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

Lizhen Xu
Scheller College of Business, Georgia Institute of Technology
800 West Peachtree Street NW, Atlanta, Georgia, 30308
Phone: (404) 894-4380; Fax: (404) 894-6030
Email: lizhen.xu@scheller.gatech.edu

Jason A. Duan
Department of Marketing, The University of Texas at Austin
1 University Station B6700, Austin, Texas, 78712
Phone: (512) 232-8323, Fax: (512) 471-1034
Email: duanj@mccombs.utexas.edu

Andrew Whinston
Department of Info, Risk, & Opr Mgt, The University of Texas at Austin
1 University Station B6500, Austin, Texas, 78712
Phone: (512) 471-7962; Fax: (512) 471-0587
Email: abw@uts.cc.utexas.edu
Abstract

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 stochastic model based on mutually exciting point processes, which model 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 propose a new metric of conversion probability to measure the conversion effects of online advertisements. Simulation algorithms for mutually exciting point processes are developed to evaluate the conversion probability and for out-of-sample prediction. Model comparison results show the proposed model outperforms the benchmark model that ignores exciting effects among advertisement clicks. 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 clicks on online advertisements.

Keywords: online advertising; purchase conversion; search advertisement; display advertisement; point process; mutually exciting
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 search engine results pages) and display advertisements (i.e., digital graphics linking to advertiser’s website embedded in web content pages) are the two leading online advertising formats (IAB and PwC, 2012). Naturally, the effectiveness of these different formats of online advertisements (ads) becomes a lasting question attracting constant academic and industrial interest. Researchers and practitioners are especially interested in the conversion effect of each type of online advertisements, 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, as existent studies on conversion effects focus solely on how non-purchase activities (e.g., advertisement clicks, website visits) affect the probability of purchasing, they usually consider the non-purchase 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 illus-
Figure 1: Illustrative Examples of the Interactions among Ad Clicks

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 afterwards, 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, ne-
glecting 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 only consider 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 re-evaluated.

This study aims to develop an innovative modeling approach that captures the exciting effects among advertisement clicks to more precisely evaluate their conversion effects. 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 non-purchase 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. As 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. While 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. 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 hierarchical Bayesian 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 which 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 the mutually exciting model better captures the complex dynamics of online advertising response and purchase processes.

We develop a new metric of conversion probability based on our proposed model, which leads to a better understanding of 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 over-emphasizing the “last click” effects and underestimates the effectiveness of display advertisements the most severely. We show that display advertisements have little direct effects 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 pattern. Because our modeling approach allows us to predict both purchase and non-purchase activities in the future, it thus furnishes a useful tool for marketing managers to target their advertising efforts.

In addition to the substantive contributions, this paper also makes several methodological
contributions. 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 ecommerce 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 for 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 the paper is organized as follows. In the next section, we survey the related literature. We then provide a brief overview of the data used for this study with some simple statistics in Section 3. In Section 4, we construct the model and explore some of its theoretical properties. In Section 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 Section 6. We conclude the paper in Section 7.

2 Literature Review

This study is related to various streams of existing literature on online advertising, consumer online 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. We will discuss the relationship of our paper to the previous literature in both domains.

Our work relates to a large volume of literature on different online advertisements and their various effects on sales (e.g., Chatterjee et al., 2003; Kulkarni et al., 2012; Mehta et al., 2008; Teixeira et al., 2012). It is particularly related to the studies on the dynamics
of online advertising exposure, website visit, webpage browsing, and purchase conversion, which is based on individual-level online clickstream data similar to our data structure (e.g., Manchanda et al., 2006; Moe and Fader, 2004; Montgomery et al., 2004). 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 focus on the conversion effects of a single type of activities (either banner advertising exposure or website visits), whereas our study considers the effects of various types of online advertisement clicks. Additionally, while they both focus on the effects of non-purchase activities on purchase conversion, we consider the dynamic interactions among non-purchase activities as well. Montgomery et al. (2004) considers 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. In addition, in our model, the occurrence of activities are stochastic and their types depend on the entire history of past behaviors.

The mutually exciting point process induces correlation among the time durations between activities in a parsimonious way. Park and Fader (2004) models the dependence of website visit durations across two different websites based on the Sarmanov family of bivariate distribution, where the overlapping durations are correlated. In our model, all the durations are correlated due to the mutually exciting properties and the correlation declines when two time intervals are further apart. Danaher (2007) models the correlated webpage views using a multivariate negative binomial model. Our model offers a new approach to induce correlation among all the random points of advertisement clicks and purchases.

In the area of statistics and probability, mutually exciting point processes are first proposed in Hawkes (1971a,b), where their theoretical properties are studied. Statistical models using the Hawkes' processes, including the simpler version of self-exciting processes, are applied in seismology (e.g., Ogata 1998), sociology (e.g., Mohler et al. 2009), and finance (e.g., Ait-Sahalia et al. 2008 and Bowscher 2007). These studies do not consider individual het-
erogeneity, 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 inter-purchase duration, which is closely related to our model. A counting process is one way of representing a point process (see 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 estimated it using the expectation-maximization (EM) algorithm. Our Bayesian inference using MCMC method not only provides an alternative and efficient way to estimate this type of stochastic models, but it facilitates straightforward simulation and out-of-sample prediction as well.

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.\footnote{We are unable to reveal the identity of the firm for the non-disclosure agreement.} The firm recorded 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 the unique cookies stored on their computers.\footnote{In this study, we consider each unique cookie ID as equivalent to an individual consumer. While this could be a strong assumption, cookie data are commonly used in the literature studying consumer online behavior (e.g., Manchanda et al., 2006)} The firm also provided the purchase data (including the time of a purchase) associated with these cookie 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 every individual.

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 include 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). Therefore, we focus on the conversion effects of online advertisements rather than the general website visits.

For this study, we use a random sample of 12,000 cookie IDs spanning over 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; we then divide the number of the ad clicks that lead to conversion by the total
Table 1: Data Description

<table>
<thead>
<tr>
<th></th>
<th>Number of Ad Clicks</th>
<th>Percentage of Ad Clicks</th>
<th>Conversion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>6,886</td>
<td>40.4%</td>
<td>.01990</td>
</tr>
<tr>
<td>Display</td>
<td>3,456</td>
<td>20.3%</td>
<td>.00203</td>
</tr>
<tr>
<td>Other</td>
<td>6,709</td>
<td>39.3%</td>
<td>.01774</td>
</tr>
</tbody>
</table>

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. Nevertheless, as is 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).

4 Model Development

To capture the interacting dynamics among different online advertising formats so as to properly account for 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 and/or geographical space. For example, in the context of this study, each click on an online advertisement or each 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 nonnegative integer-valued counting process in a one-dimensional space (i.e.,
time), 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 which are orderly (i.e., the probability that two points
occur at the same time instant is zero) can be fully characterized by the conditional intensity
function defined as follows (Daley and Vere-Jones, 2003).

\[
\lambda (t | \mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{\Pr \{ N (t + \Delta t) - N (t) > 0 | \mathcal{H}_t \}}{\Delta t},
\]  

(1)

where \( \mathcal{H}_t \) is the history of the point process up to time instant \( t \). The history \( \mathcal{H}_t \) is a set which
includes all the information and summary statistics given the realization of the stochastic
process up to \( t \). Notice that \( \mathcal{H}_{t_1} \subseteq \mathcal{H}_{t_2} \) if \( t_1 \leq t_2 \), which implies all the information given
the realization up to an earlier time instant is also contained in the history up to a later
time instant. The intensity measures the probability of instantaneous point occurrence given
the previous realization. By the definition in 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 its definition in Equation (1).

Mutually exciting point processes are a special class of point processes in which past
events affect the probability of future event occurrence and different series of events interact
with each other, as were first systematically studied by Hawkes (1971a,b). Specifically, a mu-
tually exciting point process, denoted as a vector of integers \( N (t) = [N_1 (t), ..., N_K (t)] \), is a
multivariate point process that is the superposition of multiple univariate point processes (or
marginal processes) of different types \( \{N_1 (t), ..., N_K (t)\} \), such that the conditional intensity
function for each marginal process can be written as

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

(2)

Here, \( g_{jk} (\tau) \) is the response function capturing the effect of the past occurrence of a type-\( j \)
point at time \( t - \tau \) on the probability of a type-\( k \) point occurring at time \( t \) (for \( \tau > 0 \)). The

\(^3\)Mathematically, \( \mathcal{H}_t \) is a version of \( \sigma \)-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 \( \sigma \)-Field \( \mathcal{H}_t \).
most common specification of the response function takes the form of exponential decay such that
\[ g_{jk}(\tau) = \alpha_{jk} e^{-\beta_{jk} \tau}, \quad (\alpha_{jk} > 0, \beta_{jk} > 0). \tag{3} \]
As is indicated by Equation (2), the intensity for the type-\(k\) marginal process, \(\lambda_k(t|H_t)\), 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 is 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, \(\lambda(t|H_t) \equiv \bar{\lambda}\). 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.

4.2 The Proposed Model

The mutually exciting point process provides a very 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 \(i (i = 1, \ldots, I)\), 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, \(N^i(t) = [N^i_1(t), \ldots, N^i_K(t)]\). 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. \(N^i_k(t)\) 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 \( k = 1, \ldots, K - 1 \) stand for various types of ad clicks. For our data, we consider \( K = 4 \) so that \( N_1^i(t) \) stands for purchases and \( \{ N_1^i(t), N_2^i(t), N_3^i(t) \} \) stand for clicks on search, display, and other advertisements, respectively. When individual \( i \), for example, clicked 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

\[
\lambda_k^i (t|H_t^i) = \mu_k^i \exp \left( \psi_K^i N_K^i(t) \right) + \sum_{j=1}^{K-1} \int_0^t \alpha_{jk} \exp \left( -\beta_j (t - s) \right) dN_j^i(s) \tag{4}
\]

\[
= \mu_k^i \exp \left( \psi_K^i N_K^i(t) \right) + \sum_{j=1}^{K-1} \sum_{l=1}^{N_j^i(t)} \alpha_{jk} \exp \left( -\beta_j \left( t - t_l^j(i) \right) \right) , \tag{5}
\]

for \( k = 1, \ldots, K \), where \( \mu_k^i > 0 \), \( \alpha_{jk} > 0 \), \( \beta_j > 0 \), and \( t_l^j(i) \) 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. 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 \( \lambda_k^i \) specified in Equation (4) is the baseline intensity, \( \mu_k^i \). It represents the general probability density 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 consumers’ 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 \( \mu^i = [\mu_1^i, \ldots, \mu_K^i] \) follow a multivariate log-normal distribution

\[
\mu^i \sim \text{log-MVN}_K (\theta_\mu, \Sigma_\mu) . \tag{6}
\]

The multivariate log-normal distribution facilitates the likely right-skewed distribution of \( \mu_k^i (> 0) \). In addition, the variance-covariance matrix \( \Sigma_\mu \) 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, while also capturing the dynamic change of such effects over time. It is believed that exposure to advertisements, being informed or reminded of the product information, and visiting the firm’s website will generally increase consumers’ purchase probability, albeit slightly sometimes. While the effect of a single response to an advertisement may not result in immediate conversion, such effect could accumulate over time, which in turn invites subsequent visits to the firm’s website through various advertising vehicles and hence increases the probability of future responses to advertisements. In the meantime, such interacting effects decay over time, as memory and impression fades gradually in general. Therefore, we model the effects of ad clicks in a form similar to Equation (3). For \( j = 1, \ldots, K - 1 \) and \( k = 1, \ldots, K \), \( \alpha_{jk} \) measures 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, whereas \( \beta_j \) measures how fast such effect decays over time. To keep our model parsimonious, we let \( \beta_{jk} = \beta_j \) for all \( k = 1, \ldots, K \), which implies the decaying effect only depends on the type of the advertisement click; that is, the exciting effect of a type-\( j \) point on the type-\( k \) process would decay at the same rate as that on the type-\( k' \) process.\(^4\) Therefore, a larger \( \alpha_{jk} \) indicates a greater exciting effect instantaneously, whereas a smaller \( \beta_j \) means such exciting effect is more lasting. For \( j = k \), \( \alpha_{jk} \)'s indicate the effects between the same type of points and are therefore called the self-exciting effects; for \( j \neq k \), they are mutually exciting effects between different types of points.

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

\(^4\)The model can be easily revised into different versions by allowing \( \beta_{jk} \) to take different values. In fact, we also estimated two alternative models: one allowing \( \beta_{jk} \) to be different from each other, and the other considering \( \beta_{jk} \)'s take the same value for \( k = 1, \ldots, K - 1 \) which is different from \( \beta_{jK} \). It is shown that the performance of our proposed model is superior to both alternative models: the Bayes factors of the proposed model relative to the two alternative models are \( \exp(120.24) \approx 1.7 \times 10^{52} \) and \( \exp(190.18) \approx 3.9 \times 10^{82} \), respectively.
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 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 browse 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 \( \psi_k \) can be either positive or negative. A positive \( \psi_k \) means a purchase increases the probability of future occurrence of type-\( k \) points, whereas a negative \( \psi_k \) indicates the opposite.

As is discussed earlier, the intensity \( \lambda^i = [\lambda_1^i, \ldots, \lambda_K^i] \) 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, \( \lambda^i \) 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.

Notice that Equation (4) implicitly assumes the accumulative effects from the infinite past up to time \( t = 0 \), which is unobserved in the data, equals 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) shows that the maximum likelihood estimates when omitting the history from the infinite past are consistent and efficient.

Based on the intensity function specified in Equation (5), the likelihood function for any realization of all individuals’ point processes \( \{N^i(t)\}_{i=1}^L \) can be written as (Daley and
\[ L = \prod_{i=1}^{I} \prod_{k=1}^{K} \left\{ \sum_{t=1}^{N_i(t)} \lambda_k^i \left( t_i^{k(i)} | \mathcal{H}_{i,k(t)}^i \right) \exp \left( - \int_0^T \lambda_k^i \left( t | \mathcal{H}_{i,t}^i \right) dt \right) \right\}. \]  

(7)

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 effect. Given the hierarchical nature of the model, we cast it in the hierarchical Bayesian framework. The full hierarchical model is described as follows.

\[
N^i(t) | \alpha, \beta, \psi, \mu^i \sim \lambda^i(t | H^i)
\]
\[
\mu^i | \theta_\mu, \Sigma_\mu \sim \text{log-MVN}_K(\theta_\mu, \Sigma_\mu)
\]
\[
\alpha_{jk} \sim \text{Gamma}(\bar{\alpha}, \bar{\alpha}), \beta_j \sim \text{Gamma}(\bar{\beta}, \bar{\beta}), \psi \sim \text{MVN}_K(\bar{\theta}_\psi, \bar{\Sigma}_\psi)
\]
\[
\theta_\mu \sim \text{MVN}_K(\bar{\theta}_\theta, \bar{\Sigma}_\theta), \Sigma_\mu \sim \text{IW}(\bar{S}^{-1}, \bar{\nu})
\]

where \(\alpha\) is a \((K - 1) \times K\) matrix whose \((j, k)\)th element is \(\alpha_{jk}\), and \(\beta = [\beta_1, \ldots, \beta_{K-1}]\) and \(\psi = [\psi_1, \ldots, \psi_K]\) are both vectors. The parameters to be estimated are \(\{\alpha, \beta, \psi, \{\mu^i\}, \theta_\mu, \Sigma_\mu\}\). Notice that \(\alpha, \beta, \psi, \text{and} \{\mu^i\}\) play distinct roles in the data generating process, and the model is therefore identified (Bowshers, 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 \(j, k = 1, \ldots, K - 1\). 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, as it only captures the self-exciting effects among advertisement clicks.

For model comparison purpose, we are also interested in a benchmark model in which \(\alpha_{jk} = \psi_k = 0\) for all \(j, k = 1, \ldots, K - 1\). 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. Notice that the Poisson process model is the closest benchmark to the typical
conversion models that can be used for model comparison with our proposed model. The
typical conversion models cannot be directly compared with our model because they consider
the partial likelihood only (given the occurrence of ad clicks), whereas ours considers the full
likelihood (including the likelihood of the occurrence of ad clicks).

5 Estimation

To estimate the parameters in the model, we use the Markov Chain Monte Carlo (MCMC)
method for Bayesian inference. We apply Metropolis-Hastings algorithms to sample the
parameters. 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

We first 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
2. (The estimates for 12,000 different $\mu_i$’s are omitted due to the page limit.)

The estimation results in Table 2a 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 type 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., $\mu_j$, $j = 1, 2, 3$), whose expected values ($\exp\{\theta_{\mu,j}\}$, $j = 1, 2, 3$) range from
$\exp\{-6.10\} \approx .0022$ to $\exp\{-5.39\} \approx .0046$, the values of $\alpha_{jk}$ ($j, k = 1, 2, 3$) 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 $\alpha_{jj}$ ($j = 1, 2, 3$) are greater than $\alpha_{jk}$ ($j \neq k$
Table 2: Parameter Estimates for the Mutually Exciting Model

(a) Exciting Effects (Posterior Means and Posterior Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Search</th>
<th>Display</th>
<th>Other</th>
<th>Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{11}$</td>
<td>2.8617 (0.1765)</td>
<td>0.0860 (0.0214)</td>
<td>0.5381 (0.0562)</td>
<td>0.6167 (0.0633)</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.1614 (0.0496)</td>
<td>1.7818 (0.2314)</td>
<td>0.2055 (0.0572)</td>
<td>0.0845 (0.0347)</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>0.4647 (0.0654)</td>
<td>0.1270 (0.0367)</td>
<td>8.0526 (0.4117)</td>
<td>0.8384 (0.0867)</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>-0.5664 (0.1228)</td>
<td>-0.7556 (0.2348)</td>
<td>-0.6235 (0.1229)</td>
<td>0.2787 (0.2160)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>34.0188 (1.7426)</td>
<td>46.8854 (4.9370)</td>
<td>51.5114 (2.3241)</td>
<td></td>
</tr>
</tbody>
</table>

(b) Individual Heterogeneity (Posterior Means and Posterior Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Search</th>
<th>Display</th>
<th>Other</th>
<th>Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{\mu,1}$</td>
<td>-5.3926 (0.0166)</td>
<td>-6.1027 (0.0212)</td>
<td>-5.8063 (0.0221)</td>
<td>-9.7704 (0.0762)</td>
</tr>
<tr>
<td>$\theta_{\mu,2}$</td>
<td>0.4584 (0.0246)</td>
<td>0.5934 (0.0335)</td>
<td>1.0014 (0.0365)</td>
<td>2.3914 (0.2575)</td>
</tr>
</tbody>
</table>
and \( j, k = 1, 2, 3 \). This result is consistent with the observed data pattern that it is more common for 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 \( \alpha_{21} \) being greater than \( \alpha_{12} \) is .92, and the posterior probability of \( \alpha_{23} \) being greater than \( \alpha_{32} \) is .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 towards the end, because they are more likely to excite the other two types of ad clicks than being excited by them.

Regarding the direct effects on purchase conversion, the values of \( \alpha_{j4} \) \((j = 1, 2, 3)\) are much greater in comparison with the baseline intensity for purchase occurrence (i.e., \( \mu_4^i \)), whose expected value \( \exp\{\theta_{\mu,4}\} \) is about \( \exp\{-9.77\} \approx .00006 \). 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. While all three types of advertisement clicks have direct conversion effects, display advertisement’s direct conversion effect (\( \alpha_{24} \)) 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. \( \psi_{j} \) \((j = 1, 2, 3)\) take significantly negative values, whereas the effect on repeated purchases (i.e., \( \psi_4 \)) 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, they tend to make the repeated purchases directly rather than through clicking advertising links further again.

There are also interesting results regarding the variance-covariance matrix for individual baseline intensities in Table 2b. 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 advertisements, whereas clicking this particular type of advertisements may increase the probability of clicking other types of advertisements subsequently. In addition, it is interesting to find that the individual baseline intensity for clicking display advertisements is negatively correlated with the individual baseline intensity for purchases; that is, $\Sigma_{\mu,2} < 0$. In other words, consumers who are more likely to respond to display ads often have lower initial purchase intention, which adds to the explanation of 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 deviance information criterion (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 3 shows the results of the model comparison criteria for the three models.

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>Log-Marginal Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutually Exciting</td>
<td>192,002.56</td>
<td>-93,454.69</td>
</tr>
<tr>
<td>Self-Exciting</td>
<td>193,513.08</td>
<td>-94,219.22</td>
</tr>
<tr>
<td>Poisson Process</td>
<td>206,146.80</td>
<td>-99,748.95</td>
</tr>
</tbody>
</table>

According to Table 3, the Bayes factor of the mutually exciting model relative to the self-exciting model is $\exp(-93454.69 + 94219.32) \simeq 1.0 \times 10^{332}$, and the Bayes factor of the mutually exciting model relative to the Poisson process model is $\exp(-93454.69 + 99748.95) \simeq 3.7 \times 10^{2733}$. The mutually exciting model also has the lowest DIC value. Therefore, both DIC and Bayes factor indicate that the proposed mutually exciting model outperformed the two benchmarks 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, it still underperforms substantially in comparison to the mutually exciting model due to 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.

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 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 \( t_0 \), then the conversion probability \((CP)\) for type-\(k\) advertisement in time period \( t\) given the parameters for the processes \(\{\alpha, \beta, \psi, \mu^i\}\) can be defined as

\[
CP_k (t; \mu^i, \alpha, \beta, \psi) = \Pr (N_K^i (t_0 + t) - N_K^i (t_0) > 0 | N_K^i (t_0) - N_K^i (t_0-) = 1), \tag{9}
\]

where \( k = 1, ..., K - 1 \). 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, \( p (\alpha, \beta, \psi, \theta, \Sigma | Data) \), as follows.

\[
ACP_k (t) = E \left[ CP_k (t; \mu^i, \alpha, \beta, \psi) | Data \right] = \int CP_k (t; \mu^i, \alpha, \beta, \psi) p (\mu^i | \theta, \Sigma) \cdot \]

\[
p (\alpha, \beta, \psi, \theta, \Sigma | Data) d\mu^i d\alpha d\beta d\psi d\theta d\Sigma \tag{10}
\]

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, \( p (\mu^i | \theta, \Sigma) \) as is specified in Equation (6), instead of using the posterior distribution \( p (\mu^i | Data) \) for a specific consumer.

\(^5\)In reality, an “initial click” can be approximated as long as there was no click in the recent history before \( t_0 \).
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 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 which 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 approximate the average conversion probability in Equation (10) by

$$ACP_k(t) = \int E \left[ I \left\{ N^i_k(t_0 + t) - N^i_k(t_0) > 0 \right\} | N^i_k(t_0) - N^i_k(t_0 -) = 1 \right] \cdot p(\mu^i|\theta^i, \Sigma^i) p(\alpha, \beta, \psi, \theta^i, \Sigma^i|Data) d\mu^i d\alpha d\beta d\psi d\theta^i d\Sigma^i$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} I \left\{ N^{i(r)}_K(t_0 + t) - N^{i(r)}_K(t_0) > 0 \right\},$$

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 $I \left\{ N^{i(r)}_K(t_0 + t) - N^{i(r)}_K(t_0) > 0 \right\}$ 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).\(^6\) 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. Table 4 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).

Table 4: Average Conversion Probabilities (%) of Different Advertisement Formats

<table>
<thead>
<tr>
<th>Model</th>
<th>Search</th>
<th>Display</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversion Rate</td>
<td>1.989</td>
<td>0.203</td>
<td>1.774</td>
</tr>
<tr>
<td>Mutually Exciting</td>
<td>2.030</td>
<td>0.246</td>
<td>1.978</td>
</tr>
<tr>
<td>Self-Exciting</td>
<td>2.005</td>
<td>0.238</td>
<td>1.968</td>
</tr>
<tr>
<td>Poisson Process</td>
<td>1.818</td>
<td>0.234</td>
<td>1.703</td>
</tr>
</tbody>
</table>

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 4 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

\(^6\)We sample 1,000,000 times with replacement from the 30,000 posterior samples to complete this exercise.
Such underestimation originates from the fact that although display advertisements have little direct effects on purchase conversion (recall that display advertisements have the lowest direct conversion effect, i.e., $\alpha_{24}$ is much lower than $\alpha_{14}$ and $\alpha_{34}$ according to the estimation results), it may stimulate subsequent clicks on other types of advertisements which in turn lead 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 the portfolio 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.

In order 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 4.

Comparing the conversion probabilities evaluated based on the self-exciting model (the third row of Table 4) with those based on the mutually exciting model (the second row of Table 4), 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.

If we further compare the conversion probabilities evaluated based on the Poisson process model (the fourth row of Table 4) with those based on the self-exciting model (the third row of Table 4), it is clear that the Poisson process model underestimates the conversion effect
of search advertisements more greatly than display advertisements. Recall that in comparison to 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 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 individual's future behaviors is similar to the one developed in Section 6.1. The primary difference is twofold: the baseline intensities $\mu_i$ 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.

For each of the selected 1,000 individuals, we simulate 10,000 point processes according to
the proposed model. We then calculate the average simulated numbers of purchases and clicks on each type of advertisement per individual over the fourth month for the entire predictive sample. We also construct the 90% 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 5 presents the out-of-sample validation results. Table 5 shows that the actual data all fall into the 90% predicative intervals, which indicates that the proposed model adequately captures the complex dynamics underlying consumers’ online advertisement clicking and purchasing processes.

Table 5: Average Numbers of Ad Clicks and Purchases per Customer in the Fourth Month

<table>
<thead>
<tr>
<th></th>
<th>Search</th>
<th>Display</th>
<th>Other</th>
<th>Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Prediction</td>
<td>0.22</td>
<td>0.098</td>
<td>0.27</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.078, 0.45)</td>
<td>(0.032, 0.22)</td>
<td>(0.053, 0.64)</td>
<td>(0.0082, 0.40)</td>
</tr>
<tr>
<td>Actual Data</td>
<td>0.12</td>
<td>0.070</td>
<td>0.14</td>
<td>0.013</td>
</tr>
</tbody>
</table>

As out-of-sample prediction can provide statistically corroborating evidences for the model comparison results in Table 5, we next compare the predicative 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 predicative 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 between the predicted numbers and the observed data from the holdout sample. Table 6 shows the average sum of squared errors over the 1,000 selected individuals for the three models.

Table 6: Model Comparison for Out-of-Sample Prediction

<table>
<thead>
<tr>
<th>Model</th>
<th>Mutually Exciting</th>
<th>Self-Exciting</th>
<th>Poisson Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Sum of Squared Errors</td>
<td>1.514</td>
<td>1.545</td>
<td>1.705</td>
</tr>
</tbody>
</table>
As we can see from Table 6, the out-of-sample performances confirm the model comparison results based on DIC and Bayes factors reported in Table 3. 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 predicative power and within-sample model fit. In comparison, the nested self-exciting model underperforms in predicative 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 benchmark Poisson process model demonstrates the poorest model performance in all the three criteria.

7 Conclusion

In this paper, we develop a Bayesian hierarchical model which 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 MCMC method to obtain Bayesian inference for our model. We develop a new metric of 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 sample 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. 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 in order 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 metric of 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 low direct effect on purchase conversions. Neglecting such effects and overemphasizing the “last click” effects, the commonly used measure of conversion rate is biased towards search advertisements and underestimates the relative effectiveness of display advertisements the most severely. For decision makers who are to allocate 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 non-purchase activities at the same time. The ability of predicting 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.

There are a few limitations of this study, which lead to future research directions. First,
our data only contain individuals who have clicked on any of the firm’s online advertisements at least once during the observation period. The estimated distribution of heterogeneous consumers represents the population of online consumers who possess at least a certain level of interest in the firm’s products, which are the promising prospects whom the firm is most interested in acquiring. Second, obtained based on tracking cookies, our data inherit the general limitations of cookie data, such as the inability to distinguish actual users from computers and the lack of demographic information. While we incorporate individual random effects in the model to account for consumer heterogeneity, detailed demographic data can be further incorporated into the modeling of consumer heterogeneity to deliver richer results and implications. Third, our data set does not contain the firm’s detailed marketing mix variables such as sales and promotions. Price changes and promotions can cause the baseline propensities for ad clicking and purchasing to vary over time. While consulting with the top marketing manager of the firm reveals no major variation in marketing effort during the period of study, our model can be easily extended to allow the baseline intensities to be functions of the marketing mix variables once available. In addition, when the cost information for different formats of online advertisements becomes available, the correctly estimated conversion probabilities from our study can help design a more efficient budget allocation scheme for online advertising. We believe future research should be able to easily extend our model along these directions with different data sets in various contexts.

References


IAB, PwC. 2012. IAB Internet advertising revenue report 2011 full-year results. Market research report, Interactive Advertising Bureau (IAB) and PricewaterhouseCoopers (PwC).


