

Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion

Lizhen Xu

Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308, lizhen.xu@scheller.gatech.edu

Jason A. Duan, Andrew Whinston

McCombs School of Business, The University of Texas at Austin, Austin, Texas 78712
{duanj@mcombs.utexas.edu, abw@uts.cc.utexas.edu}

This paper studies the effects of various types of online advertisements on purchase conversion by capturing the dynamic interactions among advertisement clicks themselves. It is motivated by the observation that certain advertisement clicks may not result in immediate purchases, but they stimulate subsequent clicks on other advertisements, which then lead to purchases. We develop a novel model based on mutually exciting point processes, which consider advertisement clicks and purchases as dependent random events in continuous time. We incorporate individual random effects to account for consumer heterogeneity and cast the model in the Bayesian hierarchical framework. We construct conversion probability to properly evaluate the conversion effects of online advertisements. We develop simulation algorithms for mutually exciting point processes to compute the conversion probability and for out-of-sample prediction. Model comparison results show the proposed model outperforms the benchmark models that ignore exciting effects among advertisement clicks. Using a proprietary data set, we find that display advertisements have relatively low direct effect on purchase conversion, but they are more likely to stimulate subsequent visits through other advertisement formats. We show that the commonly used measure of conversion rate is biased in favor of search advertisements and underestimates the conversion effect of display advertisements the most. Our model also furnishes a useful tool to predict future purchases and advertisement clicks for the purpose of targeted marketing and customer relationship management.

Keywords: attribution model; online advertising; conversion; mutually exciting point process; multivariate stochastic model; search advertisement; display advertisement; business analytics

History: Received September 16, 2012; accepted March 9, 2014, by Eric Bradlow, special issue on business analytics.

Published online in *Articles in Advance* April 16, 2014.

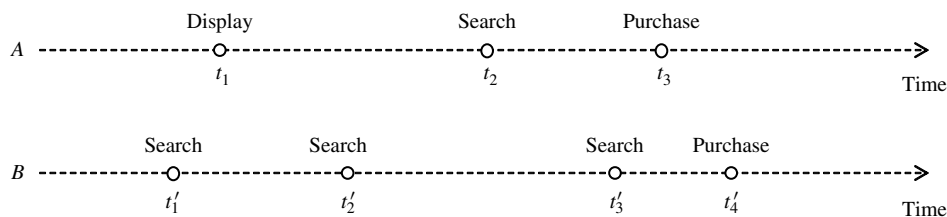
1. Introduction

As the Internet grows to become the leading advertising medium, firms invest heavily to attract consumers to visit their websites through advertising links in various formats, among which search advertisements (i.e., sponsored links displayed on the search engine results pages) and display advertisements (i.e., digital graphics linking to the advertiser's website embedded in Web content pages) are the two leading online advertising formats (Interactive Advertising Bureau and PricewaterhouseCoopers 2012). Thanks to the advancement of information technology, which makes tremendous individual-level online clickstream data available, business analytics of how to evaluate the effectiveness of these different formats of online advertisements (ads) has been attracting constant academic and industrial interest. Marketing researchers and practitioners are especially interested in the *conversion effect* of each type of online advertisement, that is, given an individual consumer clicked on a certain type of advertisement, what is the probability of her

making a purchase (or performing certain actions such as registration or subscription) thereafter.

The most common measure of conversion effects is conversion rate, which is the percentage of the advertisement clicks that directly lead to purchases among all advertisement clicks of the same type. This simple statistic provides an intuitive assessment of advertising effectiveness. However, it overemphasizes the effect of the "last click" (i.e., the advertisement click directly preceding a purchase) and completely ignores the effects of all previous advertisement clicks, which naturally leads to biased estimates. Existing literature has developed more sophisticated models to analyze the conversion effects of website visits and advertisement clicks (e.g., Moe and Fader 2004, Manchanda et al. 2006). These models account for the entire clickstream history of individual consumers and model the purchases as a result of the accumulative effects of all previous clicks, which can more precisely evaluate the conversion effects and predict the purchase probability. Nevertheless, because existent studies on conversion effects focus solely on how nonpurchase activities

Figure 1 Illustrative Examples of the Interactions Among Ad Clicks



(e.g., advertisement clicks, website visits) affect the probability of purchasing, they usually consider the nonpurchase activities as deterministic data rather than stochastic events and neglect the dynamic interactions among these activities themselves, which motivates us to fill this gap.

To illustrate the importance of capturing the dynamic interactions among advertisement clicks when studying their conversion effects, let us consider a hypothetical example illustrated in Figure 1. Suppose consumer *A* saw firm *X*'s display advertisement for its product when browsing a webpage, clicked on the ad, and was linked to the product webpage at time t_1 . Later, she searched for firm *X*'s product in a search engine and clicked on the firm's search advertisement there at time t_2 . Shortly afterward, she made a purchase at firm *X*'s website at time t_3 . In this case, how shall we attribute this purchase and evaluate the respective conversion effects of the two advertisement clicks? If we attribute the purchase solely to the search advertisement click, like how the conversion rate is computed, we ignore the fact that the search advertisement click might not have occurred without the initial click on the display advertisement. In other words, the occurrence of the display ad click at time t_1 is likely to increase the probability of the occurrence of the subsequent advertisement clicks, which eventually lead to a purchase. Without considering such an effect, we might undervalue the first click on the display ad and overvalue the next click on the search ad. Therefore, to properly evaluate the conversion effects of different types of advertisement clicks, it is imperative to account for the *exciting effects* between advertisement clicks, that is, how the occurrence of an earlier advertisement click affects the probability of occurrence of subsequent advertisement clicks. Neglecting the exciting effects between different types of advertisement clicks, the simple measurement of conversion rates might easily underestimate the conversion effects of those advertisements that tend to catch consumers' attention initially and trigger their subsequent advertisement clicks but are less likely to directly lead to a purchase, for instance, the display advertisements.

In addition to the exciting effects between different types of advertisement clicks, neglecting the exciting effects between the same type of advertisement clicks

may also lead to underestimation of their conversion effects. Consider consumer *B* in Figure 1, who clicked on search advertisements three times before making a purchase at time t'_4 . If we take the occurrence of advertisement clicks as given and consider only their accumulative effects on the probability of purchasing, like the typical conversion models, we may conclude that it takes the accumulative effects of three search advertisement clicks for consumer *B* to make the purchase decision, so each click contributes one third. Nevertheless, it is likely that the first click at t'_1 stimulates the subsequent two clicks, all of which together lead to the purchase at time t'_4 . When we consider such exciting effects, the (conditional) probability of consumer *B* making a purchase eventually given he clicked on a search advertisement at time t'_1 clearly needs to be reevaluated.

This study aims to develop an innovative modeling approach that captures the exciting effects among advertisement clicks to contribute to the attribution models for properly evaluating the effectiveness of online advertisements using individual-level online clickstream data. To properly characterize the dynamics of consumers' online behaviors, the model also needs to account for the following unique properties and patterns of online advertisement clickstream and purchase data. First, different types of online advertisements have their distinct natures and therefore differ greatly in their probabilities of being clicked, their impacts on purchase conversions, and their interactions with other types of advertisements as well. Therefore, unlike the typical univariate approach in modeling the conversion effects of website visits, to study the conversion effects of various types of online advertisements from a holistic perspective, the model needs to account for the multivariate nature of nonpurchase activities.

Second, consumers vary from individual to individual in terms of their online purchase and ad clicking behaviors, which could be affected by their inherent purchase intention, exposure to marketing communication tools, or simply preference for one advertising format over another. Because most of these factors are usually unobservable in online clickstream data, it is important to incorporate consumers' individual heterogeneity in the model.

Third, online clickstream data often contain the precise occurrence time of various activities. Although

the time data are very informative about the underlying dynamics of interest, most existing modeling approaches have yet to adequately exploit such information. Prevalent approaches to address the time effects usually involve aggregating data by an arbitrary fixed time interval or considering the activity counts only, but discarding the actual time of occurrence. It is appealing to cast the model in a continuous-time framework to duly examine the time effects between advertisement clicks and purchases. Notice that the effects of a previous ad click on later ones and purchases should decay over time. In other words, an ad click one month ago should have less direct impact on a purchase at present compared to a click several hours ago. Moreover, some advertisement formats may have more lasting effects than others, so the decaying effects may vary across different advertisement formats. Therefore, incorporating the decaying effects of different types of advertisement clicks in the model is crucial in accurately evaluating their conversion effects.

Furthermore, a close examination of the online advertisement click and purchase data set used for this study reveals noticeable clustering patterns, that is, advertisement clicks and purchases tend to concentrate in shorter time spans and there are longer time intervals without any activity, which is also termed *clumpy data* in statistics literature (e.g., Zhang et al. 2013).¹ If we are to model advertisement clicks and purchases as a stochastic process, the commonly used Poisson process model will perform poorly, because its intensity at any time is independent of its own history, and such a memoryless property implies no clustering at all (Cox and Isham 1980). For this reason, a more sophisticated model with history-dependent intensity functions is especially desirable.

In this paper, we develop a stochastic model for online purchasing and advertisement clicking that incorporates mutually exciting point processes with individual heterogeneity in a Bayesian hierarchical modeling framework. The mutually exciting point process is a multivariate stochastic process in which different types of advertisement clicks and purchases are modeled as different types of random points in continuous time. The occurrence of an earlier point affects the probability of occurrence of later points of all types so that the exciting effects among all advertisement clicks are well captured. As a result, the intensities of the point process, which can be interpreted as the instant probabilities of point occurrence, depend on the previous history of the process. Moreover, the

exciting effects are modeled to be decaying over time in a natural way. The hierarchical structure of the model allows each consumer to have her own propensity for clicking on various advertisements and purchasing so that consumers' individual processes are heterogeneous.

Our model offers a novel method to more precisely evaluate the effectiveness of various formats of online advertisements. In particular, the model manages to capture the exciting effects among advertisement clicks so that advertisement clicks, instead of being deterministic data as given, are also stochastic events dependent on the past occurrences. In this way, even for those advertisements that have little direct effect on purchase conversion but may trigger subsequent clicks on other types of advertisements that eventually lead to conversion, our model can properly account for their contributions. Compared with the benchmark model that ignores all the exciting effects among advertisement clicks, our proposed model outperforms it to a considerable degree in terms of model fit, which indicates that the mutually exciting model better captures the complex dynamics of online advertising response and purchase processes.

Based on our model and its Bayesian estimation results, we construct *conversion probability* to better evaluate the conversion effects of different types of online advertisements. We find that the commonly used measure of conversion rate is biased in favor of search advertisements by overemphasizing the "last click" effects and underestimates the effectiveness of display advertisements the most severely. We show that display advertisements have little direct effect on purchase conversion, but are likely to stimulate visits through other advertising channels. As a result, ignoring the mutually exciting effects between different types of advertisement clicks undervalues the efficacy of display advertisements the most. Likewise, ignoring the self-exciting effects leads to significant underestimation of search advertisement's conversion effects. A more accurate understanding of the effectiveness of various online advertising formats can help firms rebalance their marketing investment and optimize their portfolio of advertising spending.

Our model also better predicts individual consumers' online behavior based on their past behavioral data. Compared with the benchmark model that ignores all the exciting effects, incorporating the exciting effects among all types of online advertisements improves the model predictive power for consumers' future ad click and purchase patterns. Because our modeling approach allows us to predict both purchase and nonpurchase activities in the future, it thus furnishes a useful tool for marketing managers in targeted advertising and customer relationship management.

In addition to the substantive contributions, this paper also makes several methodological contributions.

¹ Using the clumpiness metric (entropy value) proposed by Zhang et al. (2013), we find the median clumpiness of the individuals in our data sample is 0.50. In comparison, the median clumpiness would be 0.17 if the click and purchase data were generated by memoryless Poisson processes.

We model the dynamic interactions among online advertisement clicks and their effects on purchase conversion with a mutually exciting point process. To the best of our knowledge, we are the first to apply the mutually exciting point process model in a marketing- or e-commerce-related context. We are also the first to incorporate individual random effects into the mutually exciting point process model in the applied econometric and statistic literature. This is the first study that successfully applies Bayesian inference using Markov chain Monte Carlo (MCMC) method to a mutually exciting point process model, which enables us to fit a more complex hierarchical model with random effects in correlated stochastic processes. In evaluating the conversion effects for different online advertisement formats and predicting consumers' future behaviors, we develop algorithms to simulate the point processes, which extend the thinning algorithm in Ogata (1981) to mutually exciting point processes with parameter values sampled from posterior distributions.

The rest of this paper is organized as follows. In §2, we survey the related literature. We then provide an overview of the data used for this study with summary statistics in §3. In §4, we construct the model and explore some of its theoretical properties. In §5, we discuss the inference and present the estimation results, which will be used to evaluate the conversion effects of different types of online advertisements and predict future consumer behaviors in §6. We also extend the model to incorporate additional data in §6. We conclude this paper with discussions in §7.

2. Literature Review

This study is related to various streams of existing literature on online advertising, consumer Web browsing behaviors, and their effects on purchase conversion. Our modeling approach using the mutually exciting point process also relates to existing theoretical and applied studies in statistics and probability. In addition, the rich literature on consumer behavior theory and cognitive psychology provides behavioral support for our model specifications. We next review the relevant literature in these domains.

Our work is related to the literature on the dynamics of online advertising exposure, website visit, webpage browsing, and purchase conversion. For example, Manchanda et al. (2006) study the effects of banner advertising exposure on the probability of repeated purchase using a survival model. Moe and Fader (2004) propose a model of accumulative effects of website visits to investigate their effects on purchase conversion. Both studies consider the conversion effects of a single type of activity and focus on the effects of the non-purchase activities on purchase conversion, whereas we study the effects of various types of online advertisement clicks and consider the dynamic interactions

among nonpurchase activities as well. Montgomery et al. (2004) consider the sequence of webpage views within a single site-visit session. They develop a Markov model in which, given the occurrence of a webpage view, the type of the webpage being viewed is affected by the type of the last webpage view. In contrast, we consider multiple visits over a long period of time and capture the actual time effect between different activities. To account for correlations among multivariate activities, Park and Fader (2004) apply the Sarmanov family of bivariate distributions to model the dependence of website visit durations across two different websites. Danaher and Smith (2011) further demonstrate that a more general class of copula models can be used to model multivariate distributions in various marketing applications. As we discuss in more detail in §7.1, our model based on the mutually exciting point process offers a new approach to induce correlation among all time durations between activities in a parsimonious way. Most recently, an emerging stream of research is dedicated to attribution modeling, which demonstrates that the simplistic approach of attributing conversion to the very last stop is erroneous (e.g., Li and Kannan 2014, Abhishek et al. 2013, Zantedeschi et al. 2013). Our paper enriches this increasingly vibrant stream of literature with a novel modeling framework.

In the area of statistics and probability, mutually exciting point processes were first proposed in Hawkes (1971a, b), where their theoretical properties are studied. Statistical models using Hawkes' processes, including the simpler version of self-exciting processes, are applied in seismology (e.g., Ogata 1998), sociology (e.g., Mohler et al. 2011), and finance (e.g., Ait-Sahalia et al. 2013, Bowsher 2007). These studies do not consider individual heterogeneity, and the estimation is usually conducted using method of moments or maximum likelihood estimation, whose asymptotic consistency and efficiency is studied in Ogata (1978). Our paper is thus the first to incorporate random coefficients into the mutually exciting point process model, cast it in a hierarchical framework, and obtain Bayesian inference for it. Bijwaard et al. (2006) proposes a counting process model for interpurchase duration, which is closely related to our model. A counting process is one way of representing a point process (e.g., Cox and Isham 1980). The model in Bijwaard et al. (2006) is a nonhomogeneous Poisson process where the dependence on the purchase history is introduced through covariates. Our model is not a Poisson process where the dependence on history is parsimoniously modeled by making the intensity directly as a function of the previous path of the point process itself. Bijwaard et al. (2006) also incorporates unobserved heterogeneity in the counting process model and estimates it using the expectation-maximization algorithm. Our Bayesian inference using the MCMC method not only provides

an alternative and efficient way to estimate this type of stochastic model, but it facilitates straightforward simulation and out-of-sample prediction as well.

2.1. Conceptual Background

A large volume of behavioral literature on consumer information processing and responses to advertising provides theories and evidence supporting our quantitative model formulation. First, a consumer's prior clicks on one format of online advertisement could increase the probability that she clicks on the advertisements in either the same or a different format. Studies on processing fluency show that prior exposures to advertising can enhance both perceptual and conceptual fluency (e.g., Jacoby and Dallas 1981, Shapiro 1999). Perceptual fluency refers to the ease with which consumers can identify a target stimulus on subsequent encounters and involves the processing of physical features (such as modality, shape, and color), whereas conceptual fluency refers to the ease with which the target comes to consumers' minds subsequently and involves the processing of meanings (Lee and Labroo 2004). Enhancement in these two dimensions therefore implies the increase of the likelihood of the consumer recognizing and clicking on the ads either in the same format with similar physical features or in different formats that pertain to the same product information. Additionally, extant studies find that prior ad exposure increases the probability of consideration-set membership of the advertised product (e.g., Lee 2002, Nedungadi 1990, Shapiro et al. 1997), which makes consumers more willing to consult ads that they would otherwise have ignored for informational purposes (Engel et al. 1995). Such positive effects have been shown robust for both stimulus-based and memory-based consideration-set formation (Lee 2002, Nedungadi 1990), suggesting the increase in the probability of the consumer clicking on the ads that either are physically presented or require active searches.

Furthermore, it is also well documented that enhanced perceptual and conceptual fluency positively influence consumers' affective responses (e.g., Anand and Sternthal 1991, Lee and Labroo 2004). Increased liking of the product, on one hand, will positively influence the attention and reception consumers give to marketing communications; on the other hand, it will directly increase the probability of purchase (Engel et al. 1995). Together with the above literature on the consideration-set inclusion because of prior exposure to advertisements, we argue that a consumer's successive clicks on various formats of online advertisements jointly contribute to the increase in the probability of purchase conversion.

The theories and evidence discussed above suggest that prior exposures to advertising positively influence subsequent clicks and purchases; meanwhile, such

effects are subject to decay over time. Cognitive psychology provides theories explaining why memory traces fade with the passage of time (Anderson and Milson 1989). Evidence is well documented that the probability of retrieval failure increases as a function of time, in many cases quite rapidly (e.g., Brown 1958, Muter 1980). Meanwhile, various features of the marketing communications being processed (e.g., modality of information, interrelations among components) can influence people's physiologic reactions that affect retention and retrieval (e.g., Janiszewski 1990, Rothschild and Hyun 1990), suggesting that the lasting effects could vary with different advertising formats.

Given the dynamic interaction among different formats of advertising and purchase discussed above, we hypothesize that a model of purchase conversion that fails to account for the interactions among various types of advertising will lead to biased estimation of the conversion effects of advertising and inferior predictive capability.

3. Data Overview

We obtained the data for this study from a major manufacturer and vendor of consumer electronics (e.g., computers and accessories) that sells most of its products online through its own website.² The firm records consumers' responses to its online advertisements in various formats. Every time a consumer clicks on one of the firm's online advertisements and visits the firm's website through it, the exact time of the click and the type of the online advertisement being clicked are recorded. Consumers are identified by user IDs that are primarily based on the tracking cookies stored on their computers.³ The firm also provided the purchase data (including the time of a purchase) associated with these user IDs. By combining the advertisement click and purchase data, we form a panel of individuals who have visited the firm's website through advertisements at least once, which comprises the entire history of clicking on different types of advertisements and purchasing by each individual.

² We are unable to reveal the identity of the firm because of a nondisclosure agreement.

³ The firm constructs the so-called generalized user IDs to identify users by linking the cookie IDs with other user identity information (such as user accounts and order IDs) whenever available. Although mitigated by this approach, the general limitations of cookie-based data still exist in our data set (Dreze and Zufryden 1998). Ideally, more advanced technology capable of identifying the same user across multiple devices would be preferable for more precise estimation results. Nevertheless, given that the technological reliability of cookie-based tracking is robust, and general users are increasingly receptive to the use of tracking cookies (Specific Media 2011), cookie data are commonly used in the literature studying consumer online behavior (e.g., Bucklin and Sismeiro 2003, De et al. 2010, Manchanda et al. 2006).

One unique aspect of our data is that, instead of being limited to one particular type of advertisement, our data offer a holistic view covering most major online advertising formats, which allows us to study the dynamic interactions among different types of advertisements. Because we are especially interested in the two leading formats of online advertising, namely, search and display advertisements, we categorize the advertisement clicks in our data set into three categories: *search*, *display*, and *other*. Search advertisements, also called sponsored search or paid search advertisements, refer to the sponsored links displayed by search engines on their search result pages alongside the general search results. Display advertisements, also called banner advertisements, refer to the digital graphics that are embedded in Web content pages and link to the advertiser’s website. The “other” category includes all the remaining types of online advertisements except search and display, such as classified advertisements (i.e., textual links included in specialized online listings or Web catalogs) and affiliate advertisements (i.e., referral links provided by partners in affiliate networks). Notice that our data only contain visits to the firm’s website through advertising links, and we do not have data on consumers’ direct visits (such as by typing the URL of the firm’s website directly in the Web browser) or visits through organic search results (i.e., general search results on search engine results pages). Therefore, we focus on the conversion effects of online advertisements rather than the general website visits. We will discuss data availability and limitations in more detail in §7.2.

For this study, we use a random sample of 12,000 user IDs spanning a four-month period from April 1 to July 31, 2008. We use the first three months for estimation and leave the last month as the holdout sample for out-of-sample validation. The data of the first three months contain 17,051 ad clicks and 457 purchases. Table 1 presents a detailed breakdown of different types of ad clicks. There are 2,179 individuals who have two or more ad clicks within the first three months, among whom 26.3% clicked on multiple types of advertisements.

We first perform a simple calculation of the conversion rates for different online advertisements, which are shown in Table 1. In calculating the conversion rates, we consider a certain ad click leads to a conversion if it is succeeded by a purchase of the same

individual within one day;⁴ we then divide the number of the ad clicks that lead to conversion by the total number of the ad clicks of the same type. Because of the nature of different types of advertisements, it is not surprising that their conversion rates vary significantly. The conversion rates presented in Table 1 are consistent with the general understanding in industry that search advertising leads all Internet advertising formats in terms of conversion rate, whereas display advertising has much lower conversion rates.⁵ Nevertheless, as was discussed earlier, the simple calculation of conversion rate attributes every purchase solely to the most recent ad click preceding the purchase. Naturally, it would be biased against those advertisements that are not likely to lead to immediate purchase decisions (e.g., display advertisements).

To further obtain intuition about the interacting dynamics among different types of advertisements and purchases before starting the modeling analysis, we summarize the strings of ad click and purchase sequences and perform some descriptive analysis. Because we are interested in how ad clicks could influence each other and together lead to conversion, we focus on those individuals with multiple ad clicks and purchase as well. We draw from the original data a random sample of 10,000 individuals who have at least two ad clicks and a purchase from April through June, 2008. We then summarize the strings of ad click and purchase sequences for all these individuals and obtain the counts for each unique sequence pattern. Table 2 presents the most frequent sequences.

A scrutiny of Table 2 reveals several noteworthy data patterns. First, sequences involving repeating clicks on the same types of ads before making a purchase are very common, indicating the imperative to account for the interactions among the same types of advertisements. Second, there is also a considerable portion that involves one type of ad click succeeded by another type of ad click, which indicates the necessity to capture the interactions among different types of advertisements as well. Third, display ads appear to be more likely to excite the other two types of ads than the other way around. Note that the number of display ad clicks is relatively low in absolute terms, as is shown in Table 1. When focusing on the sequences that contain display ad clicks, we can see that the sequences of “DSP” and “DOP” are more frequent than “SDP,” and “ODP” is absent from the top list. To explore further

Table 1 Data Description

	Number of ad clicks	Percentage of ad clicks (%)	Conversion rate
Search	6,886	40.4	0.01990
Display	3,456	20.3	0.00203
Other	6,709	39.3	0.01774

⁴ As we can show, choosing a different time interval longer or shorter than one day (e.g., several hours, a few days) delivers essentially the same outcomes with regard to conversion rate and conversion probability, which will be discussed in §6.1.

⁵ Industry reports show that search ads’ conversion rates vary with a median of 3.5% (MarketingSherpa 2012), whereas the average conversion rates for display ads lie between 0.15% and 0.2% (MediaMind 2010).

Table 2 Summary of Ad Click and Purchase Sequences

Rank	Sequence	Count	Rank	Sequence	Count
1	OOP	1,920	15	DOP	87
2	SSP	1,781	16	SOOP	83
3	OOOP	886	17	SDP	79
4	SSSP	635	18	OOOOOOOP	75
5	OOOOP	478	19	SSSSSP	71
6	SOP	327	20	DDDP	63
7	OOOOOP	287
8	SSSSP	274			
9	DDP	193			
10	OOOOOOP	189		Count 1	Count 2
11	OSP	166	Search	1,199	5,032
12	SSSSSP	145	Display	418	1,323
13	OOOOOOOP	119	Other	811	6,237
14	DSP	98			13.0

Notes. S, search ad click; D, display ad click; O, other ad click; P, purchase. Count 1 is the number of sequences that start with the particular type of ad click and also contain a different type of ad click. Count 2 is the number of all the sequences that contain the particular type of ad click.

along this line, we parse all sequences and calculate the percentage of the sequences that start with a particular type of ad click and contain a different type of ad click afterward among all the sequences that contain this particular type of ad click, as is shown in the lower-right panel of Table 2. The highest ratio for display ads thus suggests that ignoring the interactions among different types of ads could result in underestimating the conversion effect of display ads the most severely.

4. Model Development

To capture the interacting dynamics among different online advertising formats to properly evaluate their conversion effects, we propose a model based on mutually exciting point processes. We also account for heterogeneity among individual consumers, which casts our model in a hierarchical framework. In this section, we first provide a brief overview of mutually exciting point processes and then specify our proposed model in detail.

4.1. Mutually Exciting Point Processes

A point process is a type of stochastic process that models the occurrence of events as a series of random points in time or geographical space. For example, in the context of this study, a click on an online advertisement or a purchase can be modeled as a point occurring along the time line. We can describe such a point process by $N(t)$, which is an increasing non-negative integer-valued counting process such that $N(t_2) - N(t_1)$ is the total number of points that occurred within the time interval $(t_1, t_2]$. Most point processes can be fully characterized by the *conditional intensity function* defined as follows (Daley and Vere-Jones 2003):

$$\lambda(t | \mathcal{H}_t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{N(t + \Delta t) - N(t) > 0 | \mathcal{H}_t\}}{\Delta t}, \quad (1)$$

where \mathcal{H}_t is the *history* of the point process given the realization of the stochastic process up to time instant t , which includes all information and summary statistics before t (e.g., all points occurred and their respective occurrence times).⁶ The intensity measures the probability of instantaneous point occurrence given the previous realization. By the definition of Equation (1), given the event history \mathcal{H}_t , the probability of a point occurring within $(t, t + \Delta t]$ is $\lambda(t | \mathcal{H}_t)\Delta t$. Note that $\lambda(t | \mathcal{H}_t)$ is always positive by definition.

Hawkes (1971a, b) first systematically studied a class of point processes, namely, the mutually exciting point processes, in which past events affect the probability of future event occurrence, and different series of events interact with each other. A mutually exciting point process is a multivariate point process composed of multiple univariate point processes (or marginal processes), often denoted as $N(t) = [N_1(t), \dots, N_K(t)]$, such that the conditional intensity function for each marginal process can be written as

$$\lambda_k(t | \mathcal{H}_t) = \mu_k + \sum_{j=1}^K \int_{-\infty}^t g_{jk}(t-u) dN_j(u), \quad \mu_k > 0. \quad (2)$$

Here $g_{jk}(t-u)$ is the response function capturing the effect of the past occurrence of a type j point at time u on the probability of a type k point occurring at time t (for $u < t$). The most common specification of the response function takes the form of exponential decay such that

$$g_{jk}(\tau) = \alpha_{jk} e^{-\beta_{jk}\tau}, \quad \alpha_{jk} > 0, \beta_{jk} > 0. \quad (3)$$

As is indicated by Equation (2), the intensity for the type k marginal process, $\lambda_k(t | \mathcal{H}_t)$, is determined by the accumulative effects of the past occurrence of points of all types (not only the type k points, but also points of the other types), and meanwhile, such exciting effects decay over time, as captured by Equation (3). In other words, in a mutually exciting point process, the intensity for each marginal process at any time instant depends on the entire history of all the marginal processes. For this reason, the intensity itself is actually a random process, depending on the realization of the point process in the past.

It is worth noting that the commonly used Poisson process is a special point process such that the intensity does not depend on the history. The most common Poisson process is homogeneous, which means the intensity is constant over the entire process; that is, $\lambda(t | \mathcal{H}_t) \equiv \bar{\lambda}$. For a nonhomogeneous Poisson process, the intensity can be a deterministic function of the time but still independent of the realization of the stochastic process.

⁶ Mathematically, \mathcal{H}_t is a version of σ -field generated by the random process up to time t . Summary statistics such as how many points occurred before t or the passage of time since the most recent point are all probability events (sets) belonging to the σ -field \mathcal{H}_t .

4.2. The Proposed Model

The mutually exciting point process provides a flexible framework that well suits the nature of the research question of our interest. It allows us to model not only the effect of a particular ad click on future purchase, but also the dynamic interactions among ad clicks themselves, and all these effects can be neatly cast into a continuous-time framework to properly account for the time effect. We therefore construct our model based on mutually exciting point processes as follows.

For an individual consumer i ($i = 1, \dots, I$), we consider her interactions with the firm’s online marketing communication and her purchase actions as a multivariate point process, $N^i(t)$, which consists of K marginal processes, $N^i(t) = [N_1^i(t), \dots, N_K^i(t)]$. Each of her purchases as well as clicks on various online advertisements is viewed as a *point* occurring in one of the K marginal processes. The random variable $N_k^i(t)$ is a nonnegative integer counting the total number of type k points that occurred within the time interval $[0, t]$. We let $k = K$ stand for purchases and $k = 1, \dots, K - 1$ stand for various types of ad clicks. For our data, we consider $K = 4$ so that $N_4^i(t)$ stands for purchases and $N_1^i(t)$, $N_2^i(t)$, and $N_3^i(t)$ stand for clicks on search, display, and other advertisements, respectively. When individual i , for example, clicks on search advertisements for the second time at time t_0 , then a type 1 point occurs, and $N_1^i(t)$ jumps from 1 to 2 at $t = t_0$.

The conditional intensity function (defined by Equation (1)) for individual i ’s type k process is modeled as

$$\lambda_k^i(t | \mathcal{H}_t^i) = \mu_k^i \exp(\psi_k N_K^i(t)) + \sum_{j=1}^{K-1} \int_0^t \alpha_{jk} \exp(-\beta_j(t-s)) dN_j^i(s) \quad (4)$$

$$= \mu_k^i \exp(\psi_k N_K^i(t)) + \sum_{j=1}^{K-1} \sum_{l=1}^{N_j^i(t)} \alpha_{jk} \exp(-\beta_j(t-t_l^{(j)})), \quad (5)$$

for $k = 1, \dots, K$, where $\mu_k^i > 0$, $\alpha_{jk} > 0$, $\beta_j > 0$, and $t_l^{(j)}$ is the time instant when the l th point in individual i ’s type j process occurs. Note that time t here is continuous and measures the exact time lapse since the start of observation (using day as unit, e.g., for a time lapse of 32.5 hours, $t = 1.354$). Capable of dealing with continuous time directly, our modeling approach avoids the assumption of arbitrary fixed time intervals or the visit-by-visit analysis that merely considers visit counts and ignores the time effect.

The first component of the intensity λ_k^i specified in Equation (4) is the baseline intensity, μ_k^i . It represents the general risk of the occurrence of a particular type of event (i.e., an ad click or a purchase) for a particular individual, which can be a result of the consumer’s

inherent purchase intention, intrinsic tendency to click on certain types of online advertisements, and degree of exposure to the firm’s Internet marketing communication. Apparently, the baseline intensity varies from individual to individual. We hence model such heterogeneity among consumers by considering that $\mu^i = [\mu_1^i, \dots, \mu_K^i]$ follows a multivariate log-normal distribution,

$$\mu^i \sim \text{log-MVN}_K(\theta_\mu, \Sigma_\mu). \quad (6)$$

The multivariate log-normal distribution facilitates the likely right-skewed distribution of μ_k^i (> 0). In addition, the variance–covariance matrix Σ_μ allows for correlation between different types of baseline intensities; that is, for example, an individual having a higher tendency to click on display advertisements may also have a correlated tendency (higher or lower) to click on search advertisements.

In modeling the effects of ad clicks, we focus on their exciting effects on future purchases as well as on subsequent clicks on advertisements. As is discussed in §2, behavioral studies on consumers’ processing of advertising show that prior interactions with marketing communications can increase both perceptual and conceptual fluency, which directly contributes to the increase of the probability of consumers’ future responses to various types of advertisements. Furthermore, increased processing fluency leads to positive affective evaluation and consideration-set membership of the advertised brand, which not only increases the purchase probability directly, but also makes consumers more active to consult the focal firm’s advertisements for informational purposes. As a result, prior clicks on online advertisements could increase the probability of not only purchase conversion, but also later clicks on online ads in either the same or a different format. We hence model the effects of ad clicks in a form similar to Equation (3). We use α_{jk} ($j = 1, \dots, K - 1$ and $k = 1, \dots, K$) to measure the magnitude of increase in the intensity of type k process (i.e., ad clicks or purchase) when a type j point (i.e., a type j ad click) occurs. Hence, α_{jK} captures the direct effects of various types of ad clicks on purchase conversion, and α_{jk} ($k = 1, \dots, K - 1$) represents the exciting effects of prior ad clicks on later ad clicks in various formats. Among them, α_{jj} indicates the effect between the same type of points and is therefore called the *self-exciting* effect; for $j \neq k$, α_{jk} is the *mutually exciting* effects between different types of points.

As discussed in §2, the effects of prior ad clicks decay over time, as memory traces fade and retrieval failures occur with the passage of time. We therefore use β_j to measure how fast such effects decay over time. To keep our model parsimonious, we let $\beta_{jk} = \beta_j$ for all $k = 1, \dots, K$, which implies that the decay rates

of the exciting effects of a type j point on various types of processes would be the same.⁷ Whereas a larger α_{jk} indicates a greater exciting effect instantaneously, a smaller β_j means such exciting effect is more lasting.

The effects of purchases are different from the effects of ad clicks in at least two aspects. First, compared to a single click on an advertisement, a past purchase should have much more lasting effects on purchases and responses to advertising in the near future, especially given the nature of the products in our data (i.e., major personal electronics). With respect to the time frame of our study (i.e., three months), it is reasonable to consider such effects constant over time. Second, past purchases may impact the likelihood of future purchases and the willingness to respond to advertising in either a positive or negative way. A recent purchase may reduce the purchase need in the near future and thus lower the purchase intention and the interest in relevant ads; on the other hand, a pleasant purchase experience could eliminate purchase-related anxiety and build up brand trust, which would increase the probability of repurchase or further browsing of advertising information. Therefore, it is appropriate not to predetermine the sign of the effects of purchases. Based on these two considerations, we model the effects of purchases as a multiplicative term shifting the baseline intensity, $\exp(\psi_k N_k^i(t))$, so that each past purchase changes the baseline intensity of the type k process (i.e., purchase or one type of ad click) by $\exp(\psi_k)$, where ψ_k can be either positive or negative. A positive ψ_k means a purchase increases the probability of future occurrence of type k points, whereas a negative ψ_k indicates the opposite.

As is discussed earlier, the intensity $\lambda^i = [\lambda_1^i, \dots, \lambda_K^i]$ defined in Equation (5) is a vector random process and depends on the realization of the stochastic process $N^i(t)$ itself. As a result, λ^i keeps changing over the entire process. Figure 2 illustrates how the intensity of different marginal processes changes over time for a certain realization of the point process. It is also worth noting that the intensity function specified in Equation (5) only indicates the probability of event occurrence, whereas the actual occurrence could also be affected by many other unobservable factors, for example, unexpected incidents or impulse actions. In this sense, the model implicitly accounts for nonsystematic unobservables and idiosyncratic shocks.

⁷ The model can be easily revised into different versions by allowing β_{jk} to take different values. In fact, we also estimated two alternative models: one allowing values of β_{jk} 's to be different from each other, and the other considering that β_{jk} 's take the same value for $k = 1, \dots, K - 1$, which is different from β_{jk} . It is shown that the performance of our proposed model is superior to that of both alternative models: the Bayes factors of the proposed model relative to the two alternative models are $\exp(120.24) \simeq 1.7 \times 10^{52}$ and $\exp(190.18) \simeq 3.9 \times 10^{82}$, respectively.

Notice that Equation (4) implicitly assumes that the accumulative effects from the infinite past up to time $t = 0$, which is unobserved in the data, equal zero; that is, $\lambda_k^i(0) - \mu_k^i = 0$. In fact, the initial effect should not affect the estimates as long as the response function diminishes to zero at infinite and the study period is long enough. Ogata (1978) theoretically shows that the maximum likelihood estimates when omitting the history from the infinite past are consistent and efficient, as long as the data observation period is sufficiently long. Using simulated data, we can also show that Bayesian estimation using left-censored data can still recover true model parameters for the length of our observed period. The details of the simulation study are omitted because of page limitation and are available upon request.

Based on the intensity function specified in Equation (5), the likelihood function for any realization of all individuals' point processes $\{N^i(t)\}_{i=1}^I$ can be written as (Daley and Vere-Jones 2003)

$$L = \prod_{i=1}^I \prod_{k=1}^K \left\{ \left[\prod_{l=1}^{N_k^i(T)} \lambda_k^i(t_l^{k(i)} | \mathcal{H}_{t_l^{k(i)}}^i) \right] \cdot \exp\left(-\int_0^T \lambda_k^i(t | \mathcal{H}_t^i) dt\right) \right\}. \quad (7)$$

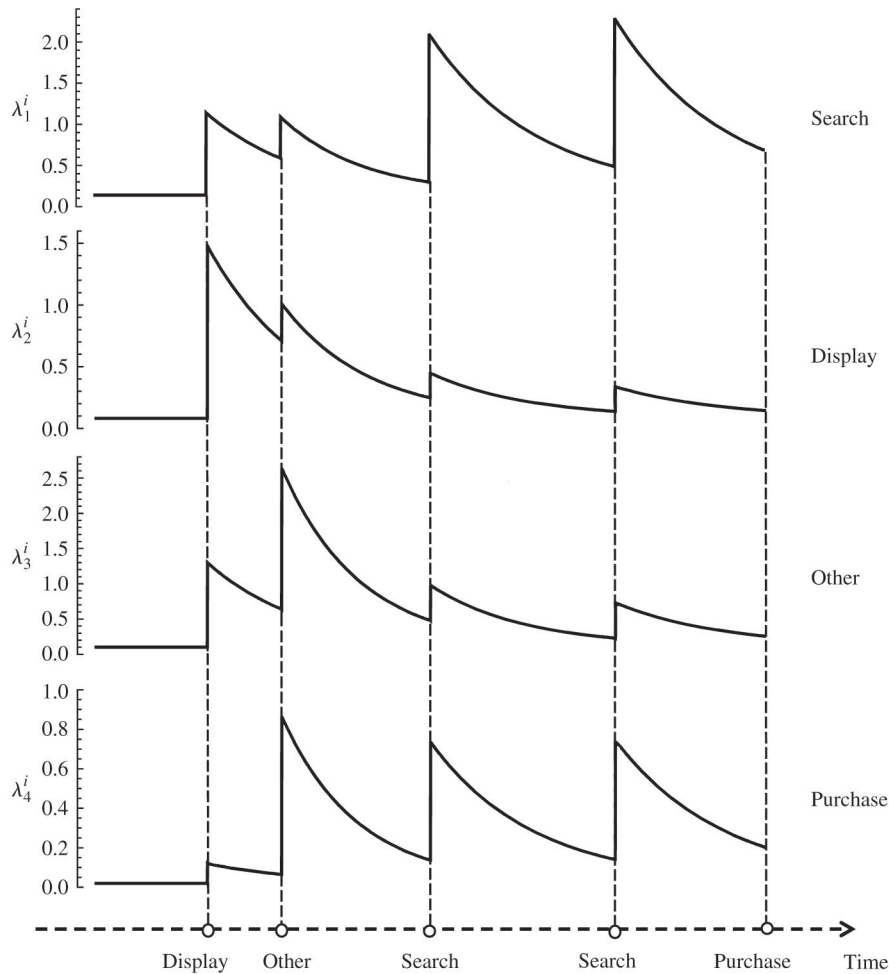
The likelihood function involving a stochastic integration can be derived in a full closed form, which is presented in detail in §A.1 of the appendix. It is worth emphasizing that unlike the typical conversion models in which advertising responses are treated only as explanatory variables for purchases, our model treats ad clicks also as random events that are impacted by the history, and hence their probability densities directly enter the likelihood function, in the same way as purchases. This fully multivariate modeling approach avoids the structure of conditional (partial) likelihood, which often arbitrarily specifies "dependent" and "independent" variables, resulting in statistically inefficient estimates for an observational study.

To summarize, we constructed a mutually exciting point process model with individual random effects. Given the hierarchical nature of the model, we cast it in the Bayesian hierarchical framework. The full hierarchical model is described as follows:

$$\begin{aligned} N^i(t) | \alpha, \beta, \psi, \mu^i &\sim \lambda^i(t | \mathcal{H}_t^i), \\ \mu^i | \theta_\mu, \Sigma_\mu &\sim \text{log-MVN}_K(\theta_\mu, \Sigma_\mu), \\ \alpha_{jk} &\sim \text{Gamma}(\bar{a}_\alpha, \bar{b}_\alpha), \quad \beta_j \sim \text{Gamma}(\bar{a}_\beta, \bar{b}_\beta), \quad \psi \sim \text{MVN}_K(\bar{\theta}_\psi, \bar{\Sigma}_\psi), \\ \theta_\mu &\sim \text{MVN}_K(\bar{\theta}_\mu, \bar{\Sigma}_\mu), \quad \Sigma_\mu \sim \text{IW}(\bar{S}^{-1}, \bar{\nu}), \end{aligned} \quad (8)$$

where α is a $(K - 1) \times K$ matrix whose (j, k) th element is α_{jk} , and $\beta = [\beta_1, \dots, \beta_{K-1}]$ and $\psi = [\psi_1, \dots, \psi_K]$ are both vectors. The parameters to be estimated are

Figure 2 Illustration of the Intensity Functions



$\{\alpha, \beta, \psi, \{\mu^i\}, \theta_\mu, \Sigma_\mu\}$. Notice that α, β, ψ , and $\{\mu^i\}$ play distinct roles in the data-generating process, and the model is therefore identified (Bowsher 2007).

4.3. Alternative and Benchmark Models

Our modeling framework is general enough to incorporate a class of nested models. We are particularly interested in a special case in which $\alpha_{jk} = 0$ for $j \neq k$ and $j, k = 1, \dots, K - 1$. It essentially ignores the exciting effects among different types of ad clicks. A past click on advertisement still has impact on the probability of future occurrence of purchases as well as ad clicks of the same type, but it will not affect the future occurrence of ad clicks of different types. Therefore, in contrast with our proposed *mutually exciting* model, we call this special case the *self-exciting* model, because it only captures the self-exciting effects among advertisement clicks.

For model comparison purpose, we are interested in different benchmark models for purchase conversion. A nested benchmark model within our point process modeling framework is letting $\alpha_{jk} = \psi_k = 0$ for all $j, k = 1, \dots, K - 1$. In other words, this benchmark

model completely ignores the exciting effects among all advertisement clicks. Ad clicks still have effects on purchases, but the occurrence of ad clicks themselves is not impacted by the history of the process (neither past ad clicks nor past purchases), and hence their intensities are taken as given and constant over time. As a result, the processes for all types of advertisement clicks are homogeneous Poisson processes, and we thus call this benchmark model the *Poisson process* model.

A commonly used model for purchase conversion is the binary response model that models the purchase conversion as a 0-or-1 variable with consumers' nonpurchase activities as explanatory variables. We are thus interested in a benchmark model in which the probability of an individual consumer (e.g., each month) is a logistic regression on the counts of her clicks on different types of online ads within the same time interval. We will refer to this benchmark model as the *logistic conversion* model. Notice that this logistic conversion model cannot be directly compared with the point process models using the model comparison criteria such as the deviance information criterion (DIC)

or log-marginal likelihood, because what constitute the random data within the likelihood function are different. The occurrence of ad clicks and the time intervals between them all enter the likelihood function as in Equation (7) in the point process models, whereas the likelihood function only accounts for the occurrence of purchase conversions in the logistic conversion model. Therefore, we compare our mutually exciting model with this benchmark model using out-of-sample prediction on purchase conversion in §6.2.

5. Estimation

To estimate the parameters in the model, we use the MCMC method for Bayesian inference. We apply the Metropolis–Hastings algorithm to sample the parameters. Section A.2 in the appendix presents the detailed steps of the MCMC algorithm. For each model, we ran the sampling chain for 50,000 iterations using the R programming language on a Windows workstation computer and discarded the first 20,000 iterations to ensure convergence.

5.1. Estimation Results

Before estimating the model using the real data, we first conduct a simulation study by estimating our model using simulated data. Results show that our model can correctly recover the true parameter values. The details of the simulation study are omitted because of page limitation and are available upon request. We

then apply the real observational data for estimation. We estimate the mutually exciting model for the data of the first three months. We report the posterior means and posterior standard deviations for major parameters in Table 3. (The estimates for 12,000 different μ^i 's are omitted due to the page limit.)

The estimation results in Table 3(a) demonstrate several interesting findings regarding the effects of online advertisement clicks. First of all, it is shown that there exist significant exciting effects between the same types of advertisement clicks as well as between different types of advertisement clicks. Compared to the baseline intensities for the occurrence of ad clicks (i.e., μ_j^i , $j = 1, 2, 3$), whose expected values ($\exp\{\theta_{\mu,j}\}$, $j = 1, 2, 3$) range from $\exp\{-6.10\} \simeq 0.0022$ to $\exp\{-5.39\} \simeq 0.0046$, the values of α_{jk} ($j, k = 1, 2, 3$) are greater by orders of magnitude. It implies that given the occurrence of a particular type of ad click, the probability of ad clicks of the same type or different types occurring in the near future is significantly increased. Therefore, the results underscore the necessity and importance of accounting for the dynamic interactions among advertisement clicks in studying their conversion effects.

Compared with the mutually exciting effects, self-exciting effects between the same type of advertisement clicks are more salient, as α_{jj} ($j = 1, 2, 3$) are greater than α_{jk} ($j \neq k$ and $j, k = 1, 2, 3$). This result is consistent with the observed data pattern that it is more common for

Table 3 Parameter Estimates for the Mutually Exciting Model

	Search	Display	Other	Purchase
(a) Exciting effects				
Search	α_{11} 2.8617 (0.1765)	α_{12} 0.0860 (0.0214)	α_{13} 0.5381 (0.0562)	α_{14} 0.6167 (0.0633)
Display	α_{21} 0.1614 (0.0496)	α_{22} 1.7818 (0.2314)	α_{23} 0.2055 (0.0572)	α_{24} 0.0845 (0.0347)
Other	α_{31} 0.4647 (0.0654)	α_{32} 0.1270 (0.0367)	α_{33} 8.0526 (0.4117)	α_{34} 0.8384 (0.0867)
Purchase	ψ_1 −0.5664 (0.1228)	ψ_2 −0.7556 (0.2348)	ψ_3 −0.6235 (0.1229)	ψ_4 0.2787 (0.2160)
Decay	β_1 34.0188 (1.7426)	β_2 46.8854 (4.9370)	β_3 51.5114 (2.3241)	
(b) Individual heterogeneity				
Mean	$\theta_{\mu,1}$ −5.3926 (0.0166)	$\theta_{\mu,2}$ −6.1027 (0.0212)	$\theta_{\mu,3}$ −5.8063 (0.0221)	$\theta_{\mu,4}$ −9.7704 (0.0762)
Covariance				
Search	$\Sigma_{\mu,11}$ 0.4584 (0.0246)			
Display	$\Sigma_{\mu,21}$ −0.1197 (0.0212)	$\Sigma_{\mu,22}$ 0.5934 (0.0335)		
Other	$\Sigma_{\mu,31}$ −0.4942 (0.0257)	$\Sigma_{\mu,32}$ −0.3380 (0.0304)	$\Sigma_{\mu,33}$ 1.0014 (0.0365)	
Purchase	$\Sigma_{\mu,41}$ 0.2256 (0.1228)	$\Sigma_{\mu,42}$ −0.4157 (0.1665)	$\Sigma_{\mu,43}$ 0.0762 (0.2440)	$\Sigma_{\mu,44}$ 2.3914 (0.2575)

Note. Posterior means and posterior standard deviations (in parentheses) are reported.

consumers to click on the same type of advertisement multiple times.

When we compare the mutually exciting effects between different types of advertisement clicks, interestingly, display advertisements tend to have greater exciting effects on the other two types of advertisement clicks than the other way round. The posterior probability of α_{21} being greater than α_{12} is 0.92, and the posterior probability of α_{23} being greater than α_{32} is 0.87. This result implies that when there is a sequence of clicks on different types of advertisements in a short time period, display ad clicks are more likely to occur at the beginning of the sequence than toward the end, because they are more likely to excite the other two types of ad clicks than to be excited by them.

Regarding the direct effects on purchase conversion, the values of α_{j4} ($j = 1, 2, 3$) are much greater compared with the baseline intensity for purchase occurrence (i.e., μ_4^i), whose expected value $\exp\{\theta_{\mu,4}\}$ is about $\exp\{-9.77\} \simeq 0.00006$. It indicates that clicking on an advertisement and visiting the firm’s website increase the probability of purchase directly, which is consistent with the previous findings in literature. Although all three types of advertisement clicks have direct conversion effects, display advertisement’s direct conversion effect (α_{24}) is much smaller, which partially explains the low conversion rate of display advertisements and the general understanding of its low conversion efficacy.

Past purchases are shown to negatively affect the probability of future clicks on advertisements. The parameters ψ_j for $j = 1, 2, 3$ take significantly negative values, whereas the effect on repeated purchases (i.e., ψ_4) is insignificant. It suggests that past purchases in general suppress consumers’ purchase need from this particular firm and thus diminish their interest in the firm’s online advertisements; although some consumers might make repeated purchases, as the positive ψ_4 suggests, they tend to make the repeated purchases directly, rather than through clicking advertising links again.

There are also interesting results regarding the variance–covariance matrix for individual baseline intensities in Table 3(b). First, notice that the covariances between the individual baseline intensities for any two types of ad clicks (i.e., $\Sigma_{\mu,21}$, $\Sigma_{\mu,31}$, $\Sigma_{\mu,32}$) are negative. In other words, a consumer having higher baseline intensity for clicking search advertisements, for example, is likely to have lower baseline intensity for clicking display advertisements. Such negative covariances imply that consumers are initially inclined to respond to one particular type of online advertisement, whereas clicking this particular type of advertisement may increase the probability of clicking other types of advertisements subsequently. In addition, it is interesting to find that the individual baseline intensity for

Table 4 Model Comparison Results

Model	DIC	Log-marginal likelihood
Mutually exciting	192,002.56	−93,454.69
Self-exciting	193,513.08	−94,219.22
Poisson process	206,146.80	−99,748.95

clicking display advertisements is negatively correlated with the individual baseline intensity for purchases; that is, $\Sigma_{\mu,42} < 0$. In other words, consumers who are more likely to respond to display ads usually have lower initial purchase intention, which partially explains the lower conversion rate of display ads.

5.2. Model Comparison

We next estimate the alternative self-exciting model and the benchmark Poisson process model and compare their goodness of fit with the mutually exciting model by computing the DIC and the log-marginal likelihood for the Bayes factor. In computing the log-marginal likelihood, we draw from the posterior distribution based on the MCMC sampling chain using the method proposed by Gelfand and Dey (1994). Table 4 shows the results of the model comparison criteria for the three models.

According to Table 4, the Bayes factor of the mutually exciting model relative to the self-exciting model is $\exp(-93,454.69 + 94,219.32) \simeq 1.0 \times 10^{332}$, and the Bayes factor of the mutually exciting model relative to the Poisson process model is $\exp(-93,454.69 + 99,748.95) \simeq 3.7 \times 10^{2,733}$. The mutually exciting model also has the lowest DIC value. Therefore, both the DIC and Bayes factor indicate that the proposed mutually exciting model outperformed the other two models by a great extent. Recall that the Poisson process model fails to capture any exciting effect among advertisement clicks at all. Given the estimation results showing that such effects do exist, it is not surprising that such a model performs poorly in terms of model fit. In contrast, the self-exciting model captures the exciting effects among the same type of advertisement clicks, which account for a considerable portion of the dynamic interactions among all advertisement clicks. Consequently, the self-exciting model improves noticeably beyond the Poisson process model. Nevertheless, its performance is still substantially inferior to that of the mutually exciting model because of its omission of the exciting effects between different types of advertisement clicks.

6. Model Applications

The existence of both mutually exciting and self-exciting effects indicated by the estimation results suggests the necessity of reassessing the effectiveness of different online advertising formats in a more proper approach. In this section, we apply our model and develop a

measure to evaluate the conversion effects of different type of online advertisements. To derive the probability of purchase occurring after clicking on a certain type of advertisement, we develop a simulation algorithm to simulate the mutually exciting processes specified in our model. The simulation approach also allows us to explore individual's future behavior, which we utilize for out-of-sample validation and prediction purposes. We also demonstrate how to generalize our model to incorporate additional information such as marketing mix variables.

6.1. Conversion Effect

By considering the occurrence of advertisement clicks as stochastic events, our modeling approach enables us to more precisely measure different advertisements' conversion effects by capturing the dynamic interactions among advertisement clicks themselves. In particular, it enables us to explicitly examine the probability of purchase occurring within a certain period of time given a click on a particular type of advertisement initially, which subsumes the cases where various subsequent advertisement clicks are triggered after the initial click and lead to the eventual purchase conversion altogether.

Formally, we define the *conversion probability (CP)* as follows. Suppose a representative consumer i clicked on a type k advertisement at time t_0 , and no click occurred in the history before t_0 .⁸ Then the conversion probability for type k advertisement in time period t given the parameters for the processes $\{\alpha, \beta, \psi, \mu^i\}$ can be defined as

$$CP_k(t; \mu^i, \alpha, \beta, \psi) \\ = \Pr(N_k^i(t_0 + t) - N_k^i(t_0) > 0 \mid N_k^i(t_0) - N_k^i(t_0^-) = 1), \quad (9)$$

where $k = 1, \dots, K - 1$. Note that $N_k^i(t_0^-)$ is defined as the limit of the type k ad click count up to but not including time instant t_0 , i.e., $N_k^i(t_0^-) = \lim_{t \uparrow t_0} N_k^i(t)$. By Equation (9), the conversion probability $CP_k(t)$ measures the probability of purchase conversion occurring within the time period t given a type k ad click occurred initially at time t_0 . Note that $CP_k(t)$ captures both the direct and the indirect effects of a type k advertisement click on purchase conversion, because the probability measure includes not only the cases in which a purchase occurs directly after the initial ad click (without any other points in between), but also those cases in which various advertisement clicks occur after the initial click and before the purchase conversion. Therefore, as a measure of the conversion effects of different types of advertisement clicks, the conversion probability defined

in Equation (9) manages to account for the exciting effects among advertisement clicks themselves.

Based on the Bayesian inference of our proposed model, we can define the *average conversion probability (ACP)* by taking the expectation over the posterior distribution of the model parameters, $p(\alpha, \beta, \psi, \theta_\mu, \Sigma_\mu \mid Data)$, as follows:

$$ACP_k(t) \\ = E[CP_k(t; \mu^i, \alpha, \beta, \psi) \mid Data] \\ = \int CP_k(t; \mu^i, \alpha, \beta, \psi) p(\mu^i \mid \theta_\mu, \Sigma_\mu) \\ \cdot p(\alpha, \beta, \psi, \theta_\mu, \Sigma_\mu \mid Data) d\mu^i d\alpha d\beta d\psi d\theta_\mu d\Sigma_\mu. \quad (10)$$

Note that we are interested in the average conversion probability of different types of advertisements for a representative consumer (i.e., a typical consumer) rather than any specific consumer in the data set. Therefore, in Equation (10), the conversion probability is averaged over the distribution of individual baseline intensities, $p(\mu^i \mid \theta_\mu, \Sigma_\mu)$, as is specified in Equation (6), instead of using the posterior distribution $p(\mu^i \mid Data)$ for a specific consumer i .

Given the complexity of the mutually exciting point processes, the conversion probability cannot be explicitly derived in a closed form. Instead, we use the Monte Carlo method to calculate such probabilities. For this purpose, we develop an algorithm to simulate the mutually exciting point processes in our proposed model. This simulation algorithm is an extension of the thinning algorithm for self-exciting point processes (Ogata 1981) to mutually exciting point processes with posterior samples of the model parameters. The algorithm details are presented in §A.3 of the appendix. Here, we provide a brief overview of this simulation algorithm. The basic idea of this algorithm is similar to the typical acceptance–rejection Monte Carlo method: we first simulate a homogeneous Poisson process with a high intensity and then drop some of the extra points probabilistically according to the actual conditional intensity function. More specifically, we first draw the model parameters from the MCMC posterior sample and draw the individual baseline intensity as well. We then find a constant intensity that dominates the aggregate intensity function of the mutually exciting point process. We can thus simulate the next point of the homogeneous Poisson process with this constant dominating intensity by generating the time interval from an exponential distribution. Next, we probabilistically reject this point according to the ratio of the aggregate intensity of the mutually exciting point process to the constant intensity of the Poisson process. Finally, we assign a type to the generated point using the intensities for different types of points as probability weights.

Applying the above algorithm to repeatedly simulate the point processes in our model, we can

⁸ In reality, an “initial click” can be approximated as long as there was no click in the recent history before t_0 .

Table 5 Average Conversion Probabilities (%) of Different Advertisement Formats

Model	Search	Display	Other
Conversion rate	1.990	0.203	1.774
Mutually exciting	2.030	0.246	1.978
Self-exciting	2.005	0.238	1.968
Poisson process	1.818	0.234	1.703

approximate the average conversion probability in Equation (10) by

$$\begin{aligned}
 ACP_k(t) &= \int E[I\{N_K^i(t_0+t) - N_K^i(t_0) > 0\} | N_K^i(t_0) - N_K^i(t_0-) = 1] \\
 &\quad \cdot p(\mu^i | \theta_\mu, \Sigma_\mu) \\
 &\quad \cdot p(\alpha, \beta, \psi, \theta_\mu, \Sigma_\mu | Data) d\mu^i d\alpha d\beta d\psi d\theta_\mu d\Sigma_\mu \\
 &\simeq \frac{1}{R} \sum_{r=1}^R I\{N_K^{i(r)}(t_0+t) - N_K^{i(r)}(t_0) > 0\}, \tag{11}
 \end{aligned}$$

where R is the total number of simulation rounds, $N^{i(r)}(t)$ is the point process simulated in the r th round, and $I\{\cdot\}$ is the indicator function such that $I\{N_K^{i(r)}(t_0+t) - N_K^{i(r)}(t_0) > 0\}$ equals 1 if there is at least one purchase point within the time interval $(t_0, t_0 + t]$ in the r th simulated point process.

We use the approach described above to compute the average conversion probabilities for search, display, and other types of advertisements based on the Bayesian inference outcome of our proposed model. For each $k \in \{1, 2, 3\}$, we run the simulation for 1,000,000 times to compute $ACP_k(t)$ according to Equation (11).⁹ We choose the time interval t equal to one day so that the average conversion probabilities for each type of advertisements are directly comparable to their conversion rates as in §3. Table 5 presents the average conversion probabilities of different advertisement formats computed based on our proposed mutually exciting model (the second row) in contrast to their conversion rates (the first row).

As a common measure of the effectiveness of various online advertisement formats, conversion rates simply attribute a purchase completely to the last advertisement click preceding it. As a result, for those advertisement formats that tend to be used as the last stop before a purchase action, such as search advertisements, their conversion effects can easily be amplified by such a measure. On the contrary, for those advertisement formats that are more likely to attract consumers' initial attention but less likely to directly lead to immediate purchase decision, such as display

advertisements, their contribution are largely ignored. The results presented in the first two rows of Table 5 confirm such bias against display advertisements by the measure of conversion rates. If we compare the ratio of the conversion rates of display advertisements versus search advertisements (i.e., 0.102) with the ratio of their average conversion probabilities based on our proposed mutually exciting model (i.e., 0.121), we find that the relative conversion effect of display advertisements is underestimated by as much as 15.7%. Such underestimation originates from the fact that although display advertisements have little direct effect on purchase conversion (recall that display advertisements have the lowest direct conversion effect, i.e., α_{24} is much lower than α_{14} and α_{34} according to the estimation results), they may stimulate subsequent clicks on other types of advertisements, which in turn leads to the purchase conversion. In contrast, the proposed measure of average conversion probability properly captures such contribution from display advertisements. Notice that the relative conversion rates often serve as an important guide for marketing managers to determine their portfolios of online advertising spending and for advertising providers to price their advertising vehicles. In this sense, our analysis results suggest that display advertisements might have long been undervalued in the online advertising practice.

To further investigate how neglecting different types of exciting effects among advertisement clicks would affect the estimation of their conversion effects, we use the same approach and compute the average conversion probabilities for the self-exciting and Poisson process models based on their respective model inference outcomes. The results are presented in the third and fourth rows of Table 5.

Comparing the conversion probabilities evaluated based on the self-exciting model (the third row of Table 5) with those based on the mutually exciting model (the second row of Table 5), we can see that the conversion effect of display advertisements is underestimated by 3.3% by the self-exciting model, whereas the conversion effects of search and other advertisements are underestimated by 1.2% and 0.5%, respectively. Notice that in comparison with our proposed mutually exciting model, the nested self-exciting model captures the exciting effects only among the same type of advertisement clicks, but ignores the exciting effects between different types of advertisement clicks. This result thus suggests that among all online advertising formats we studied, display advertisements have the most salient effects in stimulating subsequent clicks on advertisements of different types. If we ignore the mutually exciting effects among different types of advertisements, display advertisements' conversion effects would be underestimated the most severely.

⁹ We sample 1,000,000 times with replacement from the 30,000 posterior samples to complete this exercise.

If we further compare the conversion probabilities evaluated based on the Poisson process model (the fourth row of Table 5) with those based on the self-exciting model (the third row of Table 5), it is clear that the Poisson process model underestimates the conversion effect of search advertisements more greatly than display advertisements. Recall that compared with the self-exciting model, Poisson process model further ignores the self-exciting effects among the same type of advertisement clicks. This result thus indicates that search advertisements have more salient self-exciting effects; that is, a search advertisement click is more likely to be succeeded by further clicks on the same type of advertisements, which altogether lead to the purchase conversion. Therefore, ignoring such self-exciting effects would underestimate the conversion effects of search advertisements more severely than display advertisements. In conclusion, to obtain an unbiased assessment of different advertisements' conversion effects, it is important to account for the mutually exciting effects as well as the self-exciting effects among advertisement clicks.

6.2. Prediction and Validation

The simulation algorithm developed to evaluate the conversion probabilities also enables us to predict each individual's future behavior based on their historical data. It allows us to perform out-of-sample validation of our proposed model and compare model performances in terms of predictive power.

Recall that in our data set that spans a four-month period from April through July, 2008, we use the data of the first three months for model estimation and leave the fourth month's data as a holdout sample. To perform out-of-sample validation, we randomly select a sample of 1,000 individuals out of all individuals used for estimation. For each individual, we predict their advertisement clicking and purchasing behaviors for the fourth month (31 days) based on their past behaviors in the previous three months (91 days) and the Bayesian inference for model parameters obtained during the estimation step. The algorithm used to simulate individuals' future behaviors is similar to the one developed in §6.1. The primary difference is twofold: the baseline intensities μ^i no longer reflect a representative consumer, but are now individual specific and are drawn from the posterior distribution for each specific individual obtained during the estimation step; the initial effects at the beginning of the simulated processes are the accumulated effects of the actual past behavior for each specific individual over the first three months. (See §A.3 in the appendix for more details.)

For each of the selected 1,000 individuals, we simulate 10,000 point processes according to the proposed model. We then calculate the predicted average numbers of purchases and ad clicks in the fourth month over

Table 6 Average Numbers of Ad Clicks and Purchases per Customer in the Fourth Month

	Search	Display	Other	Purchase
Actual data	0.12	0.070	0.14	0.013
Model	0.16	0.074	0.13	0.013
prediction	(0.065, 0.60)	(0.027, 0.23)	(0.045, 0.82)	(0.0083, 0.056)

the 1,000 individuals by taking the median values from the 10,000 sets of simulation outcomes. We also construct the 95% interval of these numbers based on the simulation outcomes. We contrast the predicted numbers and intervals with the actual data from the holdout sample. Table 6 presents the out-of-sample validation results, which show that the actual data all fall into the 95% predictive intervals and are quite close to the predicted values, indicating that the proposed model adequately captures the complex dynamics underlying consumers' online advertisement clicking and purchasing processes.

As out-of-sample prediction can provide statistically corroborating evidence for the model comparison results in Table 4, we next compare the predictive performance across different models. For the self-exciting and the Poisson process models, we use the same simulation approach to forecast individual behaviors in the fourth month for the same predictive sample based on the Bayesian estimation outcomes from the two models. For each of the three comparative models, we calculate the predicted numbers of purchases and advertisement clicks of different types for each individual by averaging over the 10,000 simulated processes, and then we compute the sum of squared errors (SSE) between the predicted numbers and the observed data from the holdout sample. Table 7 shows the average sum of squared errors over the 1,000 selected individuals for the three models.

As we can see from Table 7, the out-of-sample performances confirm the model comparison results based on the DIC and Bayes factors reported in Table 4. The proposed mutually exciting model has the lowest average sum of squared errors and thus performs the best in terms of both out-of-sample predictive power and within-sample model fit. In comparison, the nested self-exciting model underperforms in predictive power only slightly thanks to the capture of a considerable portion of the exciting effects among advertisement clicks. Ignoring all exciting effects among advertisement clicks completely, the Poisson process model

Table 7 Model Comparison for Out-of-Sample Prediction

Model	Mutually exciting	Self-exciting	Poisson process
Average sum of squared errors	1.514	1.545	1.705

demonstrates the poorest model performance in all the three criteria.

As is noted in §4.3, we are also interested in comparing the model performance with the logistic conversion model, a benchmark representing the commonly used binary response model for purchase conversion. To compare the out-of-sample prediction performance of this model with the mutually exciting model, we use the logistic conversion model to predict the probability of purchase occurring in the fourth month for the 1,000 individuals in the predictive sample randomly selected above. The numbers of each individual's various types of ad clicks in the fourth month are simulated using Poisson distributions, whose parameters are drawn from the posterior distributions estimated by Bayesian inference given the previous three months' data. The simulated ad click numbers are then plugged into the logistic regression outcomes to calculate the predicted purchase probability in the fourth month for each individual. Out-of-sample prediction performance is compared using the SSE between the predicted purchase probability and the observed purchase occurrence. We find that the logistic conversion model has an SSE equal to 18.838, whereas the mutually exciting model has an SSE equal to 14.475. The substantial difference in prediction accuracy (over 30%) thus confirms the superior performance of our proposed model.

6.3. Incorporating Additional Information

Marketing researchers are often interested in gaining richer implications by incorporating marketing mix variables (such as prices and promotions) and consumer demographic information (such as age, gender, and income level) into the model when these data are available. Our modeling framework is general enough to facilitate such extension, as we demonstrate below.

To incorporate individual-specific and/or time-varying covariates into the model, we can allow the individual baseline intensity μ^i to take a more general form, $\mu^i(t)$, which is dependent on these newly introduced explanatory variables. Specifically, we can rewrite the conditional intensity function in Equation (5) as

$$\lambda_k^i(t | \mathcal{H}_t^i) = \mu_k^i(t) \exp(\psi_k N_k^i(t)) + \sum_{j=1}^{K-1} \sum_{l=1}^{N_j^i(t)} \alpha_{jk} \exp(-\beta_j(t - t_l^{j(i)})), \quad (12)$$

where

$$\log \mu^i(t) = \eta^i + \Phi X_i(t), \quad (13)$$

$$\eta^i \sim MVN_K(\theta_\mu + \Omega Z^i, \Sigma_\mu).$$

Here $\mu^i(t) = [\mu_1^i(t), \dots, \mu_K^i(t)]'$ and $\eta^i = [\eta_1^i, \dots, \eta_K^i]'$ are both vectors; $X_i(t)$ are M -dimensional time-varying covariates (e.g., marketing mix variables); Z^i are L -dimensional individual-specific covariates (e.g., consumer demographic information); and Φ and Ω are

$K \times M$ and $K \times L$ matrices of model parameters, respectively. We can estimate the extended model using similar Bayesian estimation strategy as the original model and applying additional Metropolis and Gibbs sampling steps for Φ and Ω . The main challenge lies in the integral over the time-varying covariates in computing the likelihood function. Complex as it might become, the likelihood function can be derived in a closed form in most cases (e.g., when $X_i(t)$ is a step function over time, such as prices). In other cases, numerical integration or Monte Carlo methods should always apply.

To illustrate how the above model extension can be implemented in the context of our study, we acquire additional data from the focal firm on its daily average markup ratio to account for the firm's marketing mix during the study period. The daily average markup ratio is calculated by averaging the ratio between the sales price and the cost over all orders placed every day from April to June, 2008. The data exhibit stable pattern and indicate no major change in the firm's sales and promotion strategy during the study period. We use these data as the $X_i(t)$ variables in Equation (13) and introduce four new parameters, $\Phi = [\phi_1, \phi_2, \phi_3, \phi_4]'$, into the original mutually exciting model. The likelihood function and the MCMC algorithm for estimation are detailed in §A.4 of the appendix. The estimation results show that our original results are robust after controlling the marketing mix. The estimates of all major parameters $\{\alpha, \beta, \psi\}$ from the extended model are very close to those from the original model, and the posterior mean values of Φ are unsurprisingly negative. Because of the page limit, we omit the detailed estimation results in this paper, which are available upon request. In addition, we also compute the fit statistics of the extended model: the DIC is 192,050.06, and the log-marginal likelihood is $-94,063.8$, which are both inferior to those of the original mutually exciting model as in Table 4. It thus indicates that incorporating this additional marketing mix component in the context of our study does not improve the model performance.

7. Discussion and Conclusion

In this paper, we develop a Bayesian hierarchical model that incorporates the mutually exciting point process and individual heterogeneity to study the conversion effects of different online advertising formats. The mutually exciting point process offers us a flexible framework to model the dynamic and stochastic interactions among online consumers' advertisement clicking and purchasing behaviors. To account for heterogeneity among consumers, our model allows them to have different propensities for ad clicking and purchasing using random effects for their baseline intensities. We successfully apply the MCMC method to obtain

Bayesian inference for our model. We construct conversion probability based on our proposed mutually exciting model to properly evaluate the conversion effects of various types of online advertisements. To compute the conversion probability and predict consumers' future behaviors, we develop a simulation algorithm by extending the existing algorithm to mutually exciting point processes with posterior sampling of parameters. Using proprietary data from a major vendor of consumer electronics, we demonstrate that our proposed mutually exciting model has superior goodness of fit and leads to proper evaluation of conversion effects by successfully capturing the exciting effects among advertisement clicks.

7.1. Related Modeling Frameworks

Empirical studies of conversion and attribution through multiple advertising channels typically require multivariate models. As we discuss below, compared with other existing multivariate modeling frameworks, the mutually exciting point process provides a flexible continuous-time model to accommodate multivariate event data with precise time-stamp records and account for the complete event history.

When the data are only recorded as counts within fixed time periods, a Poisson count model with latent multivariate linear or simultaneous equation models for the unobserved Poisson parameters has been successfully applied in marketing (e.g., Dong et al. 2011). However, if the original data are disaggregate (e.g., generated by a continuous-time process), temporal aggregation such as converting time-stamped incidences into daily or monthly counts has been known to cause biased estimation of the magnitude and duration of advertising effects (Leone 1995, Tellis and Franses 2006). Indeed, when daily aggregation is applied to our data, it is obvious that the loss of ordering in clumpy ad-click and purchase events within a day can cause erroneous estimation of the advertising effects. To test it, we aggregate our data into daily counts and test a version of the multivariate Poisson count model, where (a) the Poisson parameter for number of purchases is a log-linear model of the ad clicks in the same period and the lagged purchase counts, and (b) the Poisson parameters for the numbers of various types of ad clicks are log-linear models of the lagged ad clicks and purchases. We compare out-of-sample prediction with our model and find that its performance is substantially inferior (the prediction is inferior to that of the Poisson process model in Table 7). Therefore, instead of selecting a time interval for aggregation, the mutually exciting point process is a natural and needed development in modeling multivariate event data in continuous time.

Another approach to modeling multivariate event data is to correlate interincidence durations with

Sarmanov class distributions (e.g., Park and Fader 2004) or, more generally, copulas (e.g., Danaher and Smith 2011). A Sarmanov class model, which can be shown as a special case of copulas, does not easily extend to more than three dimensions (Danaher and Smith 2011). The number of parameters in the Sarmanov class distribution increases exponentially with the model dimension, and it often involves considerable algebraic derivation of the likelihood function. For our research context, in which past purchases and clicks on the various types of ads can all influence the probability of clicking or buying at the current time, correlating high-dimensional interincidence durations by the Sarmanov distribution could be computationally impractical. Other copula models can be more computationally tractable. But these models still have to assume that the current time hazard is correlated with a fixed number of concurrent and past durations, which amounts to the Markov property (Park and Fader 2004). For our data, in which a random number of events often occur in a clustered fashion, it is hard to justify selecting an arbitrary number of correlated durations and assuming it is a Markov process. In contrast, the mutually exciting point process does not assume Markov property by allowing dependence on all previous events and time intervals and letting their effects decay over time, which is a more natural choice in our research context.

7.2. Data Availability and Limitation

In addition to the ad clicking data, other online activity data may also be relevant in studying the conversion effects of online advertisements, for example, users' direct visit data (i.e., directly visiting the firm's website without clicking any advertising link) or ad exposure data (i.e., individual's level of exposure to various advertising broadcasting including display ads, television ads, etc.). Thanks to the abundance of digital footprints, these types of data are becoming increasingly available to researchers (e.g., Zantedeschi et al. 2013). Our modeling framework is general enough to incorporate these data once available. For example, we can incorporate direct visits into our current model as an additional marginal process (i.e., the $(K + 1)$ th process), which can be influenced by other processes and contributes to purchase conversions as well. We can further extend the model specification in §6.3 to allow $X_i(t)$ to include each individual's time-varying ad exposure levels to capture their effects on the occurrence of ad clicks and purchases.

Neither direct visit data nor ad exposure data are available in our data set. We would thus like to discuss to what extent our results are robust given such data limitation. Consumers might visit the website directly a number of times after clicking on an ad before making a purchase, and these direct visits might contribute to

the conversion as well. Without observing the direct visits, the expected conversion effects of all these unobserved events are implicitly incorporated in the model parameters formulating the conditional intensity for purchase occurrence as in Equation (4). If direct visits were observed, their effects would be separated out from the current model parameters and explicitly represented by the model parameters for the additional $(K + 1)$ th process. Therefore, in the two cases with and without direct visit data, the model parameters incorporate different effects and hence should be interpreted differently. Nevertheless, the conversion probability as we estimate in this paper, which delivers the main implications on ad effectiveness, would remain robust, because, by its definition, conversion probability takes expectation over all possible events (either explicitly modeled or implicitly incorporated) happening between an initial ad click and the final purchase conversion. Likewise, without observing the ad exposure data, we incorporate the expected ad exposure level for each individual in the individual-specific baseline intensity. The actual ad exposure data would generally vary over time if they were observed. As long as the temporal variation of the exposure data is stationary without systematic trends or disproportionately large shocks, the conversion probability will remain robust after averaging across all individuals over time. However, if there were dramatic or systematic variations in ad exposure levels (e.g., caused by major changes in the firm's advertising strategy), then missing advertising exposure data could bias the estimates. Nonetheless, to our knowledge, there was not any major change in the firm's marketing efforts during the period of our study, as is verified by the top marketing manager of the firm furnishing the data. It thus sustains our confidence in the robustness of our results.

7.3. Conclusion

This study provides valuable managerial implications for marketing managers seeking optimal online advertising strategies as well as Internet advertising providers. We underscore a new perspective in measuring the effects of online advertisement clicks on purchase conversion. We suggest that to properly assess the conversion effects of various types of online advertisements, it is inadequate to merely focus on the direct effects of advertisement clicks on purchase probabilities per se. Even though an advertisement click does not lead to immediate purchase, it may increase the probabilities of subsequent clicks through other formats of advertisements, which in turn contribute to the final conversion. Such indirect contribution should not be neglected in evaluating the conversion effects of advertisements, which calls for novel modeling methods. Our proposed mutually exciting model and the associated conversion probability provide marketing

managers and Internet advertising providers with an innovative method readily applicable to the proper measurement of the efficacy of online advertisements actualizing this particular perspective.

The results from our analysis shed new light on the understanding of the effectiveness of different types of online advertisements. We show that display advertisements are likely to stimulate subsequent visits through other online advertisement formats, such as search advertisements, though they have a low direct effect on purchase conversions. Neglecting such effects and overemphasizing the "last click" effects, the commonly used measure of conversion rate is biased toward search advertisements and underestimates the relative effectiveness of display advertisements the most severely. For decision makers who are to allocate an advertising budget among various online advertising formats, our results suggest display advertisements have not been given their due share of appreciation, and a rebalance of the advertising spending portfolio could optimize the return on investment. On the other hand, a better understanding of the effectiveness of different online advertising formats can help online advertising providers to reassess their pricing strategies for these online advertising vehicles.

In addition, our method furnishes a useful tool for Internet marketers to assess the future values of their potential customers and target their marketing efforts. We demonstrate the superior predictive power of our model in forecasting consumers' future advertisement clicking and purchasing behaviors. Beyond the typical predictive models for future purchase activities, our modeling approach also enables us to predict nonpurchase activities at the same time. The ability to predict future responses to different online advertising formats is especially important for online marketing managers to deliver targeted advertisements to potential customers in an effective manner.

This study also leads to interesting directions for future research. For instance, because we provide an approach to more properly estimate the conversion probabilities for various types of online ads, future studies that combine the cost information for different types of online advertisements with our findings can help marketers design a more efficient budget allocation scheme for online advertising. In addition, given that our data structure only involves user clicks on advertisements, the exciting effects we study focus on the effects of prior ad clicks on the probability of later ad clicks occurring. As discussed previously, when additional user online behavior data become available (e.g., advertising exposures, direct website visits), our model can be easily adapted to incorporate such data to deliver richer and deeper implications about more detailed processes of consumer responding to online advertisements. Moreover, our general

modeling framework can be applied to many other contexts involving multiple interdependent activities with individual heterogeneity in continuous time.

Acknowledgments

The authors thank Eric Bradlow, the associate editor, and two anonymous reviewers, whose insightful and constructive comments greatly improved this paper. The authors also thank Yu (Jeffrey) Hu, Szu-Chi Huang, Preston McAfee, Carl Mela, Eric Overby, Sandra Slaughter, Frenkel ter Hofstede, and D. J. Wu for generous input. They thank the seminar participants at the Georgia Institute of Technology, Joint Statistical Meetings (2011), the Atlanta–Athens Conference on Research in Information Systems (2012), the Workshop on Information Systems and Economics (2012), and the China Summer Workshop on Information Management (2013) for helpful feedback.

Appendix

A.1. Likelihood Function

The likelihood function for individual i can be written as follows:

$$L_i(\alpha, \beta, \psi, \mu^i | Data^i) = \prod_{k=1}^K \left\{ \prod_{l=1}^{N_k^i(T)} \left[\mu_k^i \exp(\psi_k N_K^i(t_l^{k(i)})) \right. \right. \\ \left. \left. + \sum_{j=1}^{K-1} \sum_{m=1}^{N_j^i(t_l^{k(i)})} \alpha_{jk} \exp(-\beta_j (t_l^{k(i)} - t_m^{j(i)})) \right] \right. \\ \left. \cdot \exp \left[-\mu_k^i \sum_{m=0}^{N_k^i(T)} \exp(\psi_k m) (t_{m+1}^{K(i)} - t_m^{K(i)}) \right. \right. \\ \left. \left. - \sum_{j=1}^{K-1} \sum_{m=1}^{N_j^i(T)} \frac{\alpha_{jk}}{\beta_j} (1 - \exp(-\beta_j (T - t_m^{j(i)}))) \right] \right\}.$$

The likelihood function for all individuals is thus

$$L(\alpha, \beta, \psi, \mu | Data) = \prod_{i=1}^I L_i(\alpha, \beta, \psi, \mu^i | Data^i),$$

where $\mu = \{\mu^i\}_{i=1}^I$.

A.2. MCMC Algorithm

Below are the details of the MCMC algorithm used in this study.

Step 1: Sample α . We consider the prior distribution of α_{jk} following $Gamma(\bar{a}_\alpha, \bar{b}_\alpha)$, where $\bar{a}_\alpha = \bar{b}_\alpha = 1/2$. We use the Metropolis–Hastings algorithm to sample α . The proposal function is a random-walk generation such that the next draw α^* is drawn from a log-normal distribution, i.e., $\alpha_{jk}^* \sim \log-N(\log(\alpha_{jk}), \bar{\sigma}_{jk}^2)$. We adjust the variance $\bar{\sigma}_{jk}^2$ adaptively such that the acceptance rate of proposed α^* is between 0.1 and 0.4. The accepting probability for α^* is given as

$$\min \left\{ \frac{L(\alpha^*, \beta, \psi, \mu | Data) \prod_{j,k} Gamma(\alpha_{jk}^* | \bar{a}_\alpha, \bar{b}_\alpha) \cdot \prod_{j,k} \alpha_{jk}^*}{L(\alpha, \beta, \psi, \mu | Data) \prod_{j,k} Gamma(\alpha_{jk} | \bar{a}_\alpha, \bar{b}_\alpha) \cdot \prod_{j,k} \alpha_{jk}}, 1 \right\},$$

where the last terms in the numerator and denominator correspond to the Jacobian in the log-normal proposal function.

Step 2: Sample β . We consider the prior distribution of β_j following $Gamma(\bar{a}_\beta, \bar{b}_\beta)$, where $\bar{a}_\beta = \bar{b}_\beta = 1/2$. We use the Metropolis–Hastings algorithm to sample β . The proposal function is a random-walk generation such that the next draw β^* is drawn from a log-normal distribution, i.e., $\beta_j^* \sim \log-N(\log(\beta_j), \bar{\sigma}_j^2)$. We adjust the variance $\bar{\sigma}_j^2$ adaptively such that the acceptance rate of proposed β^* is between 0.1 and 0.4. The accepting probability for β^* is given as

$$\min \left\{ \frac{L(\alpha, \beta^*, \psi, \mu | Data) \prod_j Gamma(\beta_j^* | \bar{a}_\beta, \bar{b}_\beta) \cdot \prod_j \beta_j^*}{L(\alpha, \beta, \psi, \mu | Data) \prod_j Gamma(\beta_j | \bar{a}_\beta, \bar{b}_\beta) \cdot \prod_j \beta_j}, 1 \right\}.$$

Step 3: Sample ψ . We consider the prior distribution of ψ following $MVN_K(\bar{\theta}_\psi, \bar{\Sigma}_\psi)$, where $\bar{\theta}_\psi = 0$, and $\bar{\Sigma}_\psi = 10^4 I_K$ (I_K is the $K \times K$ identity matrix). We use the Metropolis–Hastings algorithm to sample ψ . The proposal function is a random-walk generation such that the next draw ψ^* is drawn from a normal distribution, i.e., $\psi_k^* \sim N(\psi_k, \bar{\sigma}_k^2)$. We adjust the variance $\bar{\sigma}_k^2$ adaptively such that the acceptance rate of proposed ψ^* is between 0.1 and 0.4. The accepting probability for ψ^* is given as

$$\min \left\{ \frac{L(\alpha, \beta, \psi^*, \mu | Data) MVN_K(\psi^* | \bar{\theta}_\psi, \bar{\Sigma}_\psi)}{L(\alpha, \beta, \psi, \mu | Data) MVN_K(\psi | \bar{\theta}_\psi, \bar{\Sigma}_\psi)}, 1 \right\}.$$

Step 4: Sample μ^i . We consider the prior distribution of each μ^i following a K -dimensional multivariate log-normal distribution $\log-MVN_K(\theta_\mu, \Sigma_\mu)$. We use the Metropolis–Hastings algorithm to sample μ^i . The proposal function is a random-walk generation such that the next draw μ^{i*} is drawn from a log-normal distribution, i.e., $\mu_k^{i*} \sim \log-N(\log(\mu_k^i), \bar{\sigma}_k^2)$. We adjust the variance $\bar{\sigma}_k^2$ adaptively such that the average acceptance rate of all proposed μ^{i*} 's is between 0.1 and 0.4. The accepting probability for μ^{i*} is given as

$$\min \left\{ \frac{L_i(\alpha, \beta, \psi, \mu^{i*} | Data^i) \cdot \log-MVN_K(\mu^{i*} | \theta_\mu, \Sigma_\mu) \cdot \prod_k \mu_k^{i*}}{L_i(\alpha, \beta, \psi, \mu^i | Data^i) \cdot \log-MVN_K(\mu^i | \theta_\mu, \Sigma_\mu) \cdot \prod_k \mu_k^i}, 1 \right\}.$$

Step 5: Sample θ_μ . We consider the conjugate prior distribution for θ_μ , which follows a K -dimensional multivariate normal distribution $MVN_K(\bar{\theta}_{\theta_\mu}, \bar{\Sigma}_{\theta_\mu})$, where $\bar{\theta}_{\theta_\mu} = 0$ and $\bar{\Sigma}_{\theta_\mu} = 10^6 I_K$. The next draw θ_μ^* is drawn from a multivariate normal distribution

$$\theta_\mu^* \sim MVN_K(A, B),$$

where $A = B'((\sum_{i=1}^I \log \mu^i)' \Sigma_\mu^{-1} + \bar{\theta}_{\theta_\mu}' \bar{\Sigma}_{\theta_\mu}^{-1})'$ and $B = (I \Sigma_\mu^{-1} + \bar{\Sigma}_{\theta_\mu}^{-1})^{-1}$.

Step 6: Sample Σ_μ . We consider the conjugate prior distribution for Σ_μ , which follows a K -dimensional inverse Wishart distribution $IW(\bar{S}^{-1}, \bar{v})$, where $\bar{S} = I_K$ and $\bar{v} = 1$. The next draw Σ_μ^* is drawn from an inverse Wishart distribution,

$$\Sigma_\mu^* \sim IW \left(\left(\sum_{i=1}^I (\log \mu^i - \theta_\mu) (\log \mu^i - \theta_\mu)' + \bar{S} \right)^{-1}, I + \bar{v} \right).$$

A.3. Simulation Algorithms

To simulate the point processes according to our proposed model, we extend the thinning algorithm in Ogata (1981) to mutually exciting point processes with the parameter values drawn from the posterior distribution obtained during the

estimation process. The following is the detailed algorithm used in §6.1 to simulate the behavior of a representative consumer after clicking on a particular type of advertisement at time $t_0 = 0$.

1. Draw $\alpha, \beta, \psi, \theta_\mu, \Sigma_\mu$ with replacement from the posterior samples obtained during estimation. Generate $\mu^i \sim MVN_K(\theta_\mu, \Sigma_\mu)$.

2. Simulate a point process in $[0, T]$ given $\alpha, \beta, \psi, \mu^i$, and a realized type j_0 point at $t_0 = 0$ ($j_0 = 1, \dots, K-1$).

(a) Let $t = 0, n = 0, n_K = 0, m = \sum_{k=1}^K (\mu_k^i + \alpha_{j_0 k})$

(b) Repeat until $t > T$

(i) Simulate $s \sim \text{Exp}(m)$

(ii) Set $t = t + s$

(iii) If $t < T$, calculate

$$\lambda_k = \mu_k^i \exp(\psi_k n_K) + \alpha_{j_0 k} \exp(-\beta_{j_0} t) + \sum_{l=1, l \neq K}^n \alpha_{j_l k} \exp(-\beta_{j_l} (t - t_l))$$

and let $\lambda = \sum_{k=1}^K \lambda_k$. Generate $U \sim \text{Unif}(0, 1)$.

(A) If $U \leq \lambda/m, n = n + 1, t_n = t$. Simulate $j_n \sim \text{multinomial}(1, \lambda_1/\lambda, \dots, \lambda_K/\lambda)$.

—If $j_n = K, n_K = n_K + 1,$

$$m = \lambda - \mu_k^i \exp(\psi_k (n_K - 1)) + \mu_k^i \exp(\psi_k n_K).$$

—If $j_n \neq K,$

$$m = \lambda + \alpha_{j_n k}.$$

(B) If $U > \lambda/m, m = \lambda$.

(c) The simulation output is $\{t_1, \dots, t_n\}$ and $\{j_1, \dots, j_n\}$.

3. Repeat R times of Steps 1 and 2.

The simulation algorithm used in §6.2 to predict an individual consumer's behavior in the fourth month is similar to the above algorithm, with two differences: (1) Because the prediction is based on the activity history of a specific individual consumer i, μ^i is now supposed to be sampled from the posterior $\pi(\mu^i | \text{data})$. Therefore, in Step 1, we draw μ^i with replacement from the posterior samples obtained in the estimation step. (2) In Step 2, instead of assuming a certain type of point realized at time t_0 , now the initial effect at t_0 is the accumulative effect of the observed behavior of individual i in the data of the first three months. Consequently, the intensity we use to simulate Poisson process before thinning is specific to individual i 's history, and the calculation of m in Step 2(a) becomes

$$m = \sum_{k=1}^K \left\{ \mu_k^i \exp(\psi_k N_K^i(t_0)) + \sum_{j=1}^{K-1} \sum_{l=1}^{N_j^i(t_0)} \alpha_{j k} \exp(-\beta_j (t_0 - t_l^{j(i)})) \right\}.$$

Accordingly, λ_k in Step 2(b)(iii) also needs to be modified as

$$\lambda_k = \mu_k^i \exp(\psi_k (n_K + N_K^i(t_0))) + \sum_{j=1}^{K-1} \sum_{l=1}^{N_j^i(t_0)} \alpha_{j k} \exp(-\beta_j (t - t_l^{j(i)})) + \sum_{l=1, l \neq K}^n \alpha_{j_l k} \exp(-\beta_{j_l} (t - t_l)).$$

A.4. Model Extension

For the extended model incorporating the additional marketing mix data described in §6.3, the likelihood function for individual i can be derived as

$$\begin{aligned} \tilde{L}_i(\alpha, \beta, \psi, \phi, \mu^i | \text{Data}^i) &= \prod_{k=1}^K \left\{ \prod_{l=1}^{N_k^i(T)} \left[\mu_k^i \exp\{\phi_k \bar{p}_t^{k(i)}\} \exp(\psi_k N_K^i(t_l^{k(i)})) \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^{K-1} \sum_{m=1}^{N_j^i(t_l^{k(i)})} \alpha_{j k} \exp(-\beta_j (t_l^{k(i)} - t_m^{j(i)})) \right] \right. \\ &\quad \cdot \exp \left[-\mu_k^i \sum_{m=0}^{N_k^i(T)} \exp(\psi_k m) \int_{t_m^{k(i)}}^{t_{m+1}^{k(i)}} \exp\{\phi_k \bar{p}_t\} dt \right. \\ &\quad \left. \left. - \sum_{j=1}^{K-1} \sum_{m=1}^{N_j^i(T)} \frac{\alpha_{j k}}{\beta_j} (1 - \exp(-\beta_j (T - t_m^{j(i)}))) \right) \right\}, \end{aligned}$$

where $\{\bar{p}_t\}$ is the daily average markup ratio. The likelihood function for all individuals is thus

$$\tilde{L}(\alpha, \beta, \psi, \phi, \mu | \text{Data}) = \prod_{i=1}^I \tilde{L}_i(\alpha, \beta, \psi, \phi, \mu^i | \text{Data}^i).$$

The MCMC algorithm for estimating the extended model is similar to the one described in §A.2, where the likelihood functions are replaced with \tilde{L}_i and \tilde{L} . To sample ϕ , we need to add an additional step between Steps 3 and 4 in §A.2, as is described below.

Step for sampling ϕ . We consider the prior distribution of ϕ following $MVN_K(\bar{\theta}_\phi, \bar{\Sigma}_\phi)$, where $\bar{\theta}_\phi = 0$ and $\bar{\Sigma}_\phi = 10^4 I_K$ (I_K is the $K \times K$ identity matrix). We use Metropolis–Hastings algorithm to sample ϕ . The proposal function is a random-walk generation such that the next draw ϕ^* is drawn from a normal distribution, i.e., $\phi_k^* \sim N(\phi_k, \bar{\sigma}_k^2)$. We adjust the variance $\bar{\sigma}_k^2$ adaptively such that the acceptance rate of proposed ϕ^* is between 0.1 and 0.4. The accepting probability for ϕ^* is given as

$$\min \left\{ \frac{\tilde{L}(\alpha, \beta, \psi, \phi^*, \mu | \text{Data}) MVN_K(\phi^* | \bar{\theta}_\phi, \bar{\Sigma}_\phi)}{\tilde{L}(\alpha, \beta, \psi, \phi, \mu | \text{Data}) MVN_K(\phi | \bar{\theta}_\phi, \bar{\Sigma}_\phi)}, 1 \right\}.$$

References

- Abhishek V, Fader PS, Hosanagar K (2013) Media exposure through the funnel: A model of multi-stage attribution. Working paper, Carnegie Mellon University, Pittsburgh.
- Ait-Sahalia Y, Cacho-Diaz J, Laeven RJA (2013) Modeling financial contagion using mutually exciting jump processes. Working paper, Princeton University, Princeton, NJ.
- Anand P, Sternthal B (1991) Perceptual fluency and affect without recognition. *Memory Cognition* 19(3):293–300.
- Anderson JR, Milson R (1989) Human memory: An adaptive perspective. *Psych. Rev.* 96(4):703–719.
- Bijwaard GE, Franses PH, Paap R (2006) Modeling purchases as repeated events. *J. Bus. Econom. Statist.* 24(4):487–502.
- Bowsher CG (2007) Modeling security market events in continuous time: Intensity based, multivariate point process models. *J. Econometrics* 141(2):876–912.
- Brown J (1958) Some tests of the decay theory of immediate memory. *Quart. J. Experiment. Psych.* 10(1):12–21.

- Bucklin RE, Sismeiro C (2003) A model of Web site browsing behavior estimated on clickstream data. *J. Marketing Res.* 40(3):249–267.
- Cox DR, Isham V (1980) *Point Processes* (Chapman and Hall, New York).
- Daley DJ, Vere-Jones D (2003) *An Introduction to the Theory of Point Processes* (Springer, New York).
- Danaher PJ, Smith MS (2011) Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Sci.* 30(1):4–21.
- De P, Hu Y(J), Rahman MS (2010) Technology usage and online sales: An empirical study. *Management Sci.* 56(11):1930–1945.
- Dong X, Chintagunta PK, Manchanda P (2011) A new multivariate count data model to study multi-category physician prescription behavior. *Quant. Marketing Econom.* 9(3):301–337.
- Dreze X, Zufryden F (1998) Is internet advertising ready for prime time? *J. Advertising Res.* 38(3):7–18.
- Engel JF, Blackwell RD, Miniard PW (1995) *Consumer Behavior* (The Dryden Press, Oak Brook, IL).
- Gelfand AE, Dey DK (1994) Bayesian model choice: Asymptotics and exact calculations. *J. Royal Statist. Soc. Ser. B (Methodol.)* 56(3):501–514.
- Hawkes AG (1971a) Point spectra of some mutually exciting point processes. *J. Royal Statist. Soc. Ser. B* 33(3):438–443.
- Hawkes AG (1971b) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1):83–90.
- Interactive Advertising Bureau, PricewaterhouseCoopers (2012) IAB Internet advertising revenue report 2011 full-year results. Market research report, Interactive Advertising Bureau and PricewaterhouseCoopers, New York.
- Jacoby LL, Dallas M (1981) On the relationship between autobiographical memory and perceptual learning. *J. Experiment. Psych.: General* 110(3):306–340.
- Janiszewski C (1990) The influence of nonattended material on the processing of advertising claims. *J. Marketing Res.* 27(3):263–278.
- Lee AY (2002) Effects of implicit memory on memory-based versus stimulus-based brand choice. *J. Marketing Res.* 39(4):440–454.
- Lee AY, Labroo AA (2004) The effect of conceptual and perceptual fluency on brand evaluation. *J. Marketing Res.* 41(2):151–165.
- Leone RP (1995) Generalizing what is known about temporal aggregation and advertising carryover. *Marketing Sci.* 14(3):141–150.
- Li H(A), Kannan PK (2014) Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *J. Marketing Res.* 51(1):40–56.
- Manchanda P, Dube J-P, Goh KY, Chintagunta PK (2006) The effect of banner advertising on internet purchasing. *J. Marketing Res.* 43(1):98–108.
- MarketingSherpa (2012) 2012 search marketing benchmark report—PPC edition. Market research report, MarketingSherpa, Jacksonville, FL.
- MediaMind (2010) MediaMind global benchmark report: Standard banners—Non-standard results. Market research report, MediaMind, New York.
- Moe WW, Fader PS (2004) Dynamic conversion behavior at e-commerce sites. *Management Sci.* 50(3):326–335.
- Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE (2011) Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* 106(493):100–108.
- Montgomery AL, Li S, Srinivasan K, Liechty JC (2004) Modeling online browsing and path analysis using clickstream data. *Marketing Sci.* 23(4):579–595.
- Muter P (1980) Very rapid forgetting. *Memory Cognition* 8(2):174–179.
- Nedungadi P (1990) Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *J. Consumer Res.* 17(3):263–276.
- Ogata Y (1978) The asymptotic behavior of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* 30(2):243–261.
- Ogata Y (1981) On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* 27(1):23–31.
- Ogata Y (1998) Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* 50(2):379–402.
- Park Y-H, Fader PS (2004) Modeling browsing behavior at multiple websites. *Marketing Sci.* 23(3):280–303.
- Rothschild ML, Hyun YJ (1990) Predicting memory for components of TV commercials from EEG. *J. Consumer Res.* 16(4):472–478.
- Shapiro S (1999) When an ad's influence is beyond our conscious control: Perceptual and conceptual fluency effects caused by incidental ad exposure. *J. Consumer Res.* 26(1):16–36.
- Shapiro S, Macinnis DJ, Heckler SE (1997) The effects of incidental ad exposure on the formation of consideration sets. *J. Consumer Res.* 24(1):94–104.
- Specific Media (2011) Consumer privacy: Choice, clarity, control. Market research report, Specific Media, London.
- Tellis GJ, Franses PH (2006) Optimal data interval for estimating advertising response. *Marketing Sci.* 25(3):217–229.
- Zantedeschi D, Feit EM, Bradlow ET (2013) Measuring multi-channel advertising effectiveness using consumer-level advertising response data. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.
- Zhang Y, Bradlow ET, Small DS (2013) New measures of clumpiness for incidence data. *J. Appl. Statist.* 40(11):2533–2548.