1-1-2012

Essays in Online Advertising

Vibhanshu Abhishek
University of Pennsylvania, vibs@cmu.edu

Essays in Online Advertising

Abstract
The last several years have seen a dramatic increase in the amount of time and money consumers spend online. As a consequence, the Internet has become an important channel that firms can use to reach out and connect to consumers which has lead to the emergence of online advertising.Given the scale and novelty of online advertising, there is a growing need to understand how consumers respond to online ads and how firms should advertise using this medium. In my dissertation, I study different aspects of sponsored search and display ads which constitute the bulk of online advertising. In the first essay, I focus on the issues related to the use of aggregate data in sponsored search. I demonstrate that models commonly used in sponsored search research suffer from aggregation bias and present the implications of this aggregation bias. In the second essay, I focus on the advertiser's problem of bidding optimally in sponsored search auctions. In the third essay, I study the interactions between various forms of online advertising like banner ads, display ads and sponsored search ads and address the problem of attribution.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Operations & Information Management

First Advisor
Kartik Hosanagar

Second Advisor
Peter S. Fader

Keywords
Attribution, Auctions, Consumer behavior, Game theory, Online advertising, Sponsored search

Subject Categories
Economics | Other Economics

ESSAYS ON ONLINE ADVERTISING

Vibhanshu Abhishek

A DISSERTATION

in

Operations and Information Management

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

Kartik Hosanagar,
Associate Professor of OPIM

Co-Supervisor of Dissertation

Peter S. Fader,
Professor of Marketing

Graduate Group Chairperson

Eric Bradlow, Professor of Marketing,
Statistics, and Education

Dissertation Committee

Kartik Hosanagar, Associate Professor of Operations and Information Management
Peter S. Fader, Frances and Pei-Yuan Chia Professor, Professor of Marketing
Lorin M. Hitt, Class of 1942 Professor of Operations and Information Management
Z. John Zhang, Murrel J. Ades Professor, Professor of Marketing

ESSAYS ON ONLINE ADVERTISING

To my grandfather Dr. Raman and my mother Anshu,

who have been a tremendous influence in my life.

ACKNOWLEDGEMENT

The last five years have been the most challenging and exhilarating time of my life. I feel privileged to have wonderful advisors and caring friends who have all stood by me during this time and made this dissertation a reality.

First of all, I would like to thank Profs. Kartik Hosanagar, Pete Fader and John Zhang for being excellent mentors during the PhD programm. I feel extremely fortunate to be blessed with advisors who are extremely considerate, supportive and above all, accessible when I needed their help. I could not have hoped for better role models for my career. I still continue to learn from them everyday and sincerely wish that their support and encouragement will extend beyond Wharton. I would like to thank Lorin Hitt for being a part of my committee and having an early influence on my research agenda. I would also like to thank Kinshuk Jerath, Noah Gans, Gérard Cachon, Oded Netzer, Eric Clemons, Vadim Cherepanov, Eric Bradlow, Raghuram Iyengar and Elea Feit for many an insightful discussion.

I would like to thank the Mack Center for Technological Innovation, the Baker Retailing Center, the Wharton Customer Analytics Initiative and the Marketing Science Institute for generous financial and data support that was instrumental in the completion of this dissertation. Many thanks go to Mallory Hiatt, Kim Wartford, Andrea Nurse, Rochelle Hambeau-Miller and Karen Ressler for all the administrative help and Stan Liu and Jamie Walter for excellent IT support.

The last five years would not have been possible without the help and support my friends at Wharton. Many special thanks go out to Dan Fleder, Pnina Feldman, Jaelynn Oh, Fazil Paç, Amit Bhattacharya, Eric Schwartz and in particular, the OPIM Purification Band – Jose Guajardo, Jun Li, Toni Moreno, Alessandro Arlotto, Andreś Catalan-Cardenas, Necati Tereyağoğlu and Fangyun Tan. I hope our friendship continues to grow beyond the confines of Huntsman Hall.

I would also like to say a special thanks to my family for their unflinching support, frequent words of encouragement and understanding my physical absence during these years. Last but not the least, I would like to thank my wife Ankita for bearing with my erratic schedule and diligently proofreading *all* my papers, but most of all, for being by my side through this tumultuous process. This dissertation is as much a fruit of her dedication as it is of my labor and words can do little to express the gratitude I feel towards her.

ABSTRACT

ESSAYS ON ONLINE ADVERTISING

Vibhanshu Abhishek

Kartik Hosanagar

Peter S. Fader

In this dissertation, we study different dimensions of online advertising by focusing on research questions which are motivated by marketing and managerial considerations. The first essay highlights the inadequacies of data standards that are commonly used in sponsored search. We show that these datasets result in aggregation bias and propose alternate data standards that do not suffer from this bias. An equilibrium analysis is performed to analyze the effect of this bias on the search engine and advertisers and it is shown that the search engine tend to lose the most from this bias. The second essay focuses on an advertiser's problem of bidding optimally in sponsored search. Uncertainty in the decision-making environment, budget constraints and the presence of a large portfolio of keywords makes the bid optimization problem non-trivial. We formulate this problem mathematically and propose a "myopic" policy for one-period optimization. This policy is extended by incorporating interactions between keywords, in the form of positive spillovers from generic keywords into branded keywords. This multi-period "forward-looking" policy uses a Nerlove-Arrow model to capture the long-term interactions between these keywords. The spillovers are estimated using a dynamic linear model and used to jointly optimize the bids of the keywords using an approximate dynamic programming approach. In the third essay we discuss the problem of attribution in the context of online ads. We formulate a dynamic Hidden Markov Model to capture a consumer's behavior during the purchase process and how this process is affected by ads. This model is subsequently used to evaluate the role that each ad plays in a consumer's eventual conversion in order to solve the attribution problem.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Introduction

The last several years have seen a dramatic increase in the amount of time and money consumers spend online. A recent survey reports that consumers spend around 33% of their time online (IAB 2011). As a consequence, the Internet has become an important channel that firms can use to reach out and connect to consumers which has lead to the emergence of online advertising. Several forms of online advertising has evolved over the year, e.g. sponsored search, display advertising, email marketing, classifieds and more recently social network advertising, sponsored search and display advertising dominating the online marketing landscape. Internet advertising has become a significant component of advertisers' marketing mix. Firms spent close to US$ 35 Billion on online advertising in the year 2011 and this expense is expected to double in the next 5 years. The online advertising spend surpassed the amount spent on newspaper advertising in the year 2010 and is expected to surpass television advertising by the year 2016.

There are several factors that have contributed to this growth in online advertising – (i) measurability, (ii) scale and (iii) targetability. Firstly, it has been extremely difficult to quantify the returns from traditional advertising. Online advertising on the other hand is highly measurable, with search engines and ad networks tracking every action of consumers. This unprecedented amount of data has helped advertisersfine tune their campaigns and increasing their returns from online advertising. Secondly, a huge amount of traffic flows through channels like search engines and with the ever increasing time consumers spend on the Internet, the ad inventory available online is significantly more than one could imagine in a traditional medium like television or print. Thirdly, with development of technologies like geolocation and persistent cookies, it is very easy to track and target customers at a level of granularity which was impossible to imagine in traditional advertising.

Online advertising has lead to some fundamental changes in the field of advertising. Traditional advertising was the forte of large firms with large advertising budgets. Due to the incremental costs of online advertising, e.g. *per-click* or *per-impression*, advertising expenses can be finely controlled at a daily level. For the first time small and medium sized local firms can directly reach consumers through online advertising. The fine level of control has not only made advertising more efficient but might also lead to an increase in the amount spent on advertising, most of this growth fueled by smaller local firms. In addition, traditional advertising has primarily been *push-based* where ads are shown to consumers when they are engaged in some unrelated activity like reading or watching television. New innovations in advertising like sponsored search are primarily *pull-based* – they show ads to consumers when they are looking for something and since these ads are contextually relevant, consumers have a higher probability of engaging with these ads. The effectiveness of sponsored search vis-a-vis other channels has been fairly well documented (Ghose and Yang, 2009, Goldfarb and Tucker, 2010, Agarwal et. al., 2011). From an academic perspective online advertising presents very granular, individual level data which has eluded researchers and marketeers in traditional advertising, allowing them to gain a better insight into consumer behavior.

Given the scale and novelty of online advertising, there is a growing need to understand how consumers respond to online ads and how firms should advertise using this medium. In my dissertation, I study different aspects of sponsored search and display ads which constitute the bulk of online advertising. In the first essay, I focus on the issues related to the use of aggregate data in sponsored search. I demonstrate that models commonly used in sponsored search research suffer from aggregation bias and present the implications of this aggregation bias. In the second essay, I focus on the advertiser's problem of bidding optimally in sponsored search auctions. In the third essay, I study the interactions between various forms of online advertising like banner ads, display ads and sponsored search ads and address the problem of attribution. A detailed description of these essays is presented below.

Essay 1: Aggregation Bias in Sponsored Search Data - Existence and Implications

There has been significant recent interest in studying consumer behavior in sponsored search advertising (SSA). Researchers have typically used daily data from search engines containing measures such as average bid, average ad position, total impressions, clicks and cost for each keyword in the advertiser's campaign. A variety of random utility models have been estimated using such data and the results have helped researchers explore the factors that drive consumer click and conversion propensities. However, virtually every analysis of this kind has ignored the intra-day variation in ad position. We show that estimating random utility models on aggregated (daily) data without accounting for this variation will lead to systematically biased estimates – specifically, the impact of ad position on click-through rate (CTR) is attenuated and the predicted CTR is higher than the actual CTR. We demonstrate the existence of the bias analytically. These findings are empirically validated using a large dataset from a major search engine. We show the effect of the bias on the equilibrium of the SSA auction. Search engines are always affected adversely due to aggregation bias. In addition, we show that advertisers receive a higher payoff from SSA when they all use complete data. Finally, we empirically quantify the losses suffered by the search engine and an advertiser using aggregate data. The search engine loses over 17% of its revenues on average. We also observe that an advertiser could unilaterally increase his revenues by around 6% if he used disaggregate data.

Essay 2: Optimal Bidding in Multi-Item Multi-Slot Sponsored Search Auctions

We study optimal bidding strategies for advertisers in sponsored search auctions. In general, these auctions are run as variants of second-price auctions but have been shown to be incentive incompatible. Thus, advertisers have to be strategic about bidding. Uncertainty in the decision-making environment, budget constraints and the presence of a large portfolio of keywords makes the bid optimization problem non-trivial. We present an analytical model to compute the optimal bids for keywords in an advertiser's portfolio. To validate our approach, we estimate the parameters of the model using data from an advertiser's sponsored search campaign and use the bids proposed by the model in a field experiment. The results of the field implementation show that the proposed bidding technique is very effective in practice and leads to a 75.38% increase in the advertiser's profits. We extend our model to account for interactions between keywords, in the form of positive spillovers from generic keywords into branded keywords. We use a Nerlove-Arrow model to capture the effect of generic click and impressions on awareness and the subsequent spillovers into branded search activity. The spillovers are estimated using a dynamic linear model framework and used to jointly optimize the bids of the keywords using an approximate dynamic programming approach. This policy is slightly better than the policy that ignores interactions between keywords. It leads to an overall increase of 83.25% in the advertiser's profits.

Essay 3: The Long Road to Online Conversion: A Model of Multi-Touch Attribution

Consumers are exposed to advertisers across a number of channels. For example, a consumer may be exposed to an advertiser's display ads at multiple websites followed by sponsored search ads for a number of queries. So a conversion or a sale may be the result of a series of ads that were displayed to the consumer. This raises the key question of attribution: which ads get credit for a conversion and how much credit do each of these ads get? The issue has received considerable attention in the industry. Although the issue is well documented, current solutions are often simplistic. For example, a common practice is to attribute the sale to the most recent ad exposure. In some instances, firms apply exponentially decreasing weights based on time of ad exposure. These methods of attribution penalize prior exposures and give undue credit to ad exposures that occur just before the sale. As all the ads collectively influence the consumer's decision to make a purchase, it is difficult to disentangle the contribution of various ads towards the eventual sale. The problem of attribution is exacerbated because different entities manage different components of advertising like sponsored search and display advertising which has impeded development of a unified attribution frame work. In this paper, we address the problem of attribution using a unique data-set from a digital ad agency that managed the entire online campaign during a car launch. We present a Hidden Markov Model of an individual consumer's behavior based on the concept of a conversion funnel that captures the consumer's deliberation process. We observe that different ad formats, e.g. display and search ads, affect the consumers differently and in different states of their decision process. Display ads usually have an early impact on the consumer, moving him from a state of dormancy to a state where he is aware of the product and it might enter his consideration set. However, when the consumer actively interacts with these ads (e.g. by clicking on them), his likelihood to convert considerably increases. Secondly, we present an attribution scheme based on the proposed model that assigns credit to an ad based on the incremental impact it has the consumer's probability to convert.

CHAPTER 2 : Aggregation Bias in Sponsored Search Data: The Curse and The Cure

2.1. Introduction

Sponsored search advertising (SSA) has not only transformed the way companies conduct their marketing activities, but it has also been a tremendous resource to academic researchers who seek to better understand how consumers respond to such ads. A myriad of researchers have turned to SSA data to uncover new insights about consumer search (Ghose and Yang, 2009; Rutz and Bucklin, 2011), choice and related purchasing behaviors (Jeziorski and Segal, 2009; Yang and Ghose, 2010; Agarwal et al., 2011) and advertiser/search engine strategies (Animesh et al., 2009; Yao and Mela, 2011; Rutz et al., 2012). Many of these papers have used random utility models to study the effect of ad position, keyword length, presence or absence of brand name, etc. on the click-through and conversion rates of the ads.

Sponsored search refer to ads that are displayed alongside organic search results when a user issues a query at a search engine. The advertisers submit bids for keywords that are relevant to them, along with these ads.[1] When a user enters a query, the search engine identifies the advertisers bidding on keywords closely related to the query and uses data on bids and ad quality/performance to rank order the ads in a list of sponsored results. The most widely used pricing model is the *pay-per-click* model, in which the advertiser pays only when a user clicks on his ad. The advertiser's cost per click or CPC is determined using a generalized second price auction (GSP), i.e. whenever a user clicks on an ad at a particular position, the advertiser pays an amount equal to the minimum bid needed to secure that position.

Although SSA is a relatively new practice, it already has fairly well-established data standards associated with it. Most researchers who have modeled SSA-related issues have

[1]The term *keyword* refers to term or phrase on which an advertiser bids. *Query* (or *query term*) is the search phrase entered by the consumer when conducting the search

worked with a data structure such as the one illustrated in Table 1. Advertisers also obtain

Table 1: Sample dataset for a particular keyword

date	impressions	clicks	avg. pos	avg. bid	avg. CPC
01/11/09	180	1	19.33	1.00	0.30
01/12/09	202	0	18.42	1.00	0.00
01/13/09	212	0	17.12	1.00	0.00
01/14/09	223	5	8.19	2.00	1.24
01/15/09	166	4	7.59	2.00	1.03
01/16/09	198	3	7.94	2.00	0.89
01/17/09	197	5	8.08	2.00	1.21
01/18/09	321	21	2.00	3.00	2.12

similar datasets from search engines and analyze them to design their bidding policies. In almost all cases, these data are aggregated to the daily level and contain summary statistics for the day, such as the number of ad impressions, average position of the ad, number of clicks received and the average CPC. It should be clear how this kind of dataset lends itself to the types of models mentioned above, as well as analysis of a variety of other customer behaviors (and related firm actions). But despite the creativity and methodological prowess that has been demonstrated in this growing body of literature, we believe that these modeling efforts are plagued by a major problem: an aggregation bias due to the way that the raw (search-by-search) data are "rolled" up into Table 1.

In practice, the position of an ad can vary substantially within a day and aggregated data fail to capture this variation. It is also widely known that the impact of position on CTR is non-linear. For example, an ad at the topmost position tends to receive a disproportionately large number of clicks as compared to the other positions. The convexity in the CTR, coupled with the intra-day variation in position suggests that the daily aggregation might lead to estimation bias. The goal of this paper is to provide a thorough evaluation of the nature of this bias and to demonstrate its effects.

The paper makes the following contributions. Firstly, we show that applying logistic model to aggregated SSA data can lead to biased estimation of the parameters of a random utility model. Due to the bias, the effect of position on CTR is attenuated and the predicted CTR is higher than the actual CTR. Secondly, we study the changes in the equilibrium of the SSA auction induced by the bias. Since the advertisers use aggregate data to arrive at incorrect estimates of the ctr-position curve, one would assume that they are negatively affected by this bias. Surprisingly, if all the advertisers use aggregate data, the consequence of the bias is completely borne by the search engine. Thirdly, using a large disaggregate dataset from a major search engine, we quantify the magnitude of the bias and measure its economic impact. We extend the Hierarchical Bayesian (HB) model using a latent instrumental variable (LIV) approach to address the issue of position endogeneity which is common in sponsored search. We find that the search engine loses over 17% of its revenue on average due to the aggregation bias. Our findings raise serious concerns for SSA researchers and practitioners and also question the adequacy of the data standards that have become common in SSA. Finally, we present some data summarization techniques that can reduce or eliminate the bias.

The rest of the paper is organized as follows. Section 2 discusses related work and positions our work in the literature. In Section 3, we analytically prove the existence of the bias and build a game-theoretic model to study the economic impact of the bias. In Section 4 we analyze a large disaggregate dataset from a search engine using the HB-LIV model. We present the managerial implications of the bias in Section 5. Then we present various data summarization techniques in Section 6 and finally discuss the implications of the bias on research and practice and conclude the study in Section 7.

2.2. Related work

There has been a considerable amount of work on auction design and consumer choice models in SSA (Weber and Zheng, 2007; Liu and Whinston, 2007; Hao et al., 2009; Goldfarb and Tucker, 2007). More specifically, there are two streams of work that are closely related to our study, namely empirical research on consumer click and conversion behavior in SSA and work related to aggregation biases in choice models.

Empirical Research in Sponsored Search

There has been a lot of recent interest in trying to understand the factors driving keyword performance in SSA. Craswell et al. (2008) and Ali and Scarr (2007) propose individual keyword-level models to study how consumers navigate sponsored links. Other researchers have used logit models to measure the influence of factors like ad position and keyword characteristics on the consumer behavior in SSA (Rutz et al., 2012; Rutz and Bucklin, 2011; Ghose and Yang, 2009; Agarwal et al., 2011). Rutz et al. (2012) compare the performance of several logit models in predicting the conversions for various keywords. Their results show that keywords are heterogeneous in their conversion rate and a significant portion of this variation can be explained by the presence of brand or geographical information in the keyword. In another paper, Rutz and Bucklin (2011) measure the spill-over effect of generic keywords on branded keywords. Ghose and Yang (2009) use a random effect logit model to understand the relationship between different metrics such as CTR, conversion rates, bid prices and ad position using the advertiser's aggregate data. They show that keywords containing retailer information have a higher CTR whereas keywords that are more specific or contain brand information have a lower CTR. Recent work by Agarwal et al. (2011) uses a logit model to show that although the CTR decreases with position, the conversion rate is non-monotonic in position. They point out that the topmost position is not necessarily the revenue maximizing position.

Most of this stream of research uses aggregate data to estimate the parameters of the model. The aggregate data obfuscate the variation in ad position and research in this area has overlooked this fact. Ignoring this variation can lead to potential biases in the estimation of parameters and ultimately affect the conclusions from these studies.

Aggregation Bias

Though researchers have grappled with the issue of data aggregation for many years, there is no clear consensus on this issue. The problems associated with aggregation have been commonly encountered in spatial and demographic studies and are referred to as the Yule-Simpson effect (Good and Mittal, 1987). The drawbacks of aggregation have also been pointed out in various studies in the economics and marketing literature. Neslin and Shoemaker (1989) point out the limitations of aggregate data by refuting the claim that sales promotions undermine the consumer's repeat-purchase propensity. They show that even if the individual purchase propensities do not change before and after promotions, statistical aggregation would lead to lower average repeat probabilities for post promotional purchases. Yatchew and Griliches (1985) discuss the implications of aggregation in the context of probit models. Issues related to data aggregation in the case of logit models have been presented by Kelejian (1995). He discusses why aggregation bias might occur when logit models are estimated on aggregate data and proposes a test for the existence of this bias.

On the other hand, several researchers believe that the effect of aggregation is negligible or absent when the disaggregate model can be approximated by the aggregate model (Gupta et al., 1996; Russell and Kamakura, 1994). Using household-level panel data and store-level purchase data, Gupta et al. (1996) show that the price elasticity estimated from the two models differ by a very small amount (4.7%). Allenby and Rossi (1991) present an analytical proof for the non-existence of aggregation bias in nested logit models of consumer choice when the products are close substitutes of each other, though they assume that the micro-level consumer behavior is approximately linear in the product attributes.

The discussion reveals two themes. First, a number of recent studies have applied the logit model on aggregated SSA data to study consumer choice behavior. Second, although aggregation bias has been shown to exist in a number of environments, its non-existence has also been demonstrated in several other environments. It is not clear which of these arguments is most applicable in the SSA context, and it is thus not clear whether and to what extent aggregation bias affects SSA research. This paper uses a theoretical model to show why data aggregation might lead to biased estimates in the SSA context. An extensive disaggregate search engine dataset is used to empirically measure the extent of aggregation bias in SSA research which we believe has been done for the first time. Finally, we discuss the economic consequences of aggregation bias.

2.3. Aggregation Bias

In this section we explore the estimation bias due to the aggregation of SSA data. We begin by pointing out the distinction between the complete (disaggregate) and the summary (aggregate) data that have been referred to in the paper. Table 2 is a stylized example that presents impression level data for an ad, reflecting every search query for a particular term. Each observation contains the date on which the impression occurred, position of the ad, bid placed by the advertiser, whether the consumer clicked on the ad and finally the CPC. Search engines usually do not provide such granular data to advertisers or researchers. They provide aggregated data at the daily level as shown in Table 1 earlier that mask the intra-day variation in position.

The intra-day variation arises due to two major factors – Firstly, SSA auctions are extremely dynamic with advertisers entering and exiting the auction or changing their bids continuously. Change in the competitors' behavior lead to a change in ad position. Secondly, most of the ads, specifically *broad match* and *phrase match* ads are shown for a number of different queries.[2] As the set of competitors can be different for different queries, the position of the ad also varies across queries.

2.3.1. Analytical Proof of Aggregation Bias

Logit utility model

The logit utility model has been extensively used in economics and marketing to explain consumer choice behavior. Researchers have primarily focused on keyword-level models to analyze the effect that factors like ad position, specificity of the keyword, presence of

[2]An *exact match* occurs when the users query term exactly matches the advertisers keyword. A *phrase match* occurs when the advertisers keyword appears anywhere within the users query. Finally, a *broad match* occurs when user query is determined to be broadly similar to advertisers keyword. Broad match is commonly used by advertisers as it maximizes the number of ad impressions.

Table 2: Complete dataset for a particular keyword

impression	date	click	pos	bid	CPC
1	01/11/09	0	16	1	0.00
2	01/11/09	0	20	1	0.00
3	01/11/09	0	18	1	0.00
...	...	...	...	...	
8190	02/15/09	0	6	2	0.00
8191	02/15/09	0	5	2	0.00
8195	02/15/09	0	6	2	0.00
...	...	...	...	...	
9145	02/23/09	1	1	3	2.31

brand name and other variables have on the consumer's propensity to click on the ad. The consumer's utility has been modeled as

$$U = X'\boldsymbol{\beta} + \epsilon, \tag{2.1}$$

where X is a vector of covariates, $\boldsymbol{\beta}$ is consumer sensitivity to these attributes and $\epsilon \sim Logistic(0, 1)$. In binary choice models, this utility is not observed but constitutes a latent variable. The consumer clicks on an ad when $U > 0$, i.e.

$$Y_i = \begin{cases} 1 \text{ iff } U > 0, \\ 0 \text{ otherwise,} \end{cases}$$

where Y is a variable that denotes whether a click was made or not. In accordance with prior research, we build a keyword-level model that ignores other ad characteristics and focuses our attention on the impact of ad position on click through rate, i.e. $U = f(position)$. The simple model allows us to clearly identify the existence and direction of the bias. However, this assumption does not impose any restrictions on the model as all keyword characteristics (which do not change during the day) are subsumed in the intercept term and the we focus our attention on position which varies intra-day. Although we focus on a keyword level model in this paper, our finding are applicable for different levels of analysis.

Estimation using complete dataset

We now discuss the estimation of $\boldsymbol{\beta}$ when the model is estimated using the complete dataset. Let V_i be a random variable denoting the ad position on the i^{th} impression. We assume that V_i is independent and identically distributed and has a distribution given by $F_V(.)$ which is assumed to be constant during the period of observation. The consumer's utility is given by the following expression

$$U_i = \beta_0 + \beta_1 V_i + \epsilon_i. \tag{2.2}$$

As ϵ_i is extreme value distributed, the probability of clicking on an ad (CTR) is $p_i = 1/(1 + \exp -\{\beta_0 + \beta_1 v_i\})$. Note that p_i might vary across impressions as the ad position is varying as well. Let $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ denote the maximum likelihood estimate from the complete dataset. It can be shown that $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ satisfies the following equation[3]

$$obsctr = \frac{1}{I}\sum_{i=1}^{I} \frac{\exp\{\hat{\beta}_{0,\mathrm{c}} + \hat{\beta}_{1,\mathrm{c}} v_i\}}{1 + \exp\{\hat{\beta}_{0,\mathrm{c}} + \hat{\beta}_{1,\mathrm{c}} v_i\}}. \tag{2.3}$$

where *obsctr*, the observed click-through rate, is the fraction of the impressions that result in a click and it equals C/I, where C is the total number of clicks from I ad impressions in the entire dataset. v_i is the realization of V_i for a particular ad impression. An important property of $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ is as follows.

Lemma 1 $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ *is a consistent and unbiased estimator of* $\boldsymbol{\beta}$*.*

Estimation using aggregate dataset

Researchers do not observe V_i when aggregate data are used. They only observe the mean daily position W which is given by

$$W = \frac{V_1 + V_2 + \ldots + V_N}{N}, \tag{2.4}$$

[3]Derivations of all equations appear in the Online Appendix.

where N is the random number of ad impressions on a particular day and $V_1, \ldots, V_N$ is the ad position during each of those impressions. The distribution of W is $F_W(.)$ and it depends on the distribution of V and N. If the effect of position is estimated from aggregate data, the consumer utility from clicking the ad is effectively modeled as

$$U_d = \beta_0 + \beta_1 W_d + \epsilon_d, \tag{2.5}$$

where W_d is the average position for the day and ϵ_d is the logistically distributed error term. This formulation causes a mis-specification as the consumers do not observe the ad at a position W_d but at position V. As the variable $Z = V - W_d$ ($\mathbb{E}[Z|W] = 0$), which affects the consumer's click behavior, is not accounted for in the regression, the mis-specification is similar to omitted variables bias pointed out by Yatchew and Griliches (1985) and Wooldridge (2001). However, this issue arises primarily due to data aggregation and our approach is closely related to prior work in Marketing by Christen et al. (1997), Steenkamp et al. (2005) and Gupta et al. (1996). Further, as $Var(Z)$ is not constant (it depends on the number of impressions in a days), the findings of Yatchew and Griliches (1985) and Wooldridge (2001) are not directly applicable in this context. Hence, we derive an important relationship between W and V in order to prove the aggregation bias.

Lemma 2 *W is less than V in convex order.*[4]

$$W \leq_{\text{cx}} V \tag{2.6}$$

This relationship between W and V is very general and holds for any distribution, $F_V(.)$.

Let the number of impressions on day d be denoted by n_d and the number of clicks by c_d; $\sum_{d=1}^{D} n_d = I$ and $\sum_{d=1}^{D} c_d = C$. The maximum likelihood estimate from summary data,

[4] X is less than Y in *convex order* if $E[f(X)] \leq E[f(Y)]$ for all real convex functions f such that the expectation exists. All proofs appear in the Appendix.

$\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$, satisfies the following equation,

$$obsctr = \frac{1}{I}\sum_{d=1}^{D}\frac{n_d\exp\{\beta_{0,\mathrm{s}}+\beta_{1,\mathrm{s}}w_d\}}{1+\exp\{\beta_{0,\mathrm{s}}+\beta_{1,\mathrm{s}}w_d\}} \tag{2.7}$$

Using Lemma 2 and the preceding result, we prove an important result of this paper.

Proposition 1 $\hat{\boldsymbol{\beta}}_{\mathbf{s}}$ *is a biased estimator of* $\boldsymbol{\beta}$*.*

If $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ is equal to $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ then the convex order between W and V implies that

$$\mathbb{E}\left[\frac{\exp\{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V\}}{1+\exp\{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V\}}\right] > \mathbb{E}\left[\frac{\exp\{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W\}}{1+\exp\{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W\}}\right].$$

As both the L.H.S. and the R.H.S. equal *obsctr* (as shown in Appendix A), this inequality is incorrect and hence $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ cannot equal $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$. Since $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ is a consistent and unbiased estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ is biased. This finding is contrary to earlier work by Allenby and Rossi (1991), Gupta et al. (1996) and Russell and Kamakura (1994) which prove that aggregation bias in market or store-level scanner data is negligible. Aggregation bias is significantly reduced in these models as products are very close substitutes of each other and the consumers (or house-holds) are exposed to very similar marketing activities. However, position has a very strong effect in sponsored search (Craswell et al., 2008) and ads in different positions are perceived very differently by consumers. Coupled with variation in ad position, aggregation bias can be quite substantial in sponsored search, a hypothesis we test in Section 4.

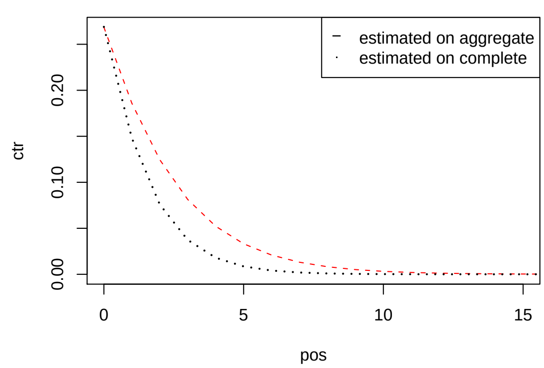

Figure 1: Logistic regression on complete and aggregate data.

Due to the convexity of the CTR-position curve and the variation in position, the mean daily CTR at every position lies above the true CTR (see Figure 1).The relationship between the estimated and the actual CTR is presented in the following proposition.

Proposition 2 *The direction of aggregation bias is such that (i) the CTR estimated from the summary data is greater than or equal to the actual CTR at any position, and (ii)* $\hat{\beta}_{1,s} > \hat{\beta}_{1,c}$.

$\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ is biased and it predicts a CTR that is higher than the actual CTR at any position. Incorrect estimation of CTR might lead advertisers to make suboptimal choices in sponsored search auctions. The second part of Proposition 3 is consistent with Wooldridge (2001) which shows that the estimates are scaled towards zero.

The omitted variable Z has the effect of increasing the disturbance term in the regression. The estimated error can be computed as a convolution of ϵ and Z, but this is analytically intractable as ϵ is logistically distributed. If Z and ϵ can be approximated closely by a normal distribution, the asymptotic behavior of $\hat{\beta}_s$ is as follows:

$$\hat{\beta}_{1,s} \xrightarrow{p} \frac{\beta_1}{\sqrt{1 + \frac{\phi Var(V)}{\pi^2/3}}}, \tag{2.8}$$

where $\phi = \mathbb{E}[1/N]$ and is bounded by $(1-(1-e^{-\lambda}))/\lambda \leq \phi \leq (1-[1-F_\lambda(1)+3(1-F_\lambda(2))/m])$ where $\lambda = \mathbb{E}[N]$. When there are a large number of daily impressions, i.e. $\lambda \to \infty$, $\phi \to 1$. Equation 2.8 shows that, if the variation in the the intra-day position is known, the actual β_1 can be approximated by multiplying $\hat{\beta}_{1,s}$ by the scaling factor. However, this approach suffers from several simplifying assumptions that are required for analytical tractability. In Section 5, we provide more general and robust empirical methods to remove the bias.

2.3.2. Effect on Equilibrium Behavior

As advertisers use estimates from historical data to bid in SSA auctions, incorrect estimation of the CTR might have a negative impact on the advertiser and search-engine profits. In this section, we analyze how the impact of the bias is shared between the advertisers and the search engine and who is affected more by aggregation.

For our analysis, we make the following assumptions – *(i)* there are K advertising slots and $K+1$ advertisers, *(ii)* the advertisers' valuation for a click, s_k are drawn from a continuous distribution with support on $[0, \infty)$, *(iii)* advertisers know their own valuation and the distribution of bids and finally, *(iv)* the click-through rate α_i decreases with the position i. In addition, we also assume that advertisers estimate α_i from historical data and are unaware of the aggregation bias. The advertisers are indexed in decreasing order of their valuations, i.e. $s_1 > s_2 > \ldots s_{K+1}$ and their bids are $b_1, \ldots, b_{K+1}$, respectively.[5] In addition, $h = (b_i, \ldots, b_{K+1})$ refers to the history of bids prior to assignment of position i.[6] The case where complete data are used is analyzed first. Here the advertisers correctly estimate α_i. Edelman et al. (2007) show that under the above assumptions, there exists a unique *envy-free* perfect Bayesian equilibrium and the optimal strategy for advertiser k under this equilibrium is to bid as follows,

$$b_k(s_k, i, h) = s_k - \frac{\alpha_i}{\alpha_{i-1}}(s_k - b_{i+1}). \tag{2.9}$$

This is the maximum CPC that the advertiser is willing to pay to move to position $i-1$ and receive more clicks. At this point, Advertiser k is indifferent between getting position $i-1$ at a CPC of $b_k(s_k, i, h)$ and position i at b_{i+1}. This is an ex-post equilibrium, i.e., it is optimal for Advertiser k to follow the equilibrium strategy for any realization of other advertisers' valuations. The search-engine revenue is $\Pi_S^C = \sum_{i=1}^{K} \alpha_i b_{i+1}$ and the payoffs for advertiser i is $\Pi_i^C = \alpha_i(s_i - b_{i+1})$. It should be noted that this equilibrium ensures an assortive match, i.e., if $s_i > s_j$ then advertiser i bids higher than advertiser j and occupies a slot above advertiser j in equilibrium. This case serves as a reference for the ensuing discussion.

[5] b_{K+1} is set to 0.
[6] Positions K through $i+1$ are assignment before bidding for position i starts.

Next, we consider the case in which aggregate data are used. Let the CTR estimated from aggregate data be denoted as $\alpha_i^{'}$.

Remark 1 *When aggregate data are used to estimate CTR, the ratio α_i/α_{i-1} is overestimated due to the presence of the aggregation bias.*[7]

Due to this overestimation, advertisers might bid incorrectly. As the equilibrium considered here is ex-post, the advertisers' bidding strategies depend neither on their beliefs about each others valuations nor on the fact that some advertisers might be using aggregate data. The bidding strategies continue to be similar to the one outlined in Equation 2.9, but the bids in this case, $b_1^{'}, \ldots, b_K^{'}$, might be different.[8] We consider two extreme cases to study the impact of aggregation bias – (i) all advertisers except one use complete data and, (ii) all advertisers use aggregate data. Let the search-engine revenue in Case I(II) be denoted by $\Pi_S^{AI(II)}$ and advertisers' profit by $\Pi_i^{AI(II)}$.

Case I: Suppose advertisers other than Advertiser j have access to complete data and can compute α_i correctly. Only Advertiser j uses aggregate data and overestimates β_1. This leads him to overestimate the ratio α_i/α_{i-1} and he bids in the following manner.

$$b_j^{'}(s_j, i, h) = s_j - \frac{\alpha_i^{'}}{\alpha_{i-1}^{'}}(s_j - b_{i+1}^{'}). \tag{2.10}$$

As Advertiser j bids lower in equilibrium and occupies a position $j^{'} \geq j$. The following proposition characterizes the equilibrium in this case (detailed analysis and proofs are provided in Appendix C).

Proposition 3 (i) *If only Advertiser j uses aggregate data, the top advertisers ($i \leq j$) bid lower, advertisers in between ($j < i \leq j^{'}$) bid higher and the remaining advertisers ($i > j^{'}$) bid the same as they would have when everyone had complete data.* (ii) *The payoffs of the*

[7]Writing the CTR in terms of the logit model we get, $\frac{\alpha_i}{\alpha_{i-1}} = \frac{\exp\{\beta_0+\beta_1 i\}}{1+\exp\{\beta_0+\beta_1 i\}} \times \frac{1+\exp\{\beta_0+\beta_1(i-1)\}}{\exp\{\beta_0+\beta_1(i-1)\}} \approx \exp\{\beta_1\}$ when CTRs are small. Since $\hat{\beta}_{1,s} > \hat{\beta}_{1,c} \Rightarrow \alpha_i/\alpha_{i-1} < \alpha_i^{'}/\alpha_{i-1}^{'}$.

[8]We continue to assume that Advertiser $K+1$ still bids 0.

search engine and Advertiser j decrease ($\Pi_S^{AI} < \Pi_S^C, \Pi_j^{AI} < \Pi_j^C$) while all other advertisers receive payoffs that are either the same or higher than payoffs they would have received if all advertisers were using complete data ($\Pi_i^{AI} \geq \Pi_i^C, i \neq j$).

Advertiser j underestimates the impact of position and incorrectly bids lower which might move him to a lower position. In turn, some advertisers who were below him move up one position. The ordering of these advertisers does not change, which is a consequence of the bidding policy (Edelman et al., 2007). As $b_j^{'} < b_{j'}$, bids required to acquire all positions above $j^{'}$ decrease. As bids for all position are (weakly) lower, the search engine loses revenue. Clearly, Advertiser j's payoff is lower, because he deviates from the optimal policy. However, the loss in revenue for the search engine is substantially higher than the loss in revenue for the advertiser using aggregate data. Interestingly, all these losses are transferred to the other advertisers ($\neq j$) as excess surplus since GSP is a zero-sum game. Hence, the search engine suffers the most due to aggregation bias and all advertisers apart from j are better off due to aggregation. In the subsequent case, we observe that the search engine internalizes all the negative impact of aggregation.

Case II: When all advertisers use aggregate data, their estimate of the CTR, $\alpha_i^{'}$ are greater than the actual CTR as shown in Proposition 3. For simplicity, we assume that all advertisers arrive at the same estimates for $\alpha_i^{'}$.[9] As a result, Advertiser k adopts the following bidding strategy,

$$b_k^{'}(s_j, i, h) = s_k - \frac{\alpha_i^{'}}{\alpha_{i-1}^{'}}(s_k - b_{i+1}^{'}).$$

It is easy to see that the bid placed by advertiser K (the last advertiser) is less than the bid he would have placed had he estimated CTR from complete data. Proceeding in an iterative fashion we can show that all advertisers place a lower bid. The equilibrium in this case is specified in the following proposition.

[9] This result continues to hold even if the advertisers arrive at different estimates of $\alpha_i^{'}$ as long as $\alpha_i/\alpha_{i-1} < \alpha_i^{'}/\alpha_{i-1}^{'}$, which always hold true due to aggregation bias as shown earlier.

Proposition 4 (i) *When all advertisers use aggregate data, the advertisers are arranged in assortive order. The resulting bids are lower than the bids when complete data are used* ($b_i^{'} < b_i, i = 1, \ldots, K$). (ii) *Search-engine revenue is lower* ($\Pi_S^{A2} < \Pi_S^C$) *and advertisers' payoff are higher* ($\Pi_i^{A2} > \Pi_i^C$) *as compared to the complete case.*

In Appendix C, we show that the advertisers bid less than what they would have had they known the actual CTR. As all the advertisers use the same incorrect estimate of the CTR, the eventual ranking remains the same as in the complete case. They receive the same number of clicks but at a lower CPC and hence their payoffs are higher. Surprisingly, the search-engine revenue suffers the most when all advertisers use aggregate data even though the advertisers make the wrong decisions. These results question the data standards that have become common in SSA and underscore the need to provide better data to advertisers. We also show that it is incentive compatible for search engine to provide richer/better data to advertisers.

Note that an advertiser always receives a higher payoff when he uses complete data as compared to aggregate data, irrespective of the fraction of advertisers using aggregate data. This intuition is formalized in the following corollary.

Corollary 1 *An advertiser can always increase his payoff from SSA by unilaterally using complete data instead of aggregate data.*

The difference in payoff between the two cases (complete v/s aggregate) can be considered as the value of complete data or alternately the disutility from aggregate data. In the following section, we analyze a large dataset from a leading search engine. The purpose of this analysis is two folds: Firstly, we quantify the magnitude of the bias and show that it is significant in the context of SSA. Secondly, we use these estimates to compute the revenue implications of the bias for the search engine and a representative advertiser.

2.4. Empirical Analysis

2.4.1. Data Description

We analyze a disaggregate dataset from a major search engine that contains 22 million unique impressions chosen randomly from all user queries between August 10, 2007 and September 25, 2007. This is a very unique dataset as search engines rarely provide impression level data to advertisers or researchers. For every impression, the dataset contains the user query, ads shown on the page and number of ads on the preceding pages. Each ad is identified by a unique ad identifier, though the dataset does not contain any ad-specific information. The dataset also contains information about clicks during this period of observation. We construct an ad-level dataset, that contains information about the keyword the advertiser was bidding on and all the impressions of the ad associated with the keyword, which is similar to the one presented in Table 2.

Table 3: Summary Statistics

Total impressions	8,142,210
Unique queries	24,235
Unique ads	229,960
Ads with more than one impression	184,481
Mean impressions for every ad	64.4
Median impressions for every ad	7.0

There is ample evidence that there is substantial variation in position and reporting the average position alone results in the loss of information on actual position as shown in Figure 2. We next investigate the impact of data aggregation.

The ad level data described earlier are aggregated at a daily level to create aggregate data. The data thus generated are similar to the campaign summaries that search engines make available to the advertisers (as presented in Table 1).

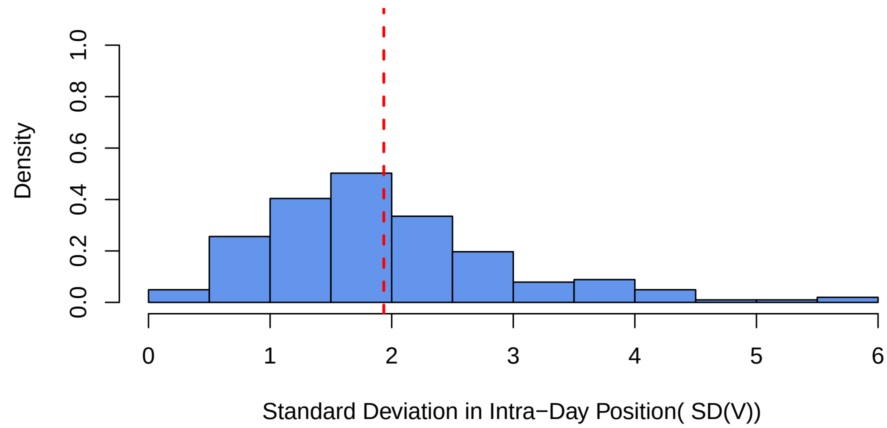

Figure 2: Mean Intra-day Variation in Position.

2.4.2. Hierarchical Bayesian Model

We estimate a random-effect logit model using Hierarchical Bayesian (HB) techniques that are commonly used in SSA. As our data do not contain any ad-specific attributes, the only covariate included in our models is position. The effect of ad characteristics is captured in the ad-specific intercept term. We demonstrate the aggregation bias for HB model.

We extend the binary choice logit model proposed earlier in Section 3 to account for multiple keywords. Under this specification, the consumer's utility from clicking on ad k during impression i is given by

$$U_{ik} = \beta_{0k} + \beta_{1k} pos_{ik} + \epsilon_{ik} \tag{2.11}$$

where ϵ_{ik} is the idiosyncratic, logistically distributed error term. $\boldsymbol{\beta_k} = (\beta_{0k}, \beta_{1k})'$ are keyword specific parameters which are assumed to be random and heterogeneous across ads. It is drawn from a multivariate normal distribution in the following manner:

$$\boldsymbol{\beta_k} \sim N_2(\boldsymbol{\mu_\beta}, V_\beta) \quad \text{where} \quad V_\beta = \begin{pmatrix} \sigma_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_0\beta_1} & \sigma_{\beta_1} \end{pmatrix}.$$

Similar models have been extensively used in prior research in SSA (Ghose and Yang, 2009; Yang and Ghose, 2010). Note that although this random coefficient model captures heterogeneity across ads, it still fails to accounts for the intra-day variation in pos_{ik} if the model is estimated on aggregate data. As a result, we expect the aggregation bias to extend to the random-coefficient model as well. To test this hypothesis, we take a random sample of ads from our data and apply the model in Equation 2.11 to both the disaggregate and aggregate datasets and compare the estimates.

The log-likelihood function for the complete data is as follows

$$LL(\boldsymbol{\beta}|\text{complete data}) = \sum_{k=1}^{K}\sum_{i=1}^{I_K} Y_{ik}\log p_{ik} + (1 - Y_{ik})\log(1 - p_{ik}), \tag{2.12}$$

where Y_{ik} is the indicator variable that denotes whether the i^{th} impression of keyword k received a click or not and p_{ik}, the click-through probability is given by

$$p_{ik} = \frac{\exp\{\beta_{0k} + \beta_{1k}v_{ik}\}}{1 + \exp\{\beta_{0k} + \beta_{1k}v_{ik}\}}. \tag{2.13}$$

The log-likelihood function for the aggregate data is as follows

$$LL(\boldsymbol{\beta}|\text{aggregate data}) = \sum_{k=1}^{K}\sum_{d=1}^{D} c_{dk}\log p_{dk} + (n_{dk} - c_{dk})\log(1 - p_{dk}), \tag{2.14}$$

where n_{dk} and c_{dk} denote the number of impressions and clicks on day d respectively and p_{dk}, the click-through probability is given by

$$p_{dk} = \frac{\exp\{\beta_{0k} + \beta_{1k}w_{dk}\}}{1 + \exp\{\beta_{0k} + \beta_{1k}w_{dk}\}}. \tag{2.15}$$

As the data on clicks are often sparse for most keywords in sponsored search, the SSA literature primarily uses Hierarchical Bayesian models. We use a similar approach and assume that the mean and variance-covariance martix for $\boldsymbol{\beta_k}$ have the following priors

$$\boldsymbol{\mu_\beta} \sim N_2(\boldsymbol{\mu}, \Sigma), \tag{2.16}$$

$$V_\beta^{-1} \sim Wishart(\nu, \Delta). \tag{2.17}$$

The parameters $\mu, \Sigma, \nu, \Delta, \boldsymbol{\mu_\beta}$ and V_β^{-1} are estimated separately from the complete and the aggregate datasets using a Markov Chain Monte Carlo (MCMC) approach. Before discussing the details of the MCMC estimation procedure, we discuss some identification issues associated with the model presented here.

2.4.3. Identification

The ad position in the previous exposition has been assumed to be exogenous. However, the position is decided by the bids placed by the advertiser. In addition, we know that past performance affects the quality score of the ad which in turn affects the position. The auction process and historical performance jointly determine the position which is one of the most important strategic variables advertisers focus on in SSA. This indicates that position is endogenous (i.e. $\mathbb{E}[pos_t \epsilon_t] \neq 0$) and the endogeneity should be explicitly incorporated in the HB model presented earlier.

Endogenity has been a major concern in the SSA literature and researchers have proposed several techniques to address this issue. Ghose and Yang (2009) and Yang and Ghose (2010) use a simultaneous equation model to address this problem. Their simultaneous model forms a triangular system of equations which can be identified without any further identification constraints. Agarwal et al. (2011) use a series of random bids to address the endogenous nature of position. In their specification, position is completely determined by the random bids and quality score which are exogenous. Recent econometric advances have lead to the development of the latent instrument variable (LIV) framework (Ebbes et al., 2005) which has been used by Rutz et al. (2012) to account for position endogeneity. The LIV framework uses a likelihood based approach, which can be easily integrated with the

HB model proposed earlier, and can be estimated using the MCMC estimator.

In a LIV formulation, the endogenous covariate is decomposed into a stochastic term that is uncorrelated with the error and another one that is possibly correlated with the error, i.e. $X = \theta + \eta$ where θ is the uncorrelated part of X such that $\mathbb{E}[\theta\epsilon] = 0$ and $\mathbb{E}[\eta\epsilon] = \sigma_{\eta\epsilon}$. As θ varies in the dataset, it is possible to identify the correlation between η and ϵ, denoted by $\sigma_{\eta\epsilon}$. The LIV approach is extended to binary choice models by introducing a latent categorical variable with M categories (Rutz et al., 2012). Position can assume any of these M categorical values with a probability $\Pi = \{\pi_1, \pi_2, \ldots, \pi_M\}$, $\sum_{m=1}^{M}$. More specifically,

$$pos_{kt} = \mathbf{\Theta}'_{kt}\gamma + \boldsymbol{Z}'_{kt}\delta + \eta_{kt}, \tag{2.18}$$

where $\mathbf{\Theta} \sim Multinomial_M(\Pi)$ is stochastic part of pos_{kt} and is exogenous and η_{kt} is endogenous. The errors $(\eta_{kt}, \epsilon_{kt})'$ are MVN distributed in the following manner

$$\begin{pmatrix} \eta_{kt} \\ \epsilon_{kt} \end{pmatrix} = MVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta & \sigma_{\eta\epsilon} \\ \sigma_{\eta\epsilon} & \sigma_\epsilon \end{pmatrix}\right]. \tag{2.19}$$

$\boldsymbol{Z}_{kt}$ represents the observed instruments, and in our analysis we use lagged position as IVs which is consistent with the approach adopted by Rutz and Bucklin (2011).

2.4.4. Estimation Results

We estimate both the HB model and HB model with LIV (HB-LIV) to draw comparisons between the two methods. A sample size of 200 ads is chosen for estimating the parameters. We make this choice primarily for computational convenience as estimating the model on disaggregate data takes a long time. The disaggregate dataset contains a large number of observations, hence the estimation on the disaggregate dataset is really slow.[10]

We start off with diffused priors ($\mu = 0, \Sigma = 100I, \nu = 5, \Delta = \nu I$) and refine them as the estimation proceeds. The exact estimation procedure is outlined in the apendix. We run

[10]We estimate the HB model on 25 different samples and the qualitative findings remain the same.

the MCMC simulation for 100,000 draws and the first 50,000 sample are discarded. The MCMC chains are stationary after the burn-in period. The MCMC chains are thinned to remove autocorrelation between draws and every 10th draw in the stationary period is used for the subsequent analysis.

The estimation results are presented in Table 4. We observe that there are significant differences in the $\boldsymbol{\mu_\beta}$ estimated on the complete and aggregate data, for both the HB and HB-LIV models. μ_1 is overestimated by 10% when the estimation is performed on aggregate data indicating that aggregation bias exists when a HB model is used. The bias in μ_1 increases to 12.9% when a HB-LIV model is used. We also observe a statistically significant difference in the estimates from the HB and the HB-LIV model, conforming the endogenity of position in sponsored search.

Table 4: Estimates of Parameters

	HB		HB-LIV	
Parameters	Complete	Aggregate	Complete	Aggregate
$\boldsymbol{\mu_\beta}$:				
$\mu_{0\beta}$	-1.495(0.000)	-1.459(0.001)	-1.672(0.301)	-1.612(0.189)
$\mu_{1\beta}$	-0.793(0.085)	-0.727(0.098)	-0.642(0.120)	-0.558(0.074)
V_β:				
σ_1	0.654(0.128)	0.753(0.170)	0.678(0.107)	0.689(0.192)
σ_2	0.153(0.037)	0.155(0.038)	0.147(0.033)	0.162(0.042)
σ_{12}	0.025(0.067)	-0.085(0.076)	0.019(0.052)	-0.090(0.076)
V_ξ:				
σ_ϵ	0.263(0.067)	0.389(0.082)	0.142(0.047)	0.196(0.065)
σ_η			1.733(0.238)	2.113(0.414)
$\sigma_{\eta\epsilon}$			0.136(0.039)	0.176(0.052)
$\boldsymbol{\Theta}$:				
θ_1			0.962 (0.083)	0.748 (0.091)
θ_2			2.838 (0.029)	2.364 (0.037)
θ_3			4.552 (0.012)	5.927 (0.045)
$\boldsymbol{\Pi}$:				
π_1			0.523 (0.255)	0.322 (0.181)
π_2			0.238 (0.173)	0.193 (0.127)
Instrument Variable:				
pos_{t-1}			0.885(0.096)	0.766(0.127)

We use the differences between the estimates from summary and complete data for a random sample of 5000 exact-match keywords to compute the empirical distribution of the error ($\boldsymbol{\varepsilon} = \hat{\boldsymbol{\beta}}_{\boldsymbol{s}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$) due to aggregation (Figure 4). This empirical distribution is used in Section 6.2 to quantify the impact of aggregation bias on search engine and advertiser revenues.

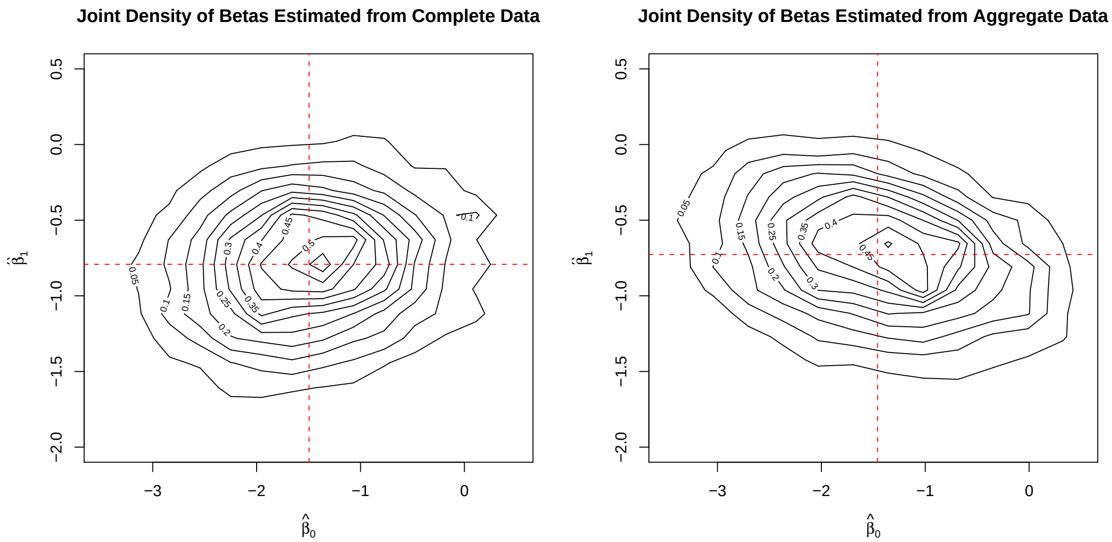

Figure 3: HB Estimation on complete and aggregate data. The red dotted lines represent the mean of the respective coefficients' distribution.

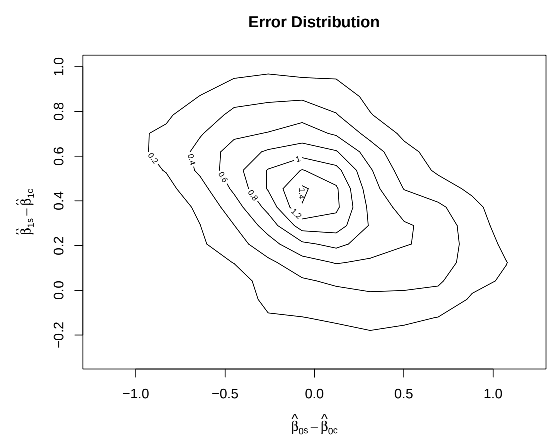

Figure 4: Error distribution estimated by computing the difference between the estimates from aggregate data and complete data.

2.4.5. Insights from Empirical Analysis

In general, we observe considerable bias in the estimates presented earlier although there is significant heterogeneity across keywords.[11] The bias is more significant in the case of broad-match ad compared to exact-match ads. Ads that have more competitors are prone to more aggregation bias than ads that have fewer competing advertisers. As the competition increases, there are more bidders in the keyword auction and frequent changes in bids by them might lead to higher intra-day variability in the ad position. Due to the increased intra-day variation, broad-match ads and ads with a large number of competitors are more prone to aggregation bias. We also observe that the bias increases as the effect of position becomes stronger. In summary, aggregation bias is significant when broad-match ads are used, there is significant competition or the estimated effect of position on CTR is large. Under such conditions, the model should explicitly account for the intra-day variation in position to accurately recover the parameters of the model. Advertisers can ignore the effects of aggregation when the impact of position on CTR is negligible.

2.5. Managerial Implications

In the previous section, we showed the existence of aggregation bias using the HB and HB-LIV models. In this section, we discuss the managerial implications of the bias. We start by quantifying the impact of aggregation bias on advertiser and search engine profits. Then, we discuss how aggregation bias might affect other managerial decisions like keyword selection.

2.5.1. Revenue implications

We now apply recent research by Ghose and Yang (2009) to quantify the impact of aggregation on advertiser and search-engine revenues. The analysis complements the theoretical investigation in Section 6.1. Given the coefficients estimated from aggregate and disaggregate data, Ghose and Yang's model helps identify an advertiser's optimal bids and

[11] M The regression is presented in Table 6 in the appendix

associated payoffs for the two cases. It does not however elaborate on the bids and payoffs of other advertisers. This precludes equilibrium analysis of all players in the market but allows us to examine whether and to what extent a unilateral switch to disaggregate data by an advertiser affects that advertiser's payoff and the search engine's revenues from that advertiser.

This analysis requires three main inputs. First, for any keyword we need to know the magnitude of the bias due to aggregation (i.e. coefficients $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$). Next, we need to compute the impact of this bias on the bids placed by the advertiser (i.e. how an advertiser's bids depend on $\hat{\boldsymbol{\beta}}$). Finally, we need to quantify how any change in the advertiser's bid affects the advertiser's and search engine's revenues. Our analysis in Section 5 provides the input for the first step. Specifically, Figure 4 identifies the distribution of the error $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\beta}}_{\boldsymbol{s}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$. Thus, given any $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$, we can compute a distribution of the true coefficient $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$. In order to compute how these coefficients translate into bid choices and eventually to advertiser/search-engine revenues, we rely on Ghose and Yang's model. They model the CTR for an ad as

$$\mathrm{CTR} = \frac{\exp\{\beta_0 + \beta_1 pos + \beta_2 Retailer + \beta_3 Brand + \beta_4 Length\}}{1 + \exp\{\beta_0 + \beta_1 pos + \beta_2 Retailer + \beta_3 Brand + \beta_4 Length\}},$$

where *Retailer*, *Brand* are dummies that equal 1 when the keyword associated with the ad contains retailer and brand information, respectively. *Length* measures the number of words in the keyword. The mean value of these variables in Ghose and Yang (2009) are as follows: $Retailer = 0.004$, $Brand = 0.427$ and $Length = 2.632$.[12] The position of the ad, pos is a function of the advertiser's bid; as the bid increases, the ad moves to a higher position. In their model, $pos = e^{1.98-1.93\mathrm{CPC}}$, where CPC is the advertiser's cost-per-click or bid. Suppose I is the number of ad impressions for a keyword and m is the expected revenue *per-click*, then the advertiser's profit is given by the number of clicks ($I \times CTR$)

[12] The CTR estimated from aggregate data $= \frac{\exp\{-1.65+1.29*0.08-0.23*0.43-0.11*2.63-0.31pos\}}{1+\exp\{-1.65+1.29*0.08-0.23*0.43-0.11*2.63-0.31pos\}} = \frac{\exp\{-1.96-0.31pos\}}{1+\exp\{-1.96-0.31pos\}}$

times the profit per click $(m - CPC)$. The advertiser chooses CPC to maximize his profits,

$$CPC^* = \underset{CPC}{\arg\max}\, I \times CTR \times (m - CPC), \tag{2.20}$$

It is easy to see the tradeoff between bidding high to get more clicks (CTR increases with CPC) and bidding low to earn greater profit per click.

The optimal bids can be computed by substituting the expression for CTR (as a function of advertiser's CPC) into the profit function. In order to compute the optimal bids, we use the mean estimates from Ghose and Yang (2009) ($(\beta_0, \beta_1) = (-1.96, -0.31)'$) and treat them as coefficients estimated from aggregate data ($\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$) for a keyword of interest.[13] Given the coefficients from the aggregate data, the coefficients from the complete data are given by $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{s}} - \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a random draw from the error distribution discussed earlier in Section 5.3.2. The error distribution is used to arrive at an empirical distribution of $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ conditional on $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$. We sample $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ from this distribution and, for each $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$, compute the optimal bid that maximizes advertiser profit in Equation (2.20). Figure 5 shows the distribution of optimal bids computed in this manner when the revenue *per-click*, m = \$1, and \$4. The optimal CPCs computed from aggregate data are shown as the dashed lines in Figure 5 and equal \$0.48 and \$1.29 when m=\$1 and \$4, respectively. The bids computed using $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ are less than almost all the bids computed using samples of $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$. This finding supports our earlier claim that aggregation bias results in the advertiser placing a bid lower than is optimal.

We next use the computed CPC to estimate the effect of aggregation bias on advertiser and search-engine revenues. Figure 6 shows the % loss suffered by the advertiser due to aggregation bias. The average loss for the representative ad is around 6% in this example. This loss is between 4-8% when the revenue *per-click* is less than \$2, which is quite typical for the ads considered in Ghose and Yang (2009). There is a great deal of variation in this

[13] The effect of factors like *Retailer* etc. is subsumed in the intercept term in accordance with the analysis in Section 5.2.

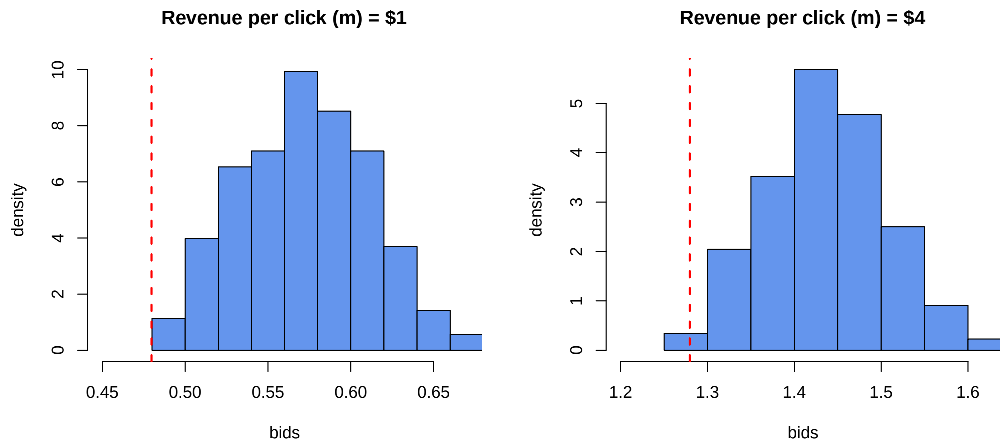

Figure 5: Optimal bids computed by maximizing Equation (2.20) for different estimates of β.

loss and in some cases the advertiser may lose more than 10-15% of his SSA profits. The impact of aggregation bias is more pronounced when the advertiser's valuation for the click is low, as in this situation, he bids lower and even small deviations from the optimal bid can lead to significant changes in the position leading to significantly lower payoffs. When the advertiser's valuation is high, he prefers to be ranked higher and his bid is correspondingly higher. Under this condition, deviations from the optimal bid are small and hence the profitability is not affected as much.

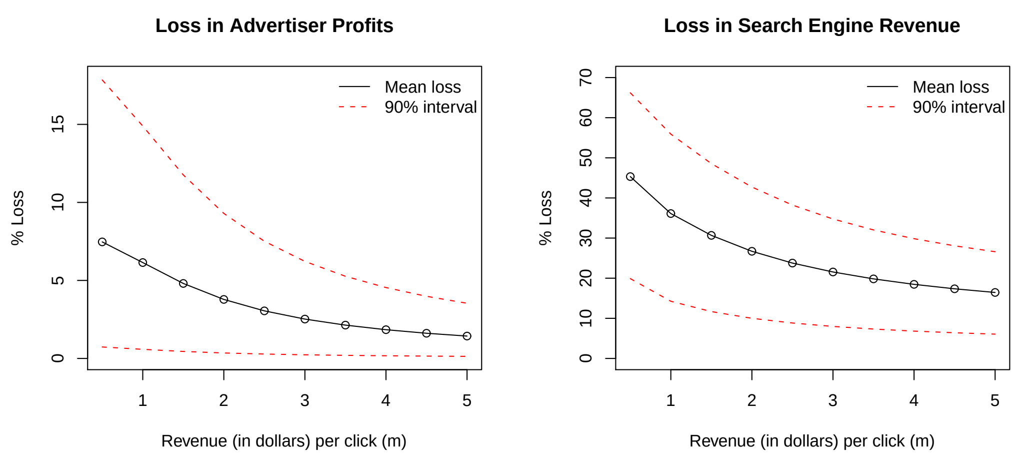

Figure 6: Loss in revenues caused due to Aggregation bias.

The effect of aggregation is considerably higher for the search engine. On average, it loses more that 20% of its payments from the advertiser as a result of aggregation bias. Lower bids, as a consequence of aggregation bias, negatively impact search-engine revenues in two ways – Firstly, lower bids imply that the search engine generates lesser revenue *per-click*. Secondly, the ad appears at a lower position due to the lower bid, which in turn leads to fewer clicks. The advertiser pays the search engine for fewer clicks and pays less for each click. In this example, we estimate that the search engine loses 1.4¢ for every impression of the representative ad. Given the large number of impressions for this advertiser, the search engine loses several thousand dollars in payment from the advertiser every week due to aggregation bias.[14]

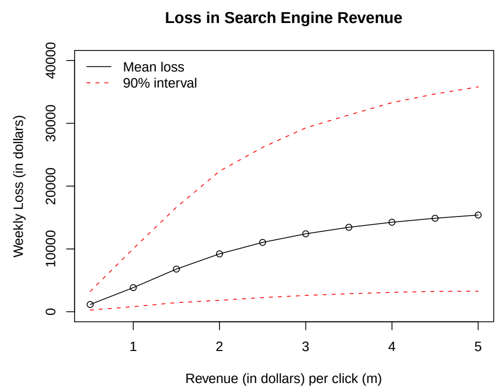

Figure 7: Loss in search-engine revenue caused due to Aggregation bias.

In the preceding analysis we considered how payments of *one advertiser* to the search engine are affected due to aggregation bias. Naturally, the dynamics are more complicated when we try to quantify the effect that the bias has on payments by other advertisers and the overall profitability of the search engine. For example, when an advertiser bids sub-optimally and moves to a lower position, another advertiser moves up to occupy the vacant position. Though the search engine loses revenue from the advertiser that moved down, it earns more from the advertiser that moved up, reducing the overall loss suffered by the search

[14]The advertiser considered in Ghose and Yang (2009) has a portfolio of 1878 unique ads and an ad on average has 411 impressions weekly. Total weekly loss $\approx$ 411 impressions $\times$1878 unique ad $\times$\$0.014 $=$ \$11,038 (compare to weekly revenues of \$19,377).

engine. It is difficult to accurately measure this effect as we do not have data from multiple advertisers, but we can estimate a lower bound for the loss suffered by the search engine. To estimate a lower bound, we assume that, even though the advertiser with aggregate data bids lower, his position does not change. As a result, the number of clicks on the ad do not change. The search engine loses revenues only because this advertiser chooses a lower CPC. The weekly loss suffered by the search engine in this case is 17% ($\sim \$3,800$/week or 0.43 ¢/impression) which is the lower bound for the overall loss experienced by the search engine.

2.5.2. Effect on Keyword Selection

Until now, we have mainly focused on how aggregation affects the estimation of the coefficient of position and subsequently, equilibrium payoffs. Aggregation bias also affects other managerial decisions like keyword selection. Given a limited budget, advertisers need to select a portfolio of keywords from several billion keywords available to them. Ghose and Yang (2009) and Rutz et al. (2012) propose several models that offer insights into keyword selection. Ghose and Yang (2009) suggest that ads associated with retailer specific keywords are more profitable than those associated with brand specific keyword. Rutz et al. (2012) provide a strategy to select optimal keywords based on factors like *wordographics*, keyword length, generic versus specific etc. It is important to note that aggregation bias also affects the estimates of these factors. As the simulations in Section 4.2 show, the impact of brand on click-through is, in fact, attenuated. Models that do not account for aggregation bias, might cause advertisers to make suboptimal tradeoffs in their keyword selection decisions.

2.6. Suggested Cures

In the previous discussion we outlined the problem of aggregation bias and some of its implications. As mentioned earlier, aggregation bias arises due to inadequate data. It might not be infeasible for search engines to store and report impression level data due to the size of such datasets and potential privacy concerns. However, the search engine can provide different data to reduce the effect of aggregation. In this section, we explore various summary statistics and measure the improvement they offer over the standard aggregate data provided by search engines. We also consider a few modeling approaches which explicitly account for the variation in position and estimate the improvement offered by them.[15] We continue to use the HB-LIV model for the subsequent analysis, but pos_t varies based on the summary statistics reported and the position model used. We first present the different datasets and then discuss the comparison between these datasets in Table 5.

Table 5: Comparative performance of various data summarization approaches.

Method	Data Requirement	MAPE(ε_0)	MAPE(ε_1)
Different Ways of Aggregation			
Arithmetic Mean	O(X)	53.4%	28.9%
Geometric Mean	O(X)	42.1%	17.5%
Harmonic Mean	O(X)	40.3%	10.2%
Poisson Model	O(X)	43.3%	22.8%
Mean and Variance	O(2X)	15.2%	7.8%
Empirical Distribution	O(2X)	9.8%	4.2%
Position-level Summary	O(NX)	0%	0%

2.6.1. Sample Mean

Different Ways of Aggregation

An important reason for bias is the non-linearity of the position-ctr curve. As a result, linear aggregation of the position does not yield the correct underlying response parameters. Christen et al. (1997) and Danaher et al. (2008) show that when the response is multiplicative, i.e. of the form $\alpha x_1^{\beta_1} x_2^{\beta_2} \ldots$, where x_i are marketing mix variables, an aggregate model should use geometric means to correctly estimate the coefficients. Unfortunately, there is no analytical analog of this result when the underlying model is logit. We use both the ge-

[15]We thank the reviewers for recommending this extension.

ometric and harmonic means and empirically compare which method of aggregation works better in sponsored search. As the position-ctr curve is convex in nature, both these aggregation methods perform better than linear aggregation, but the harmonic mean performs better than the geometric mean. This implies that if the search engine wants to provide only the mean position in the campaign reports, it should provide the harmonic mean of the position. This result also suggests that researchers who aggregate sponsored search data at a weekly or a monthly level for lack of sufficient data or computational reasons should use the harmonic mean for aggregation.

Modeling Position Variation using a Poisson

We also consider a model where V_t is drawn from a Poisson distribution with mean equal to the (arithmetic) mean u. The log-likelihood of observing the data in this case is given by

$$\begin{aligned} LL(\boldsymbol{\beta}|\text{data},\lambda) = & \sum_{k=1}^{K}\sum_{d=1}^{D} c_{dk} \log\left\{\sum_{i=0}^{\infty} P(V_{idk}=v)p_{idk}\right\} \\ & + (n_{dk}-c_{dk})\log\left\{1-\sum_{i=0}^{\infty} P(V_{idk}=v)p_{idk}\right\}, \end{aligned} \tag{2.21}$$

where λ is a $K \times D$ matrix, and every column of λ contains the scale parameter of the Poisson distribution for every day. The position of every impression V_{idk} for keyword k on the d^{th} day is drawn from a Poisson distribution with $\lambda_{kd} = \mu_{kd}$ and the probability $P(V_{idk} = v) = \lambda_{dk}^{v} e^{\lambda_{dk}}/v!$. As we see from Table 5, this approach does not perform very well because the ad position does not follow a Poisson distribution. We confirm this observation using the Neyman-Scott test.

2.6.2. Higher Order Statistics

If the search engine provided higher order moments in addition to the mean, the aggregation bias can be reduced significantly. We discuss three approaches with increasing data requirements.

Mean and Variance

In case the mean (μ_{dk}) and variance (σ_{dk}^2) of position are provided, we assume that the position has a negative binomial distribution (NBD). The NBD makes intuitive sense as the success of probability p of an NBD, can can be thought of as the probability of a competing advertiser placing a higher bid. The log-likelihood function is similar to Equation 2.21, but the distribution of ad position in this model is given by

$$P(V_{idk} = v) = \begin{pmatrix} v + r_{dk} - 1 \\ v \end{pmatrix} (1 - p_{dk})^{r_{dk}} (1 - p_{dk})^v,$$
$$\text{where } p_{dk} = \frac{\mu_{dk}}{\sigma_{dk}} \text{ and } r_{dk} = \frac{\mu_{dk}^2}{\sigma_{dk} - \mu_{dk}}.$$

We observe that using both the mean and the variance to model the variation in the position significantly improves the parameter estimates.

Empirical Distribution

If the search engine provides the empirical distribution of the ad position, the variation in position can be modeled non-parametrically. In this case, the log-likelihood is given by Equation 2.21, where $P(V_{idk} = v)$ is provided by the empirical distribution. This summarization technique performs better than all the preceding techniques.

2.6.3. Sufficient Statistics

Position-Level Summary

Although the complete dataset entirely eliminates aggregation bias, it might difficult for search engines due to privacy or technical concerns. However, the search engine can provide a position-level summary which are sufficient statistics for the logit model as we show below.

Equation 2.12 can be written as

$$LL(\boldsymbol{\beta}|\text{complete data}) = \sum_{k=1}^{K}\sum_{d=1}^{D}\sum_{v=1}^{\infty} c_{vdk}\log p_{vdk} + (n_{vdk} - c_{vdk})\log(1 - p_{vdk}), \tag{2.22}$$

where n_{vdk} and c_{vdk} are the number of impressions and clicks on the k^{th} ad at position v on the d^{th} day, respectively. Hence, the position-level summary are sufficient to correctly estimate the parameters of the model. It can observed from Table 5 that the model estimated from a position-level data does not suffer from aggregation bias.

2.6.4. Discussion

The preceding analysis presents different data summarization techniques in increasing order of the data requirement. Our analysis shows that if the search engine wishes to report only the mean position, it is best to provide the harmonic mean. The aggregation bias decreases significantly as more data is used for estimation. When both the mean and variance are used in the estimation process, not only is there a reduction in the aggregation bias but there is also a considerable decrease in the error in $\hat{\beta}_{0s}$. Using the empirical distribution marginally improves the estimation performance. When the position-level summary is used, the aggregation bias is completely eliminated. However, this technique requires substantially more data as compared to the other techniques. As we show in the preceding sections, it is in the search engine's interest to provide better data to advertisers. The appropriate dataset should be determined as a tradeoff between the loss due to aggregation and the costs associated with providing richer data to advertisers.

2.7. Conclusions

Search engine advertising is fast emerging as an important and popular medium of advertising for several firms. The medium offers rich data for advertisers on consumer click and conversion behavior. As a result, there has been considerable interest in analyzing SSA data among practitioners and researchers. Several models have been proposed to study consumer behavior and inform advertiser strategies.

This paper makes three main contributions. First, we demonstrate the existence of aggregation bias and its effect on the equilibrium of the SSA auction. We show that equilibrium bids are lower when advertisers use aggregate data. As a result the search engine's revenues are always lower due to the bias. Second, we use a large search engine dataset and quantify the magnitude of the bias and measure its economic impact. We also provide insights about the drivers of the bias and suggest conditions under which practitioners and researchers can't ignore this bias. Third, we present various summarization techniques that can be used by search engines to provide better datasets to advertisers.

These findings have important managerial and economic implications. Advertisers commonly use aggregate data provided by search engines to guide their bidding strategies. Our results suggest that advertisers might not be bidding optimally in these auctions because they overestimate the clicks obtainable at a given position. This not only impacts the advertisers negatively, but also leads to a reduction in the revenue of the advertiser. Given the size of the SSA industry, these losses can translate into several million dollars of lost revenues for the search engines. Our study points out that the current format of the data provided to advertisers is not adequate, and search engines should take steps to address this problem. We recognize that it might be infeasible for search engines to store and report impression level data due to the size of such datasets and potential privacy concerns. However, these constraints do not imply that it is infeasible to provide adequate data to advertisers. We provide guidelines to search engines about the nature of datasets that can be provided to researchers and quantify the reduction in the bias that each of these techniques can achieve.

We also find that, as a result of aggregation bias, consumer response to other ad attributes, such as ad text or branding, may also have been incorrectly estimated. Thus advertisers must be cautious in applying the biased estimates to guide key managerial decisions such as ad design and keyword selection. In the absence of adequate data from search engines, advertisers and researchers must take into account the variation in ad position within a day.

This can be determined by examining if multiple queries are matched to a single keyword (match type is broad) and if bids of competitors change considerably within a given day. If the ad position for a keyword is somewhat stable across impressions within a day, the bias is likely to be low and existing random utility models can be applied on aggregate data.

Our study focuses mainly on demonstrating the existence and direction of aggregation bias in the coefficient of position and identifying some economic consequences of this bias. An interesting and related question is how aggregation affects ad attributes like wordographics, presence of brand information, ad creative etc. and whether their coefficients also suffer from aggregation bias. In this paper, the effect of ad attributes is subsumed in the intercept term as we do not have data on ad attributes. A richer dataset that contains ad characteristics might help in a more extensive analysis of this issue. Another direction for future research is building models that endogenize the variation in position. The variation in position can be modeled using probabilistic models or structural methods. In ongoing work, we are developing a model that explicitly accounts for the intra-day variation in position and would therefore not suffer from aggregation bias.

SSA presents an exciting opportunity to understand consumer behavior and drivers of firms' advertising strategy. Through this paper we hope to inform the practitioners about the inadequacies of the data standards commonly used in SSA so that they can take steps to address these problems. We also identify issues with some common modeling techniques in SSA so that subsequent research in this emerging area is informed about these issues.

Appendix

Proofs of Equations and Propositions

Let Y_i denote an indicator variable that equals 1 if the i^{th} impression resulted in a click and zero otherwise. We assume that the clicks are independent of each other and hence Y_is are independent.The log likelihood of observing the dataset with a total of I ad impressions is given by

$$LL(\boldsymbol{\beta}|data) = \sum_{i=1}^{I} y_i \log p_i + (1 - y_i) \log(1 - p_i).$$

Proof of Equations

Proof of Equation (2.3)

The first order condition (F.O.C) for Equation (2.23) is as follows

$$\begin{aligned} \frac{\partial LL}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{I} \left\{ y_i(1 - p_i) - (1 - y_i)p_i \right\} \mathrm{x}_i' = 0, \\ &= \sum_{i=1}^{I} \left\{ y_i - p_i \right\} \mathrm{x}_i' = 0. \end{aligned}$$

where $\mathrm{x}_i = (1 \ v_i)'$. Since we know that $LL(\beta|data)$ is a convex function in β (Hayashi, 2000) this $F.O.C$ gives us the following two equations

$$C = \sum_{i=1}^{I} p_i, \tag{2.23}$$

$$\sum_{i=1}^{I} y_i v_i = \sum_{i=1}^{I} v_i p_i. \tag{2.24}$$

Dividing Equation (2.23) by I we get

$$obsctr = \frac{C}{I} = \frac{1}{I}\sum_{i=1}^{I} p_i. \tag{2.25}$$

Proof of Equation (2.7)

Let the number of impressions on day d be denoted by n_d and the number of clicks by c_d; $\sum_{d=1}^{D} n_d = I$ and $\sum_{d=1}^{D} c_d = C$. The probability of clicking on the ad on day d is given by

$$p_d = \frac{\exp\{\beta_0 + \beta_1 w_d\}}{1 + \exp\{\beta_0 + \beta_1 w_d\}}. \tag{2.26}$$

The log-likelihood of observing the aggregate data for D days is given by

$$LL(\beta|data) = \sum_{d=1}^{D} c_d \log p_d + (n_d - c_d)\log(1 - p_d) \tag{2.27}$$

Evaluating the first order condition for Equation (2.27)

$$\begin{aligned}
\frac{\partial LL}{\partial \boldsymbol{\beta}} &= \sum_{d=1}^{D} \{c_d(1 - p_d) - (n_d - c_d)p_d\}\, \overline{\mathrm{x}}'_d = 0, \\
&= \sum_{d=1}^{D} \{c_d - n_d p_d\}\, \overline{\mathrm{x}}'_d = 0,
\end{aligned}$$

where $\overline{\mathrm{x}}_d = (1\ w_d)'$, which in turn gives us

$$\sum_{d=1}^{D} c_d = C = \sum_{d=1}^{D} n_d p_d, \tag{2.28}$$

$$\sum_{d=1}^{D} w_d c_d = \sum_{d=1}^{D} n_d w_d p_d. \tag{2.29}$$

Dividing Equation (2.28) by I we get

$$obsctr = \frac{C}{I} = \sum_{d=1}^{D} \frac{n_d p_d}{I}. \tag{2.30}$$

Proof of Lemma 2

We use the following result from Muller and Stoyan (2002, page 27) to prove this result. Let $V_1, ..., V_n$ be iid random variables and $f_1, ..., f_n$ measurable real functions. Define the function $\bar{f}$ by

$$\bar{f}(v) = \frac{1}{n} \sum_{i=1}^{n} f_i(v).$$

$$\text{Then,} \qquad \sum_{i=1}^{n} \bar{f}(V_i) \leq_{\text{cx}} \sum_{i=1}^{n} f_i(V_i).$$

Using this result we now prove that $W_n \leq_{\text{cx}} V$, where W_n is the average position when there are exactly n impressions on the day. Let $f_i(v) = v/(n-1)$ for all $i = 1, \ldots, n-1$ and $f_n(v) \equiv 0$.

$$\bar{f}(v) = \frac{1}{n} \sum_{i=1}^{n-1} v/(n-1) = \frac{v}{n}.$$

$$\text{Since } \sum_{i=1}^{n} \bar{f}(V_i) = \frac{1}{n} \sum_{i=1}^{n} V_i \quad \text{and} \quad \sum_{i=1}^{n} f_i(V_i) = \frac{1}{n-1} \sum_{i=1}^{n-1} V_i \Rightarrow \frac{1}{n} \sum_{i=1}^{n} V_i \leq_{\text{cx}} \frac{1}{n-1} \sum_{i=1}^{n-1} V_i.$$

Proceeding in a recursive manner

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^{n-1} V_i &\leq_{\text{cx}} \frac{1}{n-2} \sum_{i=1}^{n-2} V_i, \\ &\cdots \\ \frac{V_1 + V_2}{2} &\leq_{\text{cx}} V. \end{aligned}$$

$$\Rightarrow W_n = \frac{1}{n} \sum_{i=1}^{n} V_i \leq_{\text{cx}} \frac{1}{n-1} \sum_{i=1}^{n-1} V_i \leq_{\text{cx}} \ldots \leq_{\text{cx}} \frac{V_1 + V_2}{2} \leq_{\text{cx}} V \qquad \blacksquare$$

Let g be any convex function

$$\begin{aligned}
\text{As } \mathbb{E}[g(W)] &= \sum_{i=1}^{\infty} g(W_n)P(n), \Rightarrow \mathbb{E}\left[\mathbb{E}[g(W)]\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} g(W_n)P(n)\right], \\
\Rightarrow \mathbb{E}\left[\mathbb{E}[g(W)]\right] &= \sum_{i=1}^{\infty} \mathbb{E}\left[g(W_n)\right] P(n) \leq \sum_{i=1}^{\infty} \mathbb{E}\left[g(V)\right] P(n) = \mathbb{E}\left[g(V)\right] \sum_{i=1}^{\infty} P(n) = \mathbb{E}\left[g(V)\right],
\end{aligned}$$

as $P(n)$ is a probability measure. Therefore $W \leq_{\text{cx}} V$. $\blacksquare$

Proof of Proposition 1

Assuming I is large we can apply Chevychev's law of large numbers to rewrite Equation (2.3) as

$$obsctr = \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}{1+e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}\right]. \tag{2.31}$$

If we have a large enough observation period, Equation (2.7) can be simplified as

$$obsctr = \sum_{d=1}^{D} \frac{n_d e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}w_n}}{I(1+e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}w_n})} = \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}{1+e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}\right].$$

As the observed *ctr*, *obsctr* is same in both the cases,

$$\mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}{1+e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}\right] = \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}{1+e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}\right]. \tag{2.32}$$

As the convex ordering in Lemma 2 holds and logit is a convex in position for $\beta_0 < 0$ (which is a reasonable assumption in SSA as the CTR on the topmost position is less that 0.2 in all cases), it follows from the definition of convex ordering that

$$\mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}{1+e^{\hat{\beta}_{0,\text{c}}+\hat{\beta}_{1,\text{c}}V}}\right] \geq \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}{1+e^{\hat{\beta}_{0,\text{s}}+\hat{\beta}_{1,\text{s}}W}}\right], \tag{2.33}$$

if $\hat{\boldsymbol{\beta}}_{\mathbf{c}} = \hat{\boldsymbol{\beta}}_{\mathbf{s}}$. The equality holds only when $F_V(.) = F_W(.)$ i.e. which hold only under few special cases (e.g. when there is exactly one impression every day or there is no intra-day variation in position). Since Equation (2.32) and Equation (2.33) cannot simultaneously be true and Equation (2.32) always holds we prove by contradiction that $\hat{\boldsymbol{\beta}}_{\mathbf{c}} \neq \hat{\boldsymbol{\beta}}_{\mathbf{s}}$. $\blacksquare$

Proof of Proposition 2

i) Since we know that

$$\mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V}}{1+e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V}}\right] = \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W}}{1+e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W}}\right],$$

and by definition of convex order

$$\mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V}}{1+e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}V}}\right] \geq \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}W}}{1+e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}W}}\right],$$

we can say that

$$\mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W}}{1+e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}W}}\right] \geq \mathbb{E}\left[\frac{e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}W}}{1+e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}W}}\right].$$

As this result hold for any distribution of W, this relation should hold pointwise for the two function.

$$\Rightarrow \frac{e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}x}}{1+e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}x}} \geq \frac{e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}x}}{1+e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}x}} \quad \blacksquare$$

ii) The preceding relationship implies that

$$\begin{aligned} e^{\hat{\beta}_{0,\mathrm{s}}+\hat{\beta}_{1,\mathrm{s}}x} &\geq e^{\hat{\beta}_{0,\mathrm{c}}+\hat{\beta}_{1,\mathrm{c}}x} \quad \forall x \geq 0 \\ \Rightarrow \hat{\beta}_{1,\mathrm{s}} &> \hat{\beta}_{1,\mathrm{c}} \quad \blacksquare \end{aligned}$$

Proof of Proposition 3

(i) We assume that advertiser j moves to position $j^{'}$ if he uses aggregate data. Since $j^{'}$ ends up higher than j in equilibrium, $b_j^{'}(s_j, j^{'}, h) < b_{j'}(s_{j'}, j', h)$ or

$$\frac{1}{\alpha_{j'-1}}((\alpha_{j'-1} - \alpha_{j'})s_{j'} + \alpha_{j'}b_{j'+1}) > \frac{1}{\alpha^{'}_{j'-1}}((\alpha^{'}_{j'-1} - \alpha^{'}_{j'})s_j + \alpha^{'}_{j'}b_{j'+1}), \tag{2.34}$$

which also implies that his bid in the equilibrium decreases, i.e. $b_j^{'} < b_{j'}$. In addition, as $b_j^{'}$ is lower than b_j, $j^{'} \geq j$. The bids for all the advertisers in this case are as follows:

$$\begin{aligned}
b_i^{'} &= \frac{1}{\alpha_{i-1}}\left(\sum_{k=i}^{K}(\alpha_k - \alpha_{k+1})s_k\right) \quad \text{for } i > j^{'}, \\
b_j^{'} &= \frac{1}{\alpha^{'}_{j'-1}}\left((\alpha^{'}_{j'-1} - \alpha^{'}_{j'})s_j + \frac{\alpha^{'}_{j'}}{\alpha_{j'}}\sum_{k=i}^{j^{'}+1}(\alpha_k - \alpha_{k+1})s_k\right), \\
b_i^{'} &= \frac{1}{\alpha^{'}_{i-2}}\left(\sum_{k=i}^{j^{'}}(\alpha_{i-2} - \alpha_{i-1})s_i + \alpha_{j'+1}b_j^{'}\right) \quad \text{for } j^{'} \geq i > j, \\
b_i^{'} &= \frac{1}{\alpha^{'}_{i-1}}\left(\sum_{k=i}^{j-1}(\alpha_{i-1} - \alpha_i)s_i + \alpha_{j-1}b^{'}_{j+1}\right) \quad \text{for } i < j.
\end{aligned}$$

As h do not change for advertisers below $j^{'}$, therefore their bids remain the same. It is easy to see that advertisers $j+1$ to $j^{'}$ end up bidding higher, i.e. $b_i^{'} > b_i$ for $j < i \leq j^{'}$ though they moves up by 1 position.

We now show that the bid associated with every position $\geq j^{'}$ is lower than when complete data is used. Lets consider the bid $b^{'}_{j'}$, placed by advertiser $j^{'}$ who occupies position $j^{'} - 1$. We start off by showing that $b^{'}_{j'} < b_{j'-1}$.

$$b^{'}_{j'} - b_{j'-1} = \frac{1}{\alpha_{j'-2}}((\alpha_{j'-2} - \alpha_{j'-1})\underbrace{(s_{j'} - s_{j'-1})}_{<0 \text{ by construction}} + \alpha_{j'-1}\underbrace{(b_j^{'} - b_{j'})}_{<0 \text{ by assumption}}) < 0.$$

So the bid for position $j^{'} - 1$ is lower than the bid in the complete case. Proceeding in a similar manner it is easy to show that bids for all positions above $j^{'} - 1$ will also be lower. This implies that $b_i^{'} < b_i$ for $i < j$ (these ads do not change position). To summarize – $b_j^{'} < b_j$, $b_i^{'} < b_i$ for $i < j$, $b_i^{'} > b_i$ for $j < i \geq j^{'}$ and $b_i^{'} = b_i$ for $i > j^{'}$.

(ii) As all the bids are either the same or lower in this case, search-engine revenue is lower ($\Pi_S^{A2} < \Pi_S^C$). The payoff of advertiser j is lower as any deviation from the optimal bidding policy results in a strictly lower payoff. Advertisers $j^{'} + 1$ onwards receive the same payoff and all other advertisers are better off due to suboptimal bid by advertiser j. Starting off with advertiser $j^{'}$,

$$\begin{aligned}
\Pi_{j^{'}}^{A2} - \Pi_{j^{'}}^{C} &= (\alpha_{j^{'}-1} - \alpha_{j^{'}})s_{j^{'}} - \alpha_{j^{'}-1}b_j^{'} + \alpha_{j^{'}}b_{j^{'}+1}^{'} \\
&= (\alpha_{j^{'}-1} - \alpha_{j^{'}})s_{j^{'}} + \alpha_{j^{'}}b_{j^{'}+1} - \frac{\alpha_{j^{'}-1}}{\alpha_{j^{'}-1}^{'}}((\alpha_{j^{'}-1}^{'} - \alpha_{j^{'}}^{'})s_j + \alpha_{j^{'}}^{'}b_{j^{'}+1}) \\
&> 0 \quad \text{(by Equation 2.34)}
\end{aligned}$$

Similarly, using induction we can show that $\Pi_i^{A2} > \Pi_i^C$ for advertiser i, s.t. $j < i \leq j^{'}$. For $i < j$, the revenues remain the same but the payment to the search engines is lower, hence their payoff are higher for these advertisers too ($\Pi_i^{A2} > \Pi_i^C$, i ¡ j).

Proof of Proposition 4

(i) If all advertisers use the same incorrect estimate of α_i, the optimal bidding policy is the one proposed by Edelman et al. (2007). They just use $\alpha_i^{'}$ instead of α_i to compute the optimal bids. We can show by induction that $b_j^{'} \leq b_j$ for all advertisers.

Step 0: Let $b_{K+1}^{'} = b_{K+1} = 0$.

Step 1: $b_K^{'} = s_K\left(1 - \frac{\alpha_K^{'}}{\alpha_{K-1}^{'}}\right) < s_K\left(1 - \frac{\alpha_K}{\alpha_{K-1}}\right) = b_K$.

Assuming $b_{j+1}^{'} < b_{j+1}$,

Step j:

$$\begin{aligned} b_j' &= s_j\left(1-\frac{\alpha_j'}{\alpha_{j-1}'}\right)+\frac{\alpha_j'}{\alpha_{j-1}'}b_{j+1}' \\ &< s_j\left(1-\frac{\alpha_j'}{\alpha_{j-1}'}\right)+\frac{\alpha_j'}{\alpha_{j-1}'}b_{j+1} \\ &< s_j\left(1-\frac{\alpha_j}{\alpha_{j-1}}\right)+\frac{\alpha_j}{\alpha_{j-1}}b_{j+1} \\ &< b_j. \end{aligned}$$

Hence, $b_j' < b_j \ \forall j \leq K$.

(ii) Since all advertisers occupy the same position as they did earlier and pay less, search engine profits are lower ($\Pi_S^{A1} < \Pi_S^C$). The advertisers' payoff in the case are higher: $\Pi_i^{A1} = \alpha_i(s_i - b_{i+1}') > \alpha_i(s_i - b_{i+1}) = \Pi_i^C$ from Proposition 5 (i).

Proof of Proposition 6

Lets assume that Advertiser j uses aggregate data and appears at a position j'. Let the equilibrium bids be denoted by $b_1', \ldots, b_K', 0$. In the equilibrium, $\alpha_{j'}(s_j - b_{j'}) > \alpha_i(s_j - b_{i+1})$ $\forall\ i \neq j'$. Now suppose that Advertiser j does not have access to aggregate data and overestimates α_i/α_{i-1}. As a result, he bids lower and moves to position $j'' \geq j'$. Following the argument in the Proof of Proposition 4, the bids for all positions i, $i \leq j''$ decrease and all *other* advertisers are better off. As the bids are (weakly) lower, the search-engine revenue is lower. The payoff to Advertiser j is $\alpha_{j''}(s_j - b_{j''}) < \alpha_{j'}(s_j - b_{j'})$ as he found it optimal to bid for position j' when he could correctly estimate the CTR. This implies that he is worse off using aggregate data.

Hierarchical Bayesian Estimation

Complete Data

The consumer's utility from clicking on an ad can be modeled as

$$U_{i,k} = \beta_{0,k} + \beta_{1,k} pos_{i,t} + \epsilon_{i,k} \tag{2.35}$$

where $\epsilon_{i,k}$ is the idiosyncratic error term. $\boldsymbol{\beta_k} \sim N_2(\boldsymbol{\mu_\beta}, V_\beta)$, with the hyper-priors $\boldsymbol{\mu_\beta} \sim N(\boldsymbol{\mu}, \Sigma)$ and $V_\beta^{-1} \sim Wishart(\nu, \Delta)$. This model can be estimated on the complete dataset using MCMC in the following way.

Step 1: Draw $\boldsymbol{\beta}|\mathbf{y}, \mathrm{X}, \boldsymbol{\mu_\beta}, V_\beta, \boldsymbol{\mu}, \Sigma, \nu, \Delta$

The posterior density of $\boldsymbol{\beta_k}$ can be expressed as

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathrm{X}, \boldsymbol{\mu_\beta}, V_\beta, \boldsymbol{\mu}, \Sigma) \propto l(\mathbf{y}|\boldsymbol{\beta}, \mathrm{X}, \boldsymbol{\mu_\beta}, V_\beta, \boldsymbol{\mu}, \Sigma) p(\boldsymbol{\beta}|\boldsymbol{\mu}, V_\beta, \boldsymbol{\mu}, \Sigma)$$

where

$$l(\mathbf{y}|\boldsymbol{\beta}, X, \boldsymbol{\mu_\beta}, V_\beta) = \prod_k \prod_i p_i^{y_i}(1-p_i)^{1-y_i}$$

and

$$p(\boldsymbol{\beta}|\boldsymbol{\mu_\beta}, V_\beta) \propto \prod_k |V_\beta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta_k} - \boldsymbol{\mu_\beta})' V_\beta^{-1} (\boldsymbol{\beta_k} - \boldsymbol{\mu_\beta})\right\}$$

is the prior density. A random walk Metropolis-Hastings algorithm is used for sampling β_k.

Step 2: Draw $\boldsymbol{\mu_\beta}|\mathbf{y}, \mathrm{X}, \boldsymbol{\beta}, V_\beta, \boldsymbol{\mu}, \Sigma, \nu, \Delta$ using Gibbs sampling

$$\begin{aligned} \boldsymbol{\mu_\beta}|\boldsymbol{\beta}, \Sigma &\sim N_2(\tilde{\mu}, \tilde{\Sigma}) \\ \text{where} \quad \tilde{\Sigma} &= \left[K V_\beta^{-1} + \Sigma^{-1}\right]^{-1} \\ \tilde{\boldsymbol{\mu}} &= \tilde{\Sigma}\left[K V_\beta^{-1} \bar{\beta} + \Sigma^{-1} \boldsymbol{\mu}\right] \end{aligned}$$

Step 3: Draw $V_\beta|\mathbf{y}, \mathrm{X}, \beta, \mu_\beta, \mu, \Sigma, \nu, \Delta$ using Gibbs sampling

$$V_\beta^{-1} \sim Wishart\left(\nu + K, \left[\Delta + \sum_{k=1}^{K} (\boldsymbol{\beta_k} - \boldsymbol{\mu_b})'(\boldsymbol{\beta_k} - \boldsymbol{\mu_b})\right]\right)$$

Initial values:

$$\begin{aligned}
\boldsymbol{\mu} &= 0 \\
\Sigma &= 100I \\
\nu &= \text{number of } \beta \text{ coefficients } + 3 \\
&= 5 \\
\Delta &= \nu I \\
\boldsymbol{\beta} &= \text{MLE Estimates from the data}
\end{aligned}$$

Aggregate Data

The estimation on the aggregate data is performed in a similar way. The consumer's utility from clicking on an ad on day d can be modeled as

$$U_{d,k} = \beta_{0,k} + \beta_{1,k} meanpos_{d,t} + \epsilon_{d,k} \tag{2.36}$$

The MCMC steps are similar except the likelihood function in Step 1 which changes to

$$l(\mathbf{y}|\boldsymbol{\beta}, \mathrm{X}, \boldsymbol{\mu_\beta}, V_\beta) = \prod_k \prod_d \begin{pmatrix} n_d \\ c_d \end{pmatrix} p_d^{c_d}(1 - p_d)^{n_d - c_d}$$

Predictors of Aggregation Bias

Table 6: Predictors of Bias

DV	$\lvert\hat{\beta}_{0,s} - \hat{\beta}_{0,c}\rvert$	$\lvert\hat{\beta}_{1,s} - \hat{\beta}_{1,c}\rvert$
(Intercept)	-0.166	-0.119^{*}
	(0.153)	(0.067)
β_0	-0.046	-0.044^{***}
	(0.030)	0.016
β_1	-0.369^{***}	
	(0.066)	
β_1^2		0.294^{**}
		(0.023)
$BROAD$	-0.002	0.048^{***}
	(0.047)	(0.021)
$\overline{U}$	0.221^{***}	0.030^{**}
	(0.023)	(0.012)
σ_U	-0.551^{***}	-0.062^{***}
	(0.072)	(0.015)
σ_U^2	0.061^{***}	
	(0.009)	
$\overline{ctr}$	-0.594	0.589^{***}
	(0.597)	(0.217)
$\overline{ctr} * \sigma_U$	1.414^{***}	
	(0.482)	
Adj. R^2	0.3815	0.4502

CHAPTER 3 : Optimal Bidding in Multi-Item Multi-Slot Sponsored Search Auctions

3.1. Introduction

With the growing popularity of search engines among consumers, advertising on search engines has also grown considerably. Search engine advertising or sponsored search has several unique characteristics in contrast to traditional advertising and other forms of online advertising. Compared to traditional advertising in print/television, sponsored search is highly measurable allowing advertisers to identify which keywords are generating clicks and which clicks are getting converted to purchases. Compared to other forms of online advertising such as banner ads, search advertising enjoys much higher click-through (CTR) and conversion rates. Search queries entered by users convey significant information about users current need and context which allow search engines to better target ads to users than is possible in other forms of online advertising. Further, search engine users, unlike users on another websites, primarily use the search engine to reach some other website. Advertising is an effective way to enable that process.

Search engines commonly use Pay Per Click (PPC) auctions to sell their available inventory of ad positions for any search query. The auction mechanism is referred to as the Generalized Second Price (GSP) auction. In these auctions, advertisers select keywords of interest, create brief text ads for the keywords and submit a bid for each keyword which indicates their willingness to pay for every click. For example, a meat seller may submit the following set of two tuples {(pork chop, $2), (fillet mignon, $5), (steak deals, $3),...} where the first element in any two-tuple is the keyword and the second element is the advertiser's bid. Large advertisers typically bid on hundreds of thousands of keywords at any instant. When a user types a query, the search engine identifies all advertisers bidding on that (or a closely related) keyword and displays their ads in an ordered list. The search engine uses the advertisers' bids along with measures of ad relevance to rank order the submitted ads. Whenever a consumer clicks on an ad in a given position, the search engine charges the corresponding advertiser a cost per click (CPC) which is the minimum bid needed to secure that position. The auctions are continuous sealed bid auctions. That is, advertisers can change their bids at any time and cannot observe the bids of their competitors. Typically advertisers are only given summary reports with details such as the total number of impressions, clicks and conversions, average rank and average CPC for each keyword on a given day. Several of these auctions are very competitive. For example, it is not uncommon to have 100 or more advertisers bidding for the same keyword. The average CPC on search engines has been continually rising over the last couple of years and search advertising is increasingly becoming a major advertising channel for a large number of firms.

The GSP auction described above differs from traditional auctions in a number of ways. First, search engines display multiple ads in response to a user query. However, the auction cannot be treated as a multi-unit auction because each ad position is different in the sense that top positions generate more clicks for the same number of ad impressions. Further, the CPC decreases as the rank of an ad increases (i.e. the CPC is higher for top ranked ad than a lower ranked ad). Thus, the advertiser has to trade-off a higher number of clicks attained at a top position against the lower margin per click. Due to this trade-off, it may sometimes be better for an advertiser to underbid and sacrifice a few clicks in order to get a higher margin per click. Indeed, several authors have demonstrated that popular second-price search auctions such as those used by Google and Yahoo are not incentive compatible (Aggarwal et al., 2006; Edelman et al., 2007). Thus, bidding one's true valuation is often suboptimal. Further, advertisers have short-term budget constraints which imply that bids cannot be submitted independently for keywords. For example, if the advertiser submits a very high bid for the keyword "fillet mignon" then it may leave a very limited portion of the budget for another keyword. The performance of the keywords may also be interdependent, wherein clicks for one keyword may help generate more searches and clicks for another. Therefore the bids for the thousands of keywords are inextricably linked. Finally, considerable uncertainty exists in the sponsored search environment. For example, the number of queries for "fillet mignon" on any given day is stochastic and is a function of the weather, special events and a variety of other unknown factors. Similarly, consumer click behavior cannot be precisely predicted and the bids of competitors are also unknown due to the sealed bid nature of the auction. The stochasticity in query arrival, consumer click behavior and competitors' bids imply that the number of clicks and total cost associated with any bid are all stochastic. All these factors - namely the incentive incompatibility of the auction, budget constraints, large portfolio of keywords with interdependent performance and uncertainty in the decision environment - make the advertiser's problem of bidding in sponsored search a non-trivial optimization problem. In this paper, we formulate and solve the advertiser's decision problem.

We propose two bidding policies in our paper. The first policy ignores the interaction between keywords and is referred to as the "myopic" policy in this paper. We extend this bidding policy to incorporate interaction between keywords, and refer to this policy as the "forward-looking" policy since it entails decision making over several time horizons. Depending on the advertiser's intent, level of sophistication and nature of the products being advertised, the advertiser might choose the myopic or the forward-looking policy. This paper makes three main contributions. The first contribution is towards improving managerial practice. Advertisers spend billions of dollars on sponsored search. An entire industry of Search Engine Marketing (SEM) firms have emerged that provide bid management services. The techniques described in the paper can help increase the Return on Investment (RoI) for advertisers and SEM firms, as demonstrated in our field implementation. The second key contribution is that our approach represents a significant step forward for the academic literature on bidding in multi-slot auctions. All the papers to date have studied the problem either in a deterministic setting or in a single-slot setting and have relied on heuristic solution techniques due to the complexity of the optimization problem. In contrast, we compute optimal bids in the more realistic stochastic multi-slot setting. The third contribution of this paper is that it is the first paper on bidding in sponsored search to incorporate the interdependence between keywords into a multi-period bidding problem. The interdependence in keyword performance, commonly referred to as spillovers, is a well-documented feature of sponsored search (Rutz et al., 2012) but has not been considered in the bidding literature.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature and position our work within the literature on sponsored search. In Section 3, we formulate the problem, derive the optimality condition for the myopic policy and discuss how it may be used to compute the optimal bids. In Section 4, we describe the dataset used for the analysis presented in this paper. In Section 5, we present the empirical analysis where we estimate the parameters of our model and run a field experiment with the bids suggested by the myopic policy. We compare the optimal bids computed by our model with those used by the firm and present results from a field implementation of the bids. We extend the myopic policy in Section 6 to incorporate interdependence between keywords. Finally, we discuss some limitations of our work in Section 7 and conclude in Section 8.

3.2. Literature review

In this section, we review three streams of active research within the field of sponsored search with a particular emphasis on prior work on bidding in sponsored search.

Mechanism Design: Search engines run PPC auctions in which they charge advertisers whenever a consumer clicks on an ad.[1] A primary area of focus in sponsored search research has been the design of the auction mechanism. Two important questions from a mechanism design perspective are the rules used to rank order the ads and the rules used to determine the amount paid by advertisers. Feng et al. (2006) compare the performance of various ad ranking mechanisms and find that a yield-optimized auction, that ranks ads based on the product of the submitted bid and ad relevance, provides the highest revenue to the search engine. In terms of payment rules, Edelman and Ostrovsky (2007) study first price

[1]Other payment rules are also feasible. These include Pay Per Action (PPA) auctions in which advertisers are charged only if the consumer performs a valid action such as a purchase. Hybrid schemes are also feasible. For example, in the context of banner ads, Kumar et al. (2007) propose a hybrid pricing model based on a combination of ad impressions and clicks.

sponsored search auctions in which advertisers pay the amount they bid and find empirical evidence of bidding cycles in such auctions. The authors indicate that a VCG-based mechanism eliminates such bidding cycles and generates higher revenues for the search engine compared to the first price auction. In a related paper, Edelman et al. (2007) demonstrate that the commonly used GSP auction, unlike Vickrey-Clarke-Groves (VCG) mechanism, is not incentive compatible. Thus, advertisers have to bid strategically even in the absence of budget constraints. Aggarwal et al. (2006) propose a "laddered" auction mechanism that is incentive compatible but the mechanism has not been adopted possibly due to the complexity of the payment rules. Mehta et al. (2007) solve the problem of matching ad slots to advertisers using a generalization of the online bipartite matching problem. Given advertisers' bids and budget constraints, Mehta et al. (2007) provide a deterministic algorithm that achieves a competitive ratio of $1 - 1/e$ for this problem.[2] Mahdian et al. (2007) extend this work and provide a solution which is nearly optimal when the frequencies of keywords are accurately known and provides a good competitive ratio even when these estimates are completely inaccurate. Aggarwal and Hartline (2006), on the other hand, model this problem as a knapsack auction. However, they consider only truthful mechanism designs and analyze various pricing schemes and the payoffs under each of these pricing schemes. Most of the above referenced papers focus on the search engine's problem and analyze how different mechanisms affect the search engine's revenues.

Consumer behavior in sponsored search: The sponsored search environment presents rich data on consumer behavior. Modeling user's propensity to click on ads and to purchase upon clicking is an important area of recent focus. Several approaches have been proposed to model clicks for individual keywords and ads (Ali and Scarr, 2007; Craswell et al., 2008; Feng et al., 2006). Ali and Scarr (2007) compare several distributions to predict *click-through* rates and suggest that Pareto-Zipf distribution is the most appropriate for explaining CTR as a function of position. Feng et al. (2006) alternately assume an exponential decay in CTR with position and demonstrate that the model fits observed data

[2] Competitive ratio is the measure for comparing online algorithms to offline algorithms where all the information is known apriori.

well. Several other papers build richer models of consumer behavior incorporating the effect that ad attributes have on *click-through* and conversion (Ghose and Yang, 2009; Yang and Ghose, 2010; Rutz et al., 2012; Agarwal et al., 2011). Rutz et al. (2012) propose a model that measures the interaction between keywords and show that there are significant positive spillovers from generic keywords to branded keywords in consumer search.

Optimal Bidding in Sponsored Search: The stream of work closely related to our paper is that on budget constrained bidding in sponsored search. Rusmevichientong and Williamson (2006) propose a model for selecting keywords from a large pool of candidates. Their model does not however address optimal bidding for these keywords and ignores the multi-slot context. Feldman et al. (2007) study the bid optimization problem and indicate that randomizing between two uniform strategies that bid equally on all keywords works well. The authors assume that all clicks have the same value independent of the keyword. Further, their results are derived in a deterministic setting where the advertisers position, clicks and the cost associated with a bid are known precisely. Borgs et al. (2007) propose a bidding heuristic that sets the same average Return on Investment (RoI) across all keywords. Their model is also derived for a deterministic setting. Finally, Muthukrishnan et al. (2007) study bidding in a stochastic environment where there is uncertainty in the number of queries for any keyword. The authors focus on a single slot auction and find that prefix bidding strategies that bid on the cheapest keywords work well in many cases. However, they find that the strategies for single slot auctions do not always extend to multi-slot auctions and that many cases are NP hard .

The prior work reveals three themes. The first is that the literature on sponsored search mechanism design has established that GSP auctions are not incentive compatible. This feature combined with the advertiser's budget constraint suggests a need to develop bidding policies. Secondly, the empirical work in sponsored search provides a variety of useful models that can be applied towards modeling consumer click behavior and the bidding behavior of advertisers. These can ultimately be used to develop data-driven optimization strategies.

Three, the issue of budget constrained bidding has received some attention. While these early papers on bid optimization have helped advance the literature, they tend to focus on deterministic settings or single slot auctions, both of which are restrictive assumptions in the sponsored search context. None of these papers account for any interdependence in keyword performance. Further, these papers develop heuristic strategies due to the complexity of the optimization problem. In contrast, we determine optimal bids in a budget-constrained multi-unit multi-slot auction under uncertainty in the decision-making environment. We also extend our basic model to incorporate interaction between keywords.

3.3. Analytical Model

Advertisers usually maintain a portfolio of thousands of keywords for a particular search engine. They submit bids for each keyword on a regular basis during a billing cycle. During each time period when bids need to be computed, the bid management system accepts a budget for that time period as an input and computes the bids for all keywords. We adopt the same framework and focus on the bid optimization problem during a specific time period in which the budget and the set of keywords have been specified.[3] Although we consider the advertiser's problem of optimizing the bid for a particular search engine in this paper, our approach can be extended to multiple search engines by treating each keyword-search engine pair as a unique keyword.

Ads placed in response to consumer search queries can play two roles for advertisers. They can help generate purchases. Or they can help build awareness, which may translate into purchases in later periods. Consumers often start their search process with generic search terms e.g. "fillet mignon". Bidding on these generic keywords might help the advertiser generate brand-specific (or retailer-specific) exposure. This in turn might enhance the awareness of a particular brand and can lead to increased branded search activity ("spillover").

[3] A common practice in the SEM industry is to use *Daily Budget = (Remaining Balance)/(Number of days left in cycle)*, where remaining balance is the initial monthly budget less the amount spent thus far. We do not focus on how the budget for a given time period is computed and treat it as an exogenous parameter in our formulation.

There is evidence of spillovers from generic to branded keywords in sponsored search ads (Rutz et al., 2012).

In this section, we ignore spillovers between keywords and assume keywords are independent. We propose a "myopic" bidding policy that solves the one-shot decision problem of the advertiser and does not factor in indirect benefits from keywords such as awareness. We relax this assumption in Section 6 and incorporate interactions between keywords. The bidding policy that incorporates interactions between keywords is referred to as the "forward-looking" policy. The forward-looking policy solves the advertising problem in a multi-period context. The motivation for developing two bidding policies is twofold. First, the myopic policy is easier to implement and also relevant in the context of commoditized products where branding is not very relevant. The complexity of a forward-looking policy may be unnecessary for many advertisers. In addition, the forward-looking policy builds on the myopic policy and it is useful for the purpose of exposition to outline the myopic policy first.

3.3.1. Notation and Setup

In this section we introduce our notation and the general framework used to study the advertiser's decision problem.

During a given time period (say a day) a keyword k is searched S_k times, where S_k is a random variable. S_k also represents the total number of impressions, i.e. the number of times the advertiser's ad is displayed by the search engine. The expected number of impressions is defined as $\mu_k = \mathbb{E}[S_k]$. We denote the bid of the advertiser for the keyword as b_k, and assume that the advertiser does not change the bid during the day. Every time the keyphrase is searched, the advertiser's ad is placed at some position in the list of all sponsored results. Let $pos_k^{(s)}$ be the position at which the ad was shown in the s^{th} search, with the topmost position denoted position 0. Let $\delta_k^{(s)}$ be an indicator of whether a person who was searching for the keyword clicked on the advertiser's link, or not: $\delta_k^{(s)} =$

$I\left(click_k^{(s)}\right)$.

The advertiser's value from a click is denoted by an independent random variable w_k. We assume that the precise value from a click is not known a priori but that it's expected value $\mathbb{E}[w_k]$ is known and equals Ew_k. In Section 5, we discuss how Ew_k is estimated from historical data. $v_k^{(s)}$ denotes the advertiser's value from the s^{th} impression. $v_k^{(s)} = \delta_k^{(s)} w_k$, i.e. it equals w_k if the user clicks on the ad or 0 otherwise. Let $\underline{b}_k^{(s)}$ be the advertiser's cost per click i.e. the bid of the advertiser at the next position $pos_k^{(s)} + 1$. The cost associated with impression s may then be expressed as $c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$.[4] Because consumers do not know the bids placed by advertisers, it seems reasonable to assume that given an ad's position in the list, the probability that a person clicks on the ad does not depend on the bid of the next advertiser. That is, conditional on the position $pos_k^{(i)}$, the vector $\left(\underline{b}_k^{(i)}, \delta_k^{(i)}\right)$ has independent components. We also assume that S_k is independent of other variables.

Besides the advertiser, there are N_k other advertisers who place their bids for keyword k. We assume that N_k is known to the advertiser. It can be observed, for example, by submitting sample queries to the search engine and observing the number of ads displayed. We note that the number of competitors may in reality vary a bit from one impression to another due to advertiser budget constraints, but we do not observe significant variation in this to warrant a random treatment for N_k.[5] The bids of the competitors cannot be directly observed because the auction is a sealed bid auction. The key assumption we make is that the competitors place their bids according to some distribution $F_k(.)$ and this does not change during the estimation period. The bids of competing advertisers are based on two factors - their intrinsic valuations for a click and their competitive responses in the GSP auction. We assume that there is an underlying valuation distribution (for clicks) which

[4] The discussion assumes that ads are ordered by bid and that the advertiser pays the bid of the next advertiser. A common practice is to use the product of bid and a quality score to rank order the advertisers, and the payment is the minimum bid needed to secure the position (e.g. the payment per click for an advertiser in position i is $bid(i{+}1) * Quality(i{+}1)/Quality(i)$. This does not affect our model. If we normalize the bid of all competitors by the ratio of their quality score relative to our advertiser ($NormalizedBid = bid * Quality_{Competitor}/Quality_{Advertiser}$), our analysis can be interpreted as based on this normalized bid.

[5] The coefficient of intra-day variation in $N_k = 0.031$ and the coefficient of inter-day varation in $N_k = 0.102$.

when combined with the advertisers' bidding strategies gives rise to the bid distribution $F_k(.)$.[6] Finally, D denotes the advertiser's budget in a given time period of interest. Table 7 summarizes our notation.

Table 7: Summary of notation

k	Variable that indexes keywords
S_k	Random variable denoting number of searches for keyword k
μ_k	Expected number of search for keyword k ($\mathbb{E}[S_k]$)
(s)	Superscript to denote s^{th} search
b_k	Bid for keyword k
$pos_k^{(s)}$	Position for keyword k in s^{th} search. $pos_k^{(s)} = 0$ denotes the top position.
$\delta_k^{(s)}$	Indicator variable for click on s^{th} search.
w_k	Random variable indicating value of a click
Ew_k	Expected value of a click on keyword k ($\mathbb{E}[w_k]$)
$v_k^{(s)}$	Value of the s^{th} search ($v_k^{(s)} = \delta_k^{(s)} w_k$)
$\underline{b}_k^{(s)}$	The bid of the next advertiser
$c_k^{(s)}$	The cost of the s^{th} search ($c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$)
N_k	Number of competitors
$F_k(.)$	Distribution of bids of competitors
D	Advertiser's budget

3.3.2. Model Formulation

The advertiser faces the following decision problem:

$$\max_{\{b_k\}} \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} | b_k\right], \quad \text{s.t. } \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} | b_k\right] \leq D. \tag{3.1}$$

The objective is to determine bids b_k in order to maximize the advertiser's expected revenues. The constraint implies that the expected cost should be less than or equal to a budget D. Note that the budget is not modeled as a hard constraint. This is a common format in which budget constraint is specified by advertisers in the SEM industry, and reflects an objective function of the form $\max_{\{b_k\}} \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} (v_k^{(s)} - \lambda c_k^{(s)}) | b_k\right]$. Thus, the objective is to maximize expected profit but the shadow price of ad dollars is specified in the form of a

[6] The proposed bids might change $F_k(.)$, but for identification purposes we assume that this competitive reaction is minimal in the short-term. Later in the paper, we discuss how the competitive reaction can be factored in by re-estimating parameters periodically and updating bids.

constraint on the expected cost. The optimization problem in Equation (3.1) always has a solution as shown in Appendix A1 (All important proofs appear in the Appendix). Solving the problem gives the following optimality condition

$$\forall k : \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} v_k^{(s)}|b_k\right] = \lambda\frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} c_k^{(s)}|b_k\right]. \tag{3.2}$$

where λ is the Lagrange multiplier. The optimality condition states that at the optimal bids the ratio of the marginal change in the advertiser's expected revenues over the marginal change in the advertiser's expected cost should be constant across keywords. An alternative way to interpret it is as follows. If we decrease the bid for keyphrase k' by ε, then the expected cost will decrease by $\varepsilon\frac{d}{db_{k'}}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[c_{k'}^{(s)}|b_{k'}\right]$ and, hence, we may increase the bid for another keyphrase k by $\varepsilon\frac{d}{db_{k'}}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[c_{k'}^{(s)}|b_{k'}\right]/\frac{d}{db_k}E\sum_{s=1}^{S_k}\left[c_k^{(s)}|b_k\right]$. In this case the expected increase in profits will be

$$\varepsilon\frac{\frac{d}{db_k}\mathbb{E}\sum_{s=1}^{S_k}\left[v_k^{(s)}|b_k\right]\frac{d}{db_{k'}}}{\frac{d}{db_k}\mathbb{E}\sum_{s=1}^{S_k}\left[c_k^{(s)}|b_k\right]}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[c_{k'}^{(s)}|b_{k'}\right] - \varepsilon\frac{d}{db_{k'}}\mathbb{E}\sum_{s=1}^{S_{k'}}\left[v_{k'}^{(s)}|b_{k'}\right] = 0.$$

We assume that consumer click behavior and competitor bidding behavior is i.i.d across ad impressions during the given time period. Hence, in Expression (3.2) we may cancel the sums over s. Therefore, the optimal vector of bids should satisfy the following condition:

$$\forall k : \frac{d}{db_k}\mathbb{E}\left[v_k|b_k\right] = \lambda\frac{d}{db_k}\mathbb{E}\left[c_k|b_k\right]. \tag{3.3}$$

3.3.3. Optimality Condition

It is hard to use the optimality condition (3.3) to compute the optimal bids. In order to apply (3.3), the advertiser needs to compute $\mathbb{E}\left[v_k|b_k\right]$ and $\mathbb{E}\left[c_k|b_k\right]$ accounting for the uncertainty in competing bids and consumer query and click behavior. In this section, we express the optimality condition in terms of parameters that can be estimated. We assume that the number of competitors N_k is known and is constant during the day. We can identify the number of competitors by performing a search on keyword k at a search engine.

Consider a specific keyword k . We tentatively drop the subscript k as we focus on an individual keyword. In order to compute $\mathbb{E}[v|b]$, we need to identify the probability of a click given the bid b, which in turn depends on the probability distribution of the ad position. Given that the competing advertisers' bids are drawn from $F(.)$, the probability of being at position i conditional on a bid b is

$$\Pr\{pos = i|b\} = \begin{pmatrix} N \\ i \end{pmatrix} (1 - F(b))^i F(b)^{N-i}. \tag{3.4}$$

The position is determined by a Bernoulli process, where the probability that a competitor bids more than b and is placed higher is equal to $1 - F(b)$. Recollect that the positions start from 0, i.e., the topmost ad has position $pos = 0$, and position i indicates that there are i other advertisers ranked above. Feng, Bhargava and Pennock's (2007) analysis of click-through data suggests that the probability that a user clicks on an ad in position pos is

$$\Pr\{\delta = 1|pos = i\} = \frac{\alpha}{\gamma^i}, \tag{3.5}$$

where α and γ are keyword specific constants. α represents the overall attractiveness of the ad and γ captures the impact of position on clicks. This functional form does not explicitly consider a number of other factors, e.g. number of words in the keyword, whether the advertiser appears in the organic results or not, presence of dominant competitors etc., that might affect CTR (Yao and Mela, 2011; Agarwal et al., 2011; Katona and Sarvary, 2010; Ghose and Yang, 2009). It focuses only on the impact of position on CTR because ad position is the primary mechanism through which bid impacts CTR. However, the parameter α captures the effect that ad/keyword-level attributes like the number of words in the keyword etc. have on the overall attractiveness of the ad. γ on the other hand captures the change in CTR with respect to position, all other factors held constant, which is consistent with prior research (Katona and Sarvary, 2010; Ghose and Yang, 2009).[7] This function also assumes that consumer behavior is i.i.d and ignores heterogeneity across consumers. We use this assumption not only for model tractability but also because search engines do not provide user-level data on impressions and clicks. Several papers that focus on keyword-level models, also assume i.i.d. consumer behavior (Agarwal et al., 2011; Yang and Ghose, 2010; Ghose and Yang, 2009). Given that the consumers click in the aforementioned manner, the probability of a click conditional on the bid b is given by

$$\begin{aligned}\Pr\{\delta = 1|b\} &= \sum_i \Pr\{\delta = 1|pos = i\} \Pr\{pos = i|b\} \\ &= \sum_i \frac{\alpha}{\gamma^i} \begin{pmatrix} N \\ i \end{pmatrix} (1 - F(b))^i F(b)^{N-i} \\ &= \alpha\gamma^{-N} (1 + (\gamma - 1) F(b))^N .\end{aligned} \tag{3.6}$$

Proposition 1: The expected value of an impression is given by

$$\mathbb{E}[v|b] = \mathbb{E}[\delta w|b] = \Pr\{\delta = 1|b\}\, \mathbb{E}[w] = \alpha\gamma^{-N} (1 + (\gamma - 1) F(b))^N Ew. \tag{3.7}$$

It follows from Proposition 1 that

$$\frac{d}{db}\mathbb{E}[v|b] = \alpha N \gamma^{-N}(\gamma - 1) f(b) (1 + (\gamma - 1)F(b))^{N-1}. \tag{3.8}$$

We now derive an expression for $\mathbb{E}[c|b]$. In order to do so, we need to characterize the probability distribution function of the bid of the next advertiser in the list of sponsored results. We first derive some intermediate results.

Lemma 1: The distribution function of the bid of the next advertiser in the list conditional

[7] However, if the presence of a dominant competitor introduces discontinuities in how position affects CTR (e.g., CTR depends on whether the advertiser is above or below the dominant competitor), the functional form fails to capture the same.

on the bid and the position is given by

$$F\left(\underline{b}|b, pos=i\right)=\left(\frac{F\left(\underline{b}\right)}{F\left(b\right)}\right)^{N-i}. \quad (3.9)$$

Applying,

$$F\left(\underline{b}|b, pos=i, \delta=1\right)=F\left(\underline{b}|b, pos=i\right)=\left(\frac{F\left(\underline{b}\right)}{F\left(b\right)}\right)^{N-i}, \quad (3.10)$$

we can derive the following lemma.

Lemma 2: The conditional distribution of the bid of the next advertiser conditional on the bid and the fact that the ad was clicked is

$$\begin{aligned} F\left(\underline{b}|b,\delta=1\right) &= \sum_{i=0}^{N} F\left(\underline{b}|b,\delta=1,pos=i\right)\times \Pr\left\{pos=i|b,\delta=1\right\} \quad (3.11) \\ &= \left(\frac{1-F\left(b\right)+\gamma F\left(\underline{b}\right)}{1+\left(\gamma-1\right)F\left(b\right)}\right)^{N}. \end{aligned}$$

When a user clicks on an ad, the advertiser has to pay the bid of the next advertiser in the list. Applying Lemma 2 and Equation (3.6) gives us

Proposition 2: The expected cost of an impression is given by

$$\begin{aligned} \mathbb{E}\left[c|b\right] &= \mathbb{E}\left[\delta\underline{b}|b\right] \quad (3.12) \\ &= \mathbb{E}\left[\underline{b}|b,\delta=1\right]\Pr\left\{\delta=1|b\right\} \\ &= \alpha\gamma^{-N}\left(b[1+(\gamma-1)F(b)]^{N}-\int_{0}^{b}[1-F(b)+\gamma F(\underline{b})]^{N}d\underline{b}\right). \end{aligned}$$

Using Proposition 2 we can derive that

$$\frac{d\mathbb{E}[c|b]}{db} = \alpha N\gamma^{-N}f(b)\left((\gamma-1)b[1+(\gamma-1)F(b)]^{N-1}+\int_{0}^{b}[1-F(b)+\gamma F(\underline{b})]^{N-1}d\underline{b}\right) \quad (3.13)$$

Substituting Expressions (3.8) and (3.13) in Equation (3.3),

$$\frac{d\mathbb{E}[v|b]}{db} = \lambda\frac{d\mathbb{E}[c|b]}{db},$$

$$\frac{1}{\lambda} = \frac{1}{Ew}\left(b + \frac{\int_0^b [1 - F(b) + \gamma F(\underline{b})]^{N-1} d\underline{b}}{(\gamma - 1)[1 + (\gamma - 1)F(b)]^{N-1}}\right).$$

Proposition 3: The optimality condition (expressed in terms of estimable parameters) is

$$\forall k : \frac{1}{Ew_k}\left(b_k + \frac{\int_0^{b_k} [1 - F_k(b_k) + \gamma_k F_k(\underline{b})]^{N_k-1} d\underline{b}}{(\gamma_k - 1)[1 + (\gamma_k - 1)F(b_k)]^{N_k-1}}\right) = const. \qquad (3.14)$$

Proposition 4: A unique bid b_k^* satisfies the optimality condition (Equation 3.14) for keyword k when

$$\gamma_k > 1 + \frac{1}{F_k(b)}\left[f_k(b)(N_k - 1)\frac{\int_0^b [1 - F_k(b) + \gamma_k F_k(x)]^{N_k-2} dx}{[1 + (\gamma_k - 1)F_k(b)]^{N_k-2}} - 1\right].$$

The optimality condition can be used in conjunction with the budget constraint to compute the optimal bids. For several common distributions and a wide range of parameters, we show in the appendix that the conditions for a unique bid (Proposition 4) are satisfied. In order to compute the optimal bids, the following keyword-specific constants need to be known: α_k (the click-through rate at the top position), γ_k (rate at which CTR decays with position), Ew_k (expected revenue per-click (RPC)), N_k (number of competing bidders), and $F_k(.)$ (distribution of competing bids). We estimate these parameters using a real-world dataset and illustrate how bids may be computed in Section 5. The optimal bids should satisfy equation (3.14) and the budget constraint,

$$\sum_k \mu_k \mathbb{E}[c_k] = D.$$

These conditions are sufficient to compute bids. The budget constraint can be rewritten as

$$\sum_k \mu_k \alpha_k \gamma_k^{-N_k}\left(b_k[1 + (\gamma_k - 1)F_k(b_k)]^{N_k} - \int_0^{b_k} [1 - F_k(b_k) + \gamma_k F_k(\underline{b})]^{N_k} d\underline{b}\right) = D. \quad (3.15)$$

For a given *const* in Equation (3.15), we compute the bid that satisfies the equation for every keyword. Then we use Equation (3.15) to calculate the expected total cost for the computed bids. If the expected cost is lower than D, we increase the constant, otherwise we decrease it. The process repeats until the expected total cost is sufficiently close to the budget.

3.4. Data Description

Our dataset is from a leading meat distributor that sells through company owned retail stores as well as online and through mail-order catalogs. This firm bids on thousands of keywords across several search engines and has a substantial online presence. Our dataset consists of daily summary records for 247 keywords that the firm uses to advertise on Google. The daily record for each keyword has the following fields,

$$(id, \quad t, \quad b, \quad i, \quad cl, \quad avgcpc, \quad avgpos)$$

where

id -	Unique identifier for each keyword
t -	date
b -	bid submitted by advertiser
i -	number of impressions during the day
cl -	number of clicks during the day
$avgcpc$ -	average cost per click on the day
$avgpos$-	average position during the day

This dataset is representative of the the type of data available to advertisers in sponsored search. Advertisers only get summary reports from search engines and do not usually have information on clicks and position for each individual ad impression. We present the summary statistics at the keyword-level for a three-month period (March 01-May 31, 2011) prior to the field implementation in Table 8.

Table 8: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Avg Bid	1.18	1.01	0.35	10.00
Avg CPC	0.73	0.59	0.00	4.42
Avg Pos	3.15	1.90	1.00	12.41
Impressions	5637.22	13106.39	1.00	98373.00
Clicks	48.37	86.76	0.00	593.00
CTR	0.03	0.07	0.00	0.60
Cost	45.95	95.11	0.00	747.43
Revenue	83.26	132.47	0.00	974.31
Gross Profit	37.31	140.78	-747.43	902.20
Avg RPC	4.33	14.30	0.00	158.96

The mean average bid across all keywords during this period is \$1.18 and the minimum and maximum average bids for any keyword during this period is 35¢and \$10, respectively. We also observe that the mean average RPC is \$4.33 where as the mean average CPC is 75¢, however there is a huge variation in the profitability across the keywords as indicated by the large standard deviation in the average RPC. These 247 keywords belong to 29 unique product categories which span frozen meats, sea foods and desserts. A comprehensive list of these product categories appears in Table 9. We *randomly* divided these 29 product categories into three distinct treatment groups. The bids for the first group continued to be controlled by the firm. This group forms the control group for our experiment. The other two groups represent the two treated groups and their bids are determined by the myopic bidding policy (Group I) outlined in Section 3 and the forward looking policy (Group II) that we outline in Section 6. The control group is used to account for any time trends that might enter the analysis due to seasonality in retail, search engine design changes and other such factors. The three groups are fairly well matched in terms of impressions, clicks, cost and revenues of their keywords. Summary statistics for the three groups are presented in Table 10.

Our dataset is divided into three distinct periods as shown in Figure 8. The first period runs from March 1-May 31, 2011. This period forms the "before" period for our analysis during which the bids for these 247 keywords were decided by the firm (summary statistics

Table 9: Product Categories

Bacon	Flat Iron	Pork
Beef	Gift Basket	Porterhouse
Beef Jerky	Gifts	Prime Rib
Beef Sirloin	Halibut	Salmon
Burgers	Ham	Shrimp
Catfish	Hot Dogs	Sole
Cheesecake	Lobster	Surf and Turf
Corned Beef	Lobster Bisque	Swordfish
Crab	London Broil	Trout
Fillet Mignon	Orange Roughy	

Table 10: Summary for the three groups of keywords.

	Control	Group I	Group II
Products Categories	8	10	11
Keywords	55	89	103
Impressions	7474.7	5336.1	4335.8
Clicks	66.6	52.2	30.6
Avg Bid	1.32	1.27	1.14
Avg CPC	0.84	0.91	0.74
Avg Pos	3.22	3.60	2.55
CTR	0.03	0.03	0.04
Avg RPC	4.80	5.36	4.35

for this period is in Table 8). During this period, there were 1.36 million impressions of the ads for the 247 keywords and they received 11,651 clicks in total. The total weekly cost of these ads was \$964 and the weekly gross revenue generated from these keywords was \$1728. We use the data from this period to compute the expected value *per-click* (Ew) and the expected daily impressions (μ) for each keyword.

The second period spans July 1-July 31, 2011 which we refer to as the "estimation period" for our analysis. We ignore the month of June from our analysis as there is a significant increase in online activity during this month due to Father's Day. The summary statistics for this period is presented in Table 11. During the estimation period, we submit random bids for the keywords in Groups I and II. The bids are uniformly drawn from $\$0.10 \times [1, 30]$ resulting in a minimum bid of 10¢to a maximum bid of \$3.00. The upper limit of \$3.00 was prescribed by the advertiser. The bids are drawn weekly which leads to four unique bids

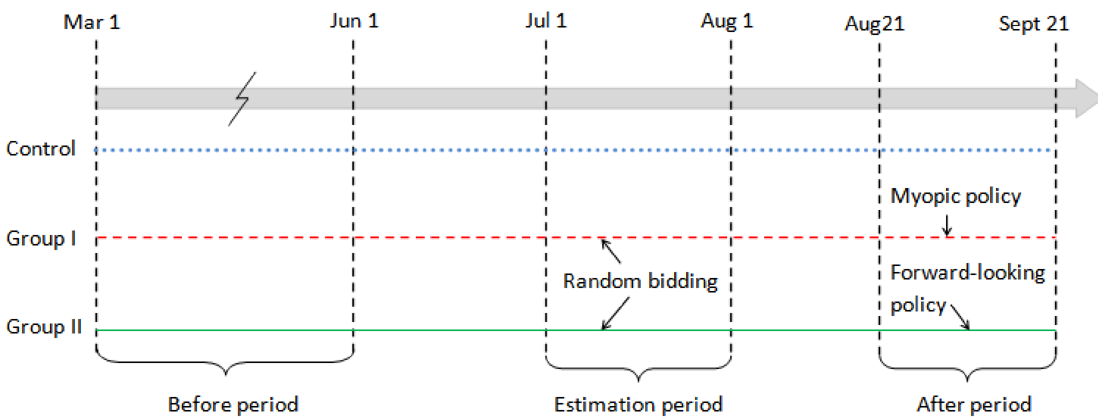

Figure 8: Illustration of the Timeline for the various data collection periods.

per keyword in the estimation period. This variation in bids leads to a significant variation in the ad position. The variation in position causes changes in the CTR and CPC and helps the identification of the parameters of our model. The exact identification strategy is discussed in Section 5. We also observe that there is decrease in the profitability of the campaign during this period as the bids for keywords in Group I and Group II are chosen randomly.

Table 11: Summary Statistics for the Random Bidding Period

	Mean	Standard Deviation	Minimum	Maximum
Bid	1.01	0.93	0.05	3.00
Avg CPC	0.77	0.58	0.00	2.89
Avg Pos	2.96	1.87	1.00	11.08
Impressions	2100.45	5115.73	1.00	43498.00
Clicks	19.00	35.78	0.00	277.00
CTR	0.03	0.06	0.00	0.50
Cost	18.29	38.51	0.00	278.97
Revenue	23.46	75.65	0.00	586.92
Gross Profit	5.18	67.51	-150.43	541.22
RPC	1.28	4.31	0.00	29.32

Finally, optimal bids are computed based on estimated parameters and deployed by the firm between August 21 and September 21, 2011. Data from the after period is used to assess the effectiveness of the bidding policies proposed in this paper. In Section 5, we discuss the estimation of parameters using data from the "estimation period". Subsequently, we discuss the results from the field implementation of the myopic policy.

3.5. Empirical Analysis

We now apply our technique to a real-world dataset of clicks and costs for several keywords and derive the optimal bids for these keywords. We then describe the results from a field implementation of the suggested bids.

3.5.1. Estimation Approach

Our data provide daily summary measures (average position, average cost per click, total clicks) but not the outcome of each individual impression. Given just these daily summary measures, it is hard to apply regression or Maximum Likelihood Estimation techniques directly on the aggregated data, hence we use the Generalized Methods of Moment (GMM) approach to estimate these parameters. Following the idea of the method of moments, we derive analytical expressions for the moments we observe empirically, namely, the expected position ($avgpos_t$), cost per click ($avgcpc_t$) and click-through rate ($ctr_t = cl_t/i_t$) given the bid for each keyword. These moments are as follows:

$$\mathbb{E}\left[pos_t|b_t\right] = N_t\left(1 - F\left(b_t\right)\right), \tag{3.16}$$

$$\mathbb{E}\left[\underline{b}_t|b_t, \delta_t = 1\right] = \int_{x<b_t} x \; d\left(\frac{1 - F\left(b_t\right) + \gamma F\left(x\right)}{1 - \left(1-\gamma\right) F\left(b_t\right)}\right)^{N_t}, \tag{3.17}$$

$$\mathbb{E}\left[\delta_t|b_t\right] = \alpha\gamma^{-N_t}\left(1 - \left(1-\gamma\right) F\left(b_t\right)\right)^{N_t}. \tag{3.18}$$

The observed moments can be expressed in terms of the analytical moments as follows:

$$\begin{aligned} avgpos_t &= \mathbb{E}\left[pos_t|b_t\right] + \xi_{1t}, \\ avgcpc_t &= \mathbb{E}\left[\underline{b}_t|b_t, \delta_t = 1\right] + \xi_{2t}, \end{aligned}$$

$$ctr_t = \mathbb{E}\left[\delta_t | b_t\right] + \xi_{3t},$$

where $\boldsymbol{\xi}_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$ are the random shocks. As the dataset contains only daily aggregates, we cannot directly estimate the distribution function $F(.)$ using nonparametric approaches since we have very few bids for each keyword. We therefore use a parametric form for $F(.)$, and estimate its parameters using the first moments associated with the position, cost per click and click-through rate. For the parametric form of the distribution $F\left(.\right)$ we choose the Weibull distribution. This choice is based on two factors. Firstly, the Weibull distribution can take on diverse shapes and offers a great deal of flexibility. Secondly, an analysis of a secondary dataset of bids submitted to a search engine for several keywords in the insurance sector (Abhishek et al., 2011) shows that the Weibull distribution is reasonably good for modeling the bids.[8] Note that we are not assuming that the distribution of bids for keywords is the same across the two datasets, rather the bids are from the same family (Weibull) and the parameters can vary across keywords. The Weibull distribution has the following cumulative distribution function

$$F\left(x; \theta, \lambda\right) = 1 - \exp\left\{-\left(\frac{x}{\lambda}\right)^{\theta}\right\}.$$

It is defined by two parameters θ and λ. Therefore, we have four unknown parameters for any keyword $(\lambda, \theta, \alpha, \gamma)$ and 3 moment conditions for every unique bid.

The estimates of the parameter $\beta = (\alpha, \gamma, \lambda, \theta)$ is given by

$$\hat{\beta} = \arg\min_{\beta \in B} \boldsymbol{\xi}(\beta)' W \boldsymbol{\xi}(\beta),$$

where $\boldsymbol{\xi}(\beta)$ is a vector of error between the observed and computed moments for a particular keyword during the observation period and W is a weighting matrix. The choice of W is critical as it determines the asymptotic properties of the estimator. Hansen et al. (1996) and

[8]The authors test several distributions such as Normal, Log-Normal, Gamma, Exponential and Logit but the Weibull distribution fits their data the best. Note however that our framework is flexible enough and other distribution can be easily accommodated.

Wooldridge (2001) suggest that the optimal weighting matrix is given by $\mathbb{E}[\boldsymbol{\xi}(\beta)'\boldsymbol{\xi}(\beta)]^{-1}$. As we do not know $\mathbb{E}[\boldsymbol{\xi}(\beta)'\boldsymbol{\xi}(\beta)]$, an iterative-GMM estimator is used (see Hansen et al., 1996) wherein the weighting matrix is iteratively re-estimated till it converges.

In order to compute the optimal bids we also need to know Ew, the expected *revenue per-click*. The expected *revenue per-click* is computed by taking the total revenues from the keyword in the "before" period and dividing it by the total number of clicks for that keyword in the same period. The advertiser attributes revenues from a purchase to the keyword that generated the session in which the purchase was made. One drawback with the approach is that it does not account for indirect benefits such as awareness. As stated above, we address that later in the paper.

3.5.2. Identification Strategy

The parameters of this model can be estimated if we have at least 2 unique bids per keyword in the data. However, there are two important reasons why data from the "before" period cannot be used to estimate the parameters of this model - (i) insufficient variation in bids, and (ii) potential endogeneity in advertiser's bids.

Limited Variation in Bids

In typical SSA campaigns advertisers change their bids infrequently, sometimes once in several months. Hence it is difficult to identify the parameters of the model. In our dataset, there are very few changes in the bids for the keywords and the average number of unique bids per keyword are 1.12. Because our model is under-identified with less than two unique bids, we use the period of random bidding to generate random bids which would lead to sufficient variability in the bids drawn for a particular keyword across days.

Endogeneity of Bids

The second concern with using historical data is the potential endogeneity of bids. In order for the GMM to provide consistent estimate we require that $\mathbb{E}[b\boldsymbol{\xi}] = 0$ or the bids and the random shock are independent of each other. However, the firm might increase the bid for a particular keyword if there is a random increase in demand, e.g. on a sunny weekend. These random shocks are observed by the advertisers but we as researchers are not aware of them. Since the firm is bidding strategically, it is very likely that the bids for a particular keyword are correlated with these random shock in the before period. We address this endogeneity issue by using random bids in the estimation period. This randomization of bids ensures that they are independent of the random shocks.

We also require that the distribution $F(.)$ does not change during the estimation period as competitive response to the random bids being set during this period. This seems like a reasonable assumption given the muted short-term competitive response in sponsored search as pointed out by Rutz et al. (2012). We revisit this assumption in Section 7.2.

3.5.3. Estimation Details and Results

In order to estimate the parameters, a nonlinear solver is used in our implementation.[9] The parameter estimates for a few representative keywords are shown in Table 12. $\overline{N}$ represents the mean number of daily competitive ads in the observation period. For brevity, we plot the distribution of the estimated parameters for all keywords in Groups I and II in Figure 9. A complete table is available from the authors upon request.

Although there is significant heterogeneity across keywords, the estimated parameter values are fairly typical in sponsored search. The mean click-through rate (α) at the topmost position is 0.026 and the mean decay parameter (γ) is 1.68 which is similar to the values reported earlier (Feng et al., 2006; Craswell et al., 2008). There is also considerable variation in the *expected revenue per-click* (Ew) and the bid distributions (λ,θ) across keywords.

[9] We use the Fletcher-Xu hybrid method provided as a part of the ClsSolve routine in TOMLAB.

Table 12: Parameter estimates for a sample subset of keywords.

keyword	λ	θ	α	γ	$\overline{N}$	$Ew(\$)$
beef sirloin steak	1.7651	0.5351	0.0266	2.1237	9.5	0.00
	(0.4927)	(0.2821)	(0.0115)	(0.2742)		
Steak Burger	0.6697	2.1944	0.0069	1.2915	5.1	1.53
	(0.4035)	(0.3731)	(0.0008)	(0.0902)		
cheesecakes	0.9736	1.3265	0.0004	1.6091	7.0	1.08
	(0.2064)	(0.4270)	(0.0000)	(0.2405)		
Porterhouse Steak	1.1413	0.8639	0.0085	1.1661	4.6	0.43
	(0.5118)	(0.1821)	(0.0015)	(0.3711)		
smoke salmon	1.3414	1.1752	0.0073	1.0255	10.1	6.62
	(0.5429)	(0.4520)	(0.0012)	(0.3989)		
corned beef	1.5368	0.7492	0.0018	1.0175	10.7	3.80
	(0.8126)	(0.5781)	(0.0004)	(0.7045)		
hot dog order	1.0769	1.0869	0.0101	1.6486	7.3	3.00
	(0.4410)	(0.7503)	(0.0036)	(0.2446)		
birthday gifts	1.1756	0.8420	0.0009	1.0659	40.2	5.74
	(0.6781)	(0.4176)	(0.0000)	(0.7850)		
birthday present	0.7524	1.3841	0.0122	1.0434	7.1	0.45
	(0.6721)	(0.4816)	(0.0057)	(0.9381)		
lobster bisque	1.311	1.0074	0.0145	1.9293	11.3	0.00
	(0.3928)	(0.5025)	(0.0037)	(0.4117)		

3.5.4. Field Implementation

Once we estimate parameters α, γ, λ and θ for all keywords, we estimate the optimal bids for these keywords. In this section we focus on the myopic policy outlined in Section 3 and discuss the results of the field implementation for keywords in Group I.

For the keywords in Group I, we use a daily budget $D = \$72.00$ based on the mean weekly spend of around \$500 during the 3 month "before" period. The bids are recomputed after two weeks when we re-estimate the parameters $(\alpha, \gamma, \lambda, \theta)$ using newly available data. The bids are recomputed to account for changes in competitor bids and consumer click behavior. However, the bids do not change much during this re-computation. Bids for a sample of keywords are below in Table 13.

The rationale for these bids can be inferred from the parameters listed in Table 12. Consider, for example, bids for keywords "smoke salmon", "hot dogs order" and "birthday gifts". Our

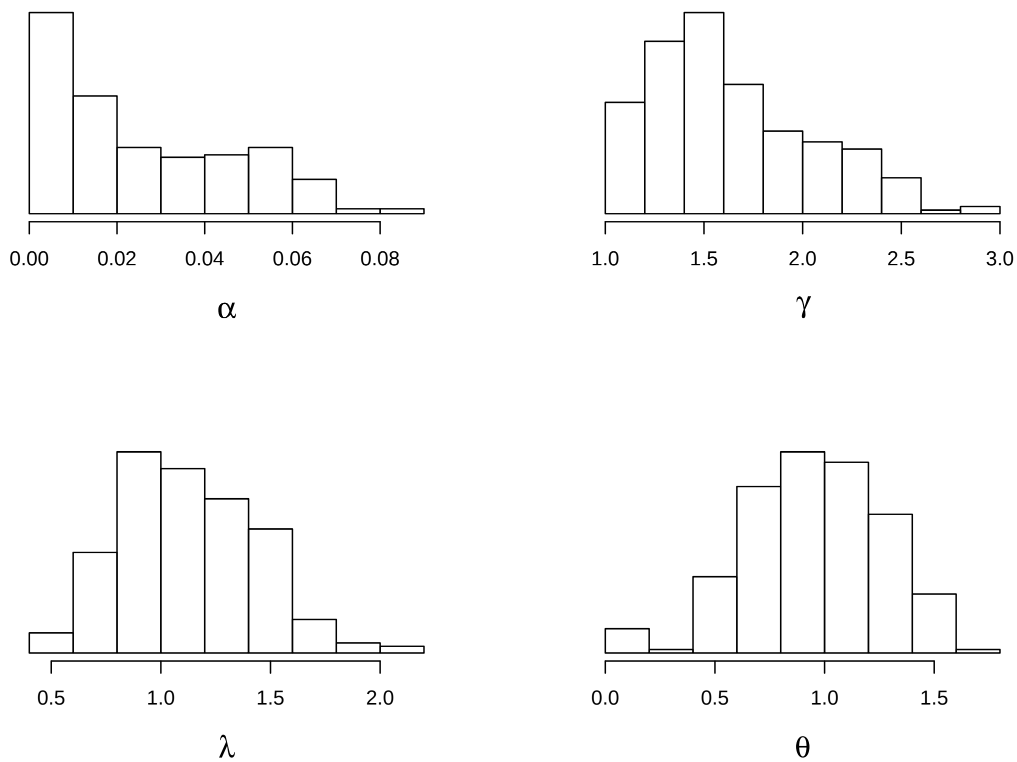

Figure 9: Distribution of estimated parameters across keywords.

algorithm suggests increasing their bids. From Table 12, we observe that their expected value per click (Ew) is high and it makes sense that the algorithm is suggesting that we increase their bids. Interestingly, the keyword "birthday gifts" has a very high Ew, yet its bid is not raised by a significant amount. This is because the keyword is very expensive (low θ) and it is very difficult to attain the top position. There are other keywords where it is worthwhile to spend the advertising dollars. This policy also decreases the bids for keywords like "beef sirloin steak", "lobster bisque" and "birthday present". The bids for "beef sirloin steak" and "lobster bisque" are decreased because they are not profitable. The bid for "birthday present" is decreased because (*i*) it is not very profitable and (*ii*) it is possible to get a similar number of clicks at a lower position (low γ) for much cheaper.

The suggested bids were deployed in the field by the advertiser for a period of 4 weeks.

Table 13: Parameter estimates for a sample subset of keywords.

keyword	Old Bids ($)	New Bids ($)
beef sirloin steak	0.82	0.00
Steak Burger	2.19	0.95
cheesecakes	0.66	0.70
Porterhouse Steak	0.76	0.30
smoke salmon	1.16	2.55
corned beef	0.31	3.00
hot dogs order	0.76	1.85
birthday gifts	0.96	1.75
birthday present	1.61	0.20
lobster bisque	0.46	0.00

During the 12 weeks in the "before" period, the firm spent a total of $5937.58 on the keywords in Group I and obtained revenues of $9776.10. In the "after" period, the total cost and total revenues associated with the keywords were $3178.82 and $4594.43 respectively. In the same period, the total cost (revenues) associated with the control keywords was $4701.52 ($9776.2) and $1667.54 ($1480.80), respectively. We use a Difference-in-Difference approach to compute the effect of our algorithm. The improvement in performance due to the algorithm is given by

$$\begin{aligned}
\tau_M &= \Delta ROI_{\text{Group I}} - \Delta ROI_{\text{Control}} \\
&= (44.53\% - 64.65\%) - (-11.20\% - 84.30\%) \\
&= 75.38\%
\end{aligned}$$

The performance of the advertising campaign increases by 75.38% on a DiD basis indicating that the myopic policy outperforms the advertisers bidding policy. In the next section we discuss some of the drivers of this performance gain. Surprisingly, we notice that there is an absolute decrease in the ROI in the campaign compared to the "before" time period and this decrease is particularly notable for the Control group. This is partly because of seasonality in meat sales. In addition, there were changes in the manner in which the search engine displayed search results. From July onwards, the search engine started highlighting the top ads by using a light pink background color, which resulted in an increase in the CTR of the top ads.[10] For the Control group we see an increase in the CTR from 0.89% to 1.4% and for the keywords in Group II we see a change from 1.04% to 1.15%. We observed that this not only resulted in an increased CTR for the keywords, but also a decrease in their performance during this time. This change negatively affects the performance of our policy as the underlying parameters that were used to compute the optimal bids have changed. However, the control group allows us to control for such changes in the search engine policy. Since the keywords in the Control Group and Group I are matched as a result of the randomization procedure, the change in the display scheme has the same effect (on average) on the keywords in either groups. Hence, the DiD approach eliminates the influence of the policy change and measure the difference in performance between the myopic policy and the policy adopted by our partner advertiser. It should be noted that, although the parameters of the model are recomputed during the experiment, the estimates did not capture the changes in the CTR. Since the data during the field experiment (1.5 week) were pooled with data from the estimation period (4 weeks), the changes during the experiment did not have a significant impact on the parameter estimates.

3.5.5. Analysis of the Field Experiment

In the previous section we presented the improvement that the myopic policy offers over the advertiser's policy. In this section, we discuss in further detail the factors that lead to the improvement in the campaign's performance. There are two main sources of improvement – Firstly, a comprehensive model that captures the effect of bid on position, position on CTR and eventually the bid on v_k and c_k, helps us in improving the bids for each keywords. Secondly, since bids for the entire portfolio are determined jointly, the advertising budget can be distributed from less profitable keywords to relatively more profitable keywords (based on on the aforementioned model). As both these approaches are concurrently applied to

[10] Several analysts suggest that the pink background for the ads is indistinguishable from the page background and users mistake these ads for organic links.
http://www.plymarketing.com/ppc/6-reasons-googles-new-ad-layout-should-really-piss-you-off/

the portfolio, it is difficult to disentangle the effect of these drivers on the campaign performance. However, we demonstrate how both these decisions affect the bids and profitability of keywords in the campaign. The following table contains a list of sample keywords and their performance in during the field experiment.

Table 14: Changes in keyword performance

keyword	Avg RPC($)	Old bids($)	New bids($)	Avg CPC	ΔGross Profit
buy orange roughy	2.95	0.35	0.85	0.29	343.33
filet mignon	5.00	1.50	2.55	1.83	210.37
buy lobster online	3.92	1.91	2.35	1.86	148.19
smoked salmon lox	5.85	1.00	3.00	2.03	38.24
bbq Beef	3.83	1.52	1.00	0.96	42.56
lobster delivery	1.62	2.25	0.95	0.87	-1.310
precooked bacon	1.57	1.50	1.05	0.68	-38.73
birthday gifts	5.74	0.96	1.75	1.73	-204.92

From the sample presented in Table 14 we observe that the advertiser was initially placing a much lower bids for keywords like "buy orange roughy", "filet mignon" etc. These keywords are not only very profitable (have a high RPC), they are considerably cheap (have a lower CPC) as they do not face intense competition. Using the analytical model presented in Section 3, we can ascertain not only the consumer response parameter (CTR v/s position) but also the competitive landscape associated with a keyword, which are subsequently incorporated in the bidding process. E.g. since there is very little competition for "buy orange roughy", the myopic policy recommends increasing the bid for this keyword. Even though the bid has been increased significantly, the CPC in the "after" period is considerably low (29¢), which further validates the low level of competition for this keyword. Interestingly, the policy reduces the bid for "bbq Beef". This decay in the CTR with position is very low for this keyword ($\gamma = 1.0338$), hence the policy choses a smaller bid to decrease the costs without affecting the revenues considerably. This complex interaction between the bid, the revenue and the cost associated with a keyword cannot be predicted in a modeless manner. Our approach explictly captures the relationship between these variable and hence outperforms the heurtistics adopted by our partner advertiser.

We also observe that the bids for several moderately profitable keywords like "lobster delivery" and "precooked bacon" have been reduced in the after period. The joint bid-optimization reduces the bids for these keywords as the budget invested in these keywords can be diverted to keywords that deliver higher profits. Since their bids are decreased, their position drops and hence they lead to fewer clicks. As a result they generate much less revenue and their gross profits are lower during the field experiment as compared to the "before" period. It should be noted that this decrease in profits is compensated for by the investment in relatively more profitable keywords. In summary, the myopic policy focuses on relatively cheap and profitable keywords and reduces the bids of other keywords to maximize the profits from these keywords. Surprising, we see that the gross profits for "birthday gifts", a high performing keyword in the "before" period, decrease considerably during the field experiment. The myopic policy increases the bid for "birthday gifts" given its high RPC. However, as the CTR increase significantly during the field experiment (due to the changes by the search engine), the value per impression v_k drops considerably and this keyword incurs a loss.

3.6. Incorporating Interdependence between Keywords

The preceding discussion assumes that keywords are independent of each other and the consumer click behavior is i.i.d.. In reality, consumers may search across several keywords before making a purchase decision and this might lead to interaction between keywords. For example, a consumer might begin his search with a generic keyword like "fillet mignon" but may eventually purchase using another keyword such as "Walmart fillet mignon". While searching for fillet mignon, he could have been exposed to ads from Walmart, causing Walmart to be part of his consideration set. Not accounting for such spillovers may cause the advertiser to undervalue "fillet mignon" and overvalue "Walmart fillet mignon". This example illustrates that there is value in accounting for these interactions while making bidding decisions. One way to capture this interaction is a full factorial design, where we consider spillovers for every possible subset of the portfolio of keywords and decide the optimal bids for keywords in this subset. However, the problem is NP hard and requires significant resources to assess the performance of each subset. In this paper we will focus on a specific kind of interaction proposed by Rutz et al. (2012). We categorize the keywords into two groups – generic and branded – and explore how these two groups of keywords interact.

A generic keyword does not contain the brand name of the firm (e.g. "fillet mignon") whereas a branded keyword does (e.g. "Walmart fillet mignon"). Advertising on generic keywords can help create awareness about the brand/product which can then increase the likelihood that the brand is a part of the consumer's consideration set and, in turn, result in greater number of branded searches. Rutz et al. (2012) show that there are considerable spillovers from generic to branded search activity in sponsored search. Methods which do not account for awareness might undervalue some keywords. E.g., in our dataset, clicks on generic keywords are usually more expensive than on branded keywords (e.g., \$0.88 v/s \$0.45) and less profitable (e.g., \$2.89 v/s \$7.80). If we just look at the RPC and CPC of the keywords, it is more profitable to invest in branded keywords as compared to generic keywords. However, as pointed out earlier, bidding on expensive generic keywords might lead to future branded search and more clicks on the profitable branded keywords. Hence, the advertiser should incorporate this spillover effect while making his bidding decisions. In the following discussion we present a model that accounts for this dynamic interaction between the generic and branded keywords while computing optimal bids.

3.6.1. Measuring Interactions

In order to incorporate the spillover effect in our decision model we first need to estimate the changes in awareness due to search activity and its effect on future search activity. We use the Nerlove-Arrow model (Rutz et al., 2012; Naik and Sawyer, 1998; Nerlove and Arrow, 1962) to capture the evolution of awareness

$$\frac{dA_t}{dt} = -(1-\eta^A)A_t + \boldsymbol{\beta}\boldsymbol{X}_t, \tag{3.19}$$

where A_t refers to the awareness level at time t, $(1-\eta^A)$ measures the decay of awareness with time, $\boldsymbol{X}_t$ is a vector of covariates that capture the search activity at time t and $\boldsymbol{\beta}$ captures the extent to which different kinds of search activity affect the level of awareness. According to the Nerlove-Arrow model, brand awareness decays over time since consumers forget about a brand as time goes by. Search activity, on the other hand, reinforces brand awareness. This increased awareness, in turn, can lead to further branded search activity. We divide the keywords into two groups – G (generic) and B (branded) – and explore how search activity related to these keywords affects the level of awareness. The two search activities that we observe in our dataset are impressions and clicks for each keyword in the campaign. Prior results suggest that ad impressions do not have a significant impact on brand awareness but clicks on ads increase brand awareness (Rutz et al., 2012). This is because an ad impression does not guarantee that the ad is seen by the consumer and, further, mere exposure to an ad may not have an impact on the consumer unless the consumer pays sufficient attention to the ad (e.g., by clicking it). We incorporate this finding in our model and assume that generic and branded clicks may increase brand awareness (which is latent in our model and cannot be directly observed). An increase in this latent awareness can lead to more search and hence more generic or branded impressions.[11] This interaction is demonstrated in Figure 10.

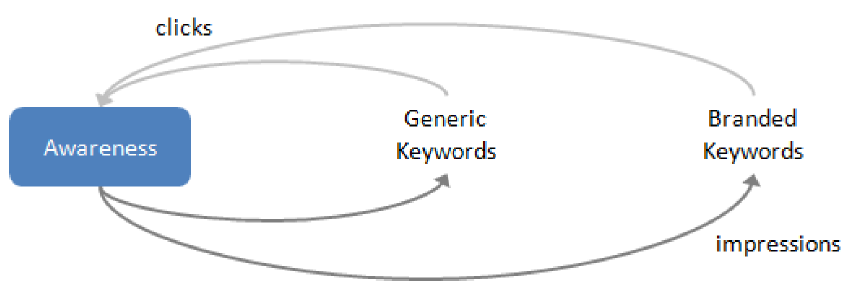

Figure 10: Interaction between search activity and latent awareness.

We first describe how generic and branded clicks affect awareness. The total number of

[11] We validate this assumption in our dataset by performing a Granger causality test and infer that impressions (both generic and branded) do not lead to more clicks but generic clicks lead to more branded impressions.

generic and branded clicks at time t are defined as $CLK_t^G = \sum_{k \in G} cl_{k,t}$ and $CLK_t^B = \sum_{k \in B} cl_{k,t}$, respectively. As we only observe daily data, we use a discrete time analogue of the model presented in Equation (3.19),

$$A_{t+1} = \eta_A A_t + \beta_G CLK_t^G + \beta_B CLK_t^B + \varepsilon_{t+1}^A, \tag{3.20}$$

where η_A captures the carry-over rate of awareness and ε_{t+1}^A is the idiosyncratic error term. Like Rutz and Bucklin (2010), we assume that the awareness at time $t+1$ is affected by the generic search activity at time t but in addition we allow for branded search activity to also impact awareness. As highlighted earlier, awareness is not observed in the data and is latent in this state-space model. Next, we outline how awareness affects both generic and branded search activity. In our model, we assume that awareness only affects the consumer's propensity to search but it has no effect on the consumer behavior after the search is executed. This implies that awareness affects the number of impressions (queries) but has no impact on the *click-through* or *conversion rates.* This assumption is in keeping with the findings of Rutz et al. (2012) who show that awareness does not have a statistically significant impact on $click-through$ and $conversion$ rates. The expected number of generic impressions at time t is defined as $\mu_t^G = \sum_{k \in G} \mu_{k,t}$ and the expected number of branded impressions at time t is defined as $\mu_t^B = \sum_{k \in B} \mu_{k,t}$, where $\mu_{k,t}$ are the expected number of impressions for keyword k at time t. The expected number of generic and branded impressions evolve with awareness in the following manner,

$$\mu_{t+1}^G = \eta_G \mu_t^G + \gamma^G A_{t+1} + \varepsilon_{t+1}^G, \tag{3.21}$$

$$\mu_{t+1}^B = \eta_B \mu_t^B + \gamma^B A_{t+1} + \varepsilon_{t+1}^B. \tag{3.22}$$

It should be noted that the effect of awareness is computed from aggregate data (not individually for each pair of generic/branded keyword).[12] In order to have a parsimonious

[12] One can conduct this analysis at a generic-branded keyword-pair level or a product level if there is sufficient data and variation in that data. Our dataset is very sparse to get statistical significance at keyword-pair or product level.

model we assume that the effect of awareness is homogeneous across all branded keywords. Similarly, the effect across generic keywords is homogeneous.

Combining Equations (3.20)-(3.22), we get a state space model whose evolution is as follows

$$\begin{bmatrix} \mu_{t+1}^{G} \\ \mu_{t+1}^{B} \\ A_{t+1} \end{bmatrix} = \begin{bmatrix} \eta_G & & \gamma_G \\ & \eta_B & \gamma_B \\ & & \eta_A \end{bmatrix} \begin{bmatrix} \mu_{t}^{G} \\ \mu_{t}^{B} \\ A_{t} \end{bmatrix} + \begin{bmatrix} & \\ & \\ \beta_G & \beta_B \end{bmatrix} \begin{bmatrix} CLK_t^G \\ CLK_t^B \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1}^{G} \\ \varepsilon_{t+1}^{B} \\ \varepsilon_{t+1}^{A} \end{bmatrix} \tag{3.23}$$

where the correlated error terms $\varepsilon_{t+1}^{\cdot\cdot}$ account for random shocks and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, V_{\varepsilon})$. The following equation represents how these latent states are linked to the observations,

$$\begin{bmatrix} IMP_{t+1}^{G} \\ IMP_{t+1}^{B} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t+1}^{G} \\ \mu_{t+1}^{B} \\ A_{t+1} \end{bmatrix} + \begin{bmatrix} \nu_{t+1}^{G} \\ \nu_{t+1}^{B} \end{bmatrix}$$

where IMP_{t+1}^{G} and IMP_{t+1}^{B} are generic and branded impressions at time $t+1$, $\nu_{t+1}^{\cdot\cdot} \sim N(0, V_{\nu})$ is the random shock. We estimate this system of equations using a Dynamic Linear Model (DLM). DLMs have been used in several situations where an important component of the model is unobserved (Rutz et al., 2012; Bass et al., 2007; Naik and Sawyer, 1998). We estimate this model using a Markov Chain Monte Carlo (MCMC) approach as proposed by West and Harrison (1997). Details of the estimation procedure are outlined in Appendix A3. The variation in the number of impressions and clicks for generic and branded keywords help us identify the parameters of the model. The estimated parameters of the model are presented in Table 15.

First, we note that there is a strong positive impact of generic clicks on awareness ($\beta_G > 0$). Second, increased awareness leads to increased branded search activity ($\gamma_B > 0$). Combining these results, we conclude that every click on a generic ad increases the number of branded impressions by $\gamma_B \beta_G$ ($= 0.38$). We also observe that the effects of branded clicks on awareness and of awareness on generic search activity are insignificant ($\beta_B \approx 0, \gamma_G \approx 0$).

Table 15: Estimated Parameters

Parameter	Mean	95% Conf. Interval
η_G	**0.9515**	[0.9735, 0.9321]
η_B	**0.8664**	[0.8411, 0.8842]
η_A	**0.2418**	[0.2297, 0.2547]
γ_G	0.0232	[-0.0006, 0.0427]
γ_B	**0.1088**	[0.0997, 0.1132]
β_G	**3.4018**	[3.2656, 3.6123]
β_B	0.0208	[-0.0105, 0.0461]

The figures in bold are statistically significant at the 95% level.

These findings are consistent with the results reported by Rutz et al. (2012). It appears reasonable that if a consumer is already aware of a brand, then clicking on a branded ad is less likely to change his awareness about that brand. Similarly, awareness about a particular brand does not affect consumer's generic search behavior.

We incorporate these estimates of spillovers into our decision theoretic model in the following manner. Given the statistically insignificant estimates of β_B and γ_G, we assume that only generic clicks affect future search behavior and this effect is limited to branded searches. We also assume that all generic clicks are identical and lead to the same relative increase in the search (or impressions) for these branded keywords. More formally,

$$\begin{aligned} \mu_{k,t+1} &= \eta_B \mu_{k,t} + \gamma_{k,B} \beta_G CLK_{G,t} \quad \forall\, k \in B, \qquad (3.24) \\ \mu_{k,t+1} &= \eta_G \mu_{k,t} \quad \forall\, k \in G. \end{aligned}$$

where $\gamma_{k,B} = \gamma_B \frac{\mu_{k,t}}{\mu_{G,t}}$ is the increase in the expected impressions of keyword $k \in B$ at time period $t+1$ for every generic click at time t. The increased impressions for branded keywords, which are usually more profitable, leads to higher revenues in future periods.

3.6.2. Forward-Looking Policy

As discussed in the previous section, bidding on keywords has two effects - current period revenues and future awareness. As a result, the advertiser faces a trade-off between maximizing current period revenues and increasing awareness (through more generic clicks) to increase revenues in the future. We consider the advertiser's problem of deciding the bids for the keywords in each time period so as to maximize the total profits for a finite time horizon. Lets denote the planning horizon by T. We assume that the budget in each time period should be less than or equal to D. The multi-period bidding problem is as follows

$$\max_{\{\bar{\boldsymbol{b}}_t\}} \sum_{t=1}^{T} r(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \quad \text{s.t.} \ \sum_k \mu_{k,t} c_k(b_{k,t}) \leq D, \ t = 1, \ldots, T$$

where $\bar{\boldsymbol{\mu}}_t = (\mu_{1,t}, \ldots \mu_{K,t})^T$ is a vector of the expected number of impressions and $\bar{\boldsymbol{b}}_t$ is a vector of bids for each keyword in period t. $r(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t)$, the expected current period profit, is computed using the formula in Equation (3.1). For ease of exposition, we define the ad spend in time period t as $C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) = \sum_k \mu_{k,t} c_k(b_{k,t})$. We formulate a finite horizon dynamic program with T periods to solve this problem.

$$\begin{aligned}
V(1, \bar{\boldsymbol{\mu}}_1) &= \max_{\{\bar{\boldsymbol{b}}_t\}} \sum_{t=1}^{T} r(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \quad \text{s.t.} \ C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \leq D, \ t = 1, \ldots, T, \\
&= \max_{\bar{\boldsymbol{b}}_1 \text{ s.t. } C(\bar{\boldsymbol{\mu}}_1, \bar{\boldsymbol{b}}_1) \leq D} \left\{ r(\bar{\boldsymbol{\mu}}_1, \bar{\boldsymbol{b}}_1) + \left(\max_{\{\bar{\boldsymbol{b}}_t \text{ s.t. } C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \leq D\}} \sum_{t=2}^{T} r(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \right) \right\}, \\
&= \max_{\bar{\boldsymbol{b}}_1 \text{ s.t. } C(\bar{\boldsymbol{\mu}}_1, \bar{\boldsymbol{b}}_1) \leq D} \left\{ r(\bar{\boldsymbol{\mu}}_1, \bar{\boldsymbol{b}}_1) + \mathbb{E}[V(2, \bar{\boldsymbol{\mu}}_2)] \right\},
\end{aligned}$$

where $V(t, \bar{\boldsymbol{\mu}}_t)$ is the value function at time t. More generally, the Bellman equation for this problem is as follows

$$V(t, \bar{\boldsymbol{\mu}}_t) = \max_{\bar{\boldsymbol{b}}_t \text{ s.t. } C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \leq D} \left\{ r(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) + \mathbb{E}[V(t+1, \bar{\boldsymbol{\mu}}_{t+1})] \right\}.$$

$\bar{\boldsymbol{\mu}}$, the vectors of mean impressions, constitute the state-space and the bids, $\bar{\boldsymbol{b}}$, are the control variable. The state evolves in a manner shown earlier in Equation (3.24). As this is a finite horizon problem, we use backward induction to solve for the optimal bids. At $t = T$, the advertiser does not care about awareness and the optimal policy in the last stage is to bid according to the "myopic" policy. In order to find the optimal bids for $t < T$, we use approximate dynamic programming. We assume that the expected number of generic clicks at time t belongs to the set $CLK = \{0, 1, \ldots, M\}$, where M is an arbitrarily large number.[13] For every $CLK \in \{0, 1, \ldots, M\}$, we evaluate the subsequent state and optimal revenues in period $t + 1$. We now solve the problem in Equation (3.1) with the additional constraint that there are exactly CLK generic clicks in period t. This problem is stated as follows

$$\max_{\{b_t\}} \sum_k \mu_{k,t} \mathbb{E}\left[v_{k,t} | b_{k,t}\right] \quad \text{s.t. } C(\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{b}}_t) \leq D \text{ and } \sum_{k \in G} \mu_{k,t} \mathbb{E}\left[\delta_{k,t} | b_{k,t}\right] = CLK.$$

The optimal policy in this period is to choose a CLK (and the associated bids, $\bar{\boldsymbol{b}}_t$) that maximize the sum of current reward and the optimal future rewards. For the field experiment, we update bids once every two weeks and there are $T = 2$ time periods in total. The optimal bids under this forward-looking policy for some keywords are shown in Table 16 below. We also present the bids that would have been placed if we had used a myopic bidding policy instead. The forward-looking policy increases the bids for some of the generic keywords if they are likely to generate clicks. Accordingly, the bids for some of the less profitable branded keywords are reduced.

Table 16: Bids under the forward-looking policy

Keyword	Ew(\$)	Myopic Bids	Forward-Looking Bids
buy barbecue	1.94	0.65	0.85
porterhouse steaks	5.50	2.30	2.40
lobster bisque	0.00	0.00	0.20
Nebraska beef	2.69	1.25	1.25
purchase hot dog	4.47	2.45	2.65
buy top sirloins online	0.00	0.00	0.05
trout fillets	0.00	0.00	0.00
beef sirloin online	0.00	0.00	0.00
BRAND-NAME lobster bisque	6.43	2.45	2.15
BRAND-NAME steak burgers	14.59	3.00	3.00

[13] $M = 200$ in our analysis.

3.6.3. Field Implementation

We apply the forward-looking policy to keywords in Group II. A daily budget $D = \$35.00$ is used based on the mean weekly spending of \$250 during the 3 month "before" period. We consider two time periods in our forward looking policy and compute the bids accordingly. The bids computed for the first period are deployed in the field for a period of 2 weeks and the bids computed for the second (last) period are deployed for two weeks thereafter.

During the 12 weeks in the "before" period, the advertiser incurred a cost of \$3052.60 and earned revenues of \$5646.03 for Group II keywords. In the "after" period the cost and revenues were \$1201.67 and \$2075.43, respectively. Using the Difference-in-Difference approach, as in Section 5.4, the improvement in performance is estimated to be

$$\begin{aligned}\tau_{FL} &= \Delta ROI_{\text{Group II}} - \Delta ROI_{\text{Control}} \\ &= (72.71\% - 84.96\%) - (-11.20\% - 84.30\%) \\ &= 83.25\%\end{aligned}$$

There is a notable increase in the performance of Group II keywords relative to the control group. Further, the forward-looking policy provides performance gains over and above that delivered by the myopic policy ($\tau_{FL} - \tau_M = 7.87\%$). The effectiveness of the forward-looking policy is likely to depend on prior brand awareness among search engine users and also on the duration of the experiment. Thus, the gains may vary in other settings based on prior brand awareness. We expect that the gains from the forward-looking policy will be greater if the field experiment is conducted over a longer duration. We were unable to experiment for an extended period of time due to limitations imposed by our partner advertiser.

3.7. Discussion

In this section, we contrast our proposed approach with policies commonly used by advertisers in sponsored search. Then we shall discuss some caveats to our models which might limit the applicability of our bidding policies.

3.7.1. Contrast with Commonly Used Strategies

Our agreement with the advertiser precludes sharing their exact bidding strategy. However, their strategy is fairly typical of strategies used by most advertisers in sponsored search. There are three main reasons why our policies perform better than the policies adopted by these advertisers. Firstly, because of the complexity of bid determination, most advertisers use simple heuristics to determine bids. One common heuristic is to simply raise bids for keywords that generate purchases at a relatively low cost and to reduce bids for keywords that do not generate purchases. While this is a reasonable heuristic, it does not account for details of the bid distribution or how the CTR decays with position. For e.g., for some keywords, reducing bids may reduce clicks significantly but it may not have a significant impact on cost-per-click. Parameters tied to competing bids and click decay have a significant impact on optimal bids. As shown in Table 13 it might not be optimal to invest heavily in a profitable *but* highly competitive ad (birthday gifts). Another challenge for advertisers is that they often manage bids for keywords individually without optimizing the portfolio as a whole. Raising and lowering bids for keywords in equal increments to manage the budget constraint is suboptimal. Optimizing the bids over the entire portfolio helps to move the advertising dollars from poorly performing keywords to profitable ones in the right increments. Thirdly, the forward-looking policy accounts for the two-fold effect that sponsored search ads have - awareness and profits. By ignoring the awareness benefits of generic keywords, advertisers often under-invest in generic keywords and over-invest in branded keywords.

3.7.2. Competitive Reaction

This paper adopts a decision-theoretic perspective of the bid problem as opposed to an equilibrium perspective. Advertisers have to submit bids based on their current beliefs and may choose to update these bids as their beliefs evolve. Our framework accommodates that by assuming that advertisers can use new data to re-estimate the model parameters and update their bids. If competing advertisers respond instantaneously to changes in bids then this may reduce the effectiveness of our bidding policies or at the very least suggest that bids need to be rapidly and continuously updated. However, current research suggests that competition in sponsored search advertising is fairly subdued (Rutz and Bucklin 2010, Steenkamp et. al. 2005). Our discussions with several managers indicates that bids for these keywords are rarely updated continuously. This is also reflected in our dataset where bids for less than 10% of the keywords were changed in the 12 week "before" period when the advertiser was deciding the bids.

To test whether rapid reaction by competitors render our computed bids ineffective, we compare the difference between the predicted and observed average (*i*) position and (*ii*) *cost per-click* in the after period (presented earlier in Equations (3.16) and (3.18)). If competitors react soon to our advertiser's new bids, it would introduce notable errors in our predictions regarding the expected position and cost-per-clicks. For most of the keywords, there is no significant difference between the predicted moments and the daily summaries reported by the search engine, which indicates that there is no significant short-term competitive reaction.[14] There can however be long-term competitive reaction and the model parameters $(\lambda, \theta, \alpha, \gamma)$ can be periodically re-estimated and the bids updated to account for these changes. This estimation would not suffer from endogeneity issues as long as the bids are determined through the proposed algorithm and are uncorrelated with random shocks. Since the issue of endogeneity no longer arises, there might not be a need for a random bidding period.

3.7.3. Spillover across Groups

While computing the effectiveness of the "forward-looking" bidding policy in Section 6, we implicitly assumed that there are no spillovers across groups. However, spillovers from key-

[14]The Mean Absolute Error (MAE) of these moments averaged across all keywords are shown in Table 17 in Appendix A2.

words in one treatment group into keywords in another group might influence the estimate of τ_{FL}. To control for this, we divide keywords into product categories and assign all keywords from a product category into the same treatment group. This experimental design is motivated by the intuition that clicks on keywords related to a particular product will not have any impact on the search behavior for other products, e.g. while clicks for "hot dogs" can spill over to branded keywords within the same product category "BRAND-NAME hot dog", it will have insignificant impact on searches for "salmon" or "BRAND-NAME salmon" . This procedure of random assignment by product categories helps ensure such that most of the spillovers are within treatment groups. This experimental design is based on Angelucci and Giorgi (2009) where they propose a methodology to measure treatment effects with spillovers. Note that the random assignment of the product categories to groups also ensures that even if there are some spillovers across groups, these effects are similar between any given pair of groups. A more sophisticated way to incorporate the spillover effect might be a multi-tier design as proposed by McConnell, Sinclair and Green (2010) but this approach would severely affect the analytical tractability of our approach and has been left as a direction for future research.

3.8. Conclusions

The presence of a large portfolio of keywords, multiple slots for each keyword and significant uncertainty in the decision environment make an advertiser's problem of bidding in sponsored search a challenging optimization problem. In this paper, we formulated the advertiser's decision problem and analytically derived the optimality condition. Our bid optimization model addresses a major gap in prior work related to incorporating multiple slots per item, uncertainty in competitor bidding behavior and consumer query and click behavior. We illustrated the technique using a real-world dataset. A field test suggests that the approach can substantially boost advertiser's RoI. We extend our basic model to account for secondary effect of these ads - awareness - and show that incorporating awareness into a multi-period bidding problem can help increase revenues further.

There are a number of interesting avenues along which our work can be extended. We discuss these below.

Exploration and Learning: Our analysis assumes that keyword-specific parameters are known or can be easily estimated based on recent historical data. If there has been sufficient bid exploration in the recent history, these parameters can be estimated as demonstrated in our empirical study. However, new keywords and keywords for which bids have settled down into a relatively narrow range present a challenge. Thus an important area of opportunity to further extend our work is to combine optimization with a suitable exploration technique. Exploration is clearly expensive but facilitates more accurate estimation of parameters. Heuristics proposed for Multi-armed Bandit and budget constrained Multi-armed Bandit problems are particularly relevant for balancing exploration and exploitation.

Modeling Advertiser Heterogeneity: The key assumption we make in this paper is that competitor bids are drawn from the same distribution. This allows us to keep the model tractable and solve the complex stochastic optimization problem faced by an advertiser but ignores heterogeneity among competitors. Modeling heterogeneity in advertisers' bidding policies is an important next step for our research. Additionally, our focus in this paper, like that of the stream of work on optimal bidding, is the operational bid determination problem faced by an advertiser at any given instant rather than an economic analysis of the long-term equilibrium that results from the bidding strategies of advertisers in a market. Equilibrium analysis is another interesting direction, albeit a complex one in this setting due to the presence of multiple keywords and a budget constraint.

Appendix

Proofs of Equations and Propositions

Solution of Equation (3.1)

The constrained optimization problem is as follows

$$\max_{\{b_k\}} \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)}\right], \quad \text{s.t.} \quad D - \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right] \geq 0.$$

The Lagrangian can be written as:

$$\mathcal{L} = \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)}\right] + \lambda \left\{ D - \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right]\right\}.$$

KKT Conditions

$$\begin{aligned}
&\forall k: \ \frac{d\mathcal{L}}{db_k} = \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} v_k^{(s)}|b_k\right] - \lambda \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} c_k^{(s)}|b_k\right] = 0 \\
&\lambda \geq 0, \\
&D - \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right] \geq 0.
\end{aligned}$$

Assuming the budget constraint is binding (i.e.$\lambda > 0$), then there exists an extremum s.t.

$$\forall k: \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} v_k^{(s)}|b_k\right] = \lambda \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} c_k^{(s)}|b_k\right]$$

As $rank\left(\frac{d(D-\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)}\right])}{db}\right) > 0$, there exists at least one local maxima, and it maximizes the objective function if it is unique.

Assuming $v_k^{(s)}, c_k^{(s)}$are i.i.d., the optimality condition reduces to

$$\mathbb{E}[S_k]\frac{d\mathbb{E}[v_k|b_k]}{db_k} \quad = \quad \lambda\mu_k \frac{d\mathbb{E}[c_k|b_k]}{db_k},$$

$$\text{or} \quad \frac{d\mathbb{E}[v_k|b_k]}{db_k} = \lambda \frac{d\mathbb{E}[c_k|b_k]}{db_k}.$$

Proof of Lemma 1

The probability that the bid of the next advertiser is less than x for some $x < b$ conditional on the bid b and the position i is equal to the probability that exactly i advertisers bid more than b and exactly $N - i$ advertisers bid less than x divided by the probability that the position is i. That is,

$$\begin{aligned}
&F\left(\underline{b} = x|b, pos = i\right) \\
&= \Pr\left\{\underline{b} < x|b, pos = i\right\}, \\
&= \frac{\Pr\left\{\underline{b} < x, pos = i|b\right\}}{\Pr\left\{pos = i|b\right\}}, \\
&= \frac{\begin{pmatrix} N \\ i \end{pmatrix} \left(1 - F\left(b\right)\right)^i F\left(x\right)^{N-i}}{\begin{pmatrix} N \\ i \end{pmatrix} \left(1 - F\left(b\right)\right)^i F\left(b\right)^{N-i}}, \\
&= \frac{F\left(x\right)^{N-i}}{F\left(b\right)^{N-i}}.
\end{aligned}$$

Proof of Lemma 2

$$\begin{aligned}
&F\left(\underline{b} = x|b, \delta = 1\right) \\
&= \Pr\left\{\underline{b} < x|b, \delta = 1\right\}, \\
&= \sum_{i=0}^{N} \Pr\left\{\underline{b} < x|b, \delta = 1, pos = i\right\} \times \\
&\quad \Pr\left\{pos = i|b, \delta = 1\right\}, \\
&= \sum_{i=0}^{N} F\left(x|b, pos = i\right) \times
\end{aligned}$$

$$
\begin{aligned}
&\frac{\Pr\{\delta=1|b,pos=i\}\Pr\{pos=i|b\}}{\Pr\{\delta=1|b\}},\\
&=\sum_{i=0}^{N}\left(\frac{F(x)}{F(b)}\right)^{N-i}\times\\
&\frac{\frac{\alpha}{\gamma^i}\begin{pmatrix}N\\ i\end{pmatrix}(1-F(b))^i F(b)^{N-i}}{\alpha\gamma^{-N}(1+(\gamma-1)F(b))^N},\\
&=\sum_{i=0}^{N}\begin{pmatrix}N\\ i\end{pmatrix}(\gamma F(x))^{N-i}\frac{(1-F(b))^i}{(1+(\gamma-1)F(b))^N},\\
&=\frac{(1-F(b)+\gamma F(x))^N}{(1+(\gamma-1)F(b))^N}.
\end{aligned}
$$

Proof of Proposition 2

$$
\begin{aligned}
\mathbb{E}[c|b] &= \mathbb{E}[\delta\underline{b}|b],\\
&= \Pr\{\delta=1|b\}\,\mathbb{E}[\underline{b}|b,\delta=1],\\
&= \alpha\gamma^{-N}[1+(\gamma-1)F(b)]^N\int_0^b \underline{b}d\left(\frac{1-F(b)+\gamma F(\underline{b})}{1+(\gamma-1)F(b)}\right)^N,\\
&= \alpha\gamma^{-N}\int_0^b \underline{b}d[1-F(b)+\gamma F(\underline{b})]^N,\\
&= \alpha\gamma^{-N}\left(b[1+(\gamma-1)F(b)]^N-\int_0^b[1-F(b)+\gamma F(\underline{b})]^N d\underline{b}\right). \quad \text{(Integrating by parts)}
\end{aligned}
$$

Proof of Equation 3.13

$$
\begin{aligned}
\frac{d\mathbb{E}[c|b]}{db} &= \alpha\gamma^{-N}\Big([1+(\gamma-1)F(b)]^N+N(\gamma-1)bf(b)[1+(\gamma-1)F(b)]^{N-1}\\
&\quad -[1-F(b)+\gamma F(b)]^N.1+Nf(b)\int_0^b[1-F(b)+\gamma F(x)]^{N-1}dx\Big),\\
&= \alpha N\gamma^{-N}f(b)\Big((\gamma-1)b[1+(\gamma-1)F(b)]^{N-1}+\int_0^b[1-F(b)+\gamma F(x)]^{N-1}dx\Big).
\end{aligned}
$$

Proof of Proposition 4

Let $h_N(b)=\int_0^b[1-F(b)+\gamma F(x)]^N dx$, $g_N(b)=[1+(\gamma-1)F(b)]^N$ and $\Psi(b)=b+h_{N-1}(b)/((\gamma-1)g_{N-1}(b))$. If $\Psi(b)$ is monotonically increasing then there is a unique b^*

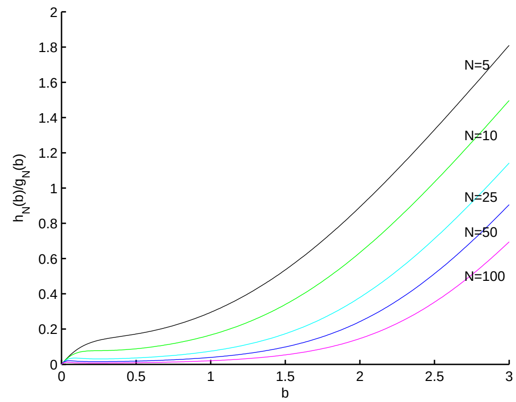

Figure 11: $h_N(b)/g_N(b)$ v/s b assuming the competitors bids are Weibull($\lambda = 1.59, \theta = 1.37, \gamma = 1.42$). The ratio $h_N(b)/g_N(b)$ decreases as N increases.

that satisfies the optimality condition (Equation 3.14).

$$\begin{aligned}
\Psi(b) &= b + \frac{h_{N-1}(b)}{(\gamma-1)g_{N-1}(b)} \\
\Psi'(b) &= 1 + \frac{h'_{N-1}(b)}{(\gamma-1)g_{N-1}(b)} - \frac{h_{N-1}(b)g'_{N-1}(b)}{(\gamma-1)g^2_{N-1}(b)}, \\
&= \frac{g_{N-1}(b)[(\gamma-1)g_{N-1}(b) + h'_{N-1}(b)] - h_{N-1}(b)g'_{N-1}(b)}{(\gamma-1)g^2_{N-1}(b)}.
\end{aligned}$$

$\Psi'(b) > 0$ if $g_{N-1}(b)[(\gamma-1)g_{N-1}(b) + h'_{N-1}(b)] - h_{N-1}(b)g'_{N-1}(b) > 0$, or

$$\begin{aligned}
\gamma[1+(\gamma-1)F(b)] &> (N-1)f(b) \times \\
&\quad \left[(\gamma-1)\frac{\int_0^b [1-F(b)+\gamma F(x)]^{N-1}dx}{[1+(\gamma-1)F(b)]^{N-1}} + \frac{\int_0^b [1-F(b)+\gamma F(x)]^{N-2}dx}{[1+(\gamma-1)F(b)]^{N-2}}\right], \\
&= (N-1)f(b)\left[(\gamma-1)\frac{h_{N-1}(b)}{g_{N-1}(b)} + \frac{h_{N-2}(b)}{g_{N-2}(b)}\right].
\end{aligned}$$

We can show that the ratio $h_N(b)/g_N(b)$ is decreasing in N implying $h_{N-2}(b)/g_{N-2}(b) \geq h_{N-1}(b)/g_{N-1}(b)$ for all $N \geq 2$. This intuition is illustrated in Figure (11) for a sample distribution. It can be seen that $h_N(b)/g_N(b)$ decreases as N is increased.

This implies that $\Psi'(b) > 0$ if (write substituting)

$$\begin{aligned}\gamma[1+(\gamma-1)F(b)] &> \gamma f(b)(N-1)\frac{h_{N-2}(b)}{g_{N-2}(b)},\\ \text{or}\quad \gamma &> 1+\frac{1}{F(b)}\left[f(b)(N-1)\frac{h_{N-2}(b)}{g_{N-2}(b)}-1\right]\end{aligned}$$

If the rate of decay of the *ctr* with respect to position (γ) is high enough, then there exists a unique b^* that satisfies the optimality condition. For some common distributions like the Weibull, Gamma and Log-Normal we numerically find that $\Psi(b)$ is always increasing in b and there exists a unique bid for every keyword k that satisfies the optimality condition. This is illustrated in Figure (12) for some sample parameters.

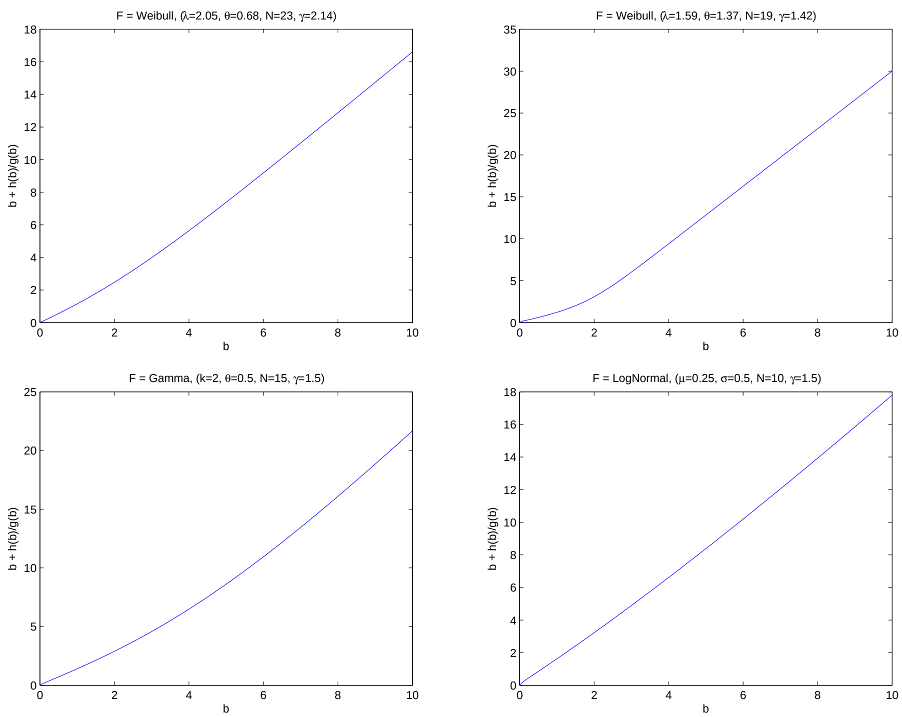

Figure 12: $\Psi(b)$ for various distributions.

Measuring Competitive Reaction

If there is competitive reaction then the predicted average position and CPC would be considerably different from the observed position or CPC as the competitors might change their bids as a response to the changes in bids by the advertiser. If the predicted and observed moments of these quantities are not very different, it suggests that the competitive reaction is subdued. In order to measure competitive reaction, we compute the difference between the predicted daily average position and *cpc* and the mean of these quantities. The MAE is reported in the table below.

Table 17: MAE between the predicted and observed moments

Quantity	MAE
position	0.141
cpc	$0.064

Given that these observed quantities are very close to the predicted values, this provides evidence to suggest that there is very weak competitive reaction during the experimental phase.

Appendix A3: Estimation of DLM parameters

This appendix provides an overview of the sampling procedure used to estimate parameters of the Dynamic Linear Model mentioned in Section 6.1. The sampling procedure mentioned here is an application of the method proposed by West and Harrison (1997). We need to estimate the parameters of the transition matrix (η_A,η_B,η_G, γ_G,γ_B), the effect of generic and branded clicks (β_G,β_B), the covariance matrices (V_ε,V_ν) and the sequence of state vectors $\Phi_T = \{\phi_1, \ldots, \phi_T\}$. We start off with non-informative Gaussian priors for these parameters Ψ =(η_A,η_B,η_G, γ_G,γ_B,β_G,β_B). We also assume that ε_t and ν_t are independent and the priors for V_ε and V_ν are assumed to be inverse Wishart. Given these assumptions, the posteriors distributions of V_ε and V_ν are inverse Wishart and the posteriors for the parameters (η_A,η_B,η_G, γ_G,γ_B,β_G,β_B) are Gaussian.

Let $\boldsymbol{D}_t = \{Y_t, \boldsymbol{D}_{t-1}\}$ denote all the information available to the researcher till time t, e.g. the clicks and impressions till time t. We use a forward-filtering and backward smoothing algorithm (e.g. Rutz and Bucklin, 2011) to sample the state spaces, $\Phi_t|\boldsymbol{D}_t$. Then we sample the parameters (η_A,η_B,η_G, γ_G,γ_B,β_G,β_B) given Φ_t and $\boldsymbol{D}_t$. These estimation steps are described below.

Step 1: Simulation for Φ_T

i) For $t = 1, \ldots, T$, compute m_t and Σ_t, the mean and the variance of the state space at time t. m_t and $\Sigma_t$are derived sequentially from the priors m_0 and Σ_0 according to the procedure outlined in West and Harrison (1997, Chapter 4).

ii) Filter-forward step: For $t = T$, sample $p(\phi_T|\boldsymbol{D}_T)$ from the posterior distribution $N(m_T, \Sigma_T)$.

iii) Backward-smoothing step: For $t = T, \ldots, 1$, sample p$(\phi_{t-1}|\phi_t, \boldsymbol{D}_T)$ conditional on the latest draw ϕ_t.

Step 2: Sampling from $p(\Psi, V_\varepsilon, V_\nu|\Phi_T, D_T)$

We sample the parameters Ψ, V_ε and V_ν sequentially. This is reasonable as the elements of the transition matrix, drift vectors and the error terms are assumed to be independent of each other. Based on these assumptions the Gibbs sampler can be used in a straight-forward manner to draw samples of η_A, η_B, η_G, γ_G, γ_B, β_G, β_B, V_ε and V_ν separately.

CHAPTER 4 : The Long Road to Online Conversion: A Model of Multi-Touch Attribution

4.1. Introduction

Online advertising has enabled advertisers to target their consumers using various channels like Hulu, Yahoo, nytimes.com, search engines or social networks. The PEW Internet report estimates the total online advertising expenditure to be US$ 40 billion by the year 2012. Although the Internet surpassed print and radio as an advertising medium in 2011 and currently accounts for 13-19% of the advertising spend, there continues to be tremendous room for growth given the amount of time consumers spend online. Online advertising has been embraced by a large number of advertisers as it not only allows for very granular targeting but is also extremely quantifiable, enabling advertisers to measure the impact of their advertising dollars. Although there are various forms of digital advertising like email marketing, social media and mobile marketing, it continues to be dominated by display and search advertising which account for almost $\frac{3}{4}$ of this expenditure. Search advertising comprises ads shown on search engines whereas display advertising consists of banner ads and emerging video formats on websites like You tube and Hulu.

A typical advertiser uses different channels and various ad formats to convey his message to consumers. This can include both traditional channels like television, newspapers, direct mailing or digital channels like sponsored search, display or social ads. In this paper, we focus primarily on the digital channels. Since the advertisers uses multiple channels for advertising, a consumer can be exposed to several different ads during his browsing sessions. These repeated interactions with an advertiser's campaign are termed "multi-touch" in popular press (Kaushik, 2012). When the user buys a product or signs up for a service (converts), his decision is influenced by prior ad exposures as shown in Figure 1. The advertiser wants to ascertain which ads across the different channels has an influence on the consumer's decision and to what extent. This problem of quantifying the influence

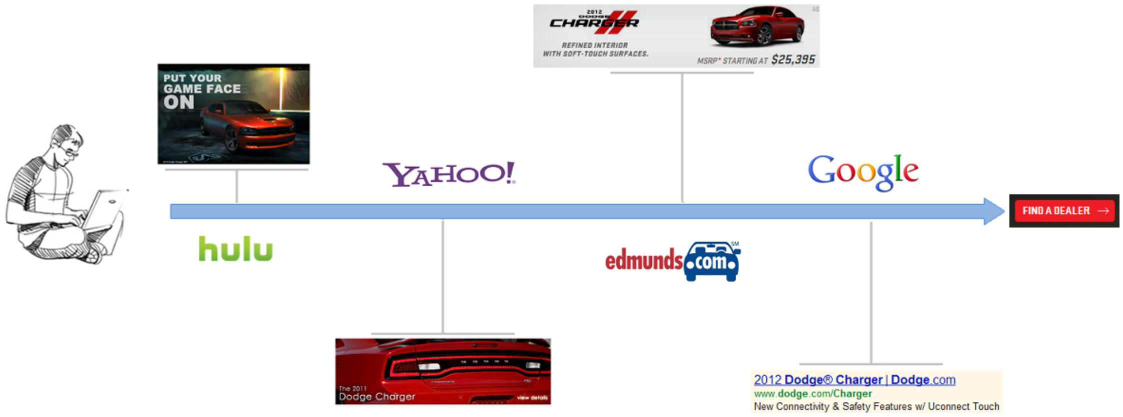

Figure 13: Multiple ad exposures across different online channels.

of each ad on the consumer's decision is referred to as the attribution problem. Once the advertiser can measure the contribution of each ad, he can use this information to optimize his ad spending.

Online channels offer a unique opportunity to address the attribution problem as advertisers have disaggregate individual level data which was not available in the case of traditional channels like television and newspapers. Given the lack of disaggregate data, the marketing literature has focused primarily on the marketing mix models (Naik et al., 2005; Ansari et al., 1995; Ramaswamy et al., 1993) which perform inter-temporal analysis of marketing channels but fail to provide insights at an individual customer level. Online advertising allows advertisers to not only observe the ads that a consumer was exposed to but also when the exposure took place. This granular data can be used to build rich models of consumer response to online ads. Unfortunately, there is very little academic research that analyzes multi-channel advertising data or addresses the problem of attribution.

In the absence of appropriate techniques, marketers have adopted rule based techniques like last-touch attribution (LTA), which assigns all the credit for a conversion to the click or impression that took place right before the conversion. Although the LTA is commonly used in the advertising industry, it completely ignores the influence that ads before the last clicked (or viewed) ad had on the consumer's decision. This causes ads that appear much earlier in the conversion funnel, e.g. display ads, to receive much less credit and ads that occur closer to the conversion event, e.g. search ads, to receive most of the credit for the conversion event. A consumer might have started down the path of conversion after being influenced by a display ad, but the LTA would suggest that the display ad had no impact on the consumer's decision. Incorrect attribution methods might move advertising dollars away from important channels and have a detrimental impact on the advertiser's revenues in the long term. This is evident in the case of display ads as the lack of appropriate attribution models have led advertisers to believe that display advertising is not very effective, thereby hindering the growth of this format of advertising. It should be noted that incorrect measurement also alters the publishers incentives (Jordan et al., 2011). If an ad displayed by the publisher is undervalued, she might be incentivized to display "seemingly" more profitable ads. This not only has an adverse effect on the advertiser but also increases the inefficiency in the marketplace.

Some heuristics have been proposed to address the problems associated with LTA, e.g. first-touch attribution or exponentially weighted attribution, but these techniques are plagued with similar problems and do not take a data-driven approach to address the issue of attribution. In the past few years, as several online channels have gained importance, most advertisers have come to realize the inadequacies associated with their current methodologies (Chandler-Pepelnjak, 2009; Kaushik, 2012). Developing an appropriate advertising attribution model is one of the biggest challenges facing the online advertising industry (Quinn, 2012; Khatibloo, 2010; New York Times, 2012; Szulc, 2012). In recent years, companies like Microsoft, Adometry and Clear Saleing have proposed heuristics that address this issue, but there is no clear consensus on which approach is the most appropriate. Surprisingly, there is very little academic research on this problem given its managerial relevance. There are some recent papers that adopt a more rigorous data-driven approach (Shao and Li, 2011; Dalessandro et al., 2012). Shao and Li (2011) propose a simple probabilistic model that solves the attribution problem using a combination of first and second order conditional probabilities. Dalessandro et al. (2012) formulate multi-touch attribution as a causal estimation problem and present a general model for multi-touch attribution. Jordan et. al. (2012) approach this problem from a mechanism designer's perspective and analytically devise an allocation and pricing rule for these ads.

In this paper we propose a model for online ad-attribution using a dynamic hidden Markov model (HMM). We present a model of individual consumer behavior based on the concept of a conversion funnel that captures a consumer's deliberation process. The conversion funnel is a model of a consumer's search and purchase process that is commonly used by marketeers (Kotler and Armstrong, 2011). A consumer moves in a staged manner from a dormant state to the state of conversion and ads affect the movement through the different stages. This model is estimated using a unique dataset from a car manufacturer that contains all the online advertising data from the beginning of a campaign. We observe that different ad formats, e.g. display and search ads, affect the consumers differently and in different states of their decision process. Display ads usually have an early impact on the consumer, moving him from a state of dormancy to a state where he is aware of the product and it might enter his consideration set. However, when the consumer actively interacts with these ads (e.g. by clicking on them), his likelihood to convert considerably increases. Secondly, we present an attribution scheme based on the proposed model that assigns credit to an ad based on the incremental impact it has on the consumer's probability to convert. This method is subsequently compared to the LTA scheme.

The paper is organized as followed. We discuss the relevant literature in Section 2 and position our research in the existing literature. In Section 3, we describe the data that we use in our empirical application. We present the dynamic HMM in Section 4 and discuss the estimation technique. The empirical results are present in Section 5. In section 6, we discuss some limitations of our model. We finally conclude in Section 7 with directions for future research.

4.2. Prior Literature

There has been significant managerial interest in the attribution problem, but the academic literature in this area has been sparse particularly due the absence of suitable multi-channel data. Access to these data has led to recent work by Shao and Li (2011) and Dalessandro et al. (2012) who adopt a data-driven strategy to address the problem of attribution. Shao and Li (2011) develop a bagged logistic regression model to predict how ads from different channels lead to a conversion. This model is further used to estimate an advertising channel's contribution towards a conversion. They also propose another probabilistic model based on a combination of first and second-order conditional probabilities to directly quantify the impact of an advertising channel on the conversion decision. In their models, an ad has the same effect whether it was the first ad that the consumer saw or the tenth ad he saw, which is clearly not a reasonable assumption. Dalessandro et al. (2012) extend this research by incorporating the sequence of ads that lead a consumer to his final decision. They use a logistic regression similar to Shao and Li (2011) to construct a mapping from advertising exposures to conversion probability, however while performing the attribution, they consider how an ad incrementally alters the transition probability conditional on all the prior exposures. These papers are statistically motivated and do not incorporate a model that underlies observed consumer behavior. As a result they might not be able to capture the different stages of a consumer's deliberation process and the varied susceptibility to advertising activities in these stages. In this paper, we try to extend this literature by incorporating well established theories from the information processing literature (Bettman et al., 1998; Howard and Sheth, 1969; Hawkins et al., 1995). This literature suggests that consumer decision making involve a multi-stage process of – (i) awareness, (ii) information search, (iii) evaluation, (iv) purchase and finally (v) post-purchase activity (Jansen and Schuster, 2011). More specifically, we base our model of consumer behavior on the conversion funnel which is commonly used in practice (Mulpuru, 2011; Court et al., 2009) and analyzed in the marketing literature (Strong, 1925; Howard and Sheth, 1969; Barry, 1987).

Our research is closely related to the literature on online advertising (Tucker, 2012; Goldfarb and Tucker, 2011; Ghose and Yang, 2009; Agarwal et al., 2011). Most of the work in this area has focused on sponsored search where researchers have tried to analyze what factors affect consumer behavior (Rutz et al., 2012; Ghose and Yang, 2009) and firm profitability (Agarwal et al., 2011; Ghose and Yang, 2009). More recently, researchers have turned to other forms of advertising like display (Goldfarb and Tucker, 2011) and facebook ads (Tucker, 2012). Goldfarb and Tucker (2011) show that matching an ad to the website content and increasing an ad's obtrusiveness independently increase purchase intent. However, in combination, these two strategies negate each other due to privacy concerns. Tucker (2012) investigates how users' perception of control over their personal information affects their likelihood to click ads on Facebook. She shows that with an increase in privacy controls, users are twice as likely to click on these ads. Although there is a lot of research on different formats of online advertising, researchers haven't looked at how these ads interacts in a multi-channel context. This paper tries to address this gap in the extant literature and proposes a model to a gain a better understanding of consumer response to different types of online ads.

From a methodological viewpoint, our research belongs to the extensive literature on HMMs in computer science (Rabiner, 1989) and more recently in marketing (Netzer et al., 2008; Montoya et al., 2010; Schwartz et al., 2011). HMM is a workhorse technique in computer science that has been applied to various application like speech recognition (Rabiner, 1989), message parsing (Molina and Pla, 2002) and facial recognition (Nefian et al., 1998) among other things. In the marketing literature, HMMs are used to capture dynamic consumer behavior when the consumer's state is unobservable (Netzer et al., 2008; Schweidel et al., 2011). HMM have been used to study physicians' prescription behavior (Montoya et al., 2010), customer relationships (Netzer et al., 2008) and online viewing behavior (Schwartz et al., 2011). Most of the papers in the literature incorporate time varying covariates to account for marketing actions, e.g. Montoya et al. (2010) analyze how detailing and sampling activities can move physicians from one state to another and alter their propensity to prescribe a newly introduced medicine. We adopt a similar approach in our paper to model the dynamics of the HMM.

4.3. Data Description

Our data is provided by a large digital advertising agency that managed the entire online campaign for a car manufacturer. This data spans a period of around 11 weeks from June 8, 2009 to August 23, 2009. The ad agency promoted display ads on several generic websites like Yahoo, MSN and Facebook and auto-specific websites like KBB and Edmunds. In addition, it also advertised on search engines like Google and Yahoo. Users are tracked across the different advertising channels and on the car manufacturer's website using cookies. The context of car sales is very relevant to the attribution problem as consumers spend a lot of time researching cars online, sometimes several weeks and as a consequence are exposed to ads in various format, across different online channels.

This dataset is unique as it contains all the display and search advertising data at an individual level since the start of the campaign. Our sample comprises a panel of 6432 randomly chosen users with a total of 146,165 observations. An observation in our dataset comprises a display ad impression or click (generic/specific), a search click or activity (page view/conversion) on the advertiser's website. We do not observe the search ads that were shown to consumers (as this data is not reported by the search engine), however when a consumer clicks on one of these ads and arrives at the advertiser's website, this click is recorded in our data and referred to as a search click. A conversion in this data is said to occur when the user performs one of the following activities on the advertiser's website - *search inventory*, *find a dealer*, *build & price* and *get a quote*. We do not focus on the different conversion activities and treat them similarly. Furthermore, as we are interested in how the ads drive the first conversion, we discard all the observations for a particular consumer after the first conversion. Since we are interested in the effect of advertising on the conversion process, we also eliminate users in our data that do not have any ad exposures. This results in a panel size of 5121 users with 112,619 observations. Summary statistics of this data at an individual level is presented in Table 18 below.

	Mean	S.D.
Generic display impressions	13.756	34.725
Generic display clicks	0.072	0.180
Generic click-through rate	0.007	0.054
Specific display impressions	4.211	10.06
Specific display clicks	0.143	0.32
Specific click-through rate	0.020	0.062
Search clicks	0.246	0.719
Web pages viewed	3.471	8.187
Conversions	0.152	0.359

Table 18: Summary Statistics

On an average there are 13.756 display impressions per customer on generic websites and 4.211 impressions on auto-specific websites. Consumers click 0.007 of these display ads on generic websites and 0.143 on auto-specific websites. We see that the click-through rate for display ads on auto-specific websites is much higher than on generic websites which indicates that context plays an important role in the consumer's click-through and decision making process. Consumers browse 3.471 pages on the car manufacturer's website in this dataset. Most of ads in this campaign are "call to action" ads, which explains the high conversion rate – 15.2% of all the consumers in this dataset end up engaging in one of the four conversion activities mentioned earlier.

4.4. Model of Multi-Touch Attribution

In this section we first present a HMM of consumer behavior and then show how this model can be used to solve the attribution problem

4.4.1. The Conversion Funnel

Our model is inspired by the idea of a conversion funnel that has been at the center of the marketing literature for several decades (Strong, 1925; Howard and Sheth, 1969; Barry, 1987). The conversion funnel is also widely adopted by practitioners and managers who frequently base their marketing decision on the conversion funnel (Mulpuru, 2011; Court et al., 2009). The conversion funnel is grounded in the information processing theory which postulates how consumers behave while taking a decision (Bettman et al., 1998). This literature suggests that consumers move through different stages of deliberation during their purchase decision process. Several marketing actions, e.g. advertising, help the user in moving closer to the end goal, i.e. an eventual purchase. This framework is also similar to the AIDA (attention, interest, desire and action) model (Kotler and Armstrong, 2011) and hierarchy of effects model (Bruner and Kumar, 2000) that are commonly used in marketing.

Several variants of the conversion funnel have been proposed, but the most commonly used funnel has following stages - *awareness, consideration* and *purchase* (Jansen and Schuster, 2011; Mulpuru, 2011; Court et al., 2009). A consumer is initially in a dormant state when he is unaware of the product or is not deliberating a purchase. When he is exposed to an ad, he might move into a state of awareness. Subsequently, if he is interested in the product, he transitions to a consideration stage where he engages in information seeking activities like visiting the website of the advertiser and reading product reviews (this is sometimes referred to as the *research* stage in the purchase funnel). Finally, based on his consideration, the consumer decides to engage in the conversion event or not. In the following discussion, we introduce a parsimonious model that captures the dynamics of the conversion funnel.

Although the conversion funnel is widely accepted and used, it has been difficult to analyze the movement of a consumer down the funnel in the context of traditional advertising. Most of the data in traditional advertising is available at an aggregate level which makes it difficult to tease apart the different stages of the consumer deliberation process outlined earlier. The individual level data presented in Section 3 offers a unique opportunity to analyze the consumer behavior at a much granular level and examine the conversion funnel using observational data.

4.4.2. Hidden Markov Model

In our data, we do not observe a consumer's underlying state and it can be inferred only through the consumer's observable actions, i.e. website visits and conversion. In this sense the consumer's state is latent and his progression through the conversion funnel is hidden. In this paper, we use a HMM to capture the user's deliberation process and his movement down the conversion funnel as a result of the different ad exposures he experiences. Several researchers have uses HMMs to model latent consumer states (Montoya et al., 2010; Netzer et al., 2008; Schwartz et al., 2011; Schweidel et al., 2011) and they are particularly suited for the problem of attribution as we explain in the next section.

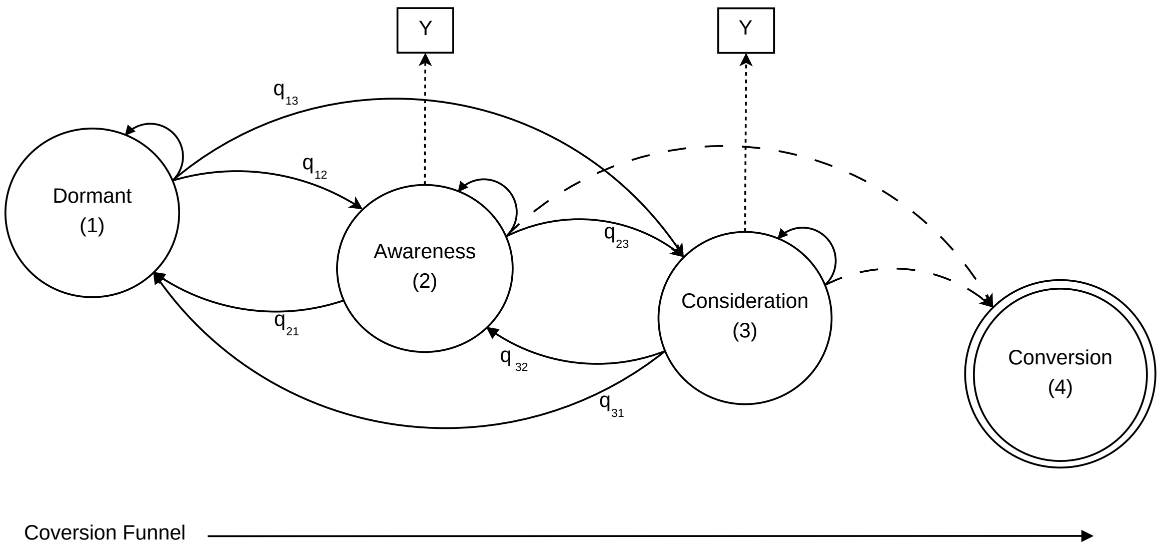

Figure 14: Diagram representing the latent states and the outcomes of the HMM. $q_{ss'}$ denote the transition probabilities from state s to state s' and Y_s is the binary random variable that captures conversion in state s.

In accordance with the conversion funnel, we construct an HMM with four states (S), where the four states are "dormant", "awareness", "consideration" and "conversion" (Figure 14). At any time t, consumer i can be in one of the four states, $S_{it} \in S$.[1] As mentioned earlier, we do not observe s_{it}, but we observe the bivariate outcome variable $Y_{it} = (N_{it},\ C_{it})$ which arises from a stochastic process conditional on the state S_{it}. N_{it} is a Poisson random variable that denotes the number of pages viewed by the consumer between time t and $t+1$ and C_{it} is a binary random variable which captures whether there was a conversion between time t and $t+1$. When the user is in a dormant state, he is unaware of the product or

[1]Variables in uppercase denote random variables and variable in lowercase denote their realizations. In addition, set notation supersedes notation for random variables unless otherwise noted.

is not deliberating a purchase. In this state, there is no activity from the consumers and the outcomes variables, page views and conversion, are set to zero. As the consumer is exposed to different ads, he might move into a state of awareness where he knows about the product and might be willing to purchase it. On further deliberation, he moves into a consideration state, where he can actively look for product related information and engage with the firm's website. Consumers can also go directly from the dormant state to the state of consideration. In this model, the *research* stage is implicitly captured by the consumer's interaction with the advertiser's website (measured through page views). As we model only the *first* conversion behavior of the consumer, the consumer moves into the "conversion" state as soon as a conversion occurs. "Conversion" is a dummy absorbing state which captures the fact that once a consumer has engaged in a conversion activity, he ceases to exist in our data.

We assume that a consumer's propensity to purchase (or convert) is zero in the dormant state and it steadily increases as he moves down the different states. We also assume that the consumer's research behavior becomes more intense as he moves down the funnel, e.g.. he is likely to visit the advertiser's website more often when he is in the consideration state as opposed to the awareness state. The transition between the states take place in a stochastic manner when an ad event a_{it} occurs and is influenced by the firm's advertising activities so far. Ads from different channels can have different effects on these transitions and these effects can be state specific. The transitions between the different states also follow a Markov process, i.e. the transitions out of a particular state depend only on the current state and not on the path that the user took to get to the state. Let $A_i = \{a_{i1}, a_{i2}, \dots, a_{iT}\}$ denote a sequence of T ad events that consumer i is exposed to, due to which the consumer ends up in states $S_i = \{S_{i1}, S_{i2}, \dots, S_{iT}\}$. $\boldsymbol{x}'_{it}$ captures the running sum of the different kinds of advertising activities till time t and contains covariates like number of display impressions at a generic website, number of display impressions at an auto-specific website and search clicks. We do not observe S_i but observe the observation vector $Y_i = \{Y_{i1}, Y_{i2}, \dots, Y_{iT}\}$. The joint probability of observing the sequence of observations $\{Y_{i1} = \boldsymbol{y}_{i1}, \dots, Y_{iT} = \boldsymbol{y}_{iT}\}$

is a function of three main components: (i) the transition probabilities between the different states – Q_{it}, (ii) the distribution of the observational variables conditional on the state – M_{it} denotes the probability of conversion and $N_{it} \sim Poisson(\lambda_{its})$, and (3) the initial state distribution – π. We describe each of these components in detail below.

Markov Chain Transition Matrix

In our model, there might be a transition from the current state s_{it} only under two conditions - (i) when a consumer is exposed to an ad event a_{it}, or (ii) when a conversion takes place and the consumer moves to the "conversion" state with certainty. If the transition occurs due to an ad event, consumer i's transition from one latent state to another is stochastically based on the transition matrix Q_{it} which is a function of the advertising activities, $\boldsymbol{x}'_{it}$ at time t. The probability that a consumer transitions to state s' at time $t+1$ conditional on him being in state s at time t is given by $P(S_{it} = s' | S_{it-1} = s) = q_{itss'}$. Let T_s be the set of states (s') that can be reached from state s. The elements of the transitions matrix specific to state s are given by

$$q_{itss'} = \frac{\exp\{\boldsymbol{x}'_{it}\boldsymbol{\beta}_{ss'}\}}{1 + \sum_{s' \in T_s} \exp\{\boldsymbol{x}'_{it}\boldsymbol{\beta}_{ss'}\}} \quad \forall\, s' \neq s, \tag{4.1}$$

$$q_{itss} = \frac{1}{1 + \sum_{s' \in T_s} \exp\{\boldsymbol{x}'_{it}\boldsymbol{\beta}_{ss'}\}}, \tag{4.2}$$

where $\boldsymbol{\beta}_{ss'}$ is the response parameter that captures how the advertising related activities affect the consumer's propensity to transition from state s to s'. $\boldsymbol{\beta}_{ss'}$ is different across states as the advertising activities $\boldsymbol{x}'_{it}$ might have different effects on the transition based on the receiving state. For e.g. display clicks might affect the transition to the "dormant" state differently than the transition to the "consideration" state.

Consumer Research and Conversion Behavior

For every consumer, the bivariate outcome variable $Y_{it} = (N_{it},\ C_{it})$ is modeled in the following manner.

Modeling page views: N_{it} is drawn from a Poisson distribution with a rate parameter λ_{its} which is a function of the current state s and advertising activity, $\boldsymbol{x}_{it}$. The probability of observing n_{it} page views is given by

$$P(N_{it} = n_{it}|S_{it} = s) = \frac{\lambda_{its}^{n_{it}} e^{-\lambda_{its}}}{n_{ist}!},$$

where $\lambda_{its} = \widetilde{\eta}_s + \boldsymbol{x}'_{it}\tau_s$, i.e. the rate parameter is the a function of the intrinsic research activity in state s and the time varying covariates $\boldsymbol{x}_{it}\tau_s$. Note that there is no research activity in the dormant state, therefore $\lambda_{it1} \equiv 0$. We also assume that the research intensity increases as the consumer moves down the conversion funnel. This constraint is enforced by setting

$$\begin{aligned}
\widetilde{\eta}_2 &= \eta_2, \\
\widetilde{\eta}_3 &= \widetilde{\eta}_2 + \exp\{\eta_3\},
\end{aligned}$$

where η_2 and η_3 are parameters to be estimated from the data.

Modeling conversions: The consumer's probability to convert depends on the state in which he is present. We follow Montoya et al. (2010) in modeling the conversion C_{it} which is binary random variable. The conditional probability $P(C_{it} = 1|S_{it} = s) = m_{its}$ is given by

$$m_{its} = \frac{\exp\{\tilde{\alpha}_s + \boldsymbol{z}'_{it}\gamma_s\}}{1 + \exp\{\tilde{\alpha}_s + \boldsymbol{z}'_{it}\gamma_s\}}.$$

$\boldsymbol{z}'_{it}$ a vector of time varying covariates which contains the advertising related activities in addition to the number of web pages the consumer has viewed on the advertiser's website. The number of page views are included in addition to the marketing activities because a consumer might be more likely to convert if he has viewed more web pages and has gathered more information about the product. γ_s captures how these covariates affect the conversion probability. We assume that there are no conversions in the dormant state ($m_{it1} = 0$) and the probability to convert, on average, increases as we move down the conversion funnel.

This assumption is operationalized in the following manner,

$$\begin{aligned}\tilde{\alpha}_2 &= \alpha_2, \\ \tilde{\alpha}_3 &= \tilde{\alpha}_2 + \exp\{\alpha_3\},\end{aligned}$$

where α_2 and α_3 are the parameters to be estimated from the data. This structure enforces that $m_{it3} \geq m_{it2}$, *ceteris paribus.* This assumption ensures the identification of the different states and is consistent with the approach adopted by Netzer et al. (2008) and Montoya et al. (2010).

Joint density: In our model we also assume that N_{it} and C_{it} are independent once the effect of N_{it} on $\boldsymbol{z}_{it}$ has been accounted for. Hence, the conditional probability of observing $\boldsymbol{y}_{it}$ is given by

$$P(Y_{it} = \boldsymbol{y}_{it}|S_{it} = s) = m_{its}^{c_{it}}(1 - m_{its})^{(1-c_{it})}P(N_{its} = n_{it}|S_{it} = s) \qquad (4.3)$$

where $\boldsymbol{y}_{it} = (n_{it},\ c_{it})'$ is the realized outcome variable.

Initial State Membership

Let π_s denote the probability that consumers are in initially in state s, where $\sum_{s \in S} \pi_s = 1$.[2] Consumers can start out in different states because of their exposure to ads on other media like television or print which can affect the initial membership probability. However, we do not have data about other forms of advertising and hence we assume that all consumers start out in the dormant state and move down the conversion funnel, i.e. initial membership probability is given by $\pi = \{1, 0, 0, 0\}$ which is an assumption we make for the identification of the model as explained in the Appendix. We think this is a reasonable assumption because the advertising campaign pertains to a new brand of cars and consumers might have been completely unaware of the product before the launch of the online campaign.

In summary, the dynamic HMM captures the consumers' behavior as they transition across

[2]Note that the subscript for the consumer is dropped because all consumers are considered homogeneous ex-ante.

the different states of the funnel and eventually convert. This model allows the ads to have an effect on consumers behavior – they affect the transition probabilities as well as the product research and conversion activities. Thus, these ads not only have an immediate impact on the consumers by changing their conversion probabilities, but they can also move consumers to different stages in the conversion funnel which can have an impact on their future conversion behavior. It should be noted that we do not incorporate consumer heterogeneity in our model for simplicity. However, there might be differences in consumers' behavior due to their prior relationship with the brand, offline advertising activity or underlying demographic variables. We extend the current model to account for observed and unobserved consumer heterogeneity in Section 6. Thus, the model allows us to attribute suitable credit to an ad even if it does not contribute to a conversion right away but helps in moving the consumer to a state with higher conversion probability. In this sense, our model differs considerably from the approach adopted by Shao and Li (2011) and Dalessandro et al. (2012) which attribute credit to an ad only when it directly leads to conversion. In the following discussion, we explain how the ad events affect these transitions and how the aforementioned model can be used to solve the attribution problem.

4.4.3. Ad Attribution

When consumer i is exposed to an ordered set of ad related activities $A_i = \{a_{i1}, a_{i2}, \ldots a_{iT}\}$, he moves through the different states of the HMM in the manner described above. a_{it} is a categorical variable that captures the ad related activity the consumer is exposed to, i.e. $a_{it} \in \{$"display impression on generic website", "display impression on auto-specific website", "click on generic website", "click on auto-specific website", "search click"$\}$. The total number of ad related events experienced by a consumer, T, can vary across customers. In our model, the ad related event a_{it} affects the customer i's underlying time varying parameters $\boldsymbol{x}'_{it}$ and $\boldsymbol{z}'_{it}$ as shown below,

$$\boldsymbol{x}_{it-1} \quad \xrightarrow{a_{it}} \quad \boldsymbol{x}_{it},$$

$$\boldsymbol{z}_{it-1} \xrightarrow{a_{it}} \boldsymbol{z}_{it}.$$

Hence, a_{it} has a two fold effect on the consumer's probability to convert which is shown below – (i) it alters the conditional conversion probability, through changes in $\boldsymbol{z}'_{it}$ and (ii) it can lead to a transition of the consumer from one state to another by affecting $\boldsymbol{x}'_{it}$. Attribution in the context of online advertising involves measuring the incremental change a_{it} provides in the conversion propensity. Keeping these factors in mind, we provide two approaches to measure the value of an ad related event. For simplicity we assume that the advertiser earns \$1 whenever a conversion occurs.

Forward-Looking Attribution

Let A_{it} be a subset of A_i which contains the ad events until time $t \leq T$. In the forward-looking approach, V_{it}, the value of an ad event a_{it}, is measured by the effect it has in expectation on the consumer i's conversion decision.

$$\begin{aligned} V_{it} &= \mathbb{E}[C_i | A_{it} = \{a_{i1}, a_{i2}, \ldots, a_{it}\}] - \mathbb{E}[C_i | A_{it-1} = \{a_{i1}, a_{i2}, \ldots, a_{it-1}\}], \qquad (4.4) \\ &= P(C_{it} = 1 | A_{it}) - P(C_{it-1} = 1 | A_{it-1}). \end{aligned}$$

The effect of an ad depends on the consumer's underlying state which in turn is affected by the ads that preceded a_{it}. The value of an ad is not only a function of the impact of the ad going forward, but also depends on the other ad exposures that took place before time t, and the attribution method presented in Equation (4.4) explicitly accounts for the effect of preceding ads. This approach is similar to Shao and Li (2011) and Dalessandro et al. (2012), but $P(C_{it} = 1 | A_{it})$ is estimated using a dynamic HMM in our case whereas they use simplistic approaches like a logistic regression and sample means to compute these probabilities. It should be noted that this method differs vastly from LTA which attributes 100% of the conversion to the last ad event and completely disregards the effects of ads that came earlier. The value ascribed to a specific type of ad event, $k \in \{$"display impression on generic website", ..., "search click"}, can be computed by summing across all ad activities of that type,

$$\Pi_k = \sum_i \sum_{t=1}^{T} \mathbb{1}_{\{a_{it}=k\}} V_{it}, \tag{4.5}$$

where $\mathbb{1}_{\{a_{it}=k\}}$ is an indicator function that equals one if ad event a_{it} is of type k.

Backward-Looking Attribution

The prior approach addresses the problem of attribution based on the marginal effect an ad has on the conversion probability. In this formulation, we measure the effect of an ad conditional on a conversion event. An ad a_{it}'s value in this case is computed by

$$\widetilde{V}_{it} = \frac{P(C_{it} = 1|A_{it}) - P(C_{it-1} = 1|A_{it-1})}{P(C_i = 1|A_i)}.$$

This method attributes *all* the proceeds from a conversion to ads that occurred before the conversion took place, although their relative contribution is weighted by $\widetilde{V}_{it}$. The value assigned to a specific type of an ad event can be computed by

$$\widetilde{\Pi}_k = \sum_i C_i \sum_{t=1}^{T} \mathbb{1}_{\{a_{it}=k\}} \widetilde{V}_{it}, \tag{4.6}$$

where C_i denotes whether we observe a conversion for customer i in the data.

In order to solve the attribution problem, we need to estimate the parameters of the HMM which is presented in the subsequent section.

4.5. Empirical Analysis

In this section, we illustrate how the HMM model can be estimated and interpreted. A richer model of consumer behavior which incorporates consumer heterogeneity is presented in Section 5. The dynamic HMM is estimated on the car campaign data presented in Section 3. We first outline the estimation procedure, briefly discuss the model validity and continue to present the estimated parameters. A sample of 4121 users is used for estimating the model and the remaining 1000 users are used for validation.

4.5.1. Estimation Procedure

Here, we outline the procedure of estimating the HMM on the data shown in Section 3. Our model differs from standard HMMs as the transition probabilities depend on the covariates that vary over time. Several techniques have been proposed to incorporate the time varying covariates in the HMM which are collectively referred to as the latent transition models. Three of the most common techniques used to estimate these models are maximum likelihood estimation (ML), expectation maximization (EM) and Markov chain Monte Carlo (MCMC) methods. In one step maximum likelihood approaches like Newton Raphson, the entire joint likelihood of observing the data is maximized to arrive at the parameter estimates (Satten and Longini, 1996; Cooper and Lipsitch, 2004). The EM techniques involves two steps - in the first step, the likelihood of observing the data is derived conditional on the parameters (expectation step), and subsequently the parameters are updated to fit the data better (maximization step) (Baum et al., 1970; Bureau et al., 2003). A third step is incorporated to estimate the effect of the time varying covariates on the transitions between the states and the observations associated with these states (Visser et al., 2009). More recently, some researchers have adopted a Bayesian approach, where the parameters are drawn from a posterior distribution using MCMC simulations. The Bayesian approach allows for easy model diagnostics using the posterior predictive checks (Berkhof et al., 2000; Schwartz et al., 2011) and also provides an easy way to compute standard errors for the parameters which might be difficult in the ML and EM techniques under certain conditions (Chung et al., 2007). MCMC techniques have also been employed extensively in the latent state literature in marketing (Montoya et al., 2010; Netzer et al., 2008). EM and MCMC techniques are generally used for their computational speed and efficiency. Computing the joint probability of the observed data can be computationally intensive for large values of T. However, maximum likelihood approaches like Newton Raphson have better convergence properties as compared to the EM and MCMC techniques which tend to converge to local minimas. As T is relatively small in our data ($\overline{T} = 18.3$) and computational complexity is not a big concern, we adopt the maximum likelihood approach to estimate our parameters.

We begin by deriving the likelihood of observing the data. Given a sequence of ad events A_i, the consumer can take several different paths $s_0 \rightarrow s_1 \rightarrow \ldots \rightarrow s_T$. The sequence of the states during this transition determines the probability of the observations $y_i = \{\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \ldots \boldsymbol{y}_{iT}\}$. The likelihood of a matrix $(2 \times T)$ of outcome variables y_i after being exposed to these actions A_i can be computed by evaluating the probabilities of each of these path $s_0 \rightarrow \ldots \rightarrow s_T$ and the conditional probability of $P(Y_{i1} = \boldsymbol{y}_{i1}, \ldots, Y_{iT} = \boldsymbol{y}_{iT}, | S_0 = s_0, \ldots, S_T = s_T)$ which is given by

$$L_i = \sum_{s_1=1}^{|S|} \sum_{s_2=1}^{|S|} \cdots \sum_{s_T=1}^{|S|} \left[\prod_{t=1}^{T} P(S_{it} = s_t | S_{it-1} = s_{t-1}) \prod_{t=1}^{T} P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t) \right], \quad (4.7)$$

where $P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t)$ can be computed as shown in Equation (4.3). This approach of summing over all possible paths has a complexity $O(|S|^T)$ and might be computationally infeasible even for moderately small values of $|S|$ and T (Cooper and Lipsitch, 2004). In order to overcome this computational complexity, McDonald and Zucchini (1997) propose an approach that significantly reduced the amount of computation required. Let

$$\Phi_{it}(y_{it}) = Diag(P(Y_{it} = \boldsymbol{y}_t | S_{it} = 1), \ldots, P(Y_{it} = \boldsymbol{y}_t | S_{it} = |S|)),$$

where $P(Y_{it} = \boldsymbol{y}_t | S_{it} = s)$ is as mentioned in Equation (4.3). The likelihood of the observed data (Equation (4.7)) can be simplified to

$$L_i = \boldsymbol{\pi'} \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}).\mathbf{1}, \quad (4.8)$$

where $\mathbf{1}$ is a $1 \times |S|$ vector of ones. This computation is significantly much faster and can be evaluated in $O(T|S|^2)$ time. The log-likelihood of observing the entire data is given by the sum of the log-likelihood across all consumers in the data,

$$LL = \sum_i \log\left[\boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\dots Q_{iT}\Phi_{iT}(y_{iT}).\mathbf{1}\right]. \tag{4.9}$$

The parameters of the model can be estimated by maximizing the log-likelihood function shown in Equation (4.9) using any non-linear optimization technique. We use the ClsSolve routine from TOMLAB to perform the unconstrained numerical optimization.

4.5.2. Model Validity

To test the validity of our model, we compare the predictive ability of the model with the commonly used logit model (Dalessandro et al., 2012; Ghose and Yang, 2009; Agarwal et al., 2011). The comparison is performed using two different approaches. First, we compare the log-likelihood of observing the outcome (conversion/no conversion) in the validation dataset. Secondly, we compute the root mean-squared error (RMSE) by calculating the difference between the observed outcome and the predicted outcome from the two models. These results are presented in Table 19 below. We observe that the HMM considerably outperforms the logit model on both these measures.

Table 19: Predictive validity

	Validation log-likelihood	RMSE
HMM	-1692.7	0.097
Logit	-2039.5	0.182

4.5.3. Parameter Estimates

Estimates of the Transition Parameters

Estimates of the transition parameters are reported in Table 20. The intercept terms are significantly negative which indicates that these states are relatively sticky and consumers do not easily transition between them. We also observe that ad related activities have a significant impact on the transition from the dormant state to the awareness state. Contrary to popular belief that display ads are ineffective (de Vries, 2012; Claburn, 2012), we see that display ads have an important effect of moving consumers from a dormant state to a state of awareness. They might not have a high conversion rate, but our model predicts that these ads significantly impact the consumer's deliberation process. This finding has significant implications for marketeers as they need to understand that display advertising has an indirect effect on conversions and they should account for this difference (as compared to search ads) in their attribution approach. In addition, display ads on auto-specific websites have a larger impact on the transition from the dormant state to the awareness state. Consumers might be more likely to notice these car related ads when they are visiting auto-specific website. Yi (1990) shows that consumers' response to ads can change significantly when they are primed by relevant context.

Table 20: Estimates of the transition parameters (β)

	β_{12}	β_{13}	β_{21}	β_{23}	β_{31}	β_{32}
(intercept)	-2.864***	-5.632***	-3.713***	-2.206***	-3.405***	-4.327***
	(0.033)	(1.073)	(0.078)	(0.384)	(0.471)	(0.732)
generic_imp	0.009***	0.003	0.102***	0.002	0.009	0.003
	(0.029)	(0.003)	(0.008)	(0.007)	(0.009)	(0.008)
specific_imp	0.014***	0.002	0.008	0.003	-0.001	0.003
	(0.009)	(0.005)	(0.010)	(0.032)	(0.001)	(0.007)
generic_clk	0.126***	0.020	-0.098	0.383***	-0.001	-0.079
	(0.048)	(0.063)	(0.150)	(0.049)	(0.038)	(0.312)
specific_clk	0.189***	0.003	0.077	0.501***	0.021*	0.000
	(0.003)	(0.048)	(0.083)	(0.077)	(0.017)	(0.000)
search_clk	0.550***	0.031*	-0.029*	0.413***	0.048	-0.002**
	(0.135)	(0.025)	(0.023)	(0.138)	(0.057)	(0.001)

***, **, * denotes coefficients that are significant at a 99%, 95%, 90% level, respectively

Although display ads have an impact on moving consumers from a dormant state to a state of awareness, they do not have a significant impact on moving consumers further down the conversion funnel, i.e. from a state of awareness to a state of consideration (β_{23}). In fact, we observe that too many display ads on generic websites can have a detrimental effect on the consumer's movement towards the conversion state. As the coefficient of generic impression is positive and significant (0.102), it suggests that if consumers are shown too many display ads on generic websites, their probability to transition back to the dormant state increases considerably. We also observe that impressions do no have an impact later on in the conversion funnel. Thus, current attribution techniques which focus mostly at the end of the funnel, give negligible credit to these ads.

Not surprisingly, we observe that clicks have a significant impact on the consumer's movement from the awareness to the consideration state with search clicks having the largest effect. Once the consumer moves to the consideration state, there is a very low probability of him transitioning out of that state. This probability is further reduced when the consumer performs more searches and clicks on search ads. When a consumer actively starts to gather information about a product (by searching for the product at a search engine), he is likely to be at the very end of the funnel, contemplating his decision just prior to the eventual conversion.

Now we analyze the effect of different ad events on the HMM transition matrix. Q_{i0} denotes the transition matrix when consumer i is not exposed to any ads. Let Q_{is}, Q_{ic} and Q_{id} represent the transition matrices when we observe exactly one search click, one display click and 10 display impression for the consumer, respectively. These matrices are presented below,

$$Q_{i0} = \begin{pmatrix} 0.95 & 0.05 & 0.00 \\ 0.02 & 0.88 & 0.10 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}, \qquad Q_{is} = \begin{pmatrix} 0.91 & 0.09 & 0.00 \\ 0.02 & 0.84 & 0.14 \\ 0.03 & 0.01 & 0.96 \end{pmatrix},$$

$$Q_{ic} = \begin{pmatrix} 0.94 & 0.06 & 0.00 \\ 0.02 & 0.84 & 0.14 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}, \qquad Q_{id} = \begin{pmatrix} 0.87 & 0.12 & 0.00 \\ 0.06 & 0.85 & 0.09 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}.$$

In the absence of any ad related activity, the states are extremely sticky and it is unlikely that consumer transitions between the different states of the HMM. When the consumer clicks on a search ad, the probability (Q_{is}) that he moves down the search funnel increases considerably (q_{i12} : $0.05 \rightarrow 0.09$ and q_{i23} : $0.10 \rightarrow 0.14$). The effect of a display click is similar but not as pronounced (Q_{ic}). We look at the effect of 10 impressions as one impression has a very small impact on the transition probabilities. Interestingly, we observe that when the consumer is exposed to too many generic display impressions his likelihood to move to the dormant state (in the opposite direction of the funnel) increases (q_{i21} : $0.02 \rightarrow 0.06$). One possible explanation for this behavior is advertising avoidance which has been documented by Goldfarb and Tucker (2011) and Johnson (2011) in the literature. A consumer might completely abandon his search if he considers these ads to be too intrusive (Goldfarb and Tucker, 2011). These transition matrices also demonstrate that consumers move down the conversion funnel in a sequential manner, e.g. from one state to another and we do not observe abrupt jumps from a dormant state to a state of consideration.

Estimates of the Response Parameters

Now we discuss the underlying parameters that affect the observations of the HMM. We first discuss the factors that affect the number of pages viewed by a customer which are presented in Table 21. We can see that consumers in the awareness and consideration states differ considerably when it comes to their browsing behavior. Consumers in the consideration state view three times as many pages on the car manufacturer's website as the consumers in the awareness state. Since the consumers in these two states behave so differently, we are certain that the model is both empirically and behaviorally identified. Advertising activities tend to increase the consumers' propensity to view more web pages but the increase is more pronounced when the consumers actively interact with the ads (e.g. by clicking on them) than when they passively enter the consumers' perception (e.g. through display impressions).

Next we consider factors that influence the consumers' conversion probability. The estimated coefficients of these factors are presented in Table 22. We notice that the probability

Table 21: Estimate of factors affecting the page views (λ)

	τ_2	τ_3
η	0.781***	0.534***
	(0.037)	(0.003)
$\widetilde{\eta}$	0.781	2.487
generic_imp	0.004**	0.008***
	(0.002)	(0.000)
specific_imp	0.004***	0.005
	(0.001)	(0.009)
generic_clk	0.089***	0.123***
	(0.008)	(0.007)
specific_clk	0.132**	0.207***
	(0.060)	(0.008)
search_clk	0.169***	0.288***
	(0.004)	(0.004)

to convert is higher in the consideration state than it is in the awareness state, *ceteris paribus*. Apart from impressions on generic websites, all advertising activities lead to an increase in the conversion probability in the state of awareness. However, conditional on being in the consideration state, impressions of any kind do not have an incremental impact on the likelihood to convert. Interestingly, the effect of a specific click in the awareness state is more prominent than the effect of a generic or a search click. We also observe that an increase in the visits to the car manufacturer's website tends to increase the conversion rate in both states. Surprisingly, this effect is stronger in the state of awareness than in the consideration state. This decrease might be attributed to the diminishing returns from further interactions with the consumer. Once the consumer is sufficiently primed to convert, increased interactions only have a small marginal effect on him.

In Table 23, we present how different activities affect the conversion probability in the awareness and the consideration states. As the consumers interact more with the advertiser (through clicks and page views), there is a substantial increase in the conversion probability. It should be kept in mind that the conversion probabilities shown here are atypical of online campaigns which usually have very few conversions following a click.

Table 22: Estimates of conversion parameters

	γ_2	γ_3
α	-4.155***	-3.087**
	(0.433)	(1.050)
$\widetilde{\alpha}$	-4.155	-3.072
generic_imp	0.015	0.008
	(0.010)	(0.019)
specific_imp	0.017**	0.020
	(0.009)	(0.019)
generic_clk	0.289***	0.318***
	(0.084)	(0.095)
specific_clk	0.607***	0.303***
	(0.090)	(0.083)
search_clk	0.146***	0.588***
	(0.027)	(0.100)
nw_activity	0.091***	0.067***
	0.005	0.007

4.5.4. Ad Attribution

In this section we compare our proposed attribution scheme with the LTA and the logit attribution method proposed by Dalessandro et al. (2012). We use the entire (training + validation) data and apply the attribution methodologies – LTA, logistic multi-touch attribution (Logit-MTA) and HMM multi-touch attribution (HMM-MTA) to compute the contribution of each ad towards the final conversions. This result is presented in Table 24. The last column labeled "%Δ " shows the % difference between the attribution computed by the HMM-MTA and the LTA.

We observe from Table 24 that all three methods attribute a significant portion of the conversions to display and search clicks, which is in agreement with the coefficients presented in Table 22. Surprisingly, we see that the HMM-MTA attributes less credit to display impressions on generic websites. In this data, generic impression occur very frequently and as a consequence they have a high chance of being the last ad activity that took place before a conversion event. Since they are likely to appear last, the LTA gives them undue credits for the conversions even through they might not have had an impact on the consumer's

Table 23: Conversion probability as a result of various factors.

	awareness	consideration
No Activity	0.016	0.046
Generic Imp	0.016	0.046
Specific Imp	0.016	0.046
Generic Click	0.021	0.064
Specific Click	0.024	0.063
Search Click	0.018	0.083
Network Activity	0.017	0.050
Generic + Specific Clicks	0.032	0.086
Generic + Search Clicks	0.024	0.115
Specific + Search Clicks	0.028	0.113
Generic + Specific + Search Clicks	0.037	0.155
Generic + Specific + Search Clicks + Network Activity	0.040	0.166

Table 24: A comparison of attribution methodologies

Ad activity	#Ads	LTA	Logit-MTA	HMM-MTA	%Δ
Generic Impression	70,444	171	152.2	124.9	-27.5
Specific Impression	21,564	78	96.5	116.2	48.7
Generic Click	369	54	84.6	75.1	37.7
Specific Click	732	150	140.7	167.6	11.3
Search Click	1,260	328	310.9	294.3	-10.3

conversion probability. These ads that get credit just due to their sheer volume have been referred to as "carpet bombers" by Dalessandro et al. (2012). We also see that the HMM-MTA increases the number of conversions attributed to display impressions on specific websites which illustrates that our attribution method rewards events that influenced the consumer's deliberation process early on in the conversion funnel. There is a marginal increase in the conversions attributed to display clicks. The HMM-MTA assigns some the conversions from the generic impression to these activities that had a positive influence on the conversions. Even though there is a slight decrease in the conversions attributed to search clicks, it continues to remain as the most important factor under all the attribution methodologies. This finding is consistent with the results reported by Dalessandro et al. (2012) who show that the Logit-MTA does not lead to significant change in the conversion attributed to search ads. In order to compute the overall contribution of a specific channel, e.g. generic display ads we need to account for the conversions attributed to generic display impressions *and* generic display clicks. In the present context, generic display ads are responsible for 200 conversions and specific display ads are responsible for 283.8 conversions.

4.6. Discussion

In this section we present a few limitations of our current model and discuss how these limitations might affect our findings. We also provide a few directions for addressing the limitations of our model.

4.6.1. Interactions with Other Advertising Media

In our model, we ignore the effect that other factors like traditional advertising might have on the consumer's behavior and eventual conversion. Here, we try to explore how ignoring these factors might bias our estimates. Let's first consider the transition probabilities. The transition probabilities arise from an underlying utility model that has the following form

$$U_{itss'} = \boldsymbol{x}'_{it}\boldsymbol{\beta}_{ss'} + x_{ait}\beta_{ass'} + \epsilon_{itss'} \quad \forall \quad s' \in S,\ s' \neq s$$

where x_{ait} captures the offline advertising activity at time t and $\epsilon_{itss'}$ is an extreme value distributed error term. If the advertising activity x_{ait} is ignored, the estimates of $\boldsymbol{\beta}_{ss'}$ will be biased. The exact nature of the bias depends on the relationship between x_{ait} and $\boldsymbol{x}'_{it}$ and the sign of $\beta_{ass'}$ (Lee, 1982). If x_{ait} and $\boldsymbol{x}'_{it}$ are independent, the estimates of $\boldsymbol{\beta}_{ss'}$ will be biased upwards if $\beta_{ass'} \geq 0$ and biased downward otherwise. On the other hand, if $x_{ait} = \boldsymbol{x}'_{it}\boldsymbol{\delta} + \varepsilon$, then the estimates of $\boldsymbol{\beta}_{ss'}$ will be further biased by $\beta_{ass'}\boldsymbol{\delta}$.

Hence, if the offline advertising activity is independent of the online advertising activity, the effect of this omission on the estimation procedure depends on how consumers react to offline ads. If offline ads have a positive (negative) impact on transition from state s to s', then $\boldsymbol{\beta}_{ss'}$ will be overestimated (underestimated). In addition, if there is a relation between offline and online advertising activities which is likely as an advertiser's activities across various channels are coordinated, the bias depends on the exact relationship between these channels. As these activities are usually positively correlated, this would lead to an overestimation of the transition parameters. The parameters of the conversion probabilities are affected in a similar manner.

4.6.2. Incorporating Consumer Heterogeneity

The model proposed in Section 4 does not incorporate heterogenity in consumer behavior. Consumers might be affected differently due to the advertising activities. This could occur for several brand specific reasons such as prior knowledge or interaction with the brand, exposure to ads from different media or due to inherent differences in underlying consumer behavior. In order to incorporate consumer heterogeneity in our model, we modify the different components of the HMM in the following manner:

Transition Probabilities

The transition probabilities between the states in presence of consumer heterogeneity can be written as

$$\begin{aligned} q_{itss'} &= \frac{\exp\{\mu_{is} + \boldsymbol{x}'_{it}\beta_{iss'}\}}{1 + \sum_{s' \in T_s} \exp\{\boldsymbol{x}'_{it}\beta_{iss'}\}} \quad \forall\, s' \neq s, \\ q_{itss} &= \frac{1}{1 + \sum_{s' \in T_s} \exp\{\mu_{is'} + \boldsymbol{x}'_{it}\beta_{iss'}\}}, \end{aligned}$$

where unlike Equation 4.1, the coefficient $\beta_{iss'}$ is specific to a consumer. This indicates that consumers can respond differently to the time-varying covariates in our model. $\boldsymbol{x}'_{it}$, in addition to the ad related activities mentioned earlier, can contain observed demographic variables and exposure to other media that vary across consumers. The parameter μ_{is} captures the time-invariant unobserved consumer heterogeneity in our model. We assume that the coefficients $(\mu_{is}\ \beta_{iss'})'$ are drawn from a multivariate normal distribution $N(\bar{\beta}_s, \Sigma_{\beta s})$.

Conversion Probability

In this extension, we assume that consumers can have different propensities to convert, *ceteris paribus*. Some consumer might be automobile aficionados who are not only more likely to convert, but are also differently affected by the different advertising activities. In order to incorporate this heterogeneity between consumers, we model the conversion probability in the following manner,

$$m_{its} = \frac{\exp\{\tilde{\alpha}_s + \alpha_{is} + \boldsymbol{z}'_{it}\gamma_{is}\}}{1 + \exp\{\tilde{\alpha}_s + \alpha_{is} + \boldsymbol{z}'_{it}\gamma_{is}\}}.$$

where $\tilde{\alpha}_s$ is the intrinsic propensity to convert in state s, α_{is} is a time-invariant consumer affinity to convert and γ_{is} is the consumer specific response to the time varying covariates.

Parameter Estimation

The likelihood for the observed data presented in Equation (4.9) can be written in terms of the heterogenous probabilities mentioned here. The estimation can be performed using a MCMC estimation approach similar to the method adopted by Netzer et al. (2008).

This model that incorporates consumer heterogeneity allows us to perform attribution for a specific consumer. Some consumers might have an higher probability to convert even without the influence of online ads. The homogenous HMM would incorrectly ascribe credit to the online ads when such consumers convert. However, the heterogenous HMM accounts for a consumer's proclivity to convert and performs the attribution based on the underlying consumer heterogeneity.

4.7. Conclusion

In this paper, we present a model that analyzes how consumers behave when they are exposed to advertising from multiple online channels. The consumer behavior is captured using a dynamic HMM which is modeled based on conversion funnel. A consumer moves through the states of the HMM in a stochastic manner when they are exposed to advertising activity. Conditional on being in a certain state, he can engage in a conversion activity with a certain probability which is a function of his current state and other time varying covariates. This model is estimated on campaign data from a car manufacturer. We show that although display ads do not have an immediate impact on conversion, they have a significant impact on the consumer behavior early on in the deliberation process. This result is contrary to the popularly held belief that display ads do not work. They work but not in the manner advertisers expect them to work. This finding has significant implications for the online advertising industry and it underscores the importance of better attribution methodologies particularly for display networks and firms like Facebook that derive most of their revenues from display advertising. We subsequently propose an attribution methodology that attributes credit to the ads based on the marginal effect they have on a consumer's conversion probability. This method not only takes into account the prior history of a consumer before being exposed to an ad, it also considers the long-term future impact the ad might have on the consumer's decision. We apply this methodology to the campaign data and show that there are considerable differences in the attribution performed by the commonly used LTA and our methodology.

To our knowledge, this is the first paper that analyzes the affect of online ads in a multi-channel context. We hope that this paper can lead to a better understanding of consumer behavior in a multi-channel context which might help researchers build better models in the future. We also believe that this research can serve as a foundation for an integrative approach to optimal budget allocation when an advertiser uses different channels for his campaign.

Appendix

Identification Assumption

For simplicity, let's assume that there are only two states where the initial membership probabilities are denoted by $\pi = \{\pi_1, \pi_2\}$. When a consumer sees an ad, he can follow one of three possible paths - (1) He starts out in the first state and transitions to the second state, (2) he starts out from the first state and stays in the same state after the ad event or (3) starts out from the second state and stays in the same state. The likelihood of observing an outcome $\boldsymbol{y}_1 = (n_1, c_1)'$ after the first ad is shown is given by

$$\begin{aligned} L &= \pi_1 q_{11} p_1^{c_1} (1-p_1)^{1-c_1} P(N_1 = n_1 | S_1 = 1) + \pi_1 q_{12} p_2^{1-c_1} (1-p_2)^{c_2} P(N_1 = n_1 | S_1 = 2) \\ &\quad + \pi_2 q_{22} p_2^{1-c_1} (1-p_2)^{c_2} P(N_1 = n_1 | S_1 = 2) \end{aligned}$$

As $p_1 = 0$ in our model, the likelihood reduces to $p_2^{c_1}(1-p_2)^{c_2} P(N_1 = n_1 | S_1 = 2)(\pi_1 q_{12} + \pi_2 q_{22})$. It can be shown that irrespective of what events follow afterward we can only identify $(\pi_1 q_{12} + \pi_2 q_{22})$ in our model. Hence we impose the restriction that $\pi_2 = 0$.

BIBLIOGRAPHY

V. Abhishek, K. Hosanagar, and P. Fader. On aggregation bias in sponsored search data: Existence and implications. *SSRN eLibrary*, 2011.

A. Agarwal, K. Hosanagar, and M. D. Smith. Location, location, location: An analysis of profitability and position in online advertising markets. *Journal of Marketing Research*, 48(6):1057–1073, 2011.

G. Aggarwal and J. D. Hartline. Knapsack auctions. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, SODA '06, pages 1083–1092, 2006.

G. Aggarwal, A. Goel, and R. Motwani. Truthful auctions for pricing search keywords. In *Proceedings of the 7th ACM conference on Electronic commerce*, EC '06, pages 1–7, 2006.

K. Ali and M. Scarr. Robust methodologies for modeling web click distributions. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 511–520, 2007.

G. M. Allenby and P. E. Rossi. There is no aggregation bias: Why macro logit models work. *Journal of Business and Economic Statistics*, pages 1–14, 1991.

M. Angelucci and G. D. Giorgi. Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, 99(1):486–508, 2009.

A. Animesh, V. Ramachandran, and S. Viswanathan. Research Note–Quality Uncertainty and the Performance of Online Sponsored Search Markets: An Empirical Investigation. *Information Systems Research*, 2009.

A. Ansari, K. Bawa, and A. Ghosh. A nested logit model of brand choice incorporating variety-seeking and marketing-mix variables. *Marketing Letters*, 6(3):199–210, 1995.

T. E. Barry. The development of the hierarchy of effects: An historical per- spective. *Current Issues and Research in Advertising*, 10:251–295, 1987.

F. Bass, N. Bruce, S. Majundar, and B. Murthi. A dynamic bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2):179–195, 2007.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

J. Berkhof, I. van Mechelen, and A. Gelman. A bayesian approach to the selection and testing of latent class models, 2000.

J. R. Bettman, M. F. Luce, and J. W. Payne. Constructive consumer choice processes. *Journal of Consumer Research*, 25(3):187–217, December 1998.

C. Borgs, J. Chayes, N. Immorlica, K. Jain, O. Etesami, and M. Mahdian. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 531–540, 2007.

G. C. Bruner and A. Kumar. Web commercials and advertising hierarchy-of-effects. *Journal of Advertising Research*, 40(1/2):35–42, 2000.

A. Bureau, S. Shiboksi, and J. P. Hughes. Applications of continuous time hidden markov models to the study of misclassified disease outcomes. *Statistics in Medicine*, 22:441–462, 2003.

J. Chandler-Pepelnjak. Measuring roi beyond the last ad: Winners and losers in the purchase funnel are different when viewd through a new lense. *Microsoft Advertising Institute*, 2009.

M. Christen, S. Gupta, J. C. Porter, R. Staelin, and D. R. Wittink. Using market-level data to understand promotion effects in a nonlinear model. *Journal of Marketing Research*, 34 (3):pp. 322–334, 1997.

H. Chung, T. Walls, and Y. Park. A latent transition model with logistic regression. *Psychometrika*, 72(3):413–435, 2007.

T. Claburn. Facebook's advertising problem. *Information Week*, May 2012.

B. Cooper and M. Lipsitch. The analysis of hospital infection data using hidden markov models. *Biostatistics*, 5(2):223–237, Apr 2004.

D. Court, D. Elzinga, S. Mulder, and O. J. Vetvik. The consumer decision journey. *McKinsey Quarterly*, June 2009.

N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining*, WSDM '08, pages 87–94, 2008.

B. Dalessandro, O. Stitelman, C. Perlich, and F. Provost. Causally motivated attribution for online advertising. *NYU Working Paper series*, 2012.

P. J. Danaher, A. Bonfrer, and S. Dhar. The effect of competitive advertising interference on sales for packaged goods. *Journal of Marketing Research*, 45(2):211–225, 2008.

G. de Vries. Online display ads: The brand awareness black hole. *Forbes*, May 2012.

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

B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. *Decision Support System*, 43:192–198, February 2007.

B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.

J. Feldman, S. Muthukrishnan, M. Pal, and C. Stein. Budget optimization in search-based advertising auctions. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 40–49, 2007.

J. Feng, H. K. Bhargava, and D. M. Pennock. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing*, 2005:137–148, 2006.

A. Ghose and S. Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.

A. Goldfarb and C. Tucker. Search Engine Advertising: Pricing Ads to Context. *SSRN eLibrary*, 2007.

A. Goldfarb and C. Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 2011.

I. J. Good and Y. Mittal. The amalgamation and geometry of two-by-two contingency tables. *The Annals of Statistics*, 15(2):pp. 694–711, 1987.

S. Gupta, P. Chintagunta, A. Kaul, and D. R. Wittink. Do household scanner data provide representative inferences from brand choices: A comparison with store data. *Journal of Marketing Research*, 33(4):383–398, 1996.

L. P. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, 14(3):262–280, 1996.

J. Hao, S. Menon, S. Raghunathan, and S. Sarkar. A comparison of rbb and rbr ranking mechanisms in sponsored search. In *Proc Conference on Information Systems and Technology*, 2009.

D. Hawkins, R. Best, and K. Coney. *Consumer behaviour: Implications for marketing strategy*. Irwin Publishing, 1995.

F. Hayashi. *Econometrics*. Princeton University Press, 2000.

J. A. Howard and J. N. Sheth. *The theory of buyer behavior*. Wiley, 1969.

B. J. Jansen and S. Schuster. Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1), 2011.

P. Jeziorski and I. Segal. What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising. *SSRN eLibrary*, 2009.

J. P. Johnson. Targeted advertising and advertising avoidance. *Cornell University Working Paper Series*, December 2011.

P. Jordan, M. Mahdian, S. Vassilvitskii, and E. Vee. The multiple attribution problem in pay-per-conversion advertising. In *Proceedings of the 4th international conference on Algorithmic game theory*, SAGT'11, pages 31–43, 2011.

Z. Katona and M. Sarvary. The race for sponsored links: Bidding patterns for search advertising. *Marketing Science*, 29(2):199–215, 2010.

A. Kaushik. Multi-channel attribution: Definitions, models and a reality check. 2012.

H. H. Kelejian. Aggregated heterogeneous depentdent data and the logit model: A suggested approach. *Economic Letters*, pages 243–248, 1995.

F. Khatibloo. Untangling the attribution web: Using cross-channel attribution to understand marketing effectiveness. *Forrester Reports*, 2010.

P. Kotler and G. Armstrong. *Principles of Marketing*. Prentice Hall, February 2011.

L.-F. Lee. Specification error in multinomial logit models : Analysis of the omitted variable bias. *Journal of Econometrics*, 20(2):197–209, November 1982.

C. J. Liu, D. and A. B. Whinston. Competing keyword auctions. In *Proc Workshop on Information Systems and Economics*, 2007.

M. Mahdian, H. Nazerzadeh, and A. Saberi. Allocating online advertisement space with unreliable estimates. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 288–294, 2007.

I. L. McDonald and W. Zucchini. *Hidden Markov and other models for discrete-valued time series*. Chapman and Hall, London, 1997.

A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54, October 2007.

A. Molina and F. Pla. Shallow parsing using specialized hmms. *J. Mach. Learn. Res.*, 2: 595–613, Mar. 2002.

R. Montoya, O. Netzer, and K. Jedidi. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science*, 29(5):909–924, 2010.

A. Muller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley, 2002.

S. Mulpuru. The purchase path of online buyers. *Forrester Report*, March 2011.

S. Muthukrishnan, M. Pál, and Z. Svitkina. Stochastic models for budget optimization in search-based advertising. In *Proceedings of the 3rd international conference on Internet and network economics*, WINE'07, 2007.

M. K. M. Naik, Prasad A. and A. G. Sawyer. Planning media schedules in the presence of dynamic advertising quality. *Marketing Science*, 17(3):214–235, 1998.

P. A. Naik, K. Raman, and R. S. Winer. Planning marketing-mix strategies in the presence of interaction effects. *Marketing Science*, 24(1):25–34, Jan. 2005.

A. V. Nefian, M. H. Hayes, and III. Hidden markov models for face recognition. In *Proc. International Conf. on Acoustics, Speech and Signal Processing*, pages 2721–2724, 1998.

M. Nerlove and K. J. Arrow. Optimal advertising policy under dynamic conditions. *Economica*, 29(114):129–142, 1962.

S. A. Neslin and R. W. Shoemaker. An alternative explanation for lower repeat rates after promotion purchases. *Journal of Marketing Research*, 26(2):205–213, 1989.

O. Netzer, J. M. Lattin, and V. Srinivasan. A hidden markov model of customer relationship dynamics. *Marketing Science*, 27(2):185–204, 2008.

New York Times. Teradata, lunexa partnership introduces digital marketing attribution solution for teradata aster analytics platforms. June 2012.

C. Quinn. Finding deeper insight into the customer journey. March 2012.

L. Rabiner. A tutorial on HMM and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, feb 1989.

V. Ramaswamy, W. S. Desarbo, D. J. Reibstein, and W. T. Robinson. An empirical pooling approach for estimating marketing mix elasticities with pims data. *Marketing Science*, 12(1):103–124, 1993.

P. Rusmevichientong and D. P. Williamson. An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Proceedings of the 7th ACM conference on Electronic commerce*, EC '06, pages 260–269, 2006.

G. J. Russell and W. A. Kamakura. Understanding brand competition using micro and macro scanner data. *Journal of Marketing Research*, 31(2):289–303, 1994.

O. J. Rutz and R. E. Bucklin. From generic to branded: A model of spillover dynamics in paid search advertising. *Journal of Marketing Research*, 48(1), 2011.

O. J. Rutz, R. E. Bucklin, and G. P. Sonnier. A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research*, 49(3), June 2012.

G. A. Satten and I. M. Longini. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(3), 1996.

E. M. Schwartz, E. Bradlow, P. Fader, and Y. Zhang. Children of the hmm: Modeling longitudinal customer behavior at hulu.com. *WCAI Working Paper Series*, August 2011.

D. A. Schweidel, E. T. Bradlow, and P. S. Fader. Portfolio dynamics for customers of a multiservice provider. *Management Science*, 57(3):471–486, 2011.

X. Shao and L. Li. Data-driven multi-touch attribution models. KDD '11, pages 258–264, 2011.

J.-B. E. M. Steenkamp, V. R. Nijs, D. M. Hanssens, and M. G. Dekimpe. Competitive reactions to advertising and promotion attacks. *Marketing Science*, 24(1):pp. 35–54, 2005.

E. K. Strong. *The Psychology of Selling Advertising*. McGraw-Hill, 1925.

C. Szulc. Why you should care about marketing attribution. *Inc.*, April 2012.

C. Tucker. Social advertising. *MIT Sloan Research Paper*, February 2012.

I. Visser, B. R. J. Jansen, and M. Speekenbrink. A framework for discrete change. In K. M. Newell and P. C. M. Molenaar, editors, *Individual Pathways of Change*. APA, 2009.

T. A. Weber and Z. E. Zheng. A model of search intermediaries and paid referrals. *Information Systems Research*, 18(4):414–436, 2007.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, New York, NY, 1997.

J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, October 2001.

S. Yang and A. Ghose. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, page mksc.1100.0552, 2010.

S. Yao and C. F. Mela. A dynamic model of sponsored search advertising. *Marketing Science*, 30:447–468, May 2011.

A. Yatchew and Z. Griliches. Specification error in probit models. *The Review of Economics and Statistics*, 67(1):134–139, February 1985.

Y. Yi. The effects of contextual priming in print advertisements. *Journal of Consumer Research*, 17(2):pp. 215–222, 1990.