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

In this paper, we address the problem of advertising 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 dataset 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 her from a disengaged state to an state in which she 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 are not useful, our results show that display ads have a significant effect on the 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 spend. 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"), her 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. 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 she can use this information to optimally allocate the advertising budget (Abhishek and Hosanagar, 2013). 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.

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. Some heuristics have been proposed to address the problems associated with LTA, e.g. first-touch attribution or exponentially weighted attribution, but

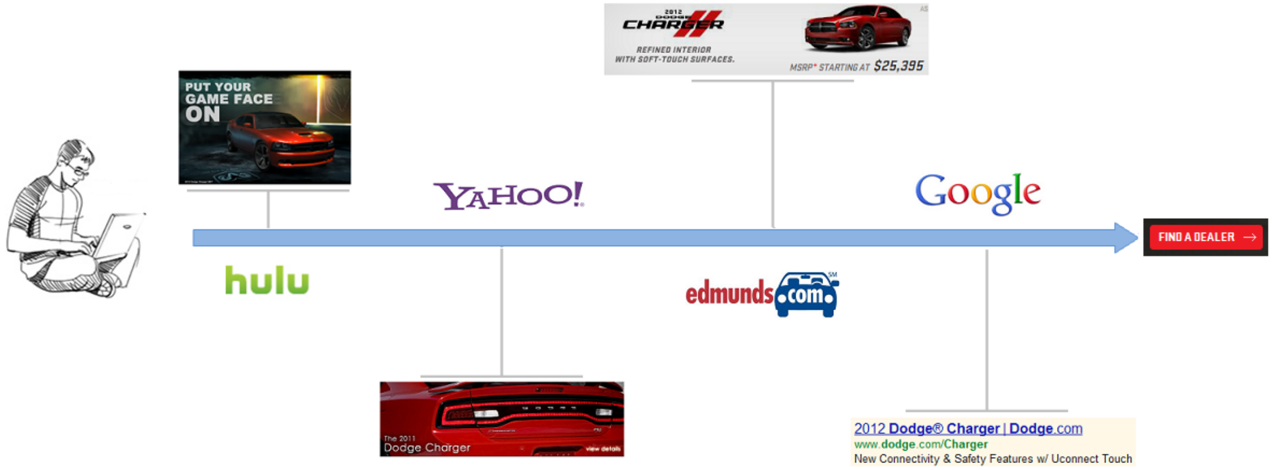

Figure 1: Multiple ad exposures across different online channels.

these techniques are plagued with similar problems and do not take a data-driven approach to the issue of attribution. 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).

Given the importance of this problem, advertising attribution has become an active area of research and several papers have tried to address the issue of attribution (Shao and Li, 2011, Dalessandro et al., 2012, Wiesel et al., 2011, Li and Kannan, 2014, Andrel et al., 2013, Tucker, 2013, Xu et al., 2014). Shao and Li (2011) have developed a bagged logistic regression model to predict how ads from different channels lead to conversions. Dalessandro et al. (2012) extend this research by incorporating the sequence of ads that lead a consumer to her 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. Xu et al. (2014) present a model using mutually exciting point process, which is used to estimate the effect of online ads. 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. In a recent paper, Hu et al. (2014) use data from Google trends to show that online advertising increases consumers interest in pre-purchase information gathering *and* sales. Using a difference-in-difference approach Ghose and Todri (2015) show that display ads can increase users propensity to search for the focal brand and product. They also show that display ads might increase the probability of a purchase. Although most of attribution research is empirically motivated, Jordan et al. (2011) and Berman (2013) propose game-theoretic approaches to devise payment rules for multi-channel ads. In this paper, we try to extend this nascent stream of research by incorporating theories from the rich marketing literature on consumer search and deliberation (Kotler and Armstrong, 2011).

Our paper differs from the prior work in two distinct ways. Firstly, we propose a novel dynamic model where consumers move through different stages on the path through purchase and this movement can be affected by ads. Secondly, our approach can identify conversions not attributable to online ads, a departure from existing literature that credits all the online conversions to online ads. The main motivation behind our research is to solve an advertiser's problem of finding out the effectiveness of different types of advertising channels and determine how much should he pay for every impression and click. To address this issue, we present a model of individual consumer behavior based on the concept of a conversion funnel that captures her deliberation process. The conversion funnel models a consumer's search and purchase process and is commonly used by marketers (Kotler and Armstrong, 2011). 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). Accordingly, a consumer moves in a staged manner from the disengaged state to the conversion state in our model, and ads affect the consumer's movement through these different stages. The consumer path to purchase is modeled using a dynamic Hidden Markov Model (HMM), which is subsequently used to solve the attribution problem. HMMs are used to capture dynamic consumer behavior when the consumer's state is unobservable (Netzer et al., 2008, Schweidel et al., 2011, Schwartz et al., 2011, Singh et al., 2011). In the extant literature, 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). Singh et al. (2011) use an HMM to understand developer learning dynamics in open source software projects. 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 where time varying covariates like ad-stock affect consumer behavior.

This model is estimated using a large 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 her from a state of disengagement to a state in which she is aware of the product, and in which the product enters her consideration set. However, when the consumer actively interacts with these ads (e.g. by clicking on them), her likelihood to convert increases considerably. In addition, 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 and latent consumer states. Secondly, the consumer model is used to support a new attribution technique that can be used by advertisers to improve upon existing techniques. Finally, from a managerial standpoint, our study informs marketers about how different ad formats influence consumers differently based on the stages of their deliberation and opens up the possibility of targeting consumers based on their (inferred) latent states. Such an approach might improve ad targeting for advertisers and allow them to target beyond standard demographic and behavioral measures.

2 A 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.

2.1 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, Bruce et al., 2012).

Several variants of the conversion funnel have been proposed, but the most commonly used funnel has the following stages - *awareness, consideration* and *purchase* (Bruce et al., 2012, Jansen and Schuster, 2011, Mulpuru, 2011, Court et al., 2009). A consumer is initially in a disengaged state when she is unaware of the product or is not deliberating a purchase. When she is exposed to an ad, she might move into a state of awareness. Subsequently, if she is interested in the product, she transitions to a consideration stage where she might engage 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 her 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. Recent work by Hu et al. (2014) and Bruce et al. (2012) models the conversion funnel using aggregate data. 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.

2.2 Hidden Markov Model

Here we propose 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 2.3. 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 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 as 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 her progression through the conversion funnel is hidden. In this paper, we use the HMM to capture the user's deliberation process and her movement down the conversion funnel as a result of the different ad exposures she 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 $|S|$ states (Figure 2) where the first state is a disengaged state. At any time t, consumer i can be in one of the $|S|$ states, $S_{it} \in S$.[1] As mentioned earlier, we do not observe s_{it}, but we observe the bivariate outcome variable $Y_{it} = (N_{it},\ C_{it})$ which arises from a stochastic process conditional on the state S_{it}. N_{it} is a Poisson random variable that denotes the number of pages viewed by the consumer between time t and $t+1$, and C_{it} is a binary random variable that captures whether there was a conversion between time t and $t+1$. When the user is in a disengaged state, she 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 her. As the consumer is exposed to

[1] 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.

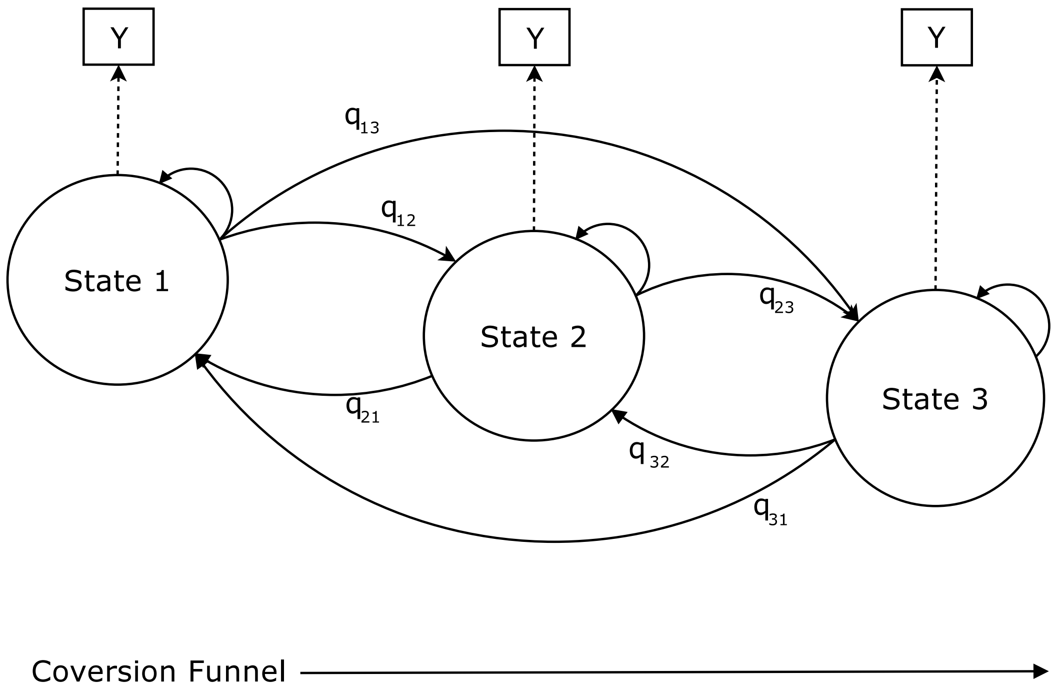

Figure 2: A representative HMM with 3 latent states. $q_{ss'}$ denote the transition probabilities from state s to state s' and Y_s is a random variable that captures the observations in state s.

different ads, she might move into a more engaged state where she has interacted with the ads or knows about the product and might be willing to purchase it. On further deliberation, she moves into an even higher state of engagement where she 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. Since we model the very $first$ conversion event, the HMM ends once a consumer has engaged in a conversion activity.

We assume that a consumer's propensity to purchase (or convert) steadily increases as she moves down the funnel. We also assume that the consumer's research behavior becomes more intense as she moves down the funnel. A transition between the states takes place in a stochastic manner at the end of the day and might be 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.

Let $A_i = \{A_{i1}, A_{i2}, \ldots, \}$ denote a sequence of ad events that consumer i is exposed to in T time periods, due to which she can ends up in states $S_i = \{S_{i1}, S_{i2}, \ldots, S_{iT}\}$. These ads can belong to any of the K distinct types of ads. X_{ikt} captures the total numbers of ads of type k during a certain period, e.g. number of display impressions on a particular day. The advertising activity leads to a change in the advertising stock based on the NerlovArrow model (Nerlove and Arrow, 1962) in the following manner,

$$O_{ikt} = \rho_k O_{i,k,t-1} + X_{ikt} + \zeta_{ikt}, \tag{1}$$

where O_{ikt} represents the advertising stock ads of type k for consumer i at time t and ζ_{ikt} represents the idiosyncratic i.i.d. shock. The NerlovArrow model has been commonly used in the marketing literature to study the effect of advertising on consumer outcomes (Zantedeschi et al., 2015, Xu et al., 2014, Rutz and Bucklin, 2011, Naik et al., 1998, Nerlove and Arrow, 1962). In this model, the ad-stock decays exponentially over time and the parameter $0 < \rho_k < 1$ controls the rate of decay. When ρ_k is large, the ad-stock decays slowly, but for smaller ρ_k it decays faster. Note that the rate at which ad-stock decays is heterogeneous across ad types. Ad-stocks of different types move the consumers through the different stages in the HMM. We can compactly represent the evolution of the ad-stock using the following expression (Zantedeschi et al., 2015),

$$\boldsymbol{O}_{it} = \boldsymbol{\rho} \boldsymbol{O}_{i,t-1} + \boldsymbol{X}_{it} + \boldsymbol{\zeta}_{it}. \tag{2}$$

Xu et al. (2014) adopt a similar approach and assume that the effect of an ad decreases exponentially over time. 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.

One of the biggest challenges in online advertising is that we do not know a consumer's underlying state at any time. Hence, S_{it} is always unobserved but the researcher can observe the consumer's activity, which is captured in 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.

2.2.1 Markov Chain Transition Matrix

We use a discrete time Markov chain in our model, where transitions occur at the end of every day. 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 ad-stock, $\boldsymbol{O}'_{it}$. 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 $P(S_{it} = s'|S_{it-1} = s) = q_{itss'}$. The elements of the transitions matrix specific to state s are given by

$$q_{itss'} = \frac{\exp\{\mu_{iss'} + \boldsymbol{O}'_{it}\boldsymbol{\beta}_{ss'}\}}{1 + \sum_{s' \neq s} \exp\{\mu_{iss'} + \boldsymbol{O}'_{it}\boldsymbol{\beta}_{ss'}\}} \quad \forall\, s' \neq s, \tag{3}$$

$$q_{itss} = \frac{1}{1 + \sum_{s' \neq s} \exp\{\mu_{iss'} + \boldsymbol{O}'_{it}\boldsymbol{\beta}_{ss'}\}}, \tag{4}$$

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 ad-stock $\boldsymbol{O}'_{it}$ might have different effects on the transition based on the receiving state. For e.g., display clicks might affect the transition to a less engaged state differently than they affect the transition to a more engaged state. In this model, the passage of time might decrease the ad-stock over time, which is captured in the evolution of the ad-stock as shown in Equation (2).

In addition to the heterogeneous effect of ads across 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 down the funnel 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 she saw online. Our approach addresses this problem by allowing for the intercept terms in the transition matrix, $\boldsymbol{\mu}_i$, to vary across consumers, which captures differences in their responses to online ads. We divide the customer heterogeneity into two distinct components as follows,

$$\boldsymbol{\mu}_i = \boldsymbol{\theta}_z + \boldsymbol{\xi}_i, \tag{5}$$

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.[2] $\boldsymbol{\xi}_i \sim MVN(\Sigma_\xi)$ captures individual specific idiosyncrasies due to factors such as brand awareness or loyalty, affinity for cars, etc. 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.[3]

2.2.2 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:

The number of page views on the advertiser's website, N_{it}, is drawn from a Poisson distribution with a rate parameter λ_{its}, which is a function of the current state s.[4] The probability of observing n_{it} page

[2] Although we try to control for factors like offline advertising effects using a DMA-specific term, $\boldsymbol{\theta}_z$, our model ignores the temporal dynamics associated with this factor. Such dynamics can be accounted for using a time varying DMA-specific intercept term, but the model identification would then require additional data like offline advertising spend that is currently unavailable to us. Due to the lack of appropriate data we leave this extension as an avenue for future research.

[3] 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.

[4] We remove page views that occur directly after an ad click through

views is given by

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

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

$$\begin{aligned} \widetilde{\eta}_1 &= \exp\{\eta_1\}, \\ \widetilde{\eta}_2 &= \widetilde{\eta}_1 + \exp\{\eta_2\}, \\ &\ldots\ldots, \\ \widetilde{\eta}_S &= \widetilde{\eta}_{S-1} + \exp\{\eta_S\}, \end{aligned} \tag{6}$$

where $\eta_1, \ldots, \eta_S$ 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 a binary random variable. The conditional probability $P(C_{it} = 1|S_{it} = s) = m_{its}$ is given by

$$m_{its} = \frac{\exp\{\tilde{\alpha}_s + \upsilon_z + \boldsymbol{z}_{it}'\gamma_s\}}{1 + \exp\{\tilde{\alpha}_s + \upsilon_z + \boldsymbol{z}_{it}'\gamma_s\}}. \tag{7}$$

Here 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 stock and 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 she 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,

$$\begin{aligned}
\tilde{\alpha}_1 &= \alpha_1, \\
\tilde{\alpha}_2 &= \tilde{\alpha}_1 + \exp\{\alpha_2\}, \\
&\ldots\ldots, \\
\tilde{\alpha}_S &= \tilde{\alpha}_{S-1} + \exp\{\alpha_S\},
\end{aligned} \tag{8}$$

where $\alpha_1, \ldots, \alpha_S$ are the parameters to be estimated from the data. This structure enforces that $m_{itS} \geq \ldots \geq m_{it1}$, *ceteris paribus*. This assumption ensures the identification of the different states and is consistent with the approach adopted by Ascarza and Hardie (2013), Montoya et al. (2010) and Netzer et al. (2008).

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

$$\begin{pmatrix} \vartheta_z \\ \upsilon_z \end{pmatrix} \sim MVN\left(0, \Omega_{\vartheta\upsilon}\right). \tag{9}$$

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

$$P(Y_{it} = \boldsymbol{y}_{it}|S_{it} = s) = \int m_{its}^{c_{it}}(1 - m_{its})^{(1-c_{it})} P(N_{its} = n_{it}|S_{it} = s, \vartheta_z, \upsilon_z) f_{iz}(\vartheta, \upsilon) d\vartheta d\upsilon, \tag{10}$$

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

2.2.3 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.

2.2.4 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 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 attribute suitable credit to an ad even if it does not contribute to a conversion immediately 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 increases the conversion probability. Our model also differs from recent work by Xu et al. (2014) 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.

2.3 Ad Attribution

When consumer i is exposed to an ordered set of ad related activities, $\boldsymbol{A}_i = \{A_{i1}, A_{i2}, \ldots A_{iJ}\}$, she moves through the different states of the HMM in the manner described above. Let A_{ij} denote a categorical random variable that captures the ad related activity the consumer is exposed to, i.e. $A_{ij} \in \{$"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, J, is random and varies across customers. In our model, the ad related event A_{ij} affects the customer i's underlying time varying parameters $\boldsymbol{O}'_{it}$ and $\boldsymbol{z}'_{it}$ in two ways – (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 the ad stock $\boldsymbol{O}'_{it}$. This approach is similar to the mutually exciting point process model proposed by Xu et al. (2014) and the three-level model of consideration, visits and purchases proposed by Li and Kannan (2014) that estimate the spillover across ad exposures to estimate the effect of an ad.

Attribution in the context of online advertising involves measuring the incremental change in revenues when a consumer is exposed to an ad. In the absence of data on revenue per conversion, we model the added value of an ad is the incremental change in the conversion probability due to the ad. However, our model can easily be generalized to incorporate different revenues upon conversion. According to Dalessandro et al. (2012) and Berman (2013), an effective attribution methodology should have the following properties: (i) *fairness* – the attribution methodology should reward different channels based on their incremental propensity to change the conversion probability, (ii) *interpretability* – it is easy to understand and justify the attribution outcomes, (iii) *efficiency* – it accounts for the total value created by the ad campaign, and (iv) it is *data driven*. Our proposed attribution scheme meets all these criteria as we explain below.

One common attribution scheme is the counter-factual methodology (Zantedeschi et al., 2015, Xu et al., 2014, Shao and Li, 2011), where the contribution of the ad A_{ij} is measured by the change in conversion rate when the ad is not shown. We adopt a similar approach and ascertain the value of ad A_{ij} when it is proceeded by ads $\boldsymbol{A}_{i,j-1}$ in the following manner.

$$\psi_{ij} = \mathbb{E}[C_i|\boldsymbol{A}_i] - \mathbb{E}[C_i|\boldsymbol{A}_i \backslash A_{ij}], \tag{11}$$

where $\boldsymbol{A}_i \backslash A_{ij}$ is the ordered sequence of all ads except ad A_{ij}. The probability that a consumer will eventually convert is derived by taking an expectation over all possible paths the consumer can take after ad A_{ij} is shown. The value of the ad unconditional of the subsequent ads is given by

$$\begin{aligned} V_{ij} &= \sum_{\boldsymbol{A}_i \supset \boldsymbol{A}_{ij}} \{\mathbb{E}[C_i|\boldsymbol{A}_i] - \mathbb{E}[C_i|\boldsymbol{A}_i \backslash A_{ij}]\} \, p(\boldsymbol{A}_i|\boldsymbol{A}_{ij}), \\ &= P(C_i = 1|\boldsymbol{A}_{ij}) - P(C_i = 1|\boldsymbol{A}_{i,j-1}). \end{aligned} \tag{12}$$

Hence, the effect of an ad depends on the consumer's underlying state, which in turn is affected by the ads that preceded A_{ij}. 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 prior to A_{ij}. 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_{ij}, would be estimated to be zero in this situation. Note that, in our formulation, V_{ij} can either be positive or negative. In cases where the ads lead to aversion as demonstrated by Goldfarb and Tucker (2011), $V_{ij} < 0$. This approach is similar to that of Shao and Li (2011) and Dalessandro et al. (2012), but $P(C_i = 1|A_{ij})$ is estimated using a dynamic HMM in our case, whereas they use simplistic approaches like logistic regression and sample means (that do not incorporate the time dynamics) to compute these probabilities. Finally, 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,

$$\Pi_k = \sum_i \sum_{j=1}^{J} \mathbb{1}_{\{a_{ij}=k\}} V_{ij}, \tag{13}$$

where $\mathbb{1}_{\{a_{ij}=k\}}$ is an indicator function that equals one if ad event a_{ij} 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. offline advertising. The overall effectiveness of the campaign is a combination of these factors. 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.

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. The advertiser ran a large-scale display campaign that was very broadly targeted at its customer demographics.[5] 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. Typically advertisers get summary statistics from different ad and search networks, which makes it infeasible to create an individual panel. Since the partner advertiser worked with exactly one ad network and search engine, we can create an individual-level panel. Our sample comprises a panel of 6432 randomly chosen users with a total of 146,165 observations. Of these, 1311 customers do not receive any ad activity and provide the baseline for our analysis. 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. When a consumer clicks on one of these

[5] Our discussions with the firm suggested that they did not engage in any behavioral targeting.

ads and arrives at the advertiser's website, this click is recorded in our data and referred to as a search click. We do not observe a search ad impressions in the data as the search engine does not report this information to the advertiser at an individual level. 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	10.952	30.985
Generic display clicks	0.057	0.161
Generic click-through rate	0.007	0.054
Specific display impressions	3.353	8.976
Specific display clicks	0.114	0.286
Specific click-through rate	0.020	0.062
Search clicks	0.196	0.642
Web pages viewed	3.770	8.185
Conversions	0.243	0.429

On average, there are 11 display impressions per customer on generic websites and 3.5 impressions on auto-specific websites. Consumers click 0.06 of these display ads on generic websites and 0.11 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 3.8 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; 24.3% of all the consumers in this dataset end up engaging in one of the four conversion activities mentioned earlier.

Although, we do not observe individual search ad impressions, our data are quite representative of the ideal data that a typical advertiser might have in the context of display and search advertising. It is extremely unlikely that search engines like Google would make the individual level search data available due to privacy limitations. Since one of the important goals of our research is applicability, we choose to work with this specific data structure, which has been used extensively in the prior literature (Li and Kannan, 2014, Xu et al., 2014).

4 Empirical Estimation

In this section, we illustrate how the HMM model can be estimated and interpreted. We first outline the estimation procedure, parameter identification and briefly discuss the model validity. The estimated parameters are presented in the subsequent section. A sample of 4932 users is used for estimating the model, and the remaining 1500 users are used for validation.

4.1 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 Singh et al. (2011), 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 $s_0 \rightarrow \ldots \rightarrow s_T$ and the conditional probability of $P(Y_{i1} = \boldsymbol{y}_{i1}, \ldots, Y_{iT} = \boldsymbol{y}_{iT}, |S_0 = s_0, \ldots, S_T = s_T)$, which is given by

$$L_i = \sum_{s_1=1}^{|S|} \sum_{s_2=1}^{|S|} \cdots \sum_{s_T=1}^{|S|} \left[\prod_{t=1}^{T} P(S_{it} = s_t | S_{it-1} = s_{t-1}) \prod_{t=1}^{T} P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t) \right], \quad (14)$$

where $P(Y_{it} = \boldsymbol{y}_{it} | S_{it} = s_t)$ can be computed as shown in Equation (10). 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 reduces 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 (10). The likelihood of the observed data (Equation (14)) can be simplified to

$$L_i = \boldsymbol{\pi}' \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}).\mathbf{1}, \tag{15}$$

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,

$$LL = \sum_i \log \left[\boldsymbol{\pi}' \Phi_{i0}(y_{i0}) Q_{i1} \Phi_{i1}(y_{i2}) Q_{i2} \ldots Q_{iT} \Phi_{iT}(y_{iT}).\mathbf{1} \right]. \tag{16}$$

The heterogeneity parameters $\boldsymbol{\Theta} = \{\bar{\boldsymbol{\theta}}, \Omega_\theta, \Sigma_\xi, \Omega_{\vartheta\upsilon}\}$ and the homogeneous HMM parameters $\boldsymbol{\Psi} = \{\beta, \tau, \gamma, \eta, \alpha, \rho, \varrho\}$ 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.

4.2 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 (16) is invariant to changes in the labels (indices). We address this issue by enforcing the identifiability constraints $\tilde{\alpha}_1 \leq \tilde{\alpha}_2 \leq \ldots \leq \tilde{\alpha}_S$ and $\tilde{\eta}_1 \leq \tilde{\eta}_2 \leq \ldots \leq \tilde{\eta}_S$, 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 (6) and (8) guarantee that states with a higher likelihood to convert are assigned higher indices.

Now we discuss how our model parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ are identified. First, we focus on *fixed-effect* parameters that are constant across consumers – $\boldsymbol{\Psi}$. 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 set by advertisers 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. Finally, 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.

Next, we consider the heterogeneity parameters, $\boldsymbol{\Theta}$. 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. We also assume that $\boldsymbol{X}_{it} \perp \boldsymbol{\theta}_z, \vartheta_z$ and υ_z.

One of the main challenges in advertising attribution is the fact that the advertising activity is not completely exogenous. Consumers who spend more time online are more likely to receive ad exposures. Advertisers might also be likely to target consumers who have shown interest in their brands. Hence, the advertising activity and the unobserved variable like customer interest (or intent) are correlated. A few researchers have tried to address this problem in the context of direct marketing by introducing exogenous variation in advertising (Zantedeschi et al., 2015, Sahni, 2015, Nair et al., 2014). Zantedeschi et al. (2015) randomize email campaigns at an individual level to measure the effect of email exposures on consumer spend. Nair et al. (2014) run a field experiment and randomly assign casino promotions to gaming customers. In the context of display and search ads, such an approach is infeasible as it is difficult to control the exogenous variation finely. Ad networks don't allow advertisers to target consumers individually but only more broadly via behavioral or demographic characteristics in display advertising or keywords in search advertising. Further, an advertiser's problem in practice is to be able to perform attribution in the presence of endogenous ad targeting. An attribution solution that assumes randomization is of little practical value to advertisers who have to take their ad performance data from ad campaigns and figure out how to assign credit to different ads. Incidentally, specific aspects of our model and the inherent randomness in the advertising process helps us resolve the identification issue as described in the following paragraph.

In the proposed model, we explicitly account for the consumer's intent through the latent states. These state represent an implicit summarization of the data observed by an advertiser. If this data is used by an advertiser to target consumers, a model that incorporates advertising activity as a function of the consumer state would not suffer from endogeneity introduced due to the advertiser's action (Zantedeschi et al., 2015, Nair et al., 2014). However, consumers might be exposed to certain types of ads because of their own actions, e.g. a consumer might see more search ads for an auto maker because she is searching for cars on Google. One might employ an instrumental variable approach to resolve this endogeneity issue. This methodology poses some challanges and for the sake of continuity, we disccus such an approach in Section 7.1 and the appendix.

In our setting, we exploit the high degree of randomness in the advertising activity. Using a model of ad exposure presented in the appendix, we can easily show that only 2-7% of the variation in the advertising activity can be explained by observable factors. Li and Kannan (2014) also report that correlations between various types of ad exposures and customer activities are quite low. This confirms our intuition that even if the advertiser intends to show an ad to a consumer, factors like competition, budget constraints and overall market trends make this an inherently noisy process at an individual level, and significantly aids our identification. Note that noisy process makes our context very different from Zantedeschi et al. (2015) and Nair et al. (2014). In their setting, the advertiser can target an individual with a great deal of certainty, making the exposure endogeneity an even bigger concern.

4.3 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. The validity of the model is tested 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 one to four and compare these models using (i) the log marginal density, (ii) validation log-likelihood, and (iii) the RMSE of conversion. The performance of the different HMMs along these three metrics is presented in Table 2. We observe that an HMM with 3 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. In the rest of the paper, we refer to the three stages in the purchase funnel as the "disengaged", "awareness" and "consideration" states consistent with prior literature (Bruce et al., 2012, Kotler and Armstrong, 2011).

Table 2: Determining the number of states

Number of States	Log Marginal Density	Validation log-likelihood	Conversion RMSE
1	-2153.2	-842.4	0.147
2	-2028.7	-742.9	0.115
3	-1712.9	-577.3	0.095
4	-1784.2	-612.8	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:

LTA: The most commonly used model in the online advertising industry.

Logit: The simplest parametric 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.

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

HMM No-Het: This model is identical to the HMM 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 to the **Full-Model** proposed in this paper. 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, while LTA has the worst performance.

Table 3: Predictive validity

	Log Marginal Density/ Log Likelihood	Validation log-likelihood	Conversion RMSE
LTA	–	–	0.318
Logit	-2139.5	-721.8	0.182
Latent-Class	-1942.2	-639.2	0.121
HMM No-Het	-1782.8	-605.8	0.113
Full-Model	-1712.9	-577.3	0.095

We also observe that although accounting for customer heterogeneity considerably improves the model fit and predictive performance, temporal dynamics is a primary driver of the performance of the HMM and accounting for it improves model performance. 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 Results

5.1 Evolution of the Ad Stock

In Section 3.2, we discussed how the advertising stock decays over time. Estimates of the decay parameters $\boldsymbol{\rho}$ for the different types of ad stocks, are presented in Table 4. These parameters are similar to the decay rates reported in prior literature (Rutz and Bucklin, 2011, Zantedeschi et al., 2015). Incidentally, the ad stock of search click decays faster than the ad stock for display clicks. This result is in agreement with the long-term branding effect of display advertising that has been discussed extensively in popular press (Adweek 2014). Search ads on the other hand have a short-term effect, perhaps due to the goal directed behavior of consumers when they are exposed to search ads. Note that the estimates presented here only reflect the rate at which these ad stocks wear out. The overall impact of the different types of ad stocks will be discussed subsequently.

Table 4: Ad stock decay rate

Ad activity	$\boldsymbol{\rho}$
Generic Impression	0.28 (0.15)
Specific Impression	**0.32** (0.09)
Generic Click	**0.57** (0.07)
Specific Click	**0.68** (0.21)
Search Click	**0.41** (0.13)

The estimates in bold are significant at a 95% level. The numbers in parentheses denote the standard deviation.

5.2 Transition Parameters

Estimates of the transition parameters are reported in Table 5. 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 awareness state. Contrary to the popular belief that display ads are ineffective (de Vries, 2012, Claburn, 2012), we see that display ads play a significant role in moving consumers from a disengaged state to an awareness 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 awareness state. Consumers are more likely to notice these car related ads when they are visiting auto-specific websites. This finding is consistent with Yi (1990), 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 web pages that are contextually more relevant.

Although display ads have an impact on moving consumers from the disengaged state to an awareness state, they do not have an impact on moving consumers further down the conversion funnel, i.e. from an awareness state to a consideration state ($\boldsymbol{\beta}_{ac}$). 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. The coefficient of generic impression is positive and significant (0.097), which suggests that if consumers are shown too many display ads on generic websites, their probability to transition back to the disengaged state increases considerably. 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 her search if he considers these ads to be too intrusive (Goldfarb and Tucker, 2011). The overall effect of the generic impression is a combination of the two effects and can be estimated using the attribution methodology presented earlier. 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, 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 awareness state, with search clicks having the largest effect. Once the consumer moves to the consideration state, there is a relatively low probability of her transitioning out of that state. This probability is further reduced when the consumer performs more searches and clicks on search ads. 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. The estimates suggest that consumers typically move down the conversion funnel

Table 5: Estimates of the transition parameters (β)

	β_{da}	β_{dc}	β_{ad}	β_{ac}	β_{cd}	β_{ca}
generic_imp	**0.032**	0.013	**0.097**	0.000	0.009	0.005
	(0.008)	(0.003)	(0.089)	(0.011)	(0.012)	(0.009)
specific_imp	**0.058**	0.002	0.008	0.045	-0.001	0.002
	(0.024)	(0.005)	(0.010)	(0.022)	(0.115)	(0.009)
generic_clk	**0.376**	0.020	0.109	**0.388**	-0.001	-0.092
	(0.048)	(0.063)	(0.150)	(0.059)	(0.038)	(0.302)
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	**0.750**	0.089	-0.029	**0.616**	0.048	**-0.027**
	(0.052)	(0.939)	(0.138)	(0.057)	(0.001)	(0.005)
Heterogeneity Parameters						
$\bar{\theta}_{ss'}$	**-3.467**	**-5.632**	**-2.918**	**-3.206**	**-5.405**	**-4.327**
	(0.067)	(1.073)	(0.778)	(0.384)	(0.471)	(0.722)
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)

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", $a =$"awareness" , $c =$ "consideration"$\}$.

in a sequential manner, e.g. from one state to another, and we do not observe abrupt jumps from the disengaged state to the consideration state.

5.3 Consumer Response Parameters

Now we discuss the underlying parameters that affect the observations of the HMM. The behavior of consumers in the disengaged, awareness and consideration states differ considerably when it comes to their browsing behavior. Consumers in the disengaged state are extremely unlikely to view any pages at

Table 6: Estimate of factors affecting the page views (λ)

	τ_1	τ_2	τ_3
η	**0.015**	**-2.408**	**-0.840**
	(0.004)	(0.745)	(0.055)
$\widetilde{\eta}$	0.015	0.101	0.533

the manufacturer's website. Consumers in the awareness state on average view 0.101 pages/day on the car manufacturer's website, while those in the consideration state view five times as many pages. Since the consumers in all these states behave so differently, we are certain that the model is both empirically and behaviorally identified.

Next we consider factors that influence the consumers' conversion probability. The estimated coefficients of these factors are presented in Table 7. We notice that the conversion probability is higher in the consideration state than it is in the awareness state, which is higher than the conversion probability 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 awareness state. This result is consistent with the common finding that generic display ads do not lead to conversions. We also observe that conditional on being in the consideration state, impressions of any kind do not have an incremental impact on the likelihood to convert. However, as we mention earlier, even though generic display impressions might not lead to conversions directly, they might move consumers down the conversion funnel. Interestingly, the effect of a specific click in the awareness 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 consideration state than in the awareness state. This decrease might be attributed to the diminishing returns from further interactions with the consumers. Once consumers are sufficiently

Table 7: Estimates of conversion parameters

	γ_1	γ_2	γ_3
α	**-7.269**	**-0.172**	**0.095**
	(0.911)	(0.004)	(0.021)
$\widetilde{\alpha}$	-7.269	-6.427	-5.327
generic_imp	0.002	0.015	0.008
	(0.005)	(0.010)	(0.019)
specific_imp	0.000	**0.041**	0.020
	(0.002)	(0.009)	(0.019)
generic_clk	**0.033**	**0.289**	**0.318**
	(0.004)	(0.084)	(0.095)
specific_clk	0.055	**0.607**	**0.503**
	(0.024)	(0.090)	(0.112)
search_clk	**0.002**	**0.446**	**0.888**
	(0.000)	(0.027)	(0.147)
nw_activity	**0.009**	**0.091**	**0.067**
	(0.004)	0.005	0.007

primed to convert, increased interactions have only a marginal effect on them.

To illustrate the effect of different factors on consumers in the awareness and the consideration states, we show how each activity affects the conversion probability (in the same time period) in Table 8. We ignore the disengaged state in this analysis, as the conversion probabilities are too low to warrant a meaningful discussion. From Table 8 we can observe that consumers are more likely to convert in the consideration state than in the awareness state. Even though the higher likelihood to convert is imposed by the identification constraints in Equation (8), the base conversion rate in the state state is thrice the conversion rate in the awareness state, which illustrates the distinct behavioral difference between 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

Table 8: Conversion rate (in %) as a result of various factors.

	awareness	consideration
Baseline	0.16%	0.49%
Generic Imp	0.16%	0.49%
Specific Imp	0.16%	0.51%
Generic Click	0.22%	0.67%
Specific Click	0.30%	0.80%
Search Click	0.25%	1.18%
Network Activity	0.18%	0.52%

depends on the latent state; e.g., the effect of a specific click is more pronounced in the awareness state than the consideration state. Similarly, a search click is more significant in the consideration state than in the awareness state. In general, as consumers interact more with the advertiser (through clicks and page views), there is a substantial increase in the conversion probability. Note that the conversion probabilities shown here are atypical of online campaigns, which usually have very few conversions following a click.

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_{ij} presented in Equation (12). Instead, we use simulations to compute the value of an ad to solve the attribution problem. To perform the attribution for a particular consumer, first, 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 $p(\Sigma_\xi|\nu, \Delta, \text{data})$. Then, 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 (12).

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 13. The column labeled "%Δ" captures the relative overestimation of an ad's effectiveness by LTA as compared to the effectiveness computed by the Full Model.

Table 9: A comparison of attribution methodologies

Ad activity	# Ads	LTA	Logit	No-HetHMM	Full-Model	%Δ
Generic Imp.	70,444	171	152.2	124.9	45.5 (8.7)	275.8
Specific Imp.	21,564	78	96.5	116.2	62.4 (10.1)	25.0
Generic Clk.	369	54	84.6	75.1	48.1 (14.5)	10.1
Specific Clk.	732	150	140.7	167.6	107.2 (20.0)	40.0
Search Clk.	1,260	328	310.9	294.3	153.4 (9.5)	113.8
Total		781	784.9	778.1	415.0 (67.1)	88.2

%Δ indicates 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.

We first compare the No-Het HMM with LTA and the logit models, and subsequently all these models are compared to the Full-Model. Focusing on the No-Het HMM helps us understand the advantages of the multi-stage model proposed in this paper whereas the comparison with the Full-Model helps bolster our understanding with the incorporation of unobserved heterogeneity. From Table 13, 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 7. 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 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 proposed 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 13 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 the 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 88%. Other methodologies, including No-Het HMM, also 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 10: Comparison of different advertising channels

Ad activity	Conversions	% Contribution
Generic Display Ads	93.6	13.0
Specific Display Ads	169.6	23.6
Search Ads	153.4	21.4
Others	301.4	42.0

This table presents the 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 10. Generic display ads are responsible for 13% of the conversions and specific display ads are responsible for 23.6%, while search ads lead to 21.4% of the conversions. Interestingly, our methodology credits other sources for 42% 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 (415 out of 781), but we believe that this result is somewhat 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. Online campaigns 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 using Latent States

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.[6] The probability distribution over a consumer's state at time t is given by

$$P(S_t = s|\mu_i, Y_1, \ldots, Y_t) = \frac{\boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it,s}P(Y_{it} = y_t|S_{it=s})}{\boldsymbol{\pi}'\Phi_{i0}(y_{i0})Q_{i1}\Phi_{i1}(y_{i2})Q_{i2}\ldots Q_{it}\Phi_{it}(y_{it}).\mathbf{1}}, \tag{17}$$

[6]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 her inferred state of deliberation.

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 all consumers at time t. For the sake of illustration, we consider this distribution 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 70% of the consumer start out in the disengaged state, 20% in the awareness state and 10% in the consideration 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 24% of the consumers have converted, 18% of them are in the awarness state and 37% of them are in the consideration state. The firm can optimally allocate the ad dollars to increase his returns from the campaign. Figure 3 also demonstrates that consumers move 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 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.

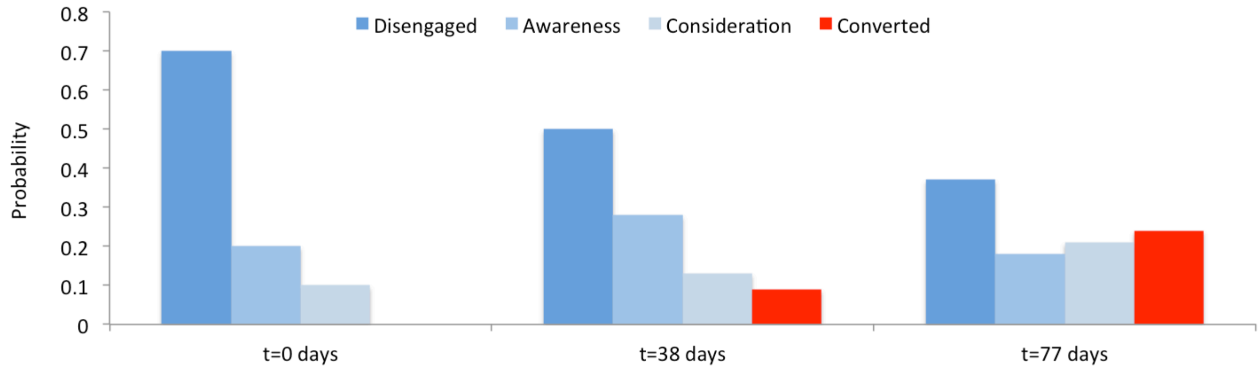

Figure 3: Distribution of consumers over time.

7 Discussion

7.1 Endogeneity of Ad Exposures

An important concern in the attribution literature is addressing the endogeneity of ad exposures. The main reason for this concern is the notion that the advertising activity might be correlated with consumer interest. For example, consider the utility of a consumer (in state s) which is given by the following equation (unrolling Equation (7))

$$U_{it} = \beta_0 + \beta X_{it} + \gamma I_{it} + \epsilon_{it}, \tag{18}$$

where X_{it} is the advertising activity and I_{it} is the customer's level of interest. If the advertising activity is correlated with the consumer's interest, i.e. $Cov(X_{it}, I_{it}) = \rho$, and we do not include I_{it} in the regression, then the estimates of β will be biased (omitted variable bias). However, in Section 5.2 we propose that our modeling choices and the inherent variation in search and display advertising might help in the empirical identification. Here, we provide further justification to bolster these claims.

Firstly, let $I_{it} = I^S_{it} + \chi_{it}$, where $I^S_{it} = \sum_s \lambda_s * \mathbf{1}_{S_{it}=s}$ and λ_s captures the mean consumer interest in state s. In our model, we assume that the advertising activity is only correlated with I^S_{it} and $Cov(X_{it}, \chi_{it}) = 0$. We believe this to be a reasonable assumption because the state summarizes all the information that the advertiser observes and can use to target a customer. If we now estimate the following model,

$$U_{it} = \beta_0 + \beta X_{it} + \gamma I^S_{it} + \epsilon_{it}, \tag{19}$$

estimate of β would be unbiased. Secondly, we adopt an instrumental variable (IV) approach where the level of ad exposures is dependent on the underlying state and supply side factors. More specifically, we use the following IVs – (i) the average number of ads seen by consumers at time t in other DMAs, and (ii) the fraction of total daily ads shown before time t. Since our attribution results don't change significantly, we do not incorporate IVs in our main analysis. Furthermore, finding valid instruments in online advertising has been a challenging problem (Abhishek et al., 2015, Rutz et al., 2012) and invalid instruments might lead to more problems than they address (Rossi, 2014). Given these issues, we rely on the exogenous variation in the advertising activity in all states to help in proper identification. Thirdly, we do not observe significant differences in the advertising activities across different states, which points to the fact that the advertising is not strongly correlated with consumers' interest. In addition, estimates from the IV model show that even when the ad exposures are explicitly modeled as a function of the underlying state, we do not see significant differences in advertising activity across states (as shown in Table 11).

Table 11: Advertising exposures

Ad activity	Disengaged	Awareness	Consideration
Generic Impression	0.130 (0.669)	0.136 (0.703)	0.125 (0.681)
Generic Clicks	0.003 (0.008)	0.003 (0.009)	0.001 (0.008)
Specific Impression	0.048 (0.193)	0.070 (0.113)	0.036 (0.049)
Specific Clicks	0.006 (0.048)	0.006 (0.067)	0.007 (0.059)
Search Clicks	0.004 (0.009)	0.004 (0.011)	0.006 (0.011)

8 Conclusion

In this paper, we present a model that analyzes how consumers behave when they are exposed to advertising from multiple online channels. 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 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 she is exposed to an ad, but it also considers the long-term future impact that the ad might have on her 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 academic papers in this area have proposed several innovative ways of estimating the causal impact of ads, but the consumer models are often simplistic. This 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.

Acknowledgements

We thank Wharton Customer Analytics Initiative for the data grant and Organic for making this data available to us. We would also like to thank Yong Tang, Oded Netzer, Elea Feit, Jun Li, Jose Guajardo and seminar participants at University of Washington, University of California – Davis and Purdue University for many an insightful comments that have helped improve this paper. This research is partially supported by the Berkman faculty award and the Mack Center grant for technological innovation.

References

V. Abhishek and K. Hosanagar. Optimal bidding in multi-item multislot sponsored search auctions. *Operations Research*, 61(4):855–873, 2013.

V. Abhishek, K. Hosanagar, and P. S. Fader. Aggregation bias in sponsored search data: The curse and the cure. *Marketing Science*, 34(1):59–77, 2015.

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.

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.

N. I. Bruce, K. Peters, and P. A. Naik. Discovering how advertising grows sales and builds brands. *Journal of Marketing Research*, 49(6):793–806, 2012.

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. *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, pages 7:1–7:9, 2012.

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

A. Ghose and V. Todri. Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly*, Forthcoming, 2015.

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

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

Y. Hu, R. Y. Du, and S. Damangir. Decomposing the impact of advertising: Augmenting sales with online search data. *Journal of Marketing Research*, 51(3):300–319, 2014.

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. 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.

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

H. Nair, S. Misra, W. J. H. IV, R. Mishra, and A. Acharya. Big data and marketing analytics in gaming: Combining empirical models and field experimentation. *Stanford University Graduate School of Business Research Paper*, 2014.

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

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.

P. E. Rossi. Even the rich can make themselves poor: A critical examination of iv methods in marketing applications. *Marketing Science*, 33(5):655–672, 2014.

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

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

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

N. S. Sahni. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Stanford University Graduate School of Business Research Paper*, 2015.

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.

P. V. Singh, Y. Tan, and N. Youn. A hidden markov model of developer learning dynamics in open source software projects. *Information Systems Research*, 22(4):790–807, 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. 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. Whinston. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, 60(6):1392–1412, 2014.

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

D. Zantedeschi, E. M. Feit, and E. T. Bradlow. Measuring multichannel advertising response. *Working paper, The Wharton School*, 2015.

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 $\mathbf{\Psi}$.

Endogeneity Correction

In order to address the endogeneity concerns, we explicitely model the advertising activity. Conditional on the state, the amount of advertising activity $\boldsymbol{X}_{it}$ is log-normally distributed according to the following equation,

$$\log(X_{it|S_{it}=s}) = \boldsymbol{\beta}_s^a + \boldsymbol{IV}_{it}'\boldsymbol{\beta}_s^{IV}, \tag{20}$$

and IV_{it} are the instrumental variables (IVs). As pointed out earlier, we want to disentagle the effect that demand side factors like consumer interest have on the advertising exposures. To achieve this objective, we choose IVs that capture supply side factors, more specifically – (i) the average number of ads seen by consumers at time t in other DMAs, and (ii) the fraction of total daily ads shown before time t. The first IV captures the advertiser's overall advertising intensity, whereas the second IV capture's the advertiser's ability to show ads. The advertiser's daily ad budget is fixed and the day progresses and the budget runs out, his ads are less likely to show up. Since these IVs affect the advertising activity directly but are not related directly to an individual's response function in the aforementioned HMM, we believe they are valid instruments.

The advertising activity as a function of the consumers's underlying state and other factors is presented in Table 12. We observe that the advertising activity is not significantly different across the different states. We also observe that the attribution based on the model with IVs is very similar to the attribution based on the Full model.

Table 12: Advertising exposures

log(ad activity)	$\boldsymbol{\beta_0^a}$	$\boldsymbol{\beta_a^a}$	$\boldsymbol{\beta_e^a}$	$\boldsymbol{\beta_{IV,1}^a}$	$\boldsymbol{\beta_{IV,2}^a}$
Generic Impression	**-0.887** (0.251)	0.019 (0.167)	-0.015 (0.080)	0.024 (0.008)	-0.188 (0.461)
Specific Impression	**-1.323** (0.224)	0.116 (0.241)	-0.126 (1.307)	–	–
Generic Click	**-3.091** (1.479)	0.093 (0.378)	0.119 (0.112)	–	–
Specific Click	**-2.788** (1.288)	-0.053 (0.290)	0.054 (0.029)	–	–
Search Click	-2.554 (1.833)	0.220 (0.177)	**0.363** (0.067)		

The estimates in bold are significant at a 95% level. The numbers in parentheses denote the standard deviation.

Table 13: A comparison of attribution methodologies

Ad activity	Full Model	Full-Model IV
Generic Imp.	45.5 (8.7)	40.7 (9.5)
Specific Imp.	62.4 (10.1)	71.2 (12.8)
Generic Clk.	48.1 (14.5)	50.9 (11.7)
Specific Clk.	107.2 (20.0)	94.2 (16.5)
Search Clk.	153.4 (9.5)	134.7 (15.1)
Total	415.0 (67.1)	372.3 (48.5)