversion</i>
<i>Covariates</i>		
Intercept	-3.17 (-3.41, -2.79)	-3.45 (-4.51, -2.26)
Position	-.24 (-.31, -.17)	-0.98 (-1.25, -.71)
<i>Semantic Word Characteristics</i>		
Generic	-1.55 (-1.92, -1.31)	-1.63 (-2.60, -.67)
Number of words	.25 (.13, .39)	-.04 (-.46, .35) ^b
U.S.	.21 (-.14, .55) ^b	.06 (-.65, .72) ^b
State	-.54 (-.92, -.17)	-2.96 (-3.60, -2.51)
City	.25 (-.02, .48) ^b	-1.36 (-2.09, -.69)
Hotel	.20 (-.07, .54) ^b	-.94 (-1.73, -.07)

^aWe report the posterior mean of the mean and 95% coverage interval. Posterior mean estimates of the unobserved heterogeneity are omitted for brevity

^b95% coverage interval spans zero.

conversion rate is, to the best of our knowledge, new to the literature. A potential explanation for this result is that hotel rooms are likely to be classified as experience goods, for which quality is difficult to ascertain before consumption. The argument that position signals quality is more likely to hold in this case than in the case of a product whose quality can more easily be ascertained before consumption (i.e., a search good). However, note that our results are for a single lodging chain. Ultimately, data on multiple categories with search and experience goods may help shed more light on this important question.

We now turn to the effect of (semantic) keyword characteristics on keyword performance, which Table 6 also reports.¹⁴ Absent these observed characteristics, keyword heterogeneity would be treated as unobserved heterogeneity, with keyword-level estimates shrunk toward a common mean. Thus, keyword characteristics help us better estimate click and conversion performance by leveraging information on observed heterogeneity. Ultimately, however, the firm can bid on multiple keywords and should do so as long as a keyword generates a profit. Apart from the “generic” characteristic (which captures differences between keywords that include the brand name of the chain from those that do not), we do not have strong expectations about the effect of the characteristics. Again, we advise caution in generalizing these results because our data are for a single regional lodging chain. As expected, “generic” has a negative effect on both click-through and conversion. Consistent with previous research, we find that keywords that include the brand name of the chain yield better click-through and conversion performance (Ghose and Yang 2009; Rutz and Bucklin 2011). Keywords with a higher number of words are associated with higher CTRs, but this does not carry over to conversion. More detailed search phrases seem to prompt consumers who may be farther down the purchase funnel to investigate served offerings through the text advertisement. However, the served offering does not seem to be able to convert the click-through at any differential rate. Notably, we find that not all detailed search is created equal. Keywords that include geographic terms (e.g., city, state) or lodging-related terms yield lower conversion rates. This result indicates that customers with well-defined search parameters in terms of product class or geography are more difficult for the firm to convert. In addition, keywords that include a state yield lower click-through. We speculate that this may be due to the regional nature of the chain.

The semantic keyword characteristics can be viewed as an important piece of the necessary toolkit for building keyword lists. The characteristics of the keyword help us gain some sense of the relative click-through and conversion performance across keywords. This enables us to explain more of the observed variance in conversion rates and to better estimate conversion rates at the keyword level, especially for sparse keywords. In summary, incorporating keyword characteristics improves the measurement of individual keyword performance. However, at this point, we are only able to make statements about relative keyword perform-

ance. We can say, for example, that based on our model, keywords that include the brand name of our chain yield higher click-through and conversion than keywords that do not include the brand name. However, generic keywords may yield sufficiently high conversion such that the cost per sale is profitable to the firm. In this case, the chain would retain such keywords on the list. Thus, inferring conversion rates is only the first step in keyword list management. We now turn our attention to a more complete treatment of this issue.

IMPLICATIONS FOR KEYWORD LIST MANAGEMENT

We now explore whether our proposed model allows managers to improve the performance of a paid search campaign at the individual keyword level going forward. Specifically, we test an approach that uses model-based conversion rates to manage the keyword list so as to improve the paid search campaign in future periods. We also compare our model-based results with commonly used model-free schemes that rely on observed conversion and click-through rates only. 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. Using the data from April 2004, we use our model-based and model-free schemes to determine which keywords should be retained in the campaign versus which keywords should be dropped. For each scheme, a potentially different list of retained keywords is generated due to differences in the keyword purging rule (e.g., according to observed or estimated conversion rates). Using the resulting keyword lists, we evaluate holdout performances in the May–June 2004 period and compare them with a status quo strategy that retains all 301 original keywords.

As the basis for keyword selection, we use a CPR threshold ($CPR_{\text{threshold}}$) to discriminate between attractive and unattractive keywords. We use the estimated keyword-level conversion rates to calculate the average monthly CPR (CPR_w^{monthly}) for each keyword w . For each model, we rank the keywords by estimated CPR_w^{monthly} and retain those keywords for which estimated $CPR_w^{\text{monthly}} \leq CPR_{\text{threshold}}$. This produces different keyword lists corresponding to each model. Note that a model-based approach is necessary to obtain CPR_w^{monthly} for all keywords used in the campaign. Without a model to estimate conversion rates, the data only allow us to calculate monthly CPR for 84 keywords. For the remaining 217 keywords, we cannot calculate a meaningful monthly CPR, because these keywords do not generate any reservations in April 2004, but costs are incurred for the clicks. We also investigate the performance of two model-free schemes for compiling the keyword list:

- *Manage according to observed conversion rate:* Calculate monthly CPR using the data from April 2004 and retain all keywords with $CPR_w^{\text{monthly}} \leq CPR_{\text{threshold}}$. Note that keywords with zero reservations in April 2004 (217 of 310 keywords) are unprofitable because the firm incurred costs (for the clicks), but no revenue in form of reservations was generated. Thus, these keywords are removed from the list. This strategy results in retaining a small number of keywords (84 of 301 keywords) and has a strong bias against infrequent keywords with few clicks.

¹⁴In the interest of brevity, we report the estimates for the best-fitting model, M3. We note, however, that the alternative models result in some different conclusions regarding the effect of keyword characteristics. This may partly explain some of the fit differences observed across the models.

•*Manage according to CTR*: Choose a CTR threshold x and keep all words with $CTR \geq x$. Managing by CTR is popular in practice, partly because CTR is available for most keywords. Note that CTR can be calculated solely on the data available from search engines; no other data need to be collected. Managing by CTR may also appeal to managers who do not take a transactional viewpoint to assess paid search but rather focus on driving traffic to their websites. However, to manage by CTR, some threshold must be chosen, and it is difficult to do so in a manner that is not arbitrary. For our application, we set the CTR threshold at the average CTR.¹⁵

Evaluating Performance in a Holdout Period

The company did not change the keyword list used in April for the remainder of the quarter. This enables us to make an assessment of the performance of the 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. However, 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 different schemes. 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 CPR 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 to \$60 and found that the comparative performance was independent of the CPR threshold. Thus, we set \$30 as the $CPR_{\text{threshold}}$ and base our discussion on that assumption.

¹⁵We tried different CTR cutoffs and evaluated the performance in our holdout sample; average CTR generates the most profit. Note, however, that this is a post hoc strategy that only works if holdout data are available. Without holdout data, a realistic scenario, the CTR cutoff can only be chosen arbitrarily.

Table 7 reports the results of our keyword list management exercise. Managing the list using the observed conversion rates selects only 84 keywords and yields lower profits than using the status quo of keeping all 301 keywords. This result confirms our intuition that managing by observed rates is likely to eliminate keywords that are ultimately profitable. Managing the list by CTR retains 158 keywords. However, the profits implied by keeping the selected 158 words are essentially the same as keeping all 301 words. The keyword list based on the proposed LIV model (M3) selects 156 keywords and increases profits over the status quo by 4% for May–June. This represents an improvement of approximately \$7 per keyword over two months. The comparisons in Table 7 also show that the LIV model performs better than the alternative model specifications. In light of many firms now having extensive keyword lists numbering in the thousands (average size in 2009 was approximately 2800; MarketingSherpa 2009), potential savings from managing keyword lists based on a model such as ours seem attractive from a managerial perspective.

CONCLUSION

Paid search campaigns have become a crucial part of the marketing budget for many firms. Our objective in this research is to develop a modeling approach that we hope will 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, CPR. However, most keywords do not lead to reservations on a regular basis. In our estimation sample, only 84 keywords of 301 led to reservations, making it impossible to calculate a meaningful CPR for the remaining 217 keywords using the data alone. Should the company immediately drop these sparse keywords? Perhaps not, but an approach for estimating the underlying conversion probabilities for those keywords is necessary for list management. This raises the question how best to produce such estimates and whether they could be used to improve campaign management.

Table 7
HOLDOUT PERFORMANCE: COMPARISON OF MODEL-BASED AND MODEL-FREE STRATEGIES

Select Keywords According to...	Number of Keywords	Number Reservations	Cost	CPR	Implied Profit
<i>Status Quo</i>					\$27,891.36
All keywords selected	301	1348	\$12,548.64	\$9.31	
<i>Observed Conversion Rates</i>					\$27,239.11
Keywords selected	74	1129	\$6,630.89	\$5.87	
Keywords not selected	227	219	\$5,917.75	\$27.02	
<i>Observed CTRs</i>					\$27,879.12
Keywords selected	158	1242	\$9,607.47	\$7.55	
Keywords not selected	143	106	\$ 2,941.17	\$ 38.70	
<i>Estimated Conversion Rates: Model M0</i>					\$28,288.75
Keywords selected	129	1214	\$7,470.39	\$6.70	
Keywords not selected	172	134	\$3,694.88	\$32.97	
<i>Estimated Conversion Rates: Model M1</i>					\$27,615.49
Keywords selected	143	1179	\$6,676.62	\$6.58	
Keywords not selected	158	169	\$4,488.65	\$28.37	
<i>Estimated Conversion Rates: Model M2</i>					\$28,827.36
Keywords selected	164	1265	\$8,643.84	\$7.21	
Keywords not selected	137	83	\$2,521.43	\$41.28	
<i>Estimated Conversion Rates: Model M3</i>					\$28,922.30
Keywords selected	156	1257	\$7,657.75	\$6.99	
Keywords not selected	145	91	\$3,507.52	\$41.33	

In developing a model to address these issues, we conceptualize 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, daily aggregated paid search data available from search engines. We model click-through and conversion rates as a function of keyword-specific intercepts, keyword position, and observed semantic characteristics, utilizing Bayesian shrinkage techniques to overcome the sparseness issues inherent to standard paid search data. An important strategic variable in the click-through and conversion models is the position of the text advertisement 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 about measurement error. An issue with modeling the auction using either a structural or reduced form approach is that competitive bid information is not part of standard paid search data. To address the issue of position endogeneity, we propose a LIV approach. Our results show that our proposed model provides the best fit to the data compared with models that treat position as exogenous or use 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. Our proposed LIV model alleviates concerns with regard to instrument availability, validity, and weakness, correcting the endogeneity bias without adversely affecting the precision of the estimates.

The practitioner community has recently argued that position has little to no effect on conversion (Ballard 2011; Van Wagner 2010); even Google itself has taken this stance (Friedman 2009). The academic literature has found both positive and negative effects of position on conversion (Agarwal, Hosanager, and Smith 2011; Ghose and Yang 2009). We find in our data that conversion rates improve as ad position improves. More significantly, we find that the elasticity of conversions with respect to position (holding constant the number of clicks) is actually higher than the elasticity of clicks with respect to position. Decomposing the increase in conversions from position improvements, we find that 35% of the increased conversions are due to the increase in clicks and 65% are due to the higher conversion rate (holding the number of clicks constant). Of substantive interest to paid search managers, we demonstrate how to use the estimated conversion rates to create attractive keyword subsets. Using holdout data, we evaluate the implied profit performance of our model-based keyword list against the status quo of maintaining the entire list as well as model-free approaches of list management through observed conversion and CTRs. Our LIV model-based list outperforms the status quo and model-free keyword lists as well as the lists generated by alternative model specifications.

We base our model on data that are readily available to paid search advertisers from the major U.S.-based search engines. Our strategy can be implemented—and the performance of a campaign measured—using such data. The ability to measure conversion rates enables managers to test different position–cost combinations and decide, according to 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 seem that such a data set 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. Second, we do not have clickstream-type data and cannot model the consumer's choice process. Overcoming these limitations are promising areas for further research. A visitor-centric panel data set, if available, could be used to investigate whether and how different keywords attract distinct visitor segments. Further research could study whether consumers' characteristics and observed search behavior enable researchers to determine the extent to which factors like position can be explained by consumer heterogeneity. Moreover, research that investigates why position is a credible signal to buyers in terms of both click-through and conversion would be valuable.

Last, a model like ours could be useful in several additional paid search applications. We could envision using a semantic text-mining approach, as Rutz and Trusov (2011) propose, to construct new keywords. A model such as ours could then be used to forecast performance of the new keywords using text alone or to evaluate the performance of the new keywords after a short test period. Our model could also be used to customize promotions according to the keyword. For example, targeted discounts can be offered after a generic search to entice consumers still early in their purchase process to buy. In addition, the rise of the Internet has enabled many new targeting schemes in which a model like ours might be useful. Social media applications allow targeting based on revealed preferences. For example, researchers could target “women who like Oprah” by serving them an advertisement. A model such as ours can easily be adapted to model the click-through on the served advertisement as well as ultimate purchase conversion among the target audience.

APPENDIX: SAMPLER

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

$$\beta_w^d | \dots \propto L^d \left[\left(\beta_w^d \right) \sim \text{MVN} \left(\mu_\beta^d, \Sigma_\beta^d \right) \right],$$

where $d \in \{cl, con\}$.

2. Generate μ_β^d and Σ_β^d :

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

$$\tilde{\mu}_\beta = \tilde{\Sigma}_\beta \left[n \left(\Sigma_\beta^d \right)^{-1} \bar{\beta} + \Sigma_b^{-1} b \right];$$

$$\Sigma_\beta^d \sim \text{IW} \left[v_1 + n, v_2 + \sum_{w=1}^n \left(\beta_w^d - \mu_\beta^d \right) \left(\beta_w^d - \mu_\beta^d \right)' \right],$$

where $d \in \{cl, con\}$, $b = 0_L$, $\Sigma_b = 10^6 \times I_L$, $v_1 = 2$, and $v_2 = v_1 I_L$.

3. Generate ξ_{wt}^{cl} and ξ_{wt}^{con} using a random-walk Metropolis-Hastings step based on the likelihood as given by the Equation 2 combined with a multivariate normal prior:

$$\begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \\ \xi_{wt}^{LIV} \end{pmatrix} | \dots \propto L^{cl} L^{con} \begin{pmatrix} \xi_{wt}^{cl} \\ \xi_{wt}^{con} \\ \xi_{wt}^{LIV} \end{pmatrix} \sim MVN(0, \Lambda),$$

where $\xi_{wt}^{LIV} = pos_{wt} - \omega \gamma_{wt}$, 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 :

$$\omega | \dots \sim MVN(\mu_w, \Sigma),$$

$$\text{where } \Sigma = \left(\Sigma_0^{-1} + \frac{1}{\sigma_{LIV,LIV} - s} \sum_{t=1}^T \gamma_{wt} \gamma'_{wt} \right)^{-1},$$

$$\mu_w = \Sigma \left(\Sigma_0 \mu_0 + \frac{1}{\sigma_{LIV,LIV} - s} \sum_{t=1}^T \gamma_{wt} pos_{wt} \right),$$

where $s = \sigma_{LIV,LIV}$

$$- \left(\sigma_{cl,LIV} \quad \sigma_{con,LIV} \right) \begin{pmatrix} \sigma_{cl,cl} & \sigma_{cl,con} \\ \sigma_{cl,con} & \sigma_{con,con} \end{pmatrix} \begin{pmatrix} \sigma_{cl,LIV} \\ \sigma_{con,LIV} \end{pmatrix},$$

$$\mu_0 = \underline{0} \text{ and } \Sigma_0 = 10^6 I_C.$$

5. Generate Λ :

$$\Lambda | \dots \sim IW \left\{ S + \sum_{w=1}^W \sum_{t=1}^T \left[\begin{pmatrix} \xi_{w,t}^{cl} \\ \xi_{w,t}^{con} \\ pos_{wt} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \omega \gamma_{wt} \end{pmatrix} \right] \left[\begin{pmatrix} \xi_{w,t}^{cl} \\ \xi_{w,t}^{con} \\ pos_{wt} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \omega \gamma_{wt} \end{pmatrix} \right]' \right\}$$

where $v = 3 + 3$ and $S = v I_3$.

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

$$\Pr(\gamma_{wt} = c) = \frac{L(\omega, \xi, \gamma_{wt}^{(c)}, \Lambda) \times \pi_c}{\sum_{i=1}^C L(\omega, \xi, \gamma_{wt}^{(i)}, \Lambda) \times \pi_i}$$

where $\gamma_{wt}^{(c)}$ denotes that keyword w at time t is assigned to class c , $L(\omega, \xi, \gamma_{wt}^{(c)}, \Lambda)$ is the likelihood function of the LIV model evaluated at $\gamma_{wt}^{(c)}$, and π_c is the prior probability of class c membership.

7. Generate π :

$\pi | \gamma \sim \text{Dirichlet}(1 + K_1, \dots, 1 + K_C)$, where the vector K_c denotes the sum of $\gamma_{wt}^{(c)}$ over all c ; that is,

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

REFERENCES

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, Location, Location: An Analysis of Profitability in Position in Online Advertising Markets," *Journal of Marketing Research*, 48 (December), 1057-73.

Angrist, Joshua and Alan Krueger (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15 (Fall), 69-85.

Ballard, Mark (2011), "Paid Search Conversion by Position," *The RKG Blog*, (accessed February 8, 2012), [available at <http://www.rimmkaufman.com/blog/paid-search-conversion-by-position/30092011/>].

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841-90.

Brooks, Nico (2004), *The Atlas Rank Report II: How Search Engine Rank Impacts Conversions*. New York: Atlas Institute.

BusinessWeek (2007), "So Many Ads, So Few Clicks," (November 12), 38.

Chen, Yongmin and Chuan He (2009), "Paid Placement: Advertising and Search on the Internet," working paper, Department of Economics, University of Colorado.

Dhar, Vasant and Anindya Ghose (2010), "Sponsored Search and Market Efficiency," *Information Systems Research*, 21 (4), 760-72.

Ebbes, Peter, Michel Wedel, and Ulf Boeckenholt (2009), "Frugal IV Alternatives to Identify the Parameter for an Endogenous Regressor," *Journal of Applied Econometrics*, 24 (3), 446-68.

_____, _____, _____, and Tom 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," *Quantitative Marketing and Economics*, 3 (4), 365-92.

Edelman, Benjamin and Michael Ostrovsky (2007), "Strategic Bidder Behavior in Sponsored Search Auctions," *Decision Support Systems*, 43 (1), 192-98.

_____, _____, and Michael Schwarz (2007), "Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97 (1), 242-59.

Eisenberg, Bryan (2004), "Benchmarking an Average Conversion Rate," *ClickZ*, (June 4), (accessed February 8, 2012), [available at <http://www.clickz.com/showPage.html?page=3362641>].

Friedman, Dan (2009), "Conversion Rates Don't Vary Much by Ad Position," *Inside Adwords*, (accessed February 8, 2012), [available at <http://adwords.blogspot.com/2009/08/conversion-rates-dont-vary-much-with-ad.html>].

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

Goldfarb, Avi and Catherine Tucker (2011), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 57 (3), 458-70.

Hausman, Jerry A. (1977), "Errors in Variables in Simultaneous Equations Models," *Journal of Econometrics*, 5 (3), 389-401.

Hsiao, Cheng (1976), "Identification and Estimation of Simultaneous Equations Models with Measurement Error," *International Economic Review*, 17 (2), 319-39.

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

MarketingSherpa (2009), *2009 Search Marketing Benchmark Guide-PPC Edition*. Warren, RI: MarketingSherpa LLC.

Rutz, Oliver J. and Randolph E. Bucklin (2011), "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, 48 (February), 87-102.

- and Michael Trusov (2011), “Zooming In on Paid Search Ads: An Individual-level Model Calibrated on Aggregated Data,” *Marketing Science*, 30 (5), 789–800.
- Spiegelhalter, David, Nicola Best, Bradley Carlin, and Angelika van der Linde (2004), “Bayesian Measures of Model Complexity and Fit,” *Journal of the Royal Statistical Society: Series B*, 64 (5), 583–639.
- Stock, James, Jonathan Wright, and Motohiro Yogo (2002), “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20 (4), 518–29.
- Van Wagner, Matt (2010), “PPC Mad Scientists Prove Google Right and Wrong,” *Search Engine Land*, (accessed February 8, 2012), [available at <http://searchengineland.com/ppc-mad-scientists-prove-google-right%E2%80%A6-and-wrong-45236>].
- Wilbur, Kenneth C. and Yi Zhu (2009), “Click Fraud,” *Marketing Science*, 28 (2), 293–308.
- Yang, Sha, Yuxin Chen, and Greg Allenby (2003), “Bayesian Analysis of Simultaneous Demand and Supply,” *Quantitative Marketing and Economics*, 1 (3), 251–75.
- Yao, Song and Carl F. Mela (2011), “A Dynamic Model of Sponsored Search Advertising,” *Marketing Science*, 30 (3), 447–68.
- Zhang, Jie, Michel Wedel, and Rik Pieters (2009), “Sales Effects of Visual Attention to Feature Ads: A Bayesian Mediation Analysis,” *Journal of Marketing Research*, 46 (June), 669–81.