Media Exposure through the Funnel: A Model of Multi-Stage Attribution

Vibhanshu Abhishek†, Peter S. Fader‡, and Kartik Hosanagar‡

{vibs@cmu.edu, faderp@wharton.upenn.edu, kartikh@wharton.upenn.edu}

†Heinz College, Carnegie Mellon University
‡Wharton School, University of Pennsylvania

Completed Research Paper
Word Count: 11,387
Media Exposure through the Funnel: A Model of Multi-Stage Attribution

Abstract

Consumers are exposed to advertisers across a number of channels. As a result, a conversion or a sale may be the result of a series of ads that were displayed to the consumer. This raises the key question of attribution: which ads get credit for a conversion and how much credit do each of these ads get? This is one of the most important issues facing the advertising industry. Although the issue is well documented, current solutions are often simplistic. Current practices apply simplistic methods like attributing the sale to the most recent ad exposure. These methods penalize prior exposures and give undue credit to ad exposures further down in the conversion funnel. In this paper, we address the problem of attribution using a unique data-set from the online campaign of a car launch. We present a Hidden Markov Model of an individual consumer’s behavior based on the concept of a conversion funnel that captures the consumer’s deliberation process. We observe that different ad formats, e.g. display and search ads, affect the consumers differently based on their state in the decision process. Display ads usually have an early impact on the consumer, moving him from a state of dormancy to a state in which he is aware of the product and even one in which the product enters his consideration set. Further, when the consumer actively interacts with these ads (e.g. by clicking on them), the likelihood of a conversion increases considerably. Finally, we show that attributing conversions based on the HMM model provides fundamentally different insights into ad effectiveness relative to the approach of attribution to last exposure.

Keywords: Online advertising, multi-channel attribution, conversion funnel, hidden Markov model
1 Introduction

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

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

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

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

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

In this paper, we propose a model for online ad-attribution using a dynamic hidden Markov model (HMM). We present a model of individual consumer behavior based on the concept of a conversion funnel that captures a consumer’s deliberation process. The conversion funnel is a model of a consumer’s search and purchase process that is commonly used by marketers (Kotler and Armstrong, 2011). A consumer moves in a staged manner from a dormant state to the state of conversion and ads affect the movement through the different stages. This model is estimated using a unique dataset from a car manufacturer that contains all the online advertising data from the beginning of a campaign. We observe that different ad formats, e.g. display and search ads, affect the consumers differently and in different states of their decision process. Display ads usually have an early impact on the consumer, moving him from a state of dormancy to a state where he is aware of the product, and it might enter his consideration set. However,
when the consumer actively interacts with these ads (e.g. by clicking on them), his likelihood to convert considerably increases. Secondly, we present an attribution scheme based on the proposed model that assigns credit to an ad based on the incremental impact it has on the consumer’s probability to convert. This method is subsequently compared to the LTA scheme. The proposed methodology gives more credit to display ads and less credit to search ads as compared to LTA. This result is contrary to the commonly held belief that display advertising is not effective. Our analysis shows that this belief has arisen due to inappropriate attribution methodology.

This paper makes several contributions. Firstly, we propose a comprehensive multi-stage model of consumer behavior in response to the advertising activity. This model is a considerable improvement over the extant literature in online advertising where consumer models are often simplistic and do not have temporal dynamics. Secondly, the consumer model is used to generate a new attribution technique which might be useful for future research in analyzing the impact of advertising on consumer behavior. This model can be easily extend to incorporate optimal advertising decisions. Thirdly, the paper uses a unique and rich individual level dataset which allows us to create and estimate this model of individual response to advertising activity. This dataset is used to create unique insights, which would not have been possible with existing techniques.

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

2 Prior Literature

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

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

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

3 Data Description

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

This dataset is unique as it contains all the display and search advertising data at an individual level since the start of the campaign. Our sample comprises a panel of 6432 randomly chosen users with a total of 146,165 observations. An observation in our dataset comprises a display ad impression or click
(generic/specific), a search click or activity (page view/conversion) on the advertiser’s website. We do not observe the search ads that were shown to consumers (as this data is not reported by the search engine) however, when a consumer clicks on one of these ads and arrives at the advertiser’s website, this click is recorded in our data and referred to as a search click. A conversion in this data is said to occur when the user performs one of the following activities on the advertiser’s website - search inventory, find a dealer, build & price and get a quote. We do not focus on the different conversion activities and treat them similarly. Furthermore, as we are interested in how the ads drive the first conversion, we discard all the observations for a particular consumer after the first conversion. Since we are interested in the effect of advertising on the conversion process, we also eliminate users in our data that do not have any ad exposures. This results in a panel size of 5121 users with 112,619 observations. Summary statistics of this data at an individual level is presented in Table 1 below.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic display impressions</td>
<td>13.75</td>
<td>34.72</td>
</tr>
<tr>
<td>Generic display clicks</td>
<td>0.07</td>
<td>0.18</td>
</tr>
<tr>
<td>Generic click-through rate</td>
<td>0.007</td>
<td>0.054</td>
</tr>
<tr>
<td>Specific display impressions</td>
<td>4.21</td>
<td>10.06</td>
</tr>
<tr>
<td>Specific display clicks</td>
<td>0.14</td>
<td>0.32</td>
</tr>
<tr>
<td>Specific click-through rate</td>
<td>0.02</td>
<td>0.062</td>
</tr>
<tr>
<td>Search clicks</td>
<td>0.24</td>
<td>0.72</td>
</tr>
<tr>
<td>Web pages viewed</td>
<td>3.47</td>
<td>8.18</td>
</tr>
<tr>
<td>Conversions</td>
<td>0.15</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics

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

4 Model of Multi-Touch Attribution

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

4.1 The Conversion Funnel

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

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

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

4.2 Hidden Markov Model

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

Figure 2: Diagram representing the latent states and the outcomes of the HMM. \( q_{ss'} \) denote the transition probabilities from state \( s \) to state \( s' \) and \( Y_s \) is the binary random variable that captures conversion in state \( s \).

In accordance with the conversion funnel, we construct an HMM with four states \( (S) \) where the four states are “dormant”, “awareness”, “consideration” and “conversion” (Figure 2). At any time \( t \), consumer \( i \) can be in one of the four states, \( S_{it} \in S \).\(^1\) As mentioned earlier, we do not observe \( s_{it} \),

\(^1\)Variables in uppercase denote random variables and variable in lowercase denote their realizations. In addition, set
but we observe the bivariate outcome variable $Y_{it} = (N_{it}, C_{it})$ which arises from a stochastic process conditional on the state $S_{it}$. $N_{it}$ is a Poisson random variable that denotes the number of pages viewed by the consumer between time $t$ and $t+1$ and $C_{it}$ is a binary random variable which captures whether there was a conversion between time $t$ and $t+1$. When the user is in a dormant state, he is unaware of the product or is not deliberating a purchase. In this state, there is no activity from the consumers and the outcomes variables, page views and conversion, are set to zero. As the consumer is exposed to different ads, he might move into a state of awareness where he knows about the product and might be willing to purchase it. On further deliberation, he moves into a consideration state where he can actively look for product related information and engage with the firm’s website. Consumers can also go directly from the dormant state to the state of consideration. In this model, the research stage is implicitly captured by the consumer’s interaction with the advertiser’s website (measured through page views). Since we model the very first conversion of the consumer, the consumer moves into the “conversion” state as soon as a conversion occurs. “Conversion” is a dummy absorbing state which captures the fact that once a consumer has engaged in a conversion activity, he ceases to exist in our data.

We assume that a consumer’s propensity to purchase (or convert) is zero in the dormant state, and it steadily increases as he moves down the different states. We also assume that the consumer’s research behavior becomes more intense as he moves down the funnel, e.g. he is likely to visit the advertiser’s website more often when he is in the consideration state as opposed to the awareness state. The transition between the states take place in a stochastic manner when an ad event $a_{it}$ occurs and is influenced by the firm’s advertising activities thus far. Ads from different channels can have different effects on these transitions and these effects can be state specific. The transitions between the different states also follow a Markov process, i.e. the transitions out of a particular state depend only on the current state and not on the path that the user took to get to the state. Let $A_i = \{a_{i1}, a_{i2}, \ldots, a_{iT}\}$ denote a sequence of $T$ ad events that consumer $i$ is exposed to, due to which the consumer ends up in states $S_i = \{S_{i1}, S_{i2}, \ldots, S_{iT}\}$. $x'_{it}$ captures the running sum of the different kinds of advertising activities till time $t$ and contains covariates like number of display impressions at a generic website, number of display impressions at an auto-specific website and search clicks. We do not observe $S_i$ but observe the observation vector $Y_i = \{Y_{i1}, Y_{i2}, \ldots, Y_{iT}\}$. The joint probability of observing the notation supersedes notation for random variables unless otherwise noted.
sequence of observations \( \{Y_i = y_{i1}, \ldots, Y_iT = y_{iT}\} \) is a function of three main components: (i) the transition probabilities between the different states – \( Q_{it} \), (ii) the distribution of the observational variables conditional on the state – \( M_{it} \) denotes the probability of conversion and \( N_{it} \) denoted as \( \text{Poisson}(\lambda_{its}) \), and (3) the initial state distribution – \( \pi \). Below, we describe each of these components in detail.

### 4.2.1 Markov Chain Transition Matrix

In our model, there might be a transition from the current state \( s_{it} \) only under two conditions - (i) when a consumer is exposed to an ad event \( a_{it} \), or (ii) when a conversion takes place and the consumer moves to the “conversion” state with certainty. If the transition occurs due to an ad event, consumer i’s transition from one latent state to another is stochastically based on the transition matrix \( Q_{it} \) which is a function of the time varying advertising activities, \( x'_{it} \) at time \( t \). The probability that a consumer transitions to the state \( s' \) at time \( t + 1 \) conditional on him being in state \( s \) at time \( t \) is given by

\[
P(S_{it} = s' | S_{it-1} = s) = q_{itss'}.
\]

Let \( T_s \) be the set of states \( (s') \) that can be reached from state \( s \). The elements of the transitions matrix specific to state \( s \) are given by

\[
q_{itss'} = \frac{\exp\{\mu_{iss'} + x'_{it}\beta_{ss'}\}}{1 + \sum_{s' \in T_s} \exp\{\mu_{iss'} + x'_{it}\beta_{ss'}\}} \quad \forall \; s' \neq s,
\]

\[
q_{itss} = \frac{1}{1 + \sum_{s' \in T_s} \exp\{\mu_{iss'} + x'_{it}\beta_{ss'}\}},
\]

where \( \beta_{ss'} \) is the response parameter that captures how the advertising related activities affect the consumer’s propensity to transition from state \( s \) to \( s' \) and \( \mu_{iss'} \) captures the consumer specific intercept term. \( \beta_{ss'} \) is different across states as the advertising activities \( x'_{it} \) might have different effects on the transition based on the receiving state. For e.g., display clicks might affect the transition to the “dormant” state differently than the transition to the “consideration” state.

### 4.2.2 Consumer Research and Conversion Behavior

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

**Modeling page views:** \( N_{it} \) is drawn from a Poisson distribution with a rate parameter \( \lambda_{its} \), which is a function of the current state \( s \), and advertising activity \( x_{it} \). The probability of observing \( n_{it} \) page views is given by
\[ P(N_{it} = n_{it}|S_{it} = s) = \frac{\lambda_{its}^{n_{its}} e^{-\lambda_{its}}}{n_{its}!}, \]

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

\[
\begin{align*}
\tilde{\eta}_1 &= 0, \\
\tilde{\eta}_2 &= \eta_2, \\
\tilde{\eta}_3 &= \tilde{\eta}_2 + \exp\{\eta_3\},
\end{align*}
\]

where \( \eta_2 \) and \( \eta_3 \) are parameters to be estimated from the data.

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

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

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

\[
\begin{align*}
\tilde{\alpha}_1 &= 0, \\
\tilde{\alpha}_2 &= \alpha_2, \\
\tilde{\alpha}_3 &= \tilde{\alpha}_2 + \exp\{\alpha_3\},
\end{align*}
\]
where $\alpha_2$ and $\alpha_3$ are the parameters to be estimated from the data. This structure enforces that $m_{it3} \geq m_{it2}$, ceteris paribus. This assumption ensures the identification of the different states and is consistent with the approach adopted by Netzer et al. (2008) and Montoya et al. (2010).

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

$$P(Y_{it} = y_{it}|S_{it} = s) = m_{its}^{c_{it}}(1 - m_{its})^{(1-c_{it})}P(N_{its} = n_{it}|S_{it} = s)$$

where $y_{it} = (n_{it}, c_{it})'$ is the realized outcome variable.

### 4.2.3 Consumer Heterogeneity

Consumers might respond differently to ads because of differences in their prior relationship with the brand, offline advertising activity or underlying demographic variables. If the unobserved consumer heterogeneity is not accounted for, it might affect the estimation of the parameters associated with the transition matrix. The following example illustrates this misspecification. Let’s assume a consumer moves from a dormant state to a state of awareness because of television ads. However, since we do not observe offline advertising or account for it, we might spuriously contribute this transition to a display or search ad he saw online. Our approach addresses this problem by allowing for the intercept terms in the transition matrix, $\mu_i = (\mu_{i12}, \mu_{i21}, \ldots, \mu_{i43})$, to vary across consumers which captures differences in their response to online ads. We divide the customer heterogeneity into two distinct components as follows:

$$\mu_i = \theta_z + \xi_i.$$  

where $\theta_z$ captures the heterogeneity due to region specific factors, e.g. offline advertising, demographic conditions, that are constant for all consumers in the same region. Here, the index $z$ denotes a specific region. The aforementioned region specific factors have an overall effect on consumers awareness or susceptibility to the brand. Since we do not observe these factors, e.g. the advertising spend for traditional media, we control for it using this random effect which varies across different regions. $\xi_i$ captures individual specific idiosyncrasies, e.g. brand awareness or loyalty, affinity for cars, etc. Furthermore, $\xi_i \sim MVN(\Sigma_z)$. We model $\mu_i$ in a Hierarchical Bayesian fashion, where $\theta_z$ and $\Sigma_z$ are DMA specific.
parameters drawn from a hyper prior distribution. The DMA specific mean and variance have the following prior distributions,

\[ \theta_z \sim MVN(\bar{\theta}, \Omega_\theta), \]
\[ \Sigma^{-1}_z \sim Wishart(\nu, \Delta). \]

The regional parameters are drawn at a DMA level because traditional advertising decisions are typically made at this level. In addition, we only observe DMA-level location information in our dataset. We incorporate heterogeneity only in the intercept term to maintain a parsimonious model.\(^2\)

4.2.4 Initial State Membership

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

4.2.5 Discussion of the HMM

In summary, the dynamic heterogeneous HMM captures the consumers’ behavior as they transition across the different states of the funnel and eventually convert. This model allows the ads to have an effect on consumers behavior – they affect the transition probabilities as well as the product research and conversion activities. Thus, these ads not only have an immediate impact on the consumers by changing their conversion probabilities, but they can also move consumers to different stages in the conversion funnel, which can have an impact on their future conversion behavior. Thus, the model

\(^2\)Our main motivation in incorporating customer heterogeneity is to prevent the unobserved heterogeneity from interfering with the estimation of the temporal dynamics. Even though our model does not estimate the differentiated response to these ads, it recovers the average affect across consumers.
allows us to attribute suitable credit to an ad even if it does not contribute to a conversion outrightly but helps in moving the consumer to a state with higher conversion probability. In this sense, our model differs considerably from the approach adopted by Shao and Li (2011) and Dallessandro et al. (2012) which attribute credit to an ad only when it directly leads to conversion. In the following discussion, we explain how the ad events affect these transitions and how the aforementioned model can be used to solve the attribution problem.

4.3 Ad Attribution

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

$$
\begin{align*}
  x_{it-1} & \rightarrow_{a_{it}} x_{it}, \\
  z_{it-1} & \rightarrow_{a_{it}} z_{it}.
\end{align*}
$$

Hence, $a_{it}$ has a two fold effect on the consumer’s probability to convert which is shown below – (i) it alters the conditional conversion probability, through changes in $z'_{it}$ and (ii) it can lead to a transition of the consumer from one state to another by affecting $x'_{it}$. Attribution in the context of online advertising involves measuring the incremental change in revenues when a consumer is exposed to $a_{it}$. For simplicity, we assume that the advertiser earns $1 whenever a conversion occurs. Hence the added value of an ad is the incremental change in the conversion probability due to the ad. However, our model can be easily generalized to incorporate different revenues upon conversion.

The ad exposures for consumer $i$ can be divided into three parts – $A_{it-1}$, $a_{it}$ and $S_{it}$. $A_{it-1} (= \{a_{i1}, a_{i2}, \ldots, a_{it-1}\})$ is the ordered set of ad events that have occurred before time $t \leq T$. $S_{it} (= \{a_{it+1}, \ldots, a_T\})$ denotes the suffix at time $t$, i.e. ad exposures that follow time $t$. To ascertain the value
of ad \(a_{it}\) when it is proceeded by ads \(A_{t-1}\) and followed by \(S_t\), we compute the difference in the expected conversion with and without the ad. More specifically,

\[
\psi_{it} = \mathbb{E}[C_i | A_{t-1}a_tS_{t-1}] - \mathbb{E}[C_i | A'_i],
\]

(7)

where \(A'_i = A_{t-1}S_{t-1}\). However, the ads that are shown after \(a_{it}\) are random and should not effect the value of \(a_{it}\). Alternatively, we can derive the probability that a consumer will eventually convert by taking an expectation over all possible paths the consumer can take after ad \(a_{it}\) is shown. The value of the ad unconditional on \(S_{t-1}\) is given by

\[
V_{it} = \mathbb{E}_{S_{it}}[\mathbb{E}[C_i | A_{t-1}a_tS_{t-1}]] - \mathbb{E}_{S_{it}}[\mathbb{E}[C_i | A_{t-1}S_{t-1}]],
\]

(8)

\[
= P(C_i = 1 | A_{it}) - P(C_i = 1 | A_{it-1}).
\]

The effect of an ad depends on the consumer’s underlying state which in turn is affected by the ads that preceded \(a_{it}\). Hence, the value of an ad is not only a function of the impact of the ad going forward, but also depends on the other ad exposures that took place before time \(t\), and the attribution method presented here explicitly accounts for the effect of preceding ads. E.g. if the preceding ads have primed the consumer to convert already, the incremental effect of an additional ad would be close to zero. This method of attribution also implicitly accounts for the effect of other factors like traditional media, e.g. television. If the consumer is likely to convert because of television ads, and not due to any online ads, the consumer heterogeneity modeled in the previous section would capture this phenomenon and the value of each online ad, \(V_{it} = 0\), in this situation. Note that, in our formulation, \(V_{it}\) can either be positive or negative. In cases where the ads lead to aversion as demonstrated by Goldfarb and Tucker (2011), \(V_{it} < 0\). This approach is similar to Shao and Li (2011) and Dalessandro et al. (2012), but \(P(C_i = 1 | A_{it})\) is estimated using a dynamic HMM in our case, whereas they use simplistic approaches like a logistic regression and sample means to compute these probabilities. It should be noted that this method differs vastly from LTA which attributes 100% of the conversion to the last ad event and completely disregards the effects of ads that came earlier.

The value ascribed to a specific type of ad event, \(k \in \{\text{“display impression on generic website”, . . . ,}

16
“search click”), can be computed by summing across all ad activities of that type,

\[ \Pi_k = \sum_i \sum_{t=1}^T \mathbb{1}_{\{a_{it}=k\}} V_{it}, \]  

(9)

where \( \mathbb{1}_{\{a_{it}=k\}} \) is an indicator function that equals one if ad event \( a_{it} \) is of type \( k \). The overall effect of the online campaign can be derived by summing across the various ad events, i.e. \( \pi = \sum_k \pi_k \). Notice that \( \pi \) can be lower or higher than the total number of online conversions. The value of the online campaign can be higher than the number of conversions because online ads might drive consumers down the conversion funnel, and although they did not convert by the end of the campaign, they might convert in the future. Conversely, the value of the campaign can be lower than the number of conversions because there are other factors beside online ads that affect consumers’ decision to convert, e.g. television advertising. The overall effectiveness of the campaign is a combination of these factors. Performing attribution on the basis of incremental change in conversions is a major departure from existing methodologies. Current attribution methodologies assign all the credit for the conversion to the online campaign, which might be erroneous, specifically when traditional advertising is an important factor affecting consumer decision making. This issue is less problematic in cases where the online campaign is the primary driver of customer traffic and other factors like offline advertising or word-of-mouth effects are absent or inconsequential.

5 Empirical Analysis

In this section, we illustrate how the HMM model can be estimated and interpreted. We first outline the estimation procedure, briefly discuss the model validity and continue to present the estimated parameters. A sample of 4121 users is used for estimating the model and the remaining 1000 users are used for validation.

5.1 Estimation Procedure

Here, we outline the procedure of estimating the HMM on the data shown in Section 3. Our model differs from standard HMMs as the transition probabilities depend on the covariates that vary over time. Several techniques have been proposed to incorporate the time varying covariates in the HMM which
are collectively referred to as the latent transition models. Three of the most common techniques used to estimate these models are maximum likelihood estimation (ML), expectation maximization (EM) and Markov chain Monte Carlo (MCMC) methods. However, since our model incorporates customer heterogeneity, we follow the MCMC approach adopted by (Montoya et al., 2010, Netzer et al., 2008) to estimate the Hierarchical Bayesian model.

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

\[
L_i = \pi' \Phi_{i0}(y_{i0}) Q_{i1}(y_{i1}) \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT}(y_{iT}) \Phi_{iT}(y_{iT}) \mathbf{1},
\]

where \( \mathbf{1} \) is a 1 \( \times \) \( |S| \) vector of ones. This computation is significantly faster and can be evaluated in \( O(T|S|^2) \) time. The log-likelihood of observing the entire data is given by the sum of the log-likelihood across all consumers in the data,
\[ LL = \sum_{i} \log \left[ \pi' \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}) \right]. \] (12)

The heterogeneity parameters \( \Psi = \{ \theta, \Omega, \nu, \Delta \} \) and the homogeneous HMM parameters \( \Lambda = \{ \beta, \tau, \gamma, \eta, \alpha \} \) are estimated using an MCMC approach. We use non informative priors and refine them as the estimation proceeds. The exact estimation procedure is outlined in Appendix. We run the MCMC simulation for 400,000 draws and the first 100,000 draws are discarded. The Raftery and Lewis test is used check for the convergence. Subsequently, the MCMC chains are thinned to remove autocorrelation between draws and every 20th draw in the stationary period is used for the subsequent analysis. Before we go into the estimation results, we briefly present our identification strategy and tests for model validity.

5.2 Identification Strategy

In an HMM, we need to identify not only the model parameters but also the states of the HMM. HMMs typically suffer from the label switching problem, i.e. the state label might change when the model is re-estimated (Jasra et al., 2005, Ryden, 2008). This problem occurs because the log-likelihood presented in Equation (12) is invariant to changes in the labels (indices). We address this issue by enforcing the identifiability constraints \( \tilde{\alpha}_1 \leq \tilde{\alpha}_2 \leq \tilde{\alpha}_3 \) and \( \tilde{\eta}_1 \leq \tilde{\eta}_2 \leq \tilde{\eta}_3 \), i.e. consumers are more likely to convert and research as they move down these states. The likelihood function take a value of 0 when the parameters lie outside the region specified by the identifiability constraints. These constraints operationalized through Equations (3) and (4) guarantee that states with a higher likelihood to convert are assigned higher indices (e.g. 3).

Now we discuss how our model parameters \( \Lambda \) and \( \Psi \) are identified. First, we focus on parameters that are constant across consumers – \( \Lambda \). The identification strategy for these parameters is fairly straightforward. These parameters are identified if there is sufficient variation in the ad exposures, conversion and network activity, which is satisfied in this context as shown in Table 1. Next, we consider the heterogeneity parameters, \( \Psi \). Consider that consumers in two different DMA locations receive the same sequence of ads, but behave differently in terms of their conversion and research behavior. The random effect \( \theta_z \) accounts for this average difference in behavior due to DMA specific unobservables like
offline advertising. $\hat{\theta}$ captures the mean DMA specific effect and $\Omega_{\theta}$ captures the variance in this random effect across different DMAs. The variation in consumer behavior at a DMA level, after accounting for all the observable factors, is used to identify $\nu$ and $\Delta$ – the parameters that capture individual-level idiosyncrasies. For our identification strategy to work, we require that conditional on the current state of the consumer, $x_{it}$ is independent of $\theta_z$ and $\xi_i$. Put differently, the advertiser should not target consumers based on their location or demographic variables otherwise the estimation results will be incorrect. This is a reasonable assumption as discussions with our partner advertiser confirmed that they did not use demographics or location based targeting for this campaign.

5.3 Model Validation

To test the validity of our model, we compare the fit and predictive ability of the model with other benchmark models presented below:

**Logit**: The simplest model we use for predicting conversions and network activity is the Logit model (Dalessandro et al., 2012). This model does not include any time dynamics or heterogeneity amongst consumers. This is an important benchmark because it helps us compare the performance of the proposed attribution method to the state-of-the-art attribution technique in literature.

**Latent-Class No Dynamics**: In this model, we introduce consumer heterogeneity by dividing consumers into three latent groups. Although this model accounts for differences in consumer behavior, there are no temporal dynamics and consumer behavior does not change over time.

**HMM No-Het**: This model is identical to the HMM model presented here, but it does not account for consumer heterogeneity. All consumers are ex-ante homogeneous, and differences in their behavior are due to the differences in the ad exposures they receive.

These benchmark models help us identify which factor, e.g. temporal dynamics or heterogeneity, is a more important predictor of their conversion behavior.

We use several different approaches to compare these benchmark models with the model proposed in this paper (**Full-Model**). We compare the Log Marginal Density (LMD) on the training sample. Then we use the validation log-likelihood of the test data. Finally, we compute the root mean-squared error (RMSE) by calculating the difference between the observed outcome and the predicted outcome from the four models. These results are presented in Table 2 below. We observe that the Full-Model
considerably outperforms the other model on all these measures.

<table>
<thead>
<tr>
<th>Table 2: Predictive validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Logit</td>
</tr>
<tr>
<td>Latent-Class No Dynamics</td>
</tr>
<tr>
<td>HMM No Heterogeneity</td>
</tr>
<tr>
<td>Full Model</td>
</tr>
</tbody>
</table>

We see that although accounting for customer heterogeneity considerably improves the model fit and predictive performance, models that incorporate temporal dynamics outperform other competing models. This clearly indicates that consumers move through several stages on their path to conversion. In the subsequent discussion, we present the parameters estimated for the Full Model and show how different factors affect the temporal dynamics.

5.4 Parameter Estimates

5.4.1 Estimates of the Transition Parameters

Estimates of the transition parameters are reported in Table 3. The intercept terms are significantly negative which indicates that these states are relatively sticky, and consumers do not easily transition between them. We also observe that ad related activities have a significant impact on the transition from the dormant state to the awareness state. Contrary to popular belief that display ads are ineffective (de Vries, 2012, Claburn, 2012), we see that display ads have an important effect of moving consumers from a dormant state to a state of awareness. They might not have a high conversion rate, but our model predicts that these ads significantly impact the consumer’s deliberation process. This finding has significant implications for marketers as they need to understand that display advertising has an indirect effect on conversions, and they should account for this difference (as compared to search ads) in their attribution approach. In addition, display ads on auto-specific websites have a larger impact on the transition from the dormant state to the awareness state. Consumers are more likely to notice these car related ads when they are visiting auto-specific websites. This finding is consistent with Yi
(1990) shows that consumers’ response to ads can change significantly when they are primed by relevant context.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{da}$</th>
<th>$\beta_{dc}$</th>
<th>$\beta_{ad}$</th>
<th>$\beta_{ac}$</th>
<th>$\beta_{cd}$</th>
<th>$\beta_{ca}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>generic_imp</td>
<td>0.006</td>
<td>0.001</td>
<td>0.097</td>
<td>0.000</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>specific_imp</td>
<td>0.014</td>
<td>0.002</td>
<td>0.008</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.010)</td>
<td>(0.032)</td>
<td>(0.001)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>generic_clk</td>
<td>0.126</td>
<td>0.020</td>
<td>-0.098</td>
<td>0.383</td>
<td>-0.001</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.063)</td>
<td>(0.150)</td>
<td>(0.049)</td>
<td>(0.038)</td>
<td>(0.312)</td>
</tr>
<tr>
<td>specific_clk</td>
<td>0.189</td>
<td>0.003</td>
<td>0.077</td>
<td>0.501</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.048)</td>
<td>(0.083)</td>
<td>(0.077)</td>
<td>(0.017)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>search_clk</td>
<td>0.550</td>
<td>0.031</td>
<td>-0.029</td>
<td>0.413</td>
<td>0.048</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.023)</td>
<td>(0.138)</td>
<td>(0.057)</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

**Heterogeneity Parameters**

$\bar{\theta}_{ss'}$  
-2.864  
-5.632  
-3.713  
-2.206  
-3.405  
-4.327  

(0.033)  
(1.073)  
(0.078)  
(0.384)  
(0.471)  
(0.732)  

$\Omega_{\theta}$  
0.941  
-5.632  
-3.713  
-2.206  
-3.405  
-4.327  

(0.033)  
(1.073)  
(0.078)  
(0.384)  
(0.471)  
(0.732)  

$\bar{\Sigma}_{ss'}$  
-2.864  
-5.632  
-3.713  
-2.206  
-3.405  
-4.327  

(0.033)  
(1.073)  
(0.078)  
(0.384)  
(0.471)  
(0.732)  

The estimates in bold are significant at a 95% level. For the sake of simplicity, the first letter of the subscript denotes the originating state and the second letter denotes the absorbing state {d = “dormant”, a = “awareness”, c = “consideration”}.

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

Not surprisingly, we observe that clicks have a significant impact on the consumer’s movement from the awareness to the consideration state with search clicks having the largest effect. Once the consumer moves to the consideration state, there is a very low probability of him transitioning out of that state. This probability is further reduced when the consumer performs more searches and clicks on search ads. When a consumer actively starts to gather information about a product (by searching for the product at a search engine), he is likely to be at the very end of the funnel, contemplating his decision just prior to the eventual conversion. We observe significant variation in the intercept parameters across DMAs which implies that consumers in different regions have a different base response to the online ads. We also observe significant within DMA heterogeneity in the intercept parameters. Some of the advertising activity considered in this analysis might be endogenous to a consumer’s behavior, e.g. search clicks. Hence, we need to be careful in the interpretations of results presented in Table 3. The results presented here measure the effect that the advertising activities have on the transition matrix, irrespective of the fact that some of these activities might be endogenous.

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

$$Q_{i0} = \begin{pmatrix}
0.95 & 0.05 & 0.00 \\
0.02 & 0.88 & 0.10 \\
0.03 & 0.01 & 0.96
\end{pmatrix}, \quad Q_{is} = \begin{pmatrix}
0.91 & 0.09 & 0.00 \\
0.02 & 0.84 & 0.14 \\
0.03 & 0.01 & 0.96
\end{pmatrix},$$
\[
Q_{ic} = \begin{pmatrix}
0.94 & 0.06 & 0.00 \\
0.02 & 0.84 & 0.14 \\
0.03 & 0.01 & 0.96
\end{pmatrix},
q_{id} = \begin{pmatrix}
0.87 & 0.12 & 0.00 \\
0.06 & 0.85 & 0.09 \\
0.03 & 0.01 & 0.96
\end{pmatrix}.
\]

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

### 5.4.2 Estimates of the Response Parameters

Now we discuss the underlying parameters that affect the observations of the HMM. We first discuss the factors that affect the number of pages viewed by a customer which are presented in Table 4. We can see that consumers in the awareness and consideration states differ considerably when it comes to their browsing behavior. Consumers in the consideration state view three times as many pages on the car manufacturer’s website as consumers in the awareness state. Since the consumers in these two states behave so differently, we are certain that the model is both empirically and behaviorally identified. Advertising activities tend to increase the consumers’ propensity to view more web pages, but the increase is more pronounced when the consumers actively interact with the ads (e.g. by clicking on them) than when they passively enter the consumers’ perception (e.g. through display impressions).
Table 4: Estimate of factors affecting the page views ($\lambda$)

<table>
<thead>
<tr>
<th></th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta$</td>
<td>0.781***</td>
<td>0.534***</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\bar{\eta}$</td>
<td>0.781</td>
<td>2.487</td>
</tr>
<tr>
<td>generic_imp</td>
<td>0.004**</td>
<td>0.008***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>specific_imp</td>
<td>0.004***</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>generic_clk</td>
<td>0.089***</td>
<td>0.123***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>specific_clk</td>
<td>0.132**</td>
<td>0.207***</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>search_clk</td>
<td>0.169***</td>
<td>0.288***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

Next we consider factors that influence the consumers’ conversion probability. The estimated coefficients of these factors are presented in Table 5. We notice that the probability to convert is higher in the consideration state than it is in the awareness state, ceteris paribus. Apart from impressions on generic websites, all advertising activities lead to an increase in the conversion probability, in the state of awareness. However, conditional on being in the consideration state, impressions of any kind do not have an incremental impact on the likelihood to convert. Interestingly, the effect of a specific click in the awareness state is more prominent than the effect of a generic or a search click. We also observe that an increase in visits to the car manufacturer’s website tends to increase the conversion rate in both states. Surprisingly, this effect is stronger in the state of awareness than in the consideration state. This decrease might be attributed to the diminishing returns from further interactions with the consumer. Once the consumer is sufficiently primed to convert, increased interactions only have a small marginal effect on him.

In Table 6, we present how different activities affect the conversion probability in the awareness and the consideration states. As pointed out earlier, consumers are more likely to convert in the
consideration state than the awareness state. Even though the higher likelihood to convert is imposed by the identification constraints in Equation (4), the base conversion rate in the consideration state is thrice the conversion rate in the awareness state, which illustrates the distinct behavioral difference in the two states. We observe that generic and specific impressions have statistically insignificant impact on the base conversion probabilities in either states. This demonstrates that display impressions only have an indirect effect on consumers propensity to convert. Our results show that the effect of different advertising activities depends on the latent state. E.g., the effect of a specific click is more pronounced in the consideration state than the awareness state. Similarly, a search click is more significant in the awareness state than in the consideration state. In the consideration state, clicks on display impressions (both generic and specific) are most likely to lead to a conversion, whereas in the consideration state search clicks are more likely to lead to conversions. As consumers interact more with the advertiser (through clicks and page views), there is a substantial increase in the conversion probability. When

Table 5: Estimates of conversion parameters

<table>
<thead>
<tr>
<th></th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-4.155***</td>
<td>-3.087**</td>
</tr>
<tr>
<td></td>
<td>(0.433)</td>
<td>(1.050)</td>
</tr>
<tr>
<td>$\tilde{\alpha}$</td>
<td>-4.155</td>
<td>-3.072</td>
</tr>
<tr>
<td>generic_imp</td>
<td>0.015</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>specific_imp</td>
<td>0.017**</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>generic_clk</td>
<td>0.289***</td>
<td>0.318***</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>specific_clk</td>
<td>0.607***</td>
<td>0.303***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>search_clk</td>
<td>0.146***</td>
<td>0.588***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>nw_activity</td>
<td>0.091***</td>
<td>0.067***</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.007</td>
</tr>
</tbody>
</table>
the consumer clicks all the different types of ads and visits the advertiser’s website, her probability to convert increases 150% in the awareness state and 261% in the consideration state. Note that all these increments have been computed keeping the underlying state of the consumer constant. The overall effect of these factors can be different once the transitions are taken into account. It should be kept in mind that the conversion probabilities shown here are atypical of online campaigns, which usually have very few conversions following a click.

<table>
<thead>
<tr>
<th></th>
<th>awareness</th>
<th>consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Activity</td>
<td>0.016</td>
<td>0.046</td>
</tr>
<tr>
<td>Generic Imp</td>
<td>0.016</td>
<td>0.046</td>
</tr>
<tr>
<td>Specific Imp</td>
<td>0.016</td>
<td>0.046</td>
</tr>
<tr>
<td>Generic Click</td>
<td>0.021</td>
<td>0.064</td>
</tr>
<tr>
<td>Specific Click</td>
<td>0.024</td>
<td>0.063</td>
</tr>
<tr>
<td>Search Click</td>
<td>0.018</td>
<td>0.083</td>
</tr>
<tr>
<td>Network Activity</td>
<td>0.017</td>
<td>0.050</td>
</tr>
<tr>
<td>Generic + Specific Clicks</td>
<td>0.032</td>
<td>0.086</td>
</tr>
<tr>
<td>Generic + Search Clicks</td>
<td>0.024</td>
<td>0.115</td>
</tr>
<tr>
<td>Specific + Search Clicks</td>
<td>0.028</td>
<td>0.113</td>
</tr>
<tr>
<td>Generic + Specific + Search Clicks</td>
<td>0.037</td>
<td>0.155</td>
</tr>
<tr>
<td>Generic + Specific + Search Clicks + Network Activity</td>
<td>0.040</td>
<td>0.166</td>
</tr>
</tbody>
</table>

The range presented in parenthesis denotes the 95% range for the posterior distribution of the estimated effect for the different factors.

6 Applications of the Model

The previous section dealt with the estimation of the parameters of our non-homogeneous HMM. Here, we use the estimates from the preceding sections to gain further insights into consumer behavior and access campaign effectiveness.
6.1 Ad Attribution

We begin our analysis by addressing the attribution issue for this campaign. Subsequently, we compare our proposed attribution scheme with the LTA and the logit attribution method proposed by Dalessandro et al. (2012).

Note that our methodology allows the advertiser to measure the effectiveness of an ad for a specific consumer at a specific time in her deliberation process. Accordingly, we perform the attribution at a consumer level and aggregate the result across the entire population to measure the effectiveness of different types of ad in the campaign. It is extremely difficult to compute the closed-form representation of the value of an ad $V_{it}$ presented in Equation (8). Instead, we use simulations to compute the value of an ad to solve the attribution problem. To perform the attribution for a particular consumer, firstly, we draw from the posterior distributions of the DMA-specific parameters $\theta_z$ and $\Sigma_z$, given by $p(\theta_z|\bar{\theta}, \Omega_\theta)$ and $p(\Sigma_z|\nu, \Delta, Data)$. Secondly, we draw the consumer-specific idiosyncrasy, $\xi_i$, from its posterior distribution, $p(\xi_i|\Sigma_z, Data)$. Given the consumer specific heterogeneity parameter $\mu_i = \theta_z + \xi_i$, we simulate the movement of the consumer according to the HMM estimated in Section 5. Consumer $i$’s decision to convert is averaged over 100 random draws to approximate Equation (8).

We use the entire (training + validation) data to compare the attribution methodologies – LTA, logistic multi-touch attribution (Logit), HMM without heterogeneity (HMM-NoHet) and the HMM with consumer heterogeneity (Full Model). The attribution results are presented in Table 7. The last column labeled “%Δ” shows the % difference between the attribution computed by the HMM-MTA and the LTA.

We observe from Table 7 that all methods attribute a significant portion of the conversions to display and search clicks, which is in agreement with the coefficients presented in Table 5. Surprisingly, we see that both HMM-NoHet and Full-Model attribute less credit to display impressions on generic websites. In this data, generic impressions occur very frequently, and as a consequence have a high chance of being the last ad activity that took place before a conversion. Since they are likely to appear last, the LTA gives them undue credits for the conversions even though they might not have had an impact on the consumer’s conversion probability. These ads that get credit just due to their sheer volume have been referred to as “carpet bombers” by Dalessandro et al. (2012). We also see that the HMM based methods increase the number of conversions attributed to display impressions on specific websites,
Table 7: A comparison of attribution methodologies

<table>
<thead>
<tr>
<th>Ad activity</th>
<th>Num. Ads</th>
<th>LTA</th>
<th>Logit</th>
<th>HMM-NoHet</th>
<th>Full-Model</th>
<th>%Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Impression</td>
<td>70,444</td>
<td>171</td>
<td>152.2</td>
<td>124.9</td>
<td>122.5 (109.7,132.6)</td>
<td>-28.4</td>
</tr>
<tr>
<td>Specific Impression</td>
<td>21,564</td>
<td>78</td>
<td>96.5</td>
<td>116.2</td>
<td>106.3 (98.2, 116.8)</td>
<td>36.3</td>
</tr>
<tr>
<td>Generic Click</td>
<td>369</td>
<td>54</td>
<td>84.6</td>
<td>75.1</td>
<td>72.1 (65.2, 82.1)</td>
<td>33.5</td>
</tr>
<tr>
<td>Specific Click</td>
<td>732</td>
<td>150</td>
<td>140.7</td>
<td>167.6</td>
<td>154.9 (65.2, 82.1)</td>
<td>3.3</td>
</tr>
<tr>
<td>Search Click</td>
<td>1,260</td>
<td>328</td>
<td>310.9</td>
<td>294.3</td>
<td>272.5 (265.7, 286.2)</td>
<td>-16.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>781</td>
<td>784.9</td>
<td>778.1</td>
<td>728.3</td>
<td>728.3 (684.7, 768.1)</td>
<td>-6.7</td>
</tr>
</tbody>
</table>

The range presented in parenthesis (for Full-Model) denotes the 95% range for the posterior distribution of the estimated effect for the different channels. For other attribution methodologies, the effect is a point estimate.

which illustrates that our attribution method rewards events that influenced the consumer’s deliberation process early on in the conversion funnel. There is a marginal increase in the conversions attributed to display clicks. HMM-NoHet and Full-Model assign some of the conversions from the generic impression to these activities that had a positive influence on the conversions. Even though there is a slight decrease in the conversions attributed to search clicks, it continues to remain as the most important factor under all the attribution methodologies. This finding is consistent with the results reported by Dalessandro et al. (2012) who show that the Logit does not lead to significant change in the conversion attributed to search ads. We also observe that the Full-Model gives less credit to the ads as compared to other methodologies. This is due to the fact that it accurately captures consumer heterogeneity that might otherwise inflate the temporal effect of ads (Netzer et al., 2008). Some consumers might have converted even without online ads and other attribution methodologies incorrectly credit the campaign for these conversions. In this context, the LTA overestimates the effect of the online campaign by 6.7%.

Table 8: Comparison of different advertising channels

<table>
<thead>
<tr>
<th>Ad activity</th>
<th>Conversions</th>
<th>% Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Display Ads</td>
<td>194.6</td>
<td>26.7</td>
</tr>
<tr>
<td>Specific Display Ads</td>
<td>261.2</td>
<td>35.9</td>
</tr>
<tr>
<td>Search Ads</td>
<td>272.5</td>
<td>37.4</td>
</tr>
<tr>
<td>Others</td>
<td>52.7</td>
<td>6.7</td>
</tr>
</tbody>
</table>

This table presents the mean number of conversions attributed to each channel by the Full Model.
To compute the overall contribution of a specific channel, e.g. generic display ads we need to account for the conversions attributed to generic display impressions and generic display clicks. The overall contributions of the various channels are presented in Table 8. Generic display ads are responsible for 194.6 conversions and specific display ads are responsible for 261.2 conversions, slightly less than search ads, which lead to 272.5 conversions. Interestingly, our methodology credits other sources for 52.7 (6.7%) online conversions. These conversions might be attributed to factors like offline advertising or brand awareness.

### 6.2 Distribution of Consumers

In the previous section, we discussed how the HMM can be used to perform attribution retroactively once the campaign is over. The HMM also allows us to infer the distribution of consumers across different states, and this insight can be used to target consumers based on their current state in the conversion funnel.\(^3\) The probability distribution over a consumer’s state at time \(t\) is given by

\[
P(S_t = s|\mu_i, Y_1, \ldots, Y_t) = \frac{\pi' \Phi_{i0}(y_0) Q_{i1} \Phi_{i1}(y_2) Q_{i2} \ldots Q_{it,s} P(Y_{it} = y_t|S_{it}=s)}{\pi' \Phi_{i0}(y_0) Q_{i1} \Phi_{i1}(y_2) Q_{i2} \ldots Q_{it} \Phi_{it}(y_{it})},
\]

where \(Q_{it,s}\) is the \(s^{th}\) column of the transition matrix \(Q_{it,s}\). Since \(\mu_i\) is a random draw for the consumer, we integrate over the posterior distribution of \(\mu_i|Data_t\), to compute the unconditional state distribution of the consumer, which is given by

\[
P(S_t = s|Y_1, \ldots, Y_t) = \frac{\int \pi' \Phi_{i0}(y_0) Q_{i1} \Phi_{i1}(y_2) Q_{i2} \ldots Q_{it,s} P(Y_{it} = y_t|S_{it}=s) f(\mu_i) \, d\mu_i}{\int \pi' \Phi_{i0}(y_0) Q_{i1} \Phi_{i1}(y_2) Q_{i2} \ldots Q_{it} \Phi_{it}(y_{it}) f(\mu_i) \, d\mu_i}.
\]

We can aggregate the \(P(S_t = s|Y_1, \ldots, Y_t)\) to compute the distribution of consumers at time \(t\). We consider the distributions at three points – (i) at the beginning of our data collection process, (ii) on day 38, the midpoint of our data collection period and (iii) at the end of our data collection period, which are presented in Figure 3. Figure 3 shows that all consumers in our model start out in the dormant state, indicated by the unit mass of consumers at 1 for the dormant state. As time goes by, they are exposed to advertising activity, and hence they transition down the conversion funnel. The distribution at the end of 77 days shows that only 15.2% of the consumers have converted, 18.2% of them are in the

---

\(^3\)Discuss targeting technology
Figure 3: Multiple ad exposures across different online channels.

awareness state and 12.8% of them are in the consideration state at the end of the campaign. The firm can optimally advertise to target these consumers and increase its ROI from the campaign.

Several advertising firms utilize behavioral targeting in their online campaigns, which targets consumers based on prior behavior like website visitation or past purchase. However, most of these methodologies rely solely on observed data. Our approach can extend the practice of behavioral targeting by inferring latent consumer states and proposing the optimal marketing intervention or advertising action conditional on the individual’s present state. E.g. the results presented in Table 3 show that too many generic impressions might be detrimental to consumers who are already aware of the campaign or the product. Hence, the firm should target them with specific impressions or search ads. The proposed methodology can also be useful in identifying customers who are more likely to convert, and targeting them with appropriate ads. Since we are limited with observational data, we postpone this discussion for future research where we can run field experiments to test the effectiveness of such an approach.

7 Discussion and Conclusion

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

In addition to the academic contribution, this paper makes several managerial contributions. Advertising attribution is one of the biggest problems facing the online advertising industry. Several approaches have been proposed in the industry, but these approaches tend to be heuristic in nature and do not model the underlying consumer behavior that drives conversion. This makes it difficult to ascertain the true impact of an ad in a meaningful manner. The paper attempts to bridge this gap in the literature by proposing a rich model of consumer behavior that captures their intrinsic deliberation process. Our proposed methodology has several advantages over existing techniques. Firstly, the model allows the advertiser to estimate the incremental impact of every ad that was shown to the consumer at an individual level. Secondly, it allows the advertiser to discern the underlying latent state of the consumer. The advertiser can thus use this information to optimally choose the subsequent advertising activity. As a consequence, advertisers can target a consumer not only based on observable characteristics but also based on unobserved factors, e.g. the consumer’s latent state. Thirdly, our model incorporates the heterogeneity between consumers within a particular DMA and across DMAs. Controlling for heterogeneity across DMAs allows advertisers to disentangle the effects of the online ads from ads in traditional advertising channels like television, radio and print. Allowing for heterogeneity within the DMA allows the model to capture intrinsic differences in consumer behavior and accurately estimate the effect of an ad on the conversion probability. Finally, our research has significant implications for ad publishers. A better attribution methodology allows better publishers to receive due credit,
thereby increasing the efficiency of the advertising market.

There are a few limitations of our research that present interesting scope for future research. Our current dataset is limited by what’s observed by the advertiser. However, there might be activities that we do not observe, e.g. visits to websites where the advertiser does not advertiser or search impressions. It is extremely difficult to collect this data because of severe limitations in cookie based tracking technology, but future tracking technologies might be able to provide richer and more holistic data to perform this analysis. Our model can be easily extended to incorporate richer data. In the present study, we only look at search and display ads, but our model can be easily extended to incorporate other forms of advertisements where individual level data is available such as email advertising and promotional mailers. One severe limitation of our dataset is the absence of offline advertising data. Accounting for DMA specific heterogeneity allows us to control for traditional advertising, but we cannot measure interactions between traditional advertising and online advertising, which is an interesting research question. Some of our modeling choices are based on our eventual goal – solving the attribution problem. In particular, we choose the proposed non-homogenous HMM with time varying covariates in lieu of the non-stationary HMM, where the transition probability also depends on the time spent in a particular state. The attribution problem becomes extremely intractable with a non-stationary HMM, hence we model the conversion funnel as a non-homogenous HMM with time varying covariates. Future research could incorporate the time dynamics in the attribution process.

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

Acknowledgements

We would like to thanks Wharton Customer Analytics Initiative for the data grant and Organic for making this data readily available to us. We would also like to thank Oded Netzer, Elea Feit, Jun Li and Jose Guajardo for many an insightful discussion. This research is partially supported by the
Berkman faculty award.

Identification Assumption

For simplicity, let’s assume that there are only two states where the initial membership probabilities are denoted by $\pi = \{\pi_1, \pi_2\}$. When a consumer sees an ad, he can follow one of three possible paths - (1) He starts out in the first state and transitions to the second state, (2) he starts out from the first state and stays in the same state after the ad event or (3) starts out from the second state and stays in the same state. The likelihood of observing an outcome $y_1 = (n_1, c_1)'$ after the first ad is shown is given by

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

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

References


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


