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

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

Abstract

Consumers are exposed to advertisers across a number of channels. As such, 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 does each of these ads get? This is one of the most important questions facing the advertising industry today. Although the issue is well documented, current solutions are often simplistic; for e.g., attributing the sale to the most recent ad exposure. In this paper, we address the problem of attribution by developing a Hidden Markov Model (HMM) of an individual consumer's behavior based on the concept of a conversion funnel. We apply the model to a unique data-set from the online campaign for the launch of a car. We observe that different ad formats, e.g. display and search ads, affect consumers differently based on their states in the decision process. Display ads usually have an early impact on the consumer, moving him from a disengaged state to an state in which he interacts with the campaign. On the other hand, search ads have a pronounced effect across all stages. Further, when the consumer 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 provides fundamentally different insights into ad effectiveness relative to the commonly used approaches for attribution. Contrary to the common belief that display ads as are not useful, our results show that display ads affect early stages of the conversion process. Furthermore, we show that only a fraction of online conversions are driven by online ads.

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

1 Introduction

Online advertising has witnessed tremendous growth over the last decade and currently accounts for around one-fifth of the overall US advertising budget. This growth has lead to several innovations in online advertising and advertisers can now reach customers through a variety of formats like search advertising, display ads and social media. Although the proliferation of these advertising formats has enabled marketers to increase their reach considerably, it has given rise to new problems. In particular, marketers have found the question of identifying the most effective online advertising formats or channels to be quite thorny. This problem arises because a typical consumer may be exposed to an advertiser across multiple formats, ranging from display advertising on various websites to sponsored ads on search engines and video advertising on websites such as YouTube. These repeated interactions with an advertiser's campaign are termed "multi-touch" in the popular press (Kaushik, 2012), and they jointly affect a customer's behavior. When a user buys a product or signs up for a service ("converts"), his decision to do so may be influenced by prior ad exposures as shown in Figure 1. Advertisers wish to ascertain how ads across these different channels influence the consumer's decision and to what extent. Quantifying the influence of each ad on a consumer's purchase decision is referred to as the attribution problem. An advertiser needs to assess the contribution of each ad so that he can use this information to optimally allocate the advertising budget. However, these ads affect consumer behavior in a complex fashion and the effect of an ad can depend on the history of prior exposures. As a consequence, solving the attribution problem is non-trivial.

The problem of attribution is not new. It arises in traditional advertising channels such as television and print. However, online channels offer a unique opportunity to address the attribution problem, as advertisers have disaggregate individual level data which were not previously available.[1] Given the lack of disaggregate data, the marketing literature has focused primarily on marketing mix models (Naik et al., 2005, Ansari et al., 1995, Ramaswamy et al., 1993), which perform inter-temporal analysis of marketing channels but fail to provide insights at an individual customer level. Granular online advertising data can be used to build rich models of consumer response to online ads. Unfortunately, until recently, there was very little academic research that analyzed multi-channel/multi-touch advertising

[1] Channels in the context of online advertising refer to, but are not limited to, sponsored search, display advertising, email marketing and social networks.

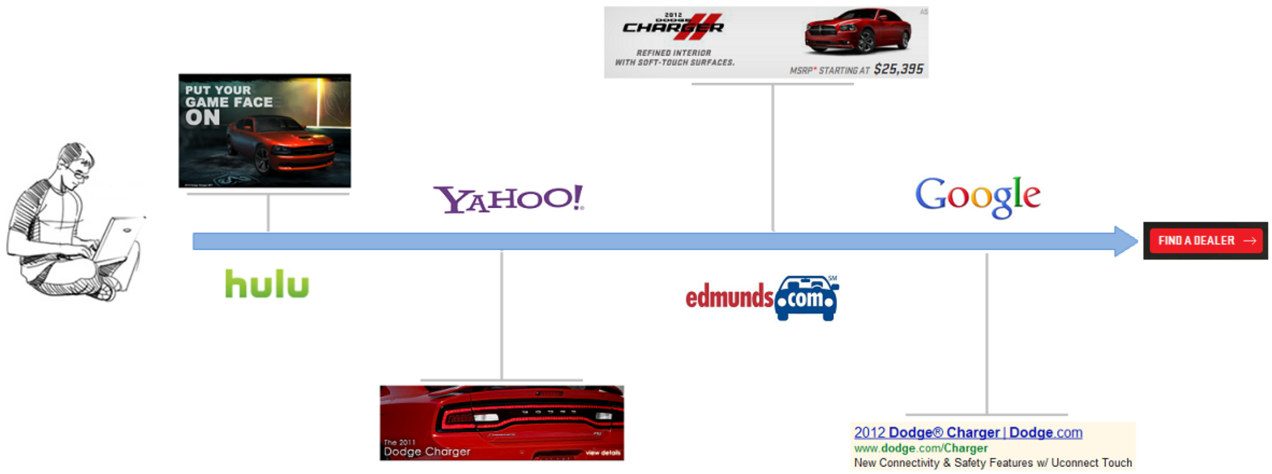

Figure 1: Multiple ad exposures across different online channels.

data to addresses the problem of attribution.

In the absence of suitable 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. LTA causes ads that appear much earlier in the conversion funnel to receive less credit and ads that occur closer to the conversion event to receive most of the credit for a conversion event. For example, a consumer might have started down the path of conversion after being influenced by a display ad, but LTA would suggest that the display ad had no impact on the consumer's decision. Incorrect attribution might move advertising dollars away from the more efficient channels and have a detrimental impact on the advertiser's profitability in the long term. It should be noted that incorrect measurement also alters the publisher's incentives, and they behave sub-optimally (Jordan et al., 2011). If a publisher is undervalues and ad, she might be incentivized to display seemingly more profitable ads. This not only has an adverse effect on the advertiser but also increases 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 the issue of attribution. Companies like Microsoft, Adometry and Clear Saleing have proposed some heuristics to address the issue, but there is no clear consensus on which approach is the most appropriate. Advertisers have come to realize the inadequacies associated with the current methodologies (Chandler-Pepelnjak, 2009, Kaushik, 2012) and, as a result they acknowledge that developing an appropriate attribution model is one of the biggest challenges facing online advertising (Quinn, 2012, Khatibloo, 2010, New York Times, 2012, Szulc, 2012). Surprisingly, there is little academic research on this problem given its managerial relevance. Shao and Li (2011) and Dalessandro et al. (2012) use simple statistical models to address this issue. More recently, Li and Kannan (2014) and Andrel et al. (2013) incorporate models of consumer behavior in their analyses. However, they do not draw from the rich marketing literature on consumer search and deliberation (Kotler and Armstrong, 2011, Barry, 1987, Bettman et al., 1998).

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 disengaged state to the state of conversion, and ads affect the consumer's movement through these different stages. This model is estimated using a unique dataset that contains all the online advertising campaign data associated with the launch of a car. We observe that different ad formats, e.g. display and search ads, affect consumers differently based on their states 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 in which it enters his consideration set. However, when the consumer actively interacts with these ads (e.g. by clicking on them), his likelihood to convert increases considerably. Secondly, we present an attribution scheme based on the HMM that assigns credit to an ad based on the incremental impact it has on the consumer's probability to convert. Compared to the LTA scheme, our proposed methodology assigns relatively greater credit to display ads and lower credit to search ads. This result is contrary to the commonly held belief that display advertising is not effective.

This paper makes three main contributions. Firstly, we propose a comprehensive multi-stage model of consumer response to advertising activity. This model is a considerable improvement over the extant literature on online advertising, in which consumer response models often lack temporal dynamics. Secondly, the consumer model is used to support a new attribution technique that improves upon existing techniques. Finally, from a managerial standpoint, our study informs marketers about how different ad formats influence consumers differently based on their stages of deliberation.

2 Prior Literature

There is significant managerial interest in the attribution problem, but the academic literature in this area has been sparse. However, access to rich multi-channel data has recently led to an increased academic interest in the attribution problem (Shao and Li, 2011, Dalessandro et al., 2012, Wiesel et al., 2011, Li and Kannan, 2014, Andrel et al., 2013, Tucker, 2013). Shao and Li (2011) have developed a bagged logistic regression model to predict how ads from different channels lead to a conversion. 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 that of Shao and Li (2011) to construct a mapping from advertising exposures to conversion probability. These papers are statistically motivated and do not incorporate a model that underlies observed consumer behavior. More recently, Li and Kannan (2014) use a Bayesian framework to understand how consumers interact with a firm using different online channels. Their analysis reveals significant carryover and spillover effects between the online channels; particularly, the effectiveness of paid search is much lower then typically estimated. Given the applied value of this literature, Wiesel et al. (2011) and Andrel et al. (2013) focus on methodologies that can easily be implemented by advertisers to perform attribution. Although most of attribution research is empirically motivated, Jordan et al. (2011) and Berman (2013) propose a game-theoretic approach to analytically devise allocation and payment rules for multi-channel ads.

In this paper, we extend this nascent 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 that 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 broadly related to the literature on online advertising (Tucker, 2012, Goldfarb and Tucker, 2011, Ghose and Yang, 2009, Agarwal et al., 2011). Much of the work in this area has focused on sponsored search where researchers have analyzed factors that 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 social ads (Tucker, 2012). Although, there is significant research on individual formats of online advertising, researchers haven't looked at how these ads interact in a multi-touch context. Notable exception are Kireyev et al. (2013) who study spillovers between display and search advertising but do not explicitly focus on developing an attribution model. 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 advertising.

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, Schweidel et al., 2011, Ascarza and Hardie, 2013). HMM is a workhorse technique in computer science that has been applied to various applications 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). HMMs 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 approximately 11 weeks from June 8, 2009 to August

23, 2009. The ad agency promoted display ads on several generic websites such as Yahoo, MSN and Facebook and auto-specific websites such as KBB and Edmunds. In addition, it also advertised on search engines such as Google and Yahoo. Users were tracked across the different advertising channels and on the car manufacturer's website using cookies. The context of car sales is relevant to the attribution problem, as consumers spend lots of time researching cars online, sometimes several weeks, and as a consequence are exposed to ads in various formats, 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* or *get a quote*. We do not differentiate between 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. Summary statistics of this data at an individual level are presented in Table 3 below.

Table 1: Summary Statistics

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

On average, there are 13.756 display impressions per customer on generic websites and 4.211 impressions on auto-specific websites. Consumers click 0.072 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 an 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 decisions 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 decisions (Bettman et al., 1998). This literature suggests that consumers move through different stages of deliberation during their purchase decision processes. Several marketing actions such as 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 disengaged 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 more granular level and examine the conversion funnel using observational data.

4.2 Hidden Markov Model

In this paper, we build a model to capture the *incremental* effect that online advertising has on the conversion process. Measuring the incremental effect of the ads is the cornerstone of our approach, which is elaborated on in Section 4.3. Based on the prior literature on the conversion funnel, we introduce a staged process through which consumers move from a state of disengagement with the online ads and the advertiser to a state of conversion. The states implicitly capture the consumers' level of engagement with the advertiser, and the level of engagement progressively increases they move along the funnel. However, we do not observe a consumer's underlying state in our data and can infer it 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 the 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 used HMMs to model latent consumer states (Montoya et al., 2010, Netzer et al., 2008, Schwartz et al., 2011, Schweidel et al., 2011). These models are particularly suited for the problem of attribution, as we explain in the next section.

In accordance with the conversion funnel, we construct an HMM with four states (S) where the four states are "disengaged", "active", "engaged" and "converted" (Figure 2). At any time t, consumer i can

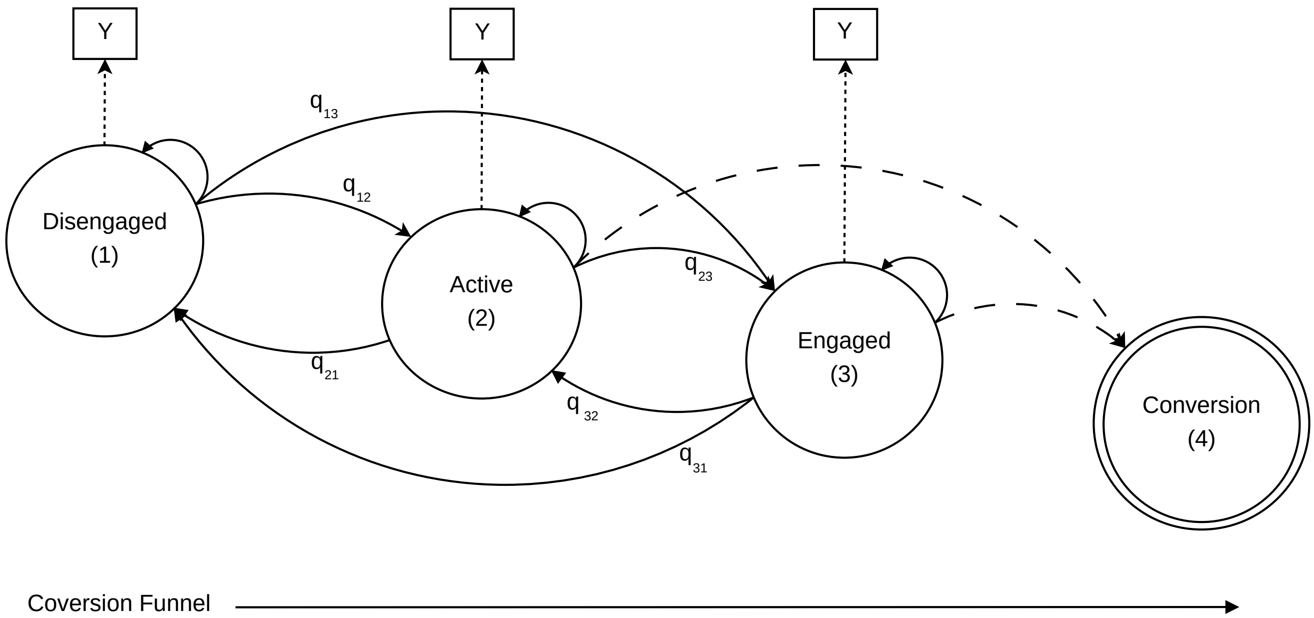

Figure 2: Diagram representing the latent states and the outcomes of the HMM. $q_{ss'}$ denote the transition probabilities from state s to state s' and Y_s is the binary random variable that captures conversion in state s.

be in one of the four states, $S_{it} \in S$.[2] As mentioned earlier, we do not observe s_{it}, but we observe the bivariate outcome variable $Y_{it} = (N_{it},\ C_{it})$ which arises from a stochastic process conditional on the state S_{it}. N_{it} is a Poisson random variable that denotes the number of pages viewed by the consumer between time t and $t+1$, and C_{it} is a binary random variable that captures whether there was a conversion between time t and $t+1$. When the user is in a disengaged state, he is not interacting with the online ads. In this state, a consumer is relatively less active, and we might not observe any online activity from him. As the consumer is exposed to different ads, he might move into an active state where he has interacted with the ads or knows about the product and might be willing to purchase it. On further deliberation, he moves into a state of engagement where he actively looks for product related information and engages with the firm's website. In our formulation, we do not restrict the transition of the customer in any manner, but allow for a flexible model such that consumers can move from any state to any other state. For example, consumers can also go directly from the disengaged state to the engaged state in the model specified here. The *research* activity mentioned in the literature on the conversion funnel is implicitly captured by a consumer's interaction with the advertiser's website,

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

measured through page views. Since we model the very *first* conversion of a consumer, the consumer moves into the "converted" state as soon as a conversion occurs. "Converted" is a dummy absorbing state that 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) 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 engaged state as opposed to the active state. A transition between the states takes 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 $S_i = \{S_{i1}, S_{i2}, \ldots, S_{iT}\}$. $\boldsymbol{x}'_{it}$ captures the advertising stock of the different kinds of advertising activities until 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 sequence of observations $\{Y_{i1} = \boldsymbol{y}_{i1}, \ldots, Y_{iT} = \boldsymbol{y}_{iT}\}$ is a function of three main components:

(i) the transition probabilities between the different states – Q_{it},

(ii) the distribution of the observational variables conditional on the state – M_{it} denotes the probability of conversion and $N_{it} \sim Poisson(\lambda_{its})$ denotes the page views, and

(iii) the initial state distribution – π.

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.[3]. 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, $\boldsymbol{x}'_{it}$, 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

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

$$q_{itss} = \frac{1}{1 + \sum_{s' \in T_s} \exp\{\mu_{iss'} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_{ss'}\}}, \tag{2}$$

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

In addition to the differential effect of ads on the different states, we also need to account for the heterogeneity in the effects of these ads across consumers. Consumers might respond differently to ads because of differences in their prior relationship with the brand, offline advertising activity or underlying demographic variables. If 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 disengaged state to an active state 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, $\boldsymbol{\mu}_i = (\mu_{i12}, \mu_{i21}, \ldots, \mu_{i43})$, to vary across consumers, which captures differences in their responses to online ads. We divide the customer heterogeneity into two distinct components as follows:

$$\boldsymbol{\mu}_i = \boldsymbol{\theta}_z + \boldsymbol{\xi}_i. \tag{3}$$

[3] We allow for transition only at ad events to simplify the attribution problem presented in Section 4.3

where $\boldsymbol{\theta}_z$ captures the heterogeneity due to region specific factors, e.g. offline advertising and 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. $\boldsymbol{\xi}_i$ captures individual specific idiosyncrasies due to factors such as brand awareness or loyalty, affinity for cars, etc. Furthermore, $\boldsymbol{\xi}_i \sim MVN(\Sigma_\xi)$. We model $\boldsymbol{\mu}_i$ in a Hierarchical Bayesian fashion, where $\boldsymbol{\theta}_z$ is a DMA specific parameter drawn from a hyper-prior distribution. The DMA specific mean has the following prior distribution,

$$\boldsymbol{\theta}_z \sim MVN(\bar{\boldsymbol{\theta}}, \Omega_\theta).$$

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.[4]

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 λ_{its}, which is a function of the current state s, and advertising activity $\boldsymbol{x}_{it}$. The probability of observing n_{it} page views is given by

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

where $\lambda_{its} = \widetilde{\eta}_s + \vartheta_z + \boldsymbol{x}'_{it}\tau_s$, i.e. the rate parameter is a function of the intrinsic research activity in state s, a DMA specific random effect and the time varying covariates $\boldsymbol{x}_{it}\tau_s$. Consumers in some regions

[4]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 effect across consumers. Although it is relatively straightforward to extend the model presented here to incorporate heterogeneity in all the coefficients, the sparseness in consumer activity prevents us from doing so.

might be more likely to visit the advertiser's website due to unobservable factors. The parameter ϑ_z is used to capture the unobserved differences between consumers across different DMA and aid in model identification. We also assume that the research intensity increases as the consumer moves down the conversion funnel. This constraint is enforced by setting

$$\begin{aligned}
\widetilde{\eta}_1 &= \exp\{\eta_1\}, \\
\widetilde{\eta}_2 &= \widetilde{\eta}_1 + \exp\{\eta_2\}, \\
\widetilde{\eta}_3 &= \widetilde{\eta}_2 + \exp\{\eta_3\},
\end{aligned} \tag{4}$$

where η_1, η_2 and η_3 are parameters to be estimated from the data.

Modeling conversions:

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

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

α_s captures the intrinsic likelihood to convert in state s, and υ_z captures the DMA specific random effect. $\boldsymbol{z}'_{it}$ denotes 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 is 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. γ_s captures how these covariates affect the conversion probability. We assume that the probability to convert, on average, increases as we move down the conversion funnel. This assumption is operationalized in the following manner,

$$\begin{aligned}
\tilde{\alpha}_1 &= \alpha_1, \\
\tilde{\alpha}_2 &= \tilde{\alpha}_1 + \exp\{\alpha_2\}, \\
\tilde{\alpha}_3 &= \tilde{\alpha}_2 + \exp\{\alpha_3\},
\end{aligned} \tag{5}$$

where $\tilde{\alpha}_1$, α_2 and α_3 are the parameters to be estimated from the data. This structure enforces that $m_{it3} \geq m_{it2} \geq m_{it1}$, *ceteris paribus.* This assumption ensures the identification of the different states and is consistent with the approach adopted by Netzer et al. (2008), Montoya et al. (2010) and Ascarza and Hardie (2013).

The customer heterogeneity in the consumer research and conversion behavior is modeled in the following manner. Similar to the approach adopted in the previous section, we assume that

$$\begin{pmatrix} \vartheta_z \\ \upsilon_z \end{pmatrix} \sim MVN\left(0, \Omega_{\vartheta\upsilon}\right). \tag{6}$$

As some unobserved factors that drive visits to the advertiser's website and conversions might be common, we propose a flexible model that allows ϑ_z and υ_z to be correlated.

Joint density of conversion and page views:

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

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

where $\boldsymbol{y}_{it} = (n_{it},\ c_{it})'$ is the realized outcome variable. Un-conditioning Y_{it} on the DMA specific random effects gives us

$$P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s) = \int m_{its}^{c_{it}}(1 - m_{its})^{(1-c_{it})} P(N_{its} = n_{it} | S_{it} = s, \vartheta_z, \upsilon_z) f_{iz}(\vartheta, \upsilon) d\vartheta d\upsilon, \tag{7}$$

where $f_{iz}(\vartheta, \upsilon)$ is the joint density of the DMA specific effects.

4.2.3 Initial State Membership

Let π_{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 of the conversion funnel because of their exposure to ads in other media such as television or print, which can affect the initial membership probability. However, in this paper, we are interested in measuring the incremental effect of online advertising on conversions. Since we have data from the beginning of the campaign, we assume that none of the consumers have interacted with the online ads and all of them start out in the disengaged state and move down the conversion funnel. Accordingly, the initial membership probability is given by $\boldsymbol{\pi}_i = \{1, 0, 0, 0\}$. We think this is a reasonable assumption because the advertising campaign pertains to a new brand of cars, and consumers might have been completely unaware of the product before the launch of the online campaign. In addition, the DMA specific random effects capture unobservables like offline advertising, etc. that might affect customer engagement with the advertiser.

4.2.4 Discussion of the HMM

In summary, the dynamic heterogeneous HMM captures 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 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 allows us to attributes 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 Dalessandro et al. (2012), which attribute credit to an ad only when it directly leads to conversion. Our model also differs from recent work by Xu et al. (Forthcoming) and Li and Kannan (2014), as we introduce a staged model of conversion unlike the approach adopted by them. 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, $A_{it} = \{a_{i1}, a_{i2}, \ldots a_{it}\}$, 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. All the ads seen by a consumer are denoted by $A_i = \{a_{i1}, a_{i2}, \ldots a_{iT}\}$. In our model, the ad related event a_{it} affects the customer i's underlying time varying parameters $\boldsymbol{x}'_{it}$ and $\boldsymbol{z}'_{it}$ as shown below,

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

Hence, a_{it} has a twofold effect on the consumer's probability to convert – (i) it alters the conditional conversion probability, through changes in $\boldsymbol{z}'_{it}$, and (ii) it can lead to a transition of the consumer from one state to another by affecting $\boldsymbol{x}'_{it}$. Attribution in the context of online advertising involves measuring the incremental change 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.

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

$$\psi_{it} = \mathbb{E}[C_i|A_i] - \mathbb{E}[C_i|A_{i\setminus t}], \tag{8}$$

where $A_{i\setminus t}$ is the ordered sequence of all ads except ad a_{it}, i.e. $A_{i\setminus t} = A_{it-1}A_{it^+}$. However, the ads that are shown after a_{it} are random and should not affect 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 A_{it^+} is given by

$$\begin{aligned} V_{it} &= \sum_{A_{it^+}} \left\{ \mathbb{E}[C_i | A_i] - \mathbb{E}[C_i | A_{i\setminus t}] \right\} p(A_{it^+}), \\ &= P(C_i = 1 | A_{it}) - P(C_i = 1 | A_{it-1}). \end{aligned} \tag{9}$$

Hence, the effect of an ad depends on the consumer's underlying state, which in turn is affected by the ads that preceded a_{it}. As a consequence, the value of an ad is a function not only of the impact of the ad going forward, but also of the other ad exposures that took place before time t. 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 effect and the value of each online ad, V_{it}, would be estimated to be zero 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 that of 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, $k \in$ {"display impression on generic website", ..., "search click"}, can be computed by summing across all ad activities of that type,

$$\Pi_k = \sum_i \sum_{t=1}^{T} \mathbb{1}_{\{a_{it}=k\}} V_{it}, \tag{10}$$

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 Π 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 a consumer's 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 4932 users is used for estimating the model, and the remaining 1500 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. Since our model incorporates customer heterogeneity, we follow the Markov chain Monte Carlo (MCMC) approach adopted by Montoya et al. (2010) and 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 $y_i = \{\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \ldots \boldsymbol{y}_{iT}\}$. The likelihood of a matrix $(2 \times T)$ of outcome variables y_i after being exposed to these actions A_i can be computed by evaluating the probabilities of each of these paths $s_0 \to \ldots \to s_T$ and the conditional probability of $P(Y_{i1} = \boldsymbol{y}_{i1}, \ldots, Y_{iT} = \boldsymbol{y}_{iT}, |S_0 = s_0, \ldots, S_T = s_T)$, which is given by

$$L_i = \sum_{s_1=1}^{|S|} \sum_{s_2=1}^{|S|} \cdots \sum_{s_T=1}^{|S|} \left[\prod_{t=1}^{T} P(S_{it} = s_t | S_{it-1} = s_{t-1}) \prod_{t=1}^{T} P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t) \right], \tag{11}$$

where $P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t)$ can be computed as shown in Equation (7). This approach of summing over all possible paths has a complexity $O(|S|^T)$ and might be computationally infeasible even for moderately small values of $|S|$ and T (Cooper and Lipsitch, 2004). In order to overcome this computational complexity, McDonald and Zucchini (1997) propose an approach that significantly reduced the amount of computation required. Let

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

where $P(Y_{it} = \boldsymbol{y}_t | S_{it} = s)$ is as mentioned in Equation (7). The likelihood of the observed data (Equation (11)) can be simplified to

$$L_i = \boldsymbol{\pi'} \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}).\mathbf{1}, \tag{12}$$

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,

$$LL = \sum_i \log \left[\boldsymbol{\pi'} \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}).\mathbf{1} \right]. \tag{13}$$

The heterogeneity parameters $\boldsymbol{\Theta} = \{\bar{\boldsymbol{\theta}}, \Omega_\theta, \Sigma_\xi, \Omega_{\vartheta\upsilon}\}$ and the homogeneous HMM parameters $\boldsymbol{\Psi} = \{\beta, \tau, \gamma, \eta, \alpha\}$ 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 the appendix. We run the MCMC simulation for 400,000 draws, and the first 200,000 draws are discarded. The Raftery and Lewis test is used to 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 (13) 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 takes a value of 0 when the parameters lie outside the region specified by the identifiability constraints. These constraints operationalized through Equations (4) and (5) guarantee that states with a higher likelihood to convert are assigned higher indices (e.g. 3).

Now we discuss how our model parameters $\mathbf{\Psi}$ and $\mathbf{\Theta}$ are identified. First, we focus on *fixed-effect* parameters that are constant across consumers – $\mathbf{\Psi}$. An important consideration in this context is the effect of unobserved factors like offline advertising, brand loyalty and brand awareness on consumer behavior. These factors might affect the response to online ads or consumers' movement down the conversion funnel. Since several of these factors, especially offline advertising, are determined at a DMA level, we try to control for these unobservables using DMA level random effects. Furthermore, we incorporate within-DMA customer level heterogeneity in the transition probabilities, which can account for customer idiosyncrasies. In addition, the behavior of consumers who are not exposed to any ads helps us empirically generate a baseline for consumer behavior, and the effects of the ads is measured in relation to this baseline. We believe that these aspects of our empirical strategy are appropriate for a clean identification of the homogeneous model parameters.

Next, we consider the heterogeneity parameters, $\mathbf{\Theta}$. 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 $\boldsymbol{\theta}_z$ accounts for this average difference in behavior due to DMA specific unobservables like offline advertising. $\bar{\boldsymbol{\theta}}$ captures the mean DMA specific effect and Ω_θ 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 Σ_ξ – the parameter that captures 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 the DMA specific parameters $\boldsymbol{\theta}_z, \vartheta_z$ and υ_z and individual level idiosyncrasies captured by $\boldsymbol{\xi}_i$. Put differently, the advertiser should not target consumers based on their locations or demographic variables, otherwise the estimation results will be biased. This is a reasonable assumption in our context, as discussions with our partner advertiser confirmed that they did not use demographics or location based targeting for this campaign.

5.3 Model Validation

As mentioned earlier, we divide our data into two parts – we use 4932 users to estimate the model and the remaining users to test the estimated model. We test the validity of the model in two ways. Firstly, we identify the appropriate number of states in the HMM. Secondly, we compare the model proposed here to several benchmark models.

In order to determine the appropriate number of hidden states, we estimate the model with varying numbers of states from two to five and compare these models using (i) the log marginal density, (ii) validation log-likelihood, and (iii) the RMSE of conversion. Note that the last state in all these models is the conversion state. The performance of the different HMMs along these three metrics is presented in Table 2. We observe that an HMM with 4 states outperforms all the competing models on log marginal density and validation log-likelihood. It also has the best out-of-sample predictive performance at 0.095, which is considerably better than other competing HMM models.

Table 2: Determining the number of states

Number of States	Log Marginal Density	Validation log-likelihood	Conversion RMSE
2	-2095.3	-820.1	0.131
3	-1891.8	-722.9	0.107
4	-1695.7	-551.7	0.095
5	-1732.4	-588.1	0.101

To test whether an HMM is the most appropriate for modeling customers in this setting, 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-Dyna: 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.

No-Het HMM: This model is identical to the HMM presented earlier, 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, 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**). First, 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 3 below. We observe that the Full-Model considerably outperforms the other models on all these measures.

Table 3: Predictive validity

	Log Marginal Density/ Log Likelihood	Validation log-likelihood	Conversion RMSE
Logit	-2139.5	-721.8	0.182
Latent-Class No-Dyna	-1942.2	-639.2	0.121
No-Het HMM	-1782.8	-605.8	0.107
Full-Model	-1695.7	-551.7	0.095

We also observe 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 4. 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 statistically significant impact on the transition from the disengaged state to the active state. Contrary to the 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 disengaged state to an active state. 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 compelling implications for marketers as they need to understand that upstream ads, e.g. display ads, have an indirect effect on conversions (as compared to downstream ads like search ads), and they should account for this difference in their attribution approach. In addition, display ads on auto-specific websites have a larger impact on the transition from the disengaged state to the active 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), who shows that consumers' response to ads can change significantly when they are primed by relevant context. Hence, advertisers should be willing to pay more for display ads on webpages that are contextually more relevant.

Although display ads have an impact on moving consumers from the disengaged state to an active state, they do not have an impact on moving consumers further down the conversion funnel, i.e. from an active state to an engaged state (β_{ae}). 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 (0.097), it suggests that if consumers are shown too many display ads on generic websites, their probability to transition back to the disengaged state increases considerably. We also observe that impressions do not have an impact later on in the conversion funnel. Thus, current attribution techniques, which focus mostly at the end of the funnel,

Table 4: Estimates of the transition parameters (β)

	β_{da}	β_{de}	β_{ad}	β_{ae}	β_{ed}	β_{ea}
generic_imp	**0.005**	0.001	**0.097**	0.000	0.009	0.005
	(0.003)	(0.003)	(0.008)	(0.011)	(0.012)	(0.009)
specific_imp	**0.012**	0.002	0.008	0.003	-0.001	0.003
	(0.009)	(0.005)	(0.010)	(0.032)	(0.001)	(0.007)
generic_clk	**0.114**	0.020	-0.098	**0.383**	-0.001	-0.079
	(0.048)	(0.063)	(0.150)	(0.049)	(0.038)	(0.312)
specific_clk	**0.191**	0.003	0.077	**0.501**	0.021	0.000
	(0.003)	(0.048)	(0.083)	(0.077)	(0.017)	(0.000)
search_clk	**0.550**	0.031	-0.029	**0.413**	0.048	**-0.002**
	(0.025)	(0.023)	(0.138)	(0.057)	(0.001)	(0.000)
Heterogeneity Parameters						
$\bar{\theta}_{ss'}$	**-2.864**	**-5.632**	**-2.188**	**-2.206**	**-3.405**	**-4.327**
	(0.033)	(1.073)	(0.078)	(0.384)	(0.471)	(0.732)
Ω_θ	**0.941**	1.2041	**0.937**	**1.006**	**0.951**	**1.518**
	(0.320)	(1.071)	(0.287)	(0.185)	(0.358)	(0.661)
Σ_ξ	1.411	**3.863**	2.137	**1.260**	1.555	**1.732**
	(1.303)	(0.523)	(1.178)	(2.384)	(0.971)	(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 $\{d =$"disengaged", $a =$"active" , $e =$ "engaged"$\}$. The range presented in parentheses denotes the standard deviation for the posterior distribution of the estimated effect for the different factors.

give negligible credit to these ads.

Not surprisingly, we observe that clicks have a significant impact on a consumer's movement from the disengaged to the active state, with search clicks having the largest effect. Once the consumer moves to the engaged 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 on 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 responses to online ads. We also observe significant within-DMA heterogeneity in the intercept parameters.

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 impressions for the consumer, respectively. These matrices are presented below,

$$Q_{i0} = \begin{pmatrix} 0.95 & 0.05 & 0.00 \\ 0.09 & 0.81 & 0.10 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}, \quad Q_{is} = \begin{pmatrix} 0.91 & 0.09 & 0.00 \\ 0.09 & 0.77 & 0.14 \\ 0.03 & 0.01 & 0.96 \end{pmatrix},$$

$$Q_{ic} = \begin{pmatrix} 0.94 & 0.06 & 0.00 \\ 0.09 & 0.77 & 0.14 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}, \quad Q_{id} = \begin{pmatrix} 0.87 & 0.12 & 0.00 \\ 0.15 & 0.76 & 0.09 \\ 0.03 & 0.01 & 0.96 \end{pmatrix}.$$

In the absence of any ad related activity, the states are extremely sticky and it is unlikely that a 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 disengaged state (in the opposite direction of the funnel) increases (q_{i21} : $0.09 \rightarrow 0.15$). 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 disengaged state to an engaged state.

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 5. We can see that consumers in the disengaged, active and engaged states differ considerably when it comes to their browsing behavior. Consumers in the disengaged state are extremely unlikely to view any pages at the manufacturer's website. Consumers in the active state on average view 0.781 pages, while those in the engaged state view three times as many pages on the car manufacturer's website as do consumers in the active state. Since the consumers in all these states behave so differently, we are certain that the model is both empirically and behaviorally identified. Advertising activities tend to increase consumers' propensity to view more web pages, but the increase is more pronounced when consumers actively interact with the ads (e.g. by clicking on them) than when they passively enter consumers' perceptions (e.g. through display impressions).

Next we consider factors that influence the consumers' conversion probability. The estimated coefficients of these factors are presented in Table 6. We notice that the probability to convert due to ads is higher in the engaged state than it is in the active state, which is higher than the conversion rate in the disengaged state, *ceteris paribus*. Ads do not play a significant role in conversion in the disengaged state, and conversions are primarily driven by unobserved customer heterogeneity. Apart from impressions on generic websites, all advertising activities lead to an increase in the conversion probability, in the active state. This result is consistent with the common finding that generic display ads do not lead to conversions. However, as we argued earlier, even though generic display impressions might not lead to conversions directly, they might move consumers down the conversion funnel. Furthermore, we also observe that – conditional on being in the engaged state – impressions of any kind do not have an incremental impact on the likelihood to convert. Interestingly, the effect of a specific

Table 5: Estimate of factors affecting the page views (λ)

	τ_1	τ_2	τ_3
η	**0.096**	**-0.378**	**0.534**
	(0.024)	(0.045)	(0.003)
$\widetilde{\eta}$	0.096	0.781	2.487
generic_imp	0.001	**0.004**	**0.008**
	(0.002)	(0.003)	(0.000)
specific_imp	0.004	**0.004**	0.005
	(0.001)	(0.013)	(0.009)
generic_clk	**0.005**	**0.089**	**0.123**
	(0.001)	(0.008)	(0.007)
specific_clk	**0.001**	**0.132**	**0.207**
	(0.000)	(0.060)	(0.008)
search_clk	0.010	**0.169**	**0.288**
	(0.014)	(0.004)	(0.004)

The estimates in bold are significant at a 95% level.

click in the active state is more prominent than the effect of a generic or a search click. One plausible explanation for this observation is the fact that consumers who are actively looking for car related information on auto-specific websites might be further along the funnel and are likely to respond to an ad that is extremely relevant to their browsing intent. 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 weaker in the engaged state than in the active state. This decrease might be attributed to the diminishing returns from further interactions with the consumers. Once consumers are sufficiently primed to convert, increased interactions have only a marginal effect on them.

In Table 7, we present how different activities affect the conversion probability in the active and the engaged states. We ignore the disengaged state in this analysis, as the conversion probabilities are too low to warrant a meaningful discussion. As pointed out earlier, consumers are more likely to convert in the engaged state than in the active state. Even though the higher likelihood to convert is imposed by the identification constraints in Equation (5), the base conversion rate in the engaged state state is

Table 6: Estimates of conversion parameters

	γ_1	γ_2	γ_3
α	**-6.982**	**1.039**	**0.078**
	(0.531)	(0.433)	(0.021)
$\widetilde{\alpha}$	-6.982	-4.155	-3.072
generic_imp	0.002	0.015	0.008
	(0.005)	(0.010)	(0.019)
specific_imp	0.000	**0.017**	0.020
	(0.002)	(0.009)	(0.019)
generic_clk	**0.010**	**0.289**	**0.318**
	(0.004)	(0.084)	(0.095)
specific_clk	0.008	**0.607**	**0.303**
	(0.024)	(0.090)	(0.083)
search_clk	**0.002**	**0.146**	**0.588**
	(0.000)	(0.027)	(0.100)
nw_activity	**0.009**	**0.091**	**0.067**
	(0.004)	0.005	0.007

The estimates in bold are significant at a 95% level.

thrice the conversion rate in the active state, which illustrates the distinct behavioral difference in the two states. We observe that generic and specific impressions have a 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. The effect of different advertising activities depends on the latent state; e.g., the effect of a specific click is more pronounced in the engaged state than the active state. Similarly, a search click is more significant in the active state than in the engaged state. In the active state, clicks on display impressions (both generic and specific) are more likely to lead to a conversion, whereas in the engaged state, search clicks are more likely to lead to conversions. In general, as consumers interact more with the advertiser (through clicks and page views), there is a substantial increase in the conversion probability. When the consumer clicks all the different types of ads and visits the advertiser's website, her probability to convert increases 183% in the active state and

192% in the engaged 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 7: Conversion probability as a result of various factors.

	active	engaged
No Activity	0.012	0.039
Generic Imp	0.012	0.039
Specific Imp	0.012	0.039
Generic Click	0.017	0.056
Specific Click	0.021	0.054
Search Click	0.014	0.073
Network Activity	0.011	0.047
Generic + Specific Clicks	0.032	0.079
Generic + Search Clicks	0.024	0.095
Specific + Search Clicks	0.028	0.091
Generic + Specific + Search Clicks	0.030	0.102
Generic + Specific + Search Clicks + Network Activity	0.034	0.114

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 campaign effectiveness.

6.1 Ad Attribution

We first address 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 ads in the campaign. It is difficult to compute the closed-form representation of the value of an ad V_{it} presented in Equation (9). 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 $\boldsymbol{\theta}_z, \vartheta_z, \upsilon_z$ and Σ_ξ, given by $p(\boldsymbol{\theta}_z|\bar{\boldsymbol{\theta}}, \Omega_\theta, \text{data})$, $p(\vartheta_z, \upsilon_z|\Sigma_{\vartheta\upsilon}, \text{data})$ and $p(\Sigma_\xi|\nu, \Delta, \text{data})$. Secondly, we draw the consumer-specific idiosyncrasy, $\boldsymbol{\xi}_i$ from its posterior distribution, $p(\boldsymbol{\xi}_i|\Sigma_\xi, \text{data})$. Given the consumer specific heterogeneity parameter $\boldsymbol{\mu}_i = \boldsymbol{\theta}_z + \boldsymbol{\xi}_i$, 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 (9).

We use the entire (training + validation) data to compare the attribution methodologies – LTA, logistic multi-touch attribution (Logit), HMM without heterogeneity (No-Het HMM) and the HMM with consumer heterogeneity (Full-Model). The attribution results are presented in Table 8. The columns labeled "%Δ1" and "%Δ2" show the % difference between the attribution computed by LTA and the No-Het HMM, and LTA and the Full-Model, respectively. Intuitively, %Δ1 captures the relative underestimation or overestimation of an ad's effectiveness by LTA assuming that all the conversions were attributed to the ads, whereas %Δ2 captures the absolute underestimation or overestimation when unobserved heterogeneity is taken into account.

We first compare the No-Het HMM with LTA and the Logit models, and subsequently we compare all these models to the Full-Model. Focusing on the No-Het HMM helps us understand the advantages of the multi-stage model proposed in this paper, and the comparison with the Full-Model helps bolster our understanding with the incorporation of unobserved heterogeneity, which might be an important factor in this context. From Table 8, we can observe 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 6. Surprisingly, we see that the No-Het HMM attributes 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 takes place before a conversion. Since they are likely to

Table 8: A comparison of attribution methodologies

Ad activity	Num. Ads	LTA	Logit	No-Het HMM	%Δ1	Full-Model	%Δ2
Generic Impression	70,444	171	152.2	124.9	36.9	80.7 [61.9, 119.2]	111.9
Specific Impression	21,564	78	96.5	116.2	-32.9	71.2 [54.2, 92.8]	9.5
Generic Click	369	54	84.6	75.1	-28.1	50.5 [20.7, 82.5]	6.5
Specific Click	732	150	140.7	167.6	-10.5	101.1 [65.2, 142.1]	48.4
Search Click	1,260	328	310.9	294.3	11.5	183.3 [164.7, 206.3]	69.2
Total		781	784.9	778.1	0.00	486.7 [371.9, 583.1]	60.4

%Δ1 and %Δ2 indicate an overestimation by LTA for positive values and underestimation for negative values. The range presented in parentheses (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.

appear last, the LTA gives them undue credit for the conversions, even though they might not have had an impact on the consumer's conversion probability. These ads that get credit 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, which illustrates that our attribution method rewards events that influenced the consumer's deliberation process early on in the conversion funnel. This result demonstrates the strength of our approach, as the effectiveness of display ads is identified due to the multi-stage model adopted by us. This finding is also consistent with the results reported by Li and Kannan (2014) and Andrel et al. (2013). There is a marginal increase in the conversions attributed to display clicks. The No-Het HMM assigns some of the conversions from the generic impression to these activities that have 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.

The most startling observation from Table 8 is that the Full-Model assigns only a fraction of the conversions to advertising activities 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 60.4%. Other methodologies, including No-Het HMM, perform poorly, as they do not account for unobserved variables like offline advertising and brand awareness that might drive online conversions. This result demonstrates another strength of our model – measuring the incremental effect of online ads, which can correctly guide advertisers in their media buying decisions.

Table 9: Comparison of different advertising channels

Ad activity	Conversions	% Contribution
Generic Display Ads	131.7	16.8
Specific Display Ads	172.3	22.1
Search Ads	183.3	23.5
Others	294.3	37.7

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 9. Generic display ads are responsible for 131.7 conversions and specific display ads are responsible for 272.3 conversions, slightly less than search ads, which lead to 183.3 conversions. Interestingly, our methodology credits other sources for 294.3 (37.7%) of the online conversions. These conversions might be attributed to factors like offline advertising or brand awareness. The fraction of conversions that can be attributed to the online campaign is fairly low in this case (486.7 out of 781), but we believe that that result is context specific. Car manufacturers run large offline campaigns, and a significant portion of the conversions can be driven by these offline ads, as we find in our analysis. However, the online campaign can be responsible for a majority of the conversions in other contexts where offline media is absent or unobserved customer heterogeneity affecting the conversion is small.

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 states in the conversion funnel.[5] The probability distribution over a consumer's state at time t is given by

$$P(S_t = s|\mu_i, Y_1, \ldots, Y_t) = \frac{\boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it,s}P(Y_{it} = y_t|S_{it=s})}{\boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it}\Phi_{it}(y_{it}).\mathbf{1}}, \tag{14}$$

where $Q_{it,s}$ is the s^{th} column of the transition matrix $Q_{it,s}$. Since μ_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

$$P(S_t = s|Y_1, \ldots, Y_t) = \frac{\int \boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it,s}P(Y_{it} = y_t|S_{it=s})\, f(\mu_i)\, \mathrm{d}\mu_i}{\int \boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it}\Phi_{it}(y_{it}).\mathbf{1}\, f(\mu_i)\, \mathrm{d}\mu_i}.$$

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 disengaged state, indicated by the unit mass of consumers at 1 for the disengaged 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, 20.1% of them are in the active state and 10.4% of them are in the engaged state at the end of the campaign. The firm can optimally advertise to target these consumers and increase its ROI from the campaign. Figure 3 also demonstrates that consumers move very slowly from one stage to another, consistent with prior findings that indicate consumers spend several weeks researching cars.

Several advertising firms utilize behavioral targeting in their online campaigns, which targets consumers based on prior behavior such as 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. For instance, the results presented in Table 3 show that too many generic impressions might be detrimental to consumers who are already aware of the

[5] Advertising networks are working on technologies that can be used to track and target customers in real-time. Such technology can use the proposed model to target a customer with an optimal ad based on his inferred state of deliberation.

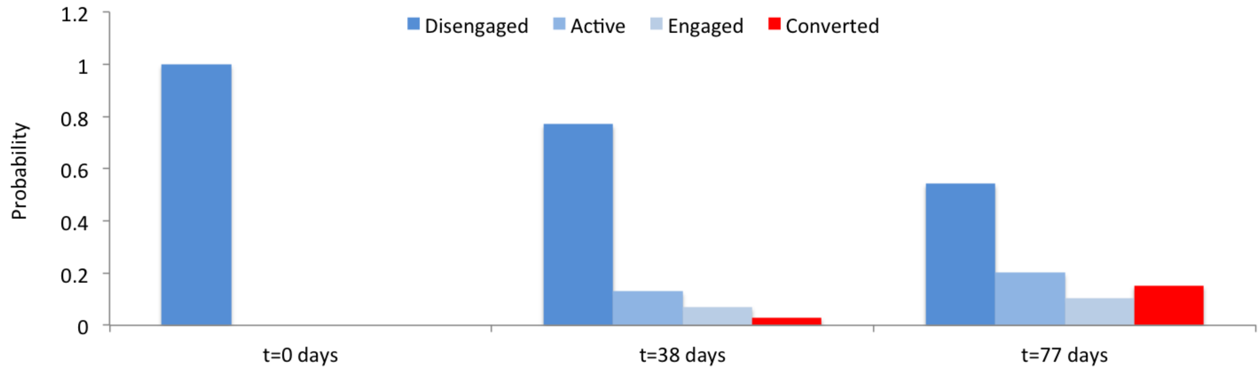

Figure 3: Multiple ad exposures across different online channels.

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. We propose this as a direction 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. Consumer behavior is modeled using a dynamic HMM which is based on the conversion funnel. A consumer moves through the states of the HMM in a stochastic manner when he is 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 he is exposed to an ad, but 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.

A few limitations of our research present interesting opportunities 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. search impression or visits to websites where the advertiser does not advertise. 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-homogeneous 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 intractable with a non-stationary HMM, hence we model the conversion funnel as a non-homogeneous HMM with time varying covariates. Future research could incorporate the time dynamics in the attribution process.

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 and the Mack Center grant for technological innovation.

References

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

E. Andrel, I. Becker, F. V. Wangenheim, and J. H. Schumann. Putting attribution to work: A graph-based framework for attribution modeling in managerial practice. *SSRN*, October 2013.

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

E. Ascarza and B. G. S. Hardie. A joint model of usage and churn in contractual settings. *Marketing Science*, 32(4):570–590, 2013.

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

R. Berman. Beyond the last touch: Attribution in online advertising. *Working paper, University of California, Berkeley*, 2013.

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

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.

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

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

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

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

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

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

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

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

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

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

A. Jasra, C. C. Holmes, and D. A. Stephens. Mcmc and the label switching problem in bayesian mixture models. *Statistical Science*, 20:50–67, 2005.

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

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

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

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

P. Kireyev, K. Pauwels, and S. Gupta. Do display ads influence search? attribution and dynamics in online advertisingi. *HBS Working Paper series*, February 2013.

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

H. Li and P. Kannan. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1):40–56, 2014.

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

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

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

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

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

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

O. Netzer, J. M. Lattin, and V. Srinivasan. A hidden markov model of customer relationship dynamics.

Marketing Science, 27(2):185–204, 2008.

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

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

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

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

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

T. Ryden. Em versus markov chain monte carlo for estimation of hidden markov models: A computational perspective. *Bayesian Analysis*, 3:659–688, 2008.

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

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

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

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

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

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

C. Tucker. The implications of improved attribution and measurability for antitrust and privacy in online advertising markets. *George Mason Law Review*, 20(4), Summer 2013.

T. Wiesel, K. Pauwels, and J. Arts. Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science*, 30(4):604–611, 2011.

L. Xu, J. A. Duan, and A. B. Whinston. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, Forthcoming.

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

APPENDIX

MCMC Estimation

This appendix describes the estimation procedure used for the parameters presented in Section 5. We use an MCMC technique to estimate the parameters. The detailed components of the MCMC estimation procedure are as follows.

Prior Distributions

The prior distribution of the various *random-effect* parameters, $\mathbf{\Theta}$, are outlined below:

1. **Individual specific idiosyncratic shock:**

$$\boldsymbol{\xi}_i \sim MVN(\mathbf{0}, \Sigma_z) \Rightarrow P(\boldsymbol{\xi}_i) \propto |\Sigma_z|^{-1} \exp\left\{-\frac{1}{2}\boldsymbol{\xi}_i' \Sigma_z^{-1} \boldsymbol{\xi}_i'\right\},$$

   where z denotes individual i's DMA.

2. **Variance matrix for idiosyncratic shock:**

$$\Sigma_\xi^{-1} \sim Wishart(\nu, \Delta).$$

3. **DMA specific intercept term:**

$$\boldsymbol{\theta}_z \sim MVN(\bar{\boldsymbol{\theta}}, \Omega_\theta) \Rightarrow P(\boldsymbol{\theta}_z) \propto |\Omega_\theta|^{-1} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_z - \bar{\boldsymbol{\theta}})' \Omega_\theta^{-1} (\boldsymbol{\theta}_z - \bar{\boldsymbol{\theta}})\right\}.$$

In addition to these random-effect parameters, we need to estimate the *fixed-effect* parameters, $\mathbf{\Psi}$. In our analysis, $\mathbf{\Psi} = \{\boldsymbol{\beta}_{ss'}, \eta_s, \boldsymbol{\tau}_s, \alpha_s, \boldsymbol{\gamma}_s\}$, where s denotes a specific state. We assume that:

$$\mathbf{\Psi} \sim MVN(\boldsymbol{\mu}_\Psi, \Sigma_\Psi) \Rightarrow P(\mathbf{\Psi}) \propto |\Sigma_\Psi|^{-1} \exp\left\{-\frac{1}{2}(\mathbf{\Psi} - \boldsymbol{\mu}_\Psi) \Sigma_\Psi^{-1} (\mathbf{\Psi} - \boldsymbol{\mu}_\Psi)\right\}.$$

Likelihood

The complete likelihood is given by

$$L(\text{data}, \{\boldsymbol{\xi}_i'\}, \{\boldsymbol{\theta}_z\}, \{\Sigma_z\}, \boldsymbol{\Psi}, \bar{\boldsymbol{\theta}}, \Omega_\theta, \nu, \Delta) = P(\text{data}|\{\boldsymbol{\xi}_i'\}, \{\boldsymbol{\theta}_z\}, \boldsymbol{\Psi})P(\{\boldsymbol{\xi}_i'\}|\{\Sigma_z\})P(\{\boldsymbol{\theta}_z\}|\bar{\boldsymbol{\theta}}, \Omega_\theta)P(\Sigma_\xi)$$
$$\times P(\boldsymbol{\Psi})P(\bar{\boldsymbol{\theta}})P(\Omega_\theta)$$

Posterior Distribution

The MCMC procedure recursively generates draws from the posterior distributions that are given by:

1. $P(\boldsymbol{\xi}_i|\boldsymbol{\theta}_z, \Sigma_\xi, \boldsymbol{\Psi}, \bar{\boldsymbol{\theta}}, \Sigma_\theta, \text{data}_i) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\xi}_i'\Sigma_\xi^{-1}\boldsymbol{\xi}_i\right\} P(\text{data}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}_z, \boldsymbol{\Psi})$,
2. $P(\boldsymbol{\theta}_z|\{\boldsymbol{\Psi}_i\}, \Sigma_\xi, \boldsymbol{\Psi}, \bar{\boldsymbol{\theta}}, \Sigma_\theta, \text{data}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_z - \bar{\boldsymbol{\theta}})'\Sigma_\theta^{-1}(\boldsymbol{\theta}_z - \bar{\boldsymbol{\theta}})\right\} P(\text{data}|\{\boldsymbol{\xi}_i\}, \boldsymbol{\theta}_z, \boldsymbol{\Psi})$,
3. $P(\boldsymbol{\Phi}|\{\boldsymbol{\xi}_i\}, \{\boldsymbol{\theta}_z\}, \Sigma_\xi, \boldsymbol{\Psi}, \bar{\boldsymbol{\theta}}, \Sigma_\theta, \text{data}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\Psi} - \boldsymbol{\mu}_\psi)'\Sigma_\Psi^{-1}(\boldsymbol{\Psi} - \boldsymbol{\mu}_\psi)\right\} P(\text{data}|\{\boldsymbol{\xi}_i\}, \{\boldsymbol{\theta}_z\}, \boldsymbol{\Psi})$,
4. $\bar{\boldsymbol{\theta}} \sim MVN(\boldsymbol{\mu}_n, V_n)$,
5. $\Omega_\theta^{-1} \sim Wishart(v_n, S_n)$,
6. $\Sigma_\xi^{-1} \sim Wishart(\nu_v, \Delta_n)$,

where

$V_n^{-1} = \left[V_0^{-1} + \Omega_\theta^{-1}\right]$,

$\boldsymbol{\mu}_n = V_n\left[\boldsymbol{\mu}_0 V_0^{-1} + \left(\sum_z \boldsymbol{\theta}_z\right)\Omega_\theta^{-1}\right]$,

$v_n = v_0 + N$,

$S_n^{-1} = \sum_z(\boldsymbol{\theta}_z - \bar{\boldsymbol{\mu}})(\boldsymbol{\theta}_z - \bar{\boldsymbol{\theta}})' + S_0^{-1}$,

$\nu_n = \nu_0 + N$, and

$\Delta_n^{-1} = \sum_z \sum_i \boldsymbol{\xi}_i\boldsymbol{\xi}_i' + \Delta_0^{-1}$.

Estimation Algorithm

Since the posterior distributions presented earlier do not have close-form analytical solutions, we use the Metropolis-Hasting (MH) algorithm to sample from the posterior distribution. To illustrate the MH algorithm, consider the parameter $\mathbf{\Theta}$. The MH algorithm proceeds as follows: Let $\mathbf{\Theta}^{(k)}$ define the k^{th} accepted draw for parameter $\mathbf{\Theta}$. The next sample $\mathbf{\Theta}^{(k+1)}$ is choosen such that

$$\mathbf{\Theta}^{(k+1)} = \mathbf{\Theta}^{(k)} + \tilde{\mathbf{\Theta}},$$

where $\tilde{\mathbf{\Theta}} \sim MVN(0, \sigma^2\Gamma)$, and σ and Γ are chosen to reduce the autocorrelation between the MCMC draws following the approach outlined in Netzer et al. (2008). The probability of accepting the $k+1^{\text{th}}$ draw is given by the ratio of the posterior probability of $k+1^{\text{th}}$ draw to the posterior probability of the k^{th} draw:

$$\Pr(\text{acceptance}) = \min\left\{\frac{P(\mathbf{\Theta}_d^{(k+1)}|\mathbf{\Theta}_{-d}^{(k)}, \mathbf{\Psi}^{(k)}, \text{data})}{P(\mathbf{\Theta}_d^{(k)}|\mathbf{\Theta}_{-d}^{(k)}, \mathbf{\Psi}^{(k)}, \text{data})}, 1\right\}.$$

A similar approach is applied to draw samples of $\mathbf{\Psi}$.