

Multi-Click Attribution in Sponsored Search Advertising: An Empirical Study in Hospitality Industry

Cornell Hospitality Quarterly
2017, Vol. 58(3) 253–262
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1938965516686112
journals.sagepub.com/home/cqx



Chris K. Anderson¹ and Ming Cheng²

Abstract

Sponsored search advertising has become a dominant form of advertising for many firms in the hospitality vertical, with Priceline and Expedia each spending in excess of US\$2 billion in online advertising in 2015. Given the competition in online advertising, it has become essential for advertisers to know how effectively to allocate financial resources to keywords. Central to budget allocation for keywords is an attribution of revenue (from converted ads) to the keywords generating consumer interest. Conventional wisdom suggests several ways to attribute revenues in the sponsored search advertising domain (e.g., last-click, first & last-click, or evenly distributed approach). We develop a multi-click attribution methodology using a unique multi-advertiser data set, which includes full advertiser and consumer-level click and purchase information. We add to the literature by developing a two-stage multi-click attribution methodology with a specific focus on sponsored search advertising in the hospitality industry with which we develop a parametric approach to calculate the value function from each stage of the estimation process. Given our multi-advertiser data set, we are able to illustrate the inefficiency of single-click attribution approaches, which undervalue assist clicks while overvaluing converted clicks.

Keywords

sponsored search advertising; attribution modeling; choice model; hospitality industry

Introduction

Sponsored search advertising, also called paid search, is one of the leading online advertising approaches in the contemporary marketplace with advertisers compensating search engines when consumers click on their ads. Search engines remove information asymmetry as they intermediate consumers and online advertisers, providing advertisers with a unique mechanism to display content to consumers in a targeted manner.

Due to its uniqueness and its potential to drive revenue through timely matched queries, sponsored search has been widely adopted by numerous advertisers across many business verticals. According to a Statista¹ report, sponsored search accounts for 43% of all online advertising with worldwide spend expected to surpass US\$86 billion in 2016. Sponsored search plays a critical role in the generation of firm-level hotel demand as illustrated by Priceline Group's annual spend of US\$2.8 billion in 2015.² Anderson (2011) indicated that more than 80% of online hotel purchases are proceeded by an online search with 11.2 searches per booking on average.

Advertisers need effective approaches to allocate appropriate resources to individual keywords, which becomes complicated as consumers have differing and complex

search processes, often clicking on several ads prior to making their purchase decision. Although search engines allow advertisers to link sales or conversions to the final ad clicked prior to purchase, they do not always link the upstream or assist clicks to purchases owing to the complex search trajectories of consumers. As such the advertisers are left on their own to somehow attribute revenues from converted clicks to these upstream assist clicks. More importantly, search engines do not inform advertisers of conversions generated by other advertisers in the event an advertiser chose not to participate in a particular advertising opportunity by either not bidding on specific keywords or by getting outbid by more aggressive advertisers. This allocation of conversions to assist clicks is what we refer to as the *attribution problem* that many advertisers face in the sponsored search advertising domain. Without appropriate attribution, converted clicks appear overly profitable and may receive an excess budget, whereas assist clicks appear less valuable and

¹Cornell University, Ithaca, NY, USA

²Suffolk University, Boston, MA, USA

Corresponding Author:

Chris K. Anderson, Cornell University, Ithaca, NY 14853, USA.

Email: cka9@cornell.edu

MCF Channel Grouping Path	Conversions	Conversion Value
1. Organic Search → Direct	264 (16.18%)	\$70,666.41 (16.08%)
2. Organic Search × 2	153 (9.38%)	\$40,440.46 (9.20%)
3. Direct × 2	100 (6.13%)	\$27,306.87 (6.22%)
4. Referral → Direct	49 (3.00%)	\$12,215.56 (2.78%)
5. Organic Search → Direct × 2	40 (2.45%)	\$12,780.00 (2.91%)
6. Direct → Organic Search	38 (2.33%)	\$8,653.25 (1.97%)
7. Organic Search → Paid Search	37 (2.27%)	\$10,997.77 (2.50%)
8. Email → Direct	36 (2.21%)	\$9,531.00 (2.17%)
9. Paid Search × 2	35 (2.14%)	\$6,542.96 (1.49%)
10. Referral → Organic Search	35 (2.14%)	\$7,442.96 (1.69%)

Figure 1.
Google Analytics Summary Data.

may be under budgeted. If assist clicks are under budgeted, advertisers bid less on these keywords, and then the advertiser’s ad may not show during searches (or show lower down the search results). If the advertiser does not participate in these assist click searches, they may lose out on some of the converted clicks later as the consumer’s research process has been altered.

Figure 1 shows a typical search journey as detailed in Google Analytics. The figure informs the advertiser of which activity proceeded a purchase decision. If multiple *Paid Search* events occurred (see line 9 in Figure 1) across multiple keywords, the advertiser somehow needs to partition or allocate revenues to these clicks in an effort to understand their profitability. The simplest and most straightforward way of attribution is to assign the converted revenue to a single click, often the first click or the last click that occurred in the customer search journey. However, if customers perform numerous searches, this may affect the profitability of some keywords. As our data set indicates, approximately 70% of all hotel purchases involving paid search have two or more ads clicked in advance of the purchase decisions. We stylize the consumer search journey with the flowchart in Figure 2. The one-click attribution approach ignores the multiple interactions prior to the purchase event and as a result may overvalue or undervalue certain clicks.

One of the challenges in developing attribution models arises from the limited information that advertisers have. Well-informed advertisers are aware of their own actions such as bids, keywords, and their match types (e.g., accurate, phrase, and broad), but unaware of many of the actions undertaken by other advertisers. However, search engines, which facilitate ad auctions, have detailed information across all advertiser actions and customer search behaviors.

Using a unique data set collected from a leading Chinese search engine, we develop a multi-click attribution methodology using the full information afforded by the search engine. We focus on hospitality, as it is one of the dominant industries for online advertising with customer search processes consisting of numerous iterations (Anderson, 2011).

The remainder of this article is laid out as follows: First, we provide a brief overview of current sponsored search advertising literature, then discuss our attribution model and the unique data set used for its development, and conclude with a summary and implications.

Literature Review

Attribution research falls into two main streams: multi-channel attribution and sponsored search advertising. Advertising is often delivered and exposed to customers through multiple platforms (e.g., email, direct mail, TV, and newspapers), whereas final purchase decisions are generally made only through one marketing channel. The attribution of marketing efforts across all customer–firm interactions, although an important issue, has only recently received detailed academic research interest. Danaher and Dagger (2013) developed a Tobit model to help a retail firm evaluate the relative advertising effectiveness through multiple platforms. Li and Kannan (2014) introduced a methodology to access the value from a variety of marketing channels such as referral, email, displayed, and paid search in an online environment. Zantedeschi, Feit, and Bradlow (2015) proposed a hierarchical Bayesian model for measuring multi-channel advertising response. Yadagiri, Saini, and Sinha (2015) developed non-parametric and semi-parametric approaches to estimate the value contributed from different marketing channels. Methodologically, we adopt and

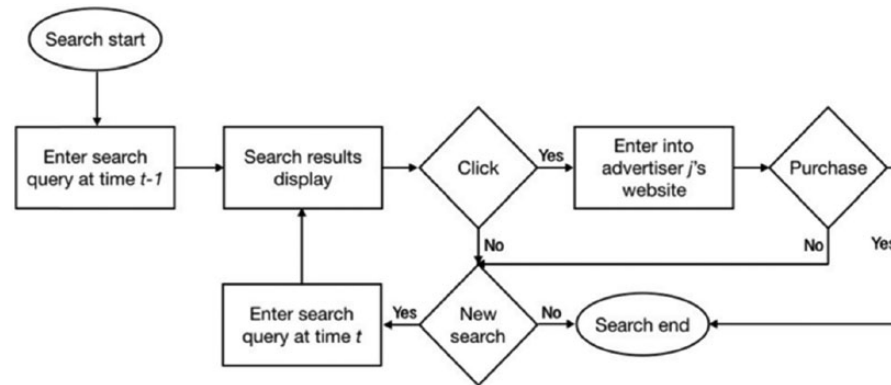


Figure 2.
Flowchart of Customer Search Loop.

extend the work developed by Yadagiri et al. (2015), whereby we develop a parametric approach to calculate the value function from each stage of the search process.

Sponsored search advertising research can be broadly categorized as either empirical or analytical. The majority of analytical research focuses on optimization of one entity (e.g., search engine or advertiser). For example, a number of studies employ game-theoretic models to study search engine mechanisms (i.e., position-based auction, bidding process) and the resulting price competition (Aggarwal, Feldman, & Muthukrishnan, 2006; Athey & Ellison, 2011; Edelman, Ostrovsky, & Schwarz, 2007; Katona & Sarvary, 2010; Varian, 2007; Weber & Zheng, 2007; Xu, Chen, & Whinston, 2011). However, customer interactions with search engines and advertisers are oftentimes overlooked due to the complexity of search behaviors (Chen & He, 2011; Yao & Mela, 2011).

Empirical sponsored search research somewhat offsets the limitations of analytical research by examining ad performance using customer clickstream data. Research has shown that paid ads displayed at more prominent positions can lead to higher click-through rates and conversions (Agarwal, Hosanagar, & Smith, 2011; Jansen, Liu, & Simon, 2013; Rutz, Bucklin, & Sonnier, 2012). The way in which advertisers phrase and configure their keywords (i.e., branded vs. generic keywords, accurate vs. phrase and broad keywords) can affect ad placement (ranking) as well as customer click-through and conversion (Jerath, Ma, & Park, 2014; Rutz & Bucklin, 2011; Yang, Lu, & Lu, 2014). Moreover, studies have shown that the cognitive process through which customers formulate searches and their stage within the purchasing/research process further affect click-through and conversion patterns (Border, 2002; Ghose & Yang, 2009; Jansen & Schuster, 2011).

Our study bridges the search and attribution space as we empirically focus on analyzing customer search and

click-through journeys in an effort to attribute revenues across multiple interactions, but with these interactions all involving paid search. Our multi-click attribution model is comprised of variables, which earlier research indicates, affect ad click-through. Consistent with the previous search advertising literature, we include variables such as rank, keyword match type, keyword character (e.g., branded vs. generic), and customer search queries as factors that affect conversion decisions. Second, we take into account temporal actions, whereby a customer's search experience at time $t - 1$ might influence his or her search behavior at time t . Finally, we introduce a set of variables that measure cumulative search history or consumer search knowledge as accumulated through prior search.

Despite the fast growth of the sponsored search literature, there is only limited research focusing on the sequential accumulation of knowledge through a sequence of clicks. In addition, Kireyev, Pauwels, and Gupta (2015) suggested that indicators that are often used for evaluating firms' effectiveness of sponsored search advertising, such as click-through rate and conversion rate, cannot provide insights into attribution as these measures do not account for search dynamics or synergy between advertisers and the search engine.

Li, Kannan, Viswanathan, and Pani (2016) focused on aspects of attribution, whereby they demonstrated the return on keyword investments vary significantly under different attribution strategies (e.g., first-click or last-click attribution) using a panel data set collected from a jewelry retailer. They also performed a simulation exercise to illustrate the effective attribution strategies that can be applied by firms. Our study builds upon that of Li et al. (2016). We propose a multi-click attribution model that is developed and examined using a data set collected from a search engine platform, rather than a single advertiser, providing a complete customer search and click-through journeys.³ We focus on

hospitality-related search behaviors owing to the complex and search-intensive nature of product discovery. Finally, we compare and contrast model estimates and fit across scenarios, whereby advertisers obtain different levels of information (e.g., full information provided by the search engine vs. limited information solely from advertiser itself).

Data Analysis

Our data set, provided by a leading Chinese search engine, contains all sponsored ad history for hotel-related searches for 50 consecutive days in January/February of 2012. The complete data sets contain data from 1,440,660 sponsored ad impressions generating 183,654 clicks (12.75% average click-through rate) across 62,253 keywords from 1,037 advertisers. We randomly select 1,000 cookieed consumers with converted clicks (purchases), about 20% of cookieed consumers in our sample. For each of these 1,000 customers, we then collect information on all prior (in advance of the purchase) paid searches. The search information compiled includes a consumer-by-consumer stream of searches and complete display information of all participating advertisers (both clicked and unclicked paid ads). Each paid search impression contains three facets of information: advertiser data (ID, name, and landing page) that are associated with all displayed paid ads (across all advertisers), keyword data (content, match type, bid and cost-per-click, ad quality, rank) that are related to the displayed ads, along with consumer click data (consumer query, click status).

For the 1,000 purchases in our sample, advertisers bid on 6,671 keywords of which 1,504 keywords were associated with clicked ads and 410 of these keywords resulted in the 1,000 conversions (purchase decision at advertiser followed the clicked ad). In general, assuming our random sample is representative, without an attribution model as developed in this article, advertisers would considerably reduce the set of keywords they were interested in bidding on—as more than 4,500 keywords failed to generate any consumer action and more than 6,000 failed to generate any sales. One could say that almost 1,100 of the keywords are generating loss. One would assume that if we increased the size of the sample, the number of keywords associated with conversions would mostly likely increase—the similar situation might also apply to the total number of keywords included in the data set, as well as the number of keywords that were clicked by customers. However, owing to our random selection, we can assume the percentages (e.g., 410 of 1,504) are indicative of typical advertiser outcomes.

We use Figure 3 to highlight the importance of multi-click attribution. Figure 3 summarizes the distribution of the number of sponsored ad clicks per conversion event. As indicated in Figure 3, only 30% of purchases were preceded by one sponsored ad click—so for these 30% of purchase occasions, a single-click attribution model would

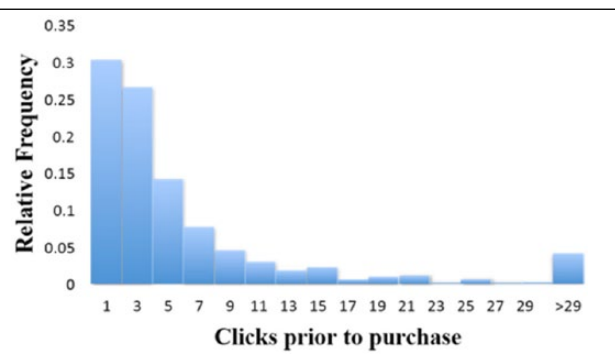


Figure 3. Distribution of the Number of Ads Clicked Prior to Purchase.

Table 1. Ads Clicked Per Advertisers as a Function of Searches Per Conversion.

Searches	Ads Clicked on Advertisers (%)				
	1	2	3	4	5
1	100.0				
2	94.3	5.7			
3	89.2	8.4	2.5		
4	91.8	6.6	1.6	0.9	
5	80.6	12.4	3.8	1.7	1.5

suffice—but, for the other 70%, some form of multi-click attribution is required.

As 70% of purchases are preceded by five clicks or less, our remaining descriptive statistics focus on these purchase events. Table 1 illustrates the impacts and limitations of single advertiser data. The rows in Table 1 indicate the number of ads clicked on prior to a conversion (one through five) with the columns summarizing the average number of advertisers clicked on during those searches. For a single-click conversion, obviously only a single advertiser is involved, but the last row of Table 1 indicates that for conversions preceded by five clicked ads, just above 80% involved only one advertiser with 1.5% of those five clicked ad conversions involving five advertisers. In Table 1, columns for advertisers two to five indicate the need to use disaggregated multi-advertiser data for attribution, as single advertiser data would exclude many consumer interactions.

Tables 2 and 3 summarize average cost-per-click (CPC) for both converted and assist clicks for different match and keyword types. Both tables clearly indicate a tendency for advertisers to bid less for assist clicks. The non-monotonic nature of CPCs as a function of the number of ads clicked as part of the consumer’s search path also indicates the possibility of other factors influencing advertiser bidding

Table 2.
Summary of Average CPC by Different Keyword Type.

Click Type	Keyword Match Type	Ads Clicked on Advertisers				
		1	2	3	4	5
Converted clicks	Broad/phrase	5.07	6.03	5.26	8.18	4.25
	Accurate	2.49	2.83	3.01	2.76	3.63
Assist clicks	Broad/phrase	3.11	2.84	2.97	2.62	2.43
	Accurate	2.56	2.53	2.27	2.23	3.17

Note. CPC = cost-per-click.

Table 3.
Summary of Average CPC by Keyword Branded Type.

Click Type	Keyword Match Type	Ads Clicked on Advertisers				
		1	2	3	4	5
Converted clicks	Generic	3.4	4.28	4.45	5.32	4.53
	Branded	1.78	1.93	1.66	2.18	1.68
Assist clicks	Generic	2.85	2.73	2.68	2.42	3.18
	Branded	1.22	1.11	1.21	1.07	0.6

Note. CPC = cost-per-click.

behavior. In the following section, we outline modeling efforts to better understand the multi-click process and build a corresponding attribution model.

Multi-Click Attribution Modeling

Multi-click attribution could be a difficult task for an individual advertiser to achieve, as advertisers only have a constrained set of observations—that is, clicks (and conversions) on their ads as a function of their bids (data summarized in the second column in Table 1). The search engine, however, has complete information at a user level—that is, all searches, impressions, and clicks across all advertisers prior to the conversion. We use the search engine's full information in our attribution approach. Although one could argue that advertisers may never have this information, it may be in the search engine's best interest to provide information to facilitate bidding by advertisers. Although they probably cannot share other advertisers' detailed actions, search engines could develop attribution models as we illustrate here. The search engine may also have incentives to provide more detail (beyond the typical sample information displayed in Figure 1) to allow more optimal bidding by advertisers—as inferred from Tables 2 and 3, many of the assist clicks may be undervalued.

In this section, we propose a multi-click attribution model and explain how individual advertisers could employ the model to allocate revenue—which is generated from converted searches—to each keyword triggered through the

search path. We start our discussion with a binary logit choice model developed to estimate a customer's conversion probability on an advertiser's displayed paid ad at a certain time. Based upon the coefficient estimates calculated from the choice model, we then introduce attribution weights and demonstrate how revenue can be attributed to individual keywords that have been triggered along the customer's search path.

Customer's Conversion Model

We use a binary logit choice model to estimate conversion decisions after clicking on a displayed sponsored ad. We assert that click-through decision made based upon

- the ranking or position of the displayed sponsored ads,
- the level of match between customer's search queries and advertiser's keyword setup, and
- the customer's search preference (i.e., preferences toward certain brands or advertisers).

In addition, we also assume that in addition to ad type, placement, and content, the conversion decisions at the current click-through level could also be a function of the knowledge collected throughout the entire search journey (the assist clicks). To capture this history, we construct three variables, total number of advertisers seen, total number of keywords searched, and total number of sponsored ads exposed, to represent the cumulative information that the customer has gathered throughout the search process. In addition, taking into account that a customer's search, click, and conversion decision at time t might be influenced by his or her search behavior at $t - 1$, we include a set of variables (e.g., keyword match, displayed ads ranking) for click ads at $t - 1$.

Combining all three aspects, we construct variables drawn from the current search and ads characteristics at time t , customer's search and displayed sponsored ads characters at time $t - 1$, and customer's cumulative information collected through the search journey as independent variables to estimate a customer's conversion at time t . The utility of a customer i 's conversion through a sponsored ad at time t can be denoted as U_{it}^{Conv} , which can be unfolded using a random utility function format and specified as $U_{it}^{Conv} = V_{it}^{Conv} + \epsilon_{it}^{Conv}$. In this random utility function, ϵ_{it}^{Conv} are the stochastic error terms assumed to be identically and independently distributed with the extreme value distribution. V_{it}^{Conv} is the deterministic portion and can be specified as a linear function of X_{it}^{Conv} . As mentioned above, X_{it}^{Conv} contains variables from three aspects: (a) customer's current search and ads characteristics at time t , (b) customer's search and displayed sponsored ads characters at time $t - 1$, and (c) customer's cumulative information collected through the search journey as independent variables to estimate a customer's conversion rate at time t . For customer's

search and sponsored ads characters at time t , we consider (a) paid ad j 's ranking position at time t , denoted as $Rank_t$; (b) paid ad j 's own characteristics, including ad j 's keyword match type, denoted as $Accurate_Keyword_t$ and $Phrase_Keyword_t$, and keyword branded characteristics, denoted as $Brandself_Keyword_t$ and $Brandother_Keyword_t$; and (c) a control variable for customer's search query (branded vs. generic), which is defined as a dummy variable BQ_t . Similarly, we could construct a set of variables to capture the customer's search and sponsored ads' characters at time $t - 1$, including $Rank_{t-1}$, $Accurate_Keyword_{t-1}$, $Phrase_Keyword_{t-1}$, $Brandself_Keyword_{t-1}$, $Brandother_Keyword_{t-1}$, and BQ_{t-1} . Finally, we introduce three variables to measure customer's cumulative information gathered in the search journey. The three variables include (a) the total number of clicks prior to the search at time t , denoted as $prior_Clicks_t$; (b) the total number of keywords searched prior to the search at time t , denoted as $prior_Keywords_t$; and (c) the total number of advertisers exposed in the search results prior to the search at time t , denoted as $prior_Advertisers_t$. Therefore, the representative utility function of V_{it}^{Conv} can be elaborated in Equation 1:

$$V_{it}^{Conv} = f \left(\begin{matrix} Rank_t, Keyword\ Match_t, \\ Keyword\ Character_t, BQ_t, \\ Rank_{t-1}, Keyword\ Match_{t-1}, \\ Keyword\ Character_{t-1}, BQ_{t-1}, \\ prior_Clicks_t, prior_ \\ Advertisers_t, prior_Keyowrds_t \end{matrix} \right). \quad (1)$$

Given the variables described above, we could further expand the utility function into Equation 2, which is elaborated as

$$\begin{aligned} V_{it}^{Conv} = & \alpha_0 + \alpha_1 Rank_t + \alpha_2 Accurate_ \\ & Keyword_t + \alpha_3 Phrase_Keyword_t \\ & + \alpha_4 Brandself_Keyword_t + \\ & \alpha_5 Brandother_Keyword_t + \alpha_6 BQ_t \\ & + \beta_1 Rank_{t-1} + \beta_2 Accurate_Keyword_{t-1} + \\ & \beta_3 Phrase_Keyword_{t-1} \\ & + \beta_4 Brandself_Keyword_{t-1} + \beta_5 Brandother_ \\ & Keyword_{t-1} + \beta_6 BQ_{t-1} \\ & + \gamma_1 prior_Clicks_t + \gamma_2 prior_ \\ & Advertisers_t + \gamma_3 prior_Keywords_t. \end{aligned} \quad (2)$$

In Equation 2, $Accurate_Keyword$ and $Phrase_Keyword$ are dummy variables representing the level of match between paid ad j 's keyword and consumer i 's query. $Brandself_Keyword$ and $Brandother_Keyword$ are dummy variables indicating how paid ad j 's keyword is branded. When the advertiser of ad j includes its own brand name in the keyword (e.g., Hilton Hotel Group chose to include

“Hilton” in its keywords campaign), $Brandself_Keyword$ is marked as 1, otherwise it is labeled as 0. Similarly, when paid ad j includes other brands or its competitor's brand name in keyword settings (e.g., Marriott Hotel selected “Hilton” as part of its keyword campaign), $Brandother_Keyword$ is set to be 1, and 0 otherwise. BQ is a dummy variable and is used to capture the branded information, imbedded in each customer search query. When a customer entered a branded query into the search box (e.g., “Hilton Hotel in Shanghai”), BQ is labeled as 1, and 0 otherwise. The customer's conversion rate can then be estimated in Equation 3:

$$P(Conversion_{it} = 1) = \frac{\exp(V_{it}^{Conv})}{1 + \exp(V_{it}^{Conv})}. \quad (3)$$

We calculate the parameter coefficients using maximum likelihood estimation approach. The coefficient estimates are summarized in Table 4.

Consider a general scenario in which a regular advertiser can only observe its own actions such as keyword settings and sponsored ads placement at time t and may not be fully aware of customer's specific search behavior from time $t - 1$, or does not have access to the detailed customer search information prior to time t . We construct two reduced models to portray these two scenarios and calculate the parameter coefficients for each reduced model (see Table 4).

In the first scenario, as illustrated in Table 4, $RM1$ (representing Reduced Model 1), we assume that the advertiser only has time t search, keyword, and ranking information. In other words, the model specification eliminates variables that are associated with customer's search at time $t - 1$ and variables that reflect the level of information that customers collected prior to time t . In the second scenario, $RM2$ (representing Reduced Model 2), we assume that an advertiser has time t and time $t - 1$ information. Although the advertiser may observe customer's search and ads performance at both times t and $t - 1$, information that reveals customer's collective knowledge gain through the search path prior to time t is still unknown.

We perform two robustness checks using likelihood ratio test technique to evaluate the full model versus two reduced models' goodness of fit. As illustrated in Table 5, the result of the robustness checks suggests that variables related to the customer's previous search attributes are not redundant and should be kept in the model settings, Likelihood Ratio (LR) = $330 > \chi^2(9) = 27.88$, and similarly those related to the search history are also relevant, $LR = 90 > \chi^2(3) = 16.28$. Results from both statistical tests suggest that the full model provides a better fit when estimating the customer's conversion probability. In addition, we conduct tests using the statistical technique proposed by Clogg, Petkova, and Haritou (1995) to examine the significance of differences in parameter estimates induced by the exclusion of attributes in the

Table 4.
Summary of Parameter Estimates of Conversion Model.

Parameter Estimates	Full Model	RMI	RM2
	M (SE)	M (SE)	M (SE)
Intercept	-2.359 (0.514)	-2.741 (0.511)	-2.372 (0.514)
Rank _t	-0.380 (0.035)	-0.524 (0.034)	-0.395 (0.035)
Accurate_Keyword _t	3.252 (0.514)	3.200 (0.510)	3.274 (0.513)
Phrase_Keyword _t	3.758 (0.519)	3.456 (0.515)	3.721 (0.518)
Brandself_Keyword _t	0.191 (0.290)	0.041 (0.270)	0.095 (0.288)
Brandother_Keyword _t	-0.517 (0.272)	-0.532 (0.258)	-0.607 (0.272)
BQ _t	0.052 (0.248)	0.077 (0.232)	0.118 (0.249)
Rank _{t-1}	-0.004 (0.031)	—	-0.074 (0.030)
Accurate_Keyword _{t-1}	-0.773 (0.118)	—	-0.881 (0.112)
Phrase_Keyword _{t-1}	-1.398 (0.155)	—	-1.582 (0.151)
Brandself_Keyword _{t-1}	-0.432 (0.341)	—	-0.619 (0.332)
Brandother_Keyword _{t-1}	-0.582 (0.297)	—	-0.716 (0.299)
BQ _{t-1}	0.471 (0.282)	—	0.494 (0.284)
Prior_Clicks _t	0.130 (0.048)	—	—
Prior_Keywords _t	-0.057 (0.041)	—	—
Prior_Advertisers _t	-0.167 (0.048)	—	—

Note. RMI = Reduced Model 1; RM2 = Reduced Model 2.

Table 5.
Comparison of Parameter Estimates Between Full Model and RMI.

Parameter Estimates	Full Model	RMI	d	t = d/s(d)
	M (SE)	M (SE)	M (SE)	
Intercept	-2.359 (0.514)	-2.741 (0.511)	-0.359 (6.098)	-0.062
Rank _t	-0.380 (0.035)	-0.524 (0.034)	-0.144 (0.052)	-2.787
Accurate_Keyword _t	3.252 (0.514)	3.200 (0.510)	-0.052 (5.129)	-0.01
Phrase_Keyword _t	3.758 (0.519)	3.456 (0.515)	-0.302 (3.256)	-0.092
Brandself_Keyword _t	0.191 (0.290)	0.041 (0.270)	-0.15 (0.408)	-0.368
Brandother_Keyword _t	-0.517 (0.272)	-0.532 (0.258)	-0.015 (0.434)	-0.034
BQ _t	0.052 (0.248)	0.077 (0.232)	0.025 (0.529)	0.047
Rank _{t-1}	-0.004 (0.031)	—	—	—
Accurate_Keyword _{t-1}	-0.773 (0.118)	—	—	—
Phrase_Keyword _{t-1}	-1.398 (0.155)	—	—	—
Brandself_Keyword _{t-1}	-0.432 (0.341)	—	—	—
Brandother_Keyword _{t-1}	-0.582 (0.297)	—	—	—
BQ _{t-1}	0.471 (0.282)	—	—	—
Prior_Clicks _t	0.130 (0.048)	—	—	—
Prior_Keywords _t	-0.057 (0.041)	—	—	—
Prior_Advertisers _t	-0.167 (0.048)	—	—	—

Note. RMI = Reduced Model 1.

reduced-form models. As shown in Table 5, the impact of *Rank* on customer conversion rate estimation at time *t* is significantly decreased (from -0.524 to -0.380), when introducing variables that capture customer search and ads performance at time *t* - 1 and variables that show cumulative information customer collected from the search path.

Similarly, the comparison between estimates of *Rank*_{*t*-1} from the full model and the RM2 in Table 6 indicates that there is a significant decrease on the impact of *Rank*_{*t*-1} on customer's conversion rate estimation at time *t* with inclusion of variables describing the cumulative aspects of the search journey.

Table 6.
Comparison of Parameter Estimates Between Full Model and RM2.

Parameter Estimates	Full Model	RM2	<i>d</i>	<i>t</i> = <i>d</i> / <i>s</i> (<i>d</i>)
	<i>M</i> (SE)	<i>M</i> (SE)	<i>M</i> (SE)	
Intercept	-2.359 (0.514)	-2.372 (0.514)	-0.013 (6.163)	-0.002
Rank _{<i>t</i>}	-0.380 (0.035)	-0.395 (0.035)	-0.015 (0.054)	-0.278
Accurate_Keyword	3.252 (0.514)	3.274 (0.513)	0.022 (5.189)	0.004
Phrase_Keyword	3.758 (0.519)	3.721 (0.518)	-0.037 (3.301)	-0.011
Brandself_Keyword	0.191 (0.290)	0.095 (0.288)	-0.096 (0.463)	-0.208
Brandother_Keyword _{<i>t</i>}	-0.517 (0.272)	-0.607 (0.272)	-0.09 (0.481)	-0.188
BQ _{<i>t</i>}	0.052 (0.248)	0.118 (0.249)	0.066 (0.608)	0.107
Rank _{<i>t-1</i>}	-0.004 (0.031)	-0.074 (0.030)	-0.07 (0.035)	-2.00
Accurate_Keyword _{<i>t-1</i>}	-0.773 (0.118)	-0.881 (0.112)	-0.108 (0.156)	-0.689
Phrase_Keyword _{<i>t-1</i>}	-1.398 (0.155)	-1.582 (0.151)	-0.184 (0.129)	-1.427
Brandself_Keyword _{<i>t-1</i>}	-0.432 (0.341)	-0.619 (0.332)	-0.187 (0.395)	-0.475
Brandother_Keyword _{<i>t-1</i>}	-0.582 (0.297)	-0.716 (0.299)	-0.134 (0.459)	-0.293
BQ _{<i>t-1</i>}	0.471 (0.282)	0.494 (0.284)	0.023 (0.574)	0.042
Prior_Clicks _{<i>t</i>}	0.130 (0.048)	—	—	—
Prior_Keywords _{<i>t</i>}	-0.057 (0.041)	—	—	—
Prior_Advertisers _{<i>t</i>}	-0.167 (0.048)	—	—	—

Note. RM2 = Reduced Model 2.

Attribution Model

Given the availability of the parameter coefficients in the customer’s conversion model, we can now use the conversion probability to create attribution weights. Because the full model provides a better fit to the conversion estimation, our discussion is based upon the full model. Assume that a customer clicked on five paid ads before making a purchase. For each of these clicks, we can estimate the customer’s conversion probability using the parameters calculated from the customer conversion model, as illustrated in Table 4. We then scale each conversion probability to calculate the attribution weight for each clicked paid ad. Assume that the conversion probabilities for the five clicks in the search and conversion path are $P_1 = .5, P_2 = .4, P_3 = .7, P_4 = .1, P_5 = .3$, where the summation of all the conversion probabilities equals to 2.0, denoted as $\sum P_i = 2.0$. Given the estimated conversion probabilities, we then derive the attribution weights using $w_i = P_i / \sum P_i$, therefore, the attribution weight for each clicked paid ads is specified as $w_1 = 0.25, w_2 = 0.2, w_3 = 0.35, w_4 = 0.05, w_5 = 0.15$, where $\sum w_i = 1$. As of now, the advertiser would be able to allocate its revenue to each keyword, $w_i \times R$, where R represents the revenue from the conversion.

In Figure 4, we compare the weighted attributions with conversions for all 1,504 clicked keywords within our sample of 1,000 conversions.

As illustrated in Figure 4, the *x*-axis is the sum of conversions for each of the 1,504 keywords, whereas the *y*-axis is the sum of the weighted attributions for each of these

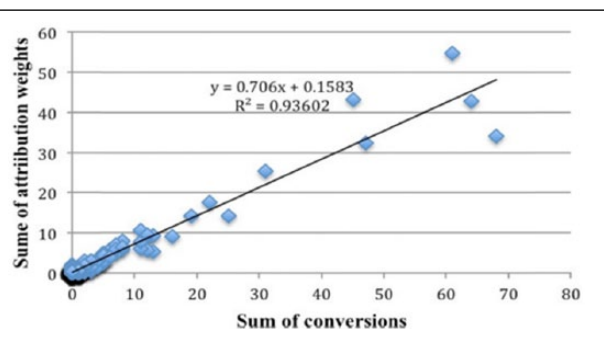


Figure 4.
Weighted Keyword Attributions Versus Conversions.

keywords. As indicated earlier, only 410 of 1,504 keywords converted—In other words, there are 1,094 keywords (1,504 – 410 = 1,094) that were clicked by consumers but there was no direct conversion generated, therefore, they have conversion sum of zero but a non-zero weighted attribution. Figure 5 presents the distribution of the sum of attribution weights for these assist clicks.

As illustrated in Figure 5, a large proportion of the assist clicks have less than 0.1 attribution weight, indicating that keywords that are associated with these assist clicks receive less than 10% of revenue allocation. Figures 4 and 5 simply serve as validity checks indicating that although our model generates non-zero attribution weights for assist clicks, these weights are consistent with total conversions (Figure 4) and are fairly small in nature (Figure 5). The non-zero weights of

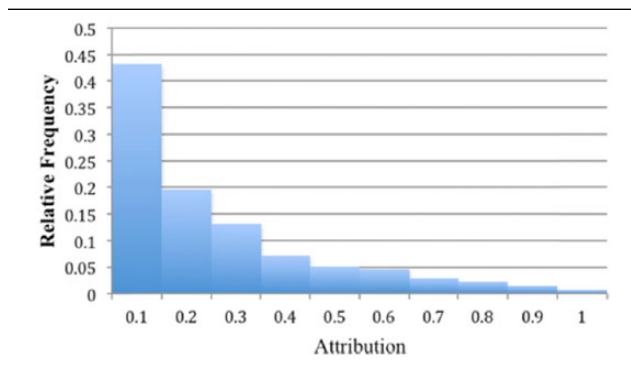


Figure 5.
Attribution Weights for Assist Clicks.

all assist clicks, even though some are small in nature, indicate that all keywords involved in a converted path will have measurable return on investment and non-zero revenues, whereas under a last-click framework, approximately two thirds of these keywords would have zero attributed revenue.

Managerial Implication and Conclusion

That advertisers use search engines to promote their products or services has been a predominant phenomena in the online advertising field for more than a decade and largely indicative of Google's market capitalization. Advertiser ability to allocate resources (e.g., budget) more effectively and efficiently in sponsored search advertising is of critical importance to both researchers and practitioners. In this study, we propose an attribution model to address aspects of this issue and illustrate how advertisers can attribute converted click revenue to each of the keywords involved in the search journey. Using a unique data set that captures the disaggregate query-level customer search and click behavior, as well as the associated advertisers' sponsored search information, we demonstrate the entire process of computing the attribution weights—which is the key to the attribution model, and further highlight the need to move beyond a single-click attribution methodology.

We consider three scenarios in the study. In the first scenario, an advertiser has full information, which not only includes its clicked paid ads information at the current time t but also contains the same customer's search and ad clicked information at $t - 1$ as well as aggregate-level information summarizing the amount of information collected throughout his or her search path before initiating the search at time t . In the other two scenarios, we consider advertisers with partial search information. We show via those reduced models that exclusion of search history both reduces model fit and adds bias to parameter estimates common across all models.

Sponsored search advertising enables an individual advertiser to link conversions/sales to all customer online

search activities, but the insight from this linkage maybe be biased owing to last-click or single-click attribution. Solely depending upon the last click may lead to a consequence where an advertiser overvalues the contribution of the keyword, which is associated with the converted click and undervalues the rest of the keywords that are associated with the assist clicks along the search lines. Although it is arguable that a customer's search depth might be quite limited for certain industries, our empirical findings have shown that more than two thirds of the hotel-related purchases involve a multi-ad click process. Rather than devoting all financial resources to converted keywords, advertisers now have a set of keywords (converted and assist clicks) that revenue can be probabilistically attributed across.

Second, our empirical findings suggest that there might be great opportunities for search engines to collaborate with advertisers by providing more detailed disaggregate-level ads and customer search information. We find that the impact of *Rank* is significantly reduced on conversion estimation in a full-information model (vs. partial-information models). As a result, advertisers deploying single-click attribution approaches might overestimate the impact of ad rank on customer conversion rate when the available information used to calculate customer conversion is limited. In this situation, advertisers may unnecessarily elevate their bids in an effort to push their ads closer to the top of the search results. Aside from providing aggregate-level paid ads performance reports, search engines could provide advertisers with more disaggregate-level information to aid advertisers throughout the keyword selection and bid process and help switch advertiser's focus from rankings to keyword allocations. Although potentially resulting in decreased bids for converted keywords, more informed attribution would result in increased budget allocations across assist clicks.

Our findings are limited as we focus only on clicked ads and not looked at the impact of non-paid clicks or impressions without a click. This last point is probably the largest limitation of our study as clearly some consumers may choose not to click on ads (by principle) but still use the ad information or web address in their (future) research process. Several online advertising platforms (e.g., Facebook) have reinforced this fact as advertisers can now bid on a cost-per-impression versus a cost-per-click basis. Extensions of our approach would naturally focus on all impressions at the consumer level and the interplay upon paid search and organic search as inclusion of organic clicks may change the distribution of attribution weights as well as their eventual ROI estimates. Finally, given the disaggregate level of our data, another fruitful research stream may be an examination of approaches to estimate attribution weights in a constrained single advertiser setting and comparison with those weights in a full-information setting.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, or publication of this article.

Notes

1. <http://www.statista.com/statistics/267056/paid-search-advertising-expenditure-worldwide/>, accessed March 2016
2. Priceline Group 2015 Annual Form 10-K.
3. Information from one single firm may not help rebuild the complete customer sponsored ad search and click-through journeys. Assuming that the firm in question called Advertiser A, one scenario could be that a customer searched for a keyword and clicked on Advertiser B's ad (instead of Advertiser A's). In this situation, the triggered keyword will be labeled as unclicked in the data set, and therefore, this keyword (which is supposed to be a "touch point") is neglected from the customer's search path.

References

- Agarwal, A., Hosanagar, K., & Smith, M. D. (2011). Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of Marketing Research*, 48, 1057-1073.
- Aggarwal, G., Feldman, J., & Muthukrishnan, S. (2006). *Bidding to the top: VCG and equilibria of position-based auctions*. Berlin, Germany: Springer-Verlag.
- Anderson, C. K. (2011). Search, OTAs, and online booking: An expanded analysis of the billboard effect. *Cornell Hospitality Report*, 11(8), 6-10.
- Athey, S., & Ellison, G. (2011). Position auctions with consumer search. *The Quarterly Journal of Economics*, 126, 1213-1270.
- Border, A. (2002). A taxonomy of web search. *Association for Computing Machinery Special Interest Group on Information Retrieval Forum*, 36(2), 3-10.
- Chen, H., & He, C. (2011). Paid placement: Advertising and search on the Internet. *The Economic Journal*, 121, F309-F328.
- Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *The American Journal of Sociology*, 100, 1261-1293.
- Danaher, P. J., & Dagger, T. S. (2013). Comparing the relative effectiveness of advertising channels: A case study of a multimedia Blitz campaign. *Journal of Marketing Research*, 50, 517-534.
- Edelman, B., Ostrovsky, M., & Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review*, 97, 242-259.
- Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55, 1605-1622.
- Jansen, B. J., Liu, Z., & Simon, Z. (2013). The effect of ad rank on the performance of keyword advertising campaigns. *Journal of the Association for Information Science and Technology*, 64, 2115-2132.
- Jansen, B. J., & Schuster, S. (2011). Bidding on the buying funnel for sponsored search campaigns. *Journal of Electronic Commerce Research*, 12(1), 1-18.
- Jerath, K., Ma, L., & Park, Y. H. (2014). Consumer click behavior at search engine: The role of keyword popularity. *Journal of Marketing Research*, 51, 480-486.
- Katona, Z., & Sarvary, M. (2010). The race for sponsored links: Bidding patterns for search advertising. *Marketing Science*, 29, 199-215.
- Kireyev, P., Pauwels, K., & Gupta, S. (2015). Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing*, 33, 475-490.
- Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51, 40-56.
- Li, H., Kannan, P. K., Viswanathan, S., & Pani, A. (2016). Attribution strategies and return on keyword investment in paid search advertising. *Marketing Science*, 35(6) 831-848.
- Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, 48, 87-102.
- Rutz, O. J., Bucklin, R. E., & Sonmier, G. P. (2012). A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research*, 49, 306-319.
- Varian, H. R. (2007). Position auctions. *International Journal of Industrial Organization*, 25, 1163-1178.
- Weber, T. A., & Zheng, Z. (2007). A model of search intermediaries and paid referrals. *Information Systems Research*, 18, 414-436.
- Xu, L., Chen, J., & Whinston, A. (2011). Price competition and endogenous valuation in search advertising. *Journal of Marketing Research*, 48, 566-586.
- Yadagiri, M. M., Saini, S. K., & Sinha, R. (2015). A non-parametric approach to the multi-channel attribution problem. In *Web Information Systems Engineering—WISE 2015* (pp. 338-352). Springer International Publishing, Switzerland.
- Yang, S., Lu, S., & Lu, X. (2014). Modeling competition and its impact on paid-search advertising. *Marketing Science*, 33, 134-153.
- Yao, S., & Mela, C. F. (2011). A dynamic model of sponsored search advertising. *Marketing Science*, 30, 447-468.
- Zantedeschi, D., Feit, E. M., & Bradlow, E. (2015). Measuring multi-channel advertising response. *Management Science*.

Author Biographies

Chris K. Anderson is an Associate Professor at Cornell University's School of Hotel Administration and Director of Cornell's Center for Hospitality Research. He teaches and conducts research in data analytics, pricing, distribution and revenue management.

Ming Cheng is an assistant professor of Marketing at Sawyer Business School, Suffolk University. Her research expertise lies in the area of digital, social media marketing and database marketing.