2016

Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data

Daniel Zantedeschi

Eleanor M. Feit

Eric T. Bradlow
University of Pennsylvania

Recommended Citation

Zantedeschi, D., Feit, E. M., & Bradlow, E. T. (2016). Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data. *Management Science, 63* (8), 2706-2728. http://dx.doi.org/10.1287/mnsc.2016.2451

Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data

Abstract

Advances in data collection have made it increasingly easy to collect information on advertising exposures. However, translating this seemingly rich data into measures of advertising response has proven difficult, largely because of concerns that advertisers target customers with a higher propensity to buy or increase advertising during periods of peak demand. We show how this problem can be addressed by studying a setting where a firm randomly held out customers from each campaign, creating a sequence of randomized field experiments that mitigates (many) potential endogeneity problems. Exploratory analysis of individual holdout experiments shows positive effects for both email and catalog; however, the estimated effect for any individual campaign is imprecise, because of the small size of the holdout. To pool data across campaigns, we develop a hierarchical Bayesian model for advertising response that allows us to account for individual differences in purchase propensity and marketing response. Building on the traditional ad-stock framework, we are able to estimate separate decay rates for each advertising medium, allowing us to predict channel-specific short- and long-term effects of advertising and use these predictions to inform marketing strategy. We find that catalogs have substantially longer-lasting impact on customer purchase than emails. We show how the model can be used to score and target individual customers based on their advertising responsiveness, and we find that targeting the most responsive customers increases the predicted returns on advertising by approximately 70% versus traditional recency, frequency, and monetary value-based targeting.

Keywords

advertising response, media mix, multichannel, randomized holdouts, dynamic linear model, Tobit model, hierarchical Bayes, single-source data

Disciplines

Advertising and Promotion Management | Business | Business Administration, Management, and Operations | Business Analytics | Management Sciences and Quantitative Methods | Marketing | Sales and Merchandising

Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data*

Daniel Zantedeschi
Eleanor McDonnell Feit
Eric T. Bradlow

The Wharton School
University of Pennsylvania

PRELIMINARY
PLEASE DO NOT QUOTE
DO NOT CIRCULATE WITHOUT PERMISSION

November 27, 2013

*We thank the Wharton Customer Analytics Initiative and three of its corporate partners who provided data, including the digital ad agency Organic. WCAI staff Melissa Hartz, Ben Adams and Ankur Roy all assisted with data preparation.

Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data

Abstract

Advances in data collection have made it increasingly easy to collect data on consumer-level purchases that are linked to those same customers' advertising exposures. However, advances in advertising response modeling (i.e., marketing mix modeling) have lagged behind the availability of this granular data. Extending extant models to multi-channel consumer-level data, we develop a Bayesian Tobit model that can be used to measure the effectiveness of advertising exposures, while accommodating the typical sparsity of consumer-level response data. Building on the traditional ad-stock framework, we are able to differentially determine separate decay rates for each advertising medium. This allows us to estimate channel-specific short- and long-term effects of advertising.

We demonstrate the applicability of our model on six data sets covering a variety of marketing platforms including television, online display, social media, e-mail and catalog. In each case, we show how the results provide insights into the interplay between different marketing channels in driving consumer response. We also illustrate how the model can be used to score and target individual customers on the basis of their advertising responsiveness, finding, in one case, that targeting the most responsive customers nearly doubles returns versus targeting the heavy purchasers.

Keywords advertising response, media mix, multi-channel, dynamic linear model, tobit model, hierarchical Bayes, single-source data

1 Introduction

Today's advertisers find themselves allocating budgets among an expanding array of advertising platforms including television, social media, online display, catalog, e-mail and many others. In this increasingly multi-channel environment, advertisers need tools to gauge advertising response across multiple channels in order to understand which channels produce the best sales returns and to make budgeting decisions about which channels to advertise on and which customers to target.

The upside to the increase in available advertising channels is that most of the new advertising platforms regularly collect consumer-level exposure data. Marketers now have the opportunity to know not just how much they spend on each advertising platform, but also exactly which viewers were exposed to that advertising at each point in time. When this advertising exposure data can be linked to data on individuals' purchases, the long-wished-for consumer-level advertising response data, i.e., "single-source data", becomes available, at a low cost.

In light of these changes in the advertising industry, we return to a classic question in marketing science: how do you measure the relationship between advertising exposures and consumer response? While this is a long-studied problem, our work is different from most prior work in that we focus on modeling the advertising process at the customer-level, which allows us to inform our estimates of advertising response not only from temporal variation in sales and aggregate advertising, but also by comparing subsequent purchases for customers who are exposed to an advertisement at a particularly point in time versus those who are not.

Specifically, we utilize an individual-level advertising response function in the tradition of the ad-stock literature (Nerlove and Arrow 1962, Jorgenson 1966). Importantly, we consider advertising response functions that are specified at the customer-level and allow for each customer to have a different response to each advertising channel. The ad-stock response function allows us to relate ad exposures to subsequent sales with just two parameters for each channel: the first characterizing the contemporaneous effect of advertising and the second related to the carry-over to the next period. By allowing for different pairs of parameters for each channel, we can gauge the differential contribution of different forms of advertising. We also propose a straightforward extension which allows for the possibility of interactions among the different channels, i.e., "synergies" (Naik, Raman and Winer 2005, Naik and Peters 2009), and we find substantial interactions in the data sets we explore. Consistent with the literature on dynamic choice models (McAlister et al. 1991, Seetharaman 2004), we also control for state dependence in the response behavior. Without controlling for state-dependence, as we demonstrate, we risk biasing the estimates of advertising response, particularly the carry-over of advertising. Finally, using a hierarchical Bayesian framework, advertising response, interaction and state-dependence parameters are all estimated for each consumer, allowing us to score and target consumers based on their predicted response across different channels.

While it should be intuitive that using more detailed data provides additional statistical power, there are two additional well-understood advantages to using consumer-level advertising response data to measure advertising effectiveness (see Tellis 2004, p. 61–66). First, multicollinearity between different types of advertising is substantially reduced when the data is at the individual level. Second, the well-known endogeneity problem that is induced by advertisers choosing to advertise more during periods when sales are unusually high is (partially) mitigated when response is measured at the consumer-level. Both of these advantages stem from the fact that unlike aggregate advertising spending, individual-level exposures are not often directly under the control of the advertiser. Consumers' media consumption choices, and therefore the opportunities to advertise to an individual, are typically under the control of the consumer (and sometimes the publisher) and not the advertiser. The resulting variation across individuals in their exposure patterns can be exploited to measure advertising effectiveness much more accurately than with traditional aggregate media-mix models.

However, while multicollinearity and endogeneity are less of a concern, consumer-level data at finer time scales is characterized by sparsity: a large fraction of "zeros" in the observed response variable, which poses significant modeling challenges. For example, daily or weekly purchase data for individual consumers typically include a large number of days without any purchases. This renders classic time series methods, often used to model aggregate advertising response (see Dekimpe and Hanssens 2000) less applicable. To tackle sparsity, we propose an adaptation of a classic Tobit I selection model. Similarly, advertising exposures are also sometimes sparse at the same time scale, leading to weak information about each individual's advertising response function. We overcome this second problem by specifying a hierarchical model that allows for shrinkage across the population.

The resulting model can be applied to any data set that includes marketing response, e.g., purchases, for each consumer in each time period, as well as the amount of advertising exposure on each channel. As a decision support tool, the output can be used to assess which channels are most effective overall and how advertising response plays out over time for each channel. For example, for a direct marketing data set, we show that e-mails and catalogs have about the same *overall* advertising response, but catalog response is spread out over a more extended period of time, which has important implications for whether and how often you send catalogs or emails. The model can also be used to target individual customers based on their responsiveness to advertising. For example, in that same application we show how targeting the most responsive customers can lead to substantial increases in advertising ROI versus the common practice of targeting customers based on their past purchase history, irrespective of their prior advertising response.

To summarize, building on the traditional ad-stock framework, we develop a model that can be used to exploit consumer-level advertising response data to gauge advertising response by channel for each consumer. The model can be applied across the wide variety of rapidly emerging sources of consumer-level advertising response data and the resulting estimated model can be used to inform the marketing mix and target individual consumers. To help readers better understand the broad data opportunity, we next provide a brief taxonomy of potential data sources that could be used to evaluate advertising response using our model. We then conclude the introduction by placing this work within the broader literature on advertising response.

1.1 Consumer-Level Advertising Response Data

In this section, we provide a brief overview of several readily-available sources of consumer-level advertising response data, both to illustrate the broad applicability of the data opportunity and to provide guidance to practitioners on assembling appropriate data to evaluate their own advertising. Table 1 summarizes the available data sources, which we describe in more detail next.

Advertising Data	**Response Data**	**Matching Mechanism**
online advertising	online sales online signups or other online response	cookie-tracking
direct marketing	direct purchases	CRM system
direct marketing	multi-channel purchases	name/address match or loyalty card
advertising panel	purchase panel	user-side instrumentation
television set-top box	loyalty card or credit card	name/address match

Table 1: Readily-available sources of data that match consumer-level advertising exposures to purchases.

Online Advertising and Purchases. Digital advertisers have pioneered the practice of tracking advertising exposures at the individual level. Driven largely by the structure of online environments, which deliver content directly to individual devices, online advertisers found it very natural to track each advertising exposure to a specific user. Today, nearly all the major web-based advertising platforms, e.g., Google DoubleClick, track each user with a cookie and record advertising exposures. This data is often provided to the advertiser in the normal course of doing business, e.g., through the DoubleClick Dynamic Advertising Reporting & Tracking (DART) system. In fact, online advertising is often priced according to the number of consumer impressions delivered, typically measured in cost per thousand impressions (CPM), and providing exposure data to the advertiser provides some verification that the impressions were actually served. For businesses where purchases are also made online, the data on advertising exposures can easily be linked via user cookies to purchases that are tracked as part of the clickstream data collected on the advertiser's online store. So, for advertisers that advertise exclusively online, and where all transactions occur online, it is relatively easy to assemble linked consumer-level advertising and response data. Manchanda et al. (2006) and Braun and Moe (2013) were among early researchers to use online advertising and purchase data to gauge advertising response. The web analytics practitioner community was also quick to recognize the potential value of this data source and has coined the term "attribution modeling" to describe several *ad-hoc* approaches to allocate credit for a customer's purchase back to a particular advertisement. In one of our applications, we illustrate the use of our model using a typical DART data set, and find that advertisements for a car manufacturer are more effective when hosted on automotive shopping sites rather than general media sites.

Direct Marketing. By nature, direct marketing focuses on providing advertisements directly to a particular physical address, phone number or e-mail, and these appeals often direct the customer to respond directly back to the company. Service providers such as insurance companies and media companies have long kept detailed records on direct marketing to each customer and subsequent subscriptions and renewals in their CRM systems. Further, multi-channel retailers like Target, Kroger and Macy's, are beginning to maintain large customer databases so that they can send promotions directly to their customers. To match their records of direct marketing exposures to subsequent purchases, these multi-channel companies often invest in strategies to get consumers to identify themselves when they make an in-store purchase such as loyalty card programs and store credit-cards. Companies that invest in this tracking can pull consumer-level advertising response data out of their CRM systems at nearly no additional cost. We obtained this type of data from a specialty retailer to assess the effectiveness of its e-mail and catalog campaigns. Our model is particularly useful in this context, because the the company can re-target customers and we illustrate how our model can be used to identify and target customers who have been highly responsive to advertising.

Single-source Panel Data. For a traditional manufacturing or consumer products company, like Coca-Cola, that advertises primarily through broadcast media, and distributes products through multiple retailers, it is more challenging to assemble advertising data at the consumer level. Traditional broadcast and cable systems historically do not track advertising exposures at the individual level and even if exposures were tracked, it would be incredibly challenging to assemble data on purchases from across the many outlets where Coke is sold. Faced with no alternative, in the 1980's, many advertisers turned to suppliers who could collect data from the consumer side, enrolling panelists who agree to have their exposures and purchases tracked. For example, by finding panelists who are enrolled in Nielsen National People Meter, where their TV advertising exposures are monitored via a device connected to their TVs, and also in Nielsen Homescan, where their CPG purchases are tracked by having the panelists scan the UPCs of the products they purchase, you can construct single-source panel data for television. Companies such as Nielsen, IRI, GfK and Kantar all have built single-source panels and such data is available in most developed countries, albeit at a high cost, due to the expense of enrolling and monitoring panelists. In a third application, we illustrate the use of our model using single-source panel data that includes television, online and Facebook brand exposures for two fast-moving consumer goods brands in a Western European country. For one brand, we find that television exposures have a much longer-lasting effect than Facebook exposures. Our research linking Facebook exposures with other forms of advertising is one of the first such studies to do so.

We note that for single-source panel data (versus the other types of data described in Table 1) it is relatively easy to collect advertising and response data for all competitors using the same panel. This data can be exploited to estimate a brand choice model and possibly pool information about advertising response across brands (Bollinger, Cohen and Lai 2013). However, in the interest of developing a general tool that can be applied across other types of consumer-level advertising response data such as online and direct marketing data, where competitive advertising and response information is not available, we do not attempt to exploit competitor data in this research; however, we do show how it can be incorporated in our mathematical framework.

Set-top Box and Loyalty Card Data. While traditional television does not track media consumption automatically, cable systems are rapidly installing newer "set-top box" systems that allow cable providers like Time Warner, Comcast and Cablevision to track and potentially target advertising exposures to individual households (c.f. Kent and Schweidel 2011). One can easily foresee a future where this data (appropriately anonymized) is provided back to their advertisers, just as is common practice in digital advertising today. To connect this data to consumers' purchases, companies like Tivo Research & Analytics are working to match set-top box data from multiple cable and satellite providers with purchase data collected through the loyalty card programs of several major retailers (TRA 2013), bringing the consumer-level focus pioneered in digital advertising to traditional media. Credit card companies are also a potential source of purchase data that could be matched (by name and address) to cable viewing data, albeit only at the level of the retailer and not the specific product. As tracking technologies evolve and companies form alliances to match consumers' advertising exposures to their purchases, the ultimate dream of consumer-level advertising response data for multiple channels collected in the normal course of business at low cost (and not with a sample of consumers enrolled in a panel) will soon be here. While we do not analyze set-top box data matched to purchase data in this research, the model we propose could easily be applied to this emerging data source to help advertisers gauge the effectiveness of their television advertising at relatively low cost.

In summary, there are several rapidly-emerging, low-cost sources of consumer-level advertising response data which have several common features. First, the data provides the date and time of all advertising exposures tracked to an individual “consumer,” which might be operationalized as a household with cable service, a cookie-tracked user or a particular e-mail address, depending on the type of data. Second, the data on advertising exposures can be mapped at the customer level to data on some type of response, which might be operationalized as purchases, subscriptions, website visits, etc. Third, with the exception of the more expensive single-source panel data, the data generally only provides information about the advertisers own brands, and not competitive advertising or response. Conceptually, any data with this structure can be used to more accurately measure advertising response and provide guidance to advertisers on their media mix and customer targeting. Our goal with this paper is to develop a broadly-applicable model that can be used to analyze this data structure and provide direction to advertisers.

1.2 Related Literature

Given the long history of advertising response models in marketing, we do not attempt to provide a comprehensive overview of the literature. (See Tellis 2004 and Tellis and Ambler 2007 for comprehensive reviews.) Instead, we focus on explaining how our work relates to the literature. To frame our contribution to advertising response modeling, we position our work at the intersection between consumer-level modeling and multi-channel advertising response (see Table 2), noting the sparsity of research that combines these two features.

Aggregate advertising response models are one of the earliest tools proposed in marketing science and very shortly thereafter, researchers extended the models to account for advertising across multiple-channels establishing marketing response modeling and marketing mix optimization as a key tool for advertisers (see Bowman and Gatignon 2010). Research in this literature (typically) employs a variety of time-series methods suitable for aggregate advertising spending and aggregate sales data and has explored many questions about “how advertising works” such as whether there are long-term effects of advertising (Ataman, van Heerde and Mela 2010) and whether there are advertising synergies (i.e., interactions) between channels (Naik and Ramam 2003, Danaher, Bonfrer and Dhar 2008, Naik and Peters 2009). While these are just a few representative papers, there is a long literature on understanding multi-channel advertising response with aggregate data.

Researchers as early as Little (1979) recognized that there are advantages to modeling advertising response at the consumer-level, but were stymied by the lack of data. Research with consumer-level data first emerged when single-source panel data became available in the late 1980's and this work tended to focus on just one channel: television (Pedrick and Zufryden 1991, Deighton, Henderson and Neslin 1994, Tellis 1998). Collecting single-source data for multiple channels was cost prohibitive. More recently researchers have turned to using single-channel data on display advertising and online purchases (Manchanda et al. 2006, Braun and Moe 2013), where, as we discussed above, advertising exposures and purchases are regularly tracked at the cookie level. So, most of the early work on advertising response with individual-level data has focused on a single channel, largely due to data limitations.

There are only two recent papers that we are aware of that propose multi-channel advertising response models with consumer-level data: Danaher and Dagger (2013) and Bollinger, Cohen and Lai (2013). Danaher and Dagger (2013) develop a clever measurement strategy for collecting multi-channel advertising exposure data by surveying consumers and asking them to recall which media channels they watched or read and linking that back to the media plan. Because their application focused on advertising for a two-week clearance sale at a retailer, they focused on same-period response to advertising and did not address the issue of how to relate exposures to purchases in subsequent time periods as we do here. More closely related to our work is that of Bollinger, Cohen and Lai (2013), who also develop an hierarchical ad-stock framework and estimate different decay parameters for each channel. In comparison to their work, our paper has a number of important methodological differences: (i) we consider a broader set of data sets and advertising channels, both to show the general applicability and the statistical robustness of the approach;

(ii) our paper includes both a Tobit I selection model (to handle sparsity) and ad-stock error terms that, as described in Section 4, have significant desirable properties; and (iii) we develop decision support tools to target channels and customers based on their predicted overall responsiveness to advertising. Moreover, while Bollinger, Cohen and Lai (2013) focus on the competitive implications of multiple advertising channels by modeling consumer's choice among competitors and pooling advertising response parameters across brands within a category, our model supports the decision process of an advertiser who does not have access to competitors individual-level data and wishes to properly measure advertising effectiveness for a *single* brand. We will discuss possible extensions of our work, that allow for competitive effects in Section 4. In the next section we formally lay out the model.

	Aggregate Expenditures & Sales	**Consumer Exposures & Purchases**
Single Channel	Little 1979 Broadbent 1984*	Tellis 1998 Pedrick and Zufryden 1991 Deighton, Henderson and Neslin 1994 Manchanda, Rossi and Chintagunta 2004 Braun and Moe 2013
Multi-channel	Tellis 2004; Raman and Naik 2006 Naik and Peters 2009 Danaher, Bonfrer and Dhar 2008	**This Research** Bollinger, Cohen and Lai 2013 Danaher and Dagger 2013

Table 2: A categorization of prior work on advertising response models along two key dimensions: number of channels and level of aggregation. Our paper is among the first investigating individual-level exposures and purchases in a multi-channel environment.

2 Model Development and Computation

The foundation of our model is the discrete-time exponentially decaying ad-stock model that was first introduced by Koyck (1954) and Jorgenson (1966). The key construct in the Koyck model is that at time t each individual $i = 1, \ldots, N$ accumulates a latent stock variable, W_{ikt}, for each channel as a result of current and previous periods exposures to advertising, X_{ikt}, on that channel k:

$$W_{ikt} = \sum_{j=0}^{t} \rho_{ik}^{j} \left(X_{i,k,t-j} + \varepsilon_{i,k,t-j} \right) \tag{1a}$$

$$= \rho_{ik} W_{i,k,t-1} + X_{ikt} + \varepsilon_{ikt} \tag{1b}$$

where ρ_{ik} is a consumer- and channel-specific decay parameter constrained to $[0, 1)$ and ε_{ikt} is a series of *iid* shocks which are also consumer- and channel-specific. The formula in (1a) presents an exponentially decaying stock variable which has a long tradition in marketing (see the survey by Huang, Leng and Liang 2012). However there are practical difficulties with working with this representation since it extends over possibly many time periods. Fortunately, it can be easily shown to be equivalent to the autoregressive process in (1b).

The inclusion of the error term in equation (1b) is of significant note, as there is some disagreement in the literature about whether the shocks should be included and there are theoretical and computational motivations for doing so. Dubé, Hitsch and Manchanda (2005) justify the error term on the basis that the data (in their case aggregate exposures and in our case individual-level exposures) does not fully capture all aspects of the advertising, such as the quality of the copy. In our case, there may also be variations in how much the customer attended to the ad during a particular exposure or how persuasive the copy is to a particular consumer, magnifying the need for a stochastic term in the ad-stock. The error term is also important in the computation of the stocks since it allows us to transform the ad-stock model to a state-space form as we describe below.

We extend the single latent stock formulation given in equation (1b) to a multi-channel environment by introducing K stock variables as in (1a):

$$W_{i1t} = \sum_{j=0}^{t} \rho_{i1}^{j} \left(X_{i,1,t-j} + \varepsilon_{i,1,t-j} \right) \tag{2a}$$

$$W_{i2t} = \sum_{j=0}^{t} \rho_{i2}^{j} \left(X_{i,2,t-j} + \varepsilon_{i,2,t-j} \right) \tag{2b}$$

$$\dots$$

$$W_{iKt} = \sum_{j=0}^{t} \rho_{iK}^{j} \left(X_{i,K,t-j} + \varepsilon_{i,K,t-j} \right) \tag{2c}$$

We also include additional stocks for the interactions among the different channels, a topic of significant current empirical interest (see Wilbur 2008, Naik, Raman and Winer 2005). For illustration, consider the interaction between exposures on channel 1 and channel 2. We can form an additional stock variable with a separate decay and error term as follows:

$$W_{i,1:2,t} = \sum_{j=0}^{t} \rho_{i,1:2}^{j} \left(X_{i,1,t-j} \cdot X_{i,2,t-j} + \varepsilon_{i,1:2,t-j} \right). \tag{3}$$

The specification in equation (3) allows for the possibility that a exposure on channel 1 and 2 simultaneously in the same period gives rise to a contemporaneous interaction effect given by $X_{i,1,t} \cdot X_{i,2,t}$ and a carry-over effect at lag j given by $\rho_{i,1:2}^{j} X_{i,1,t-j} \cdot X_{i,2,t-j}$. These, together with the series of *iid* error terms $\varepsilon_{i,1:2,t-j}$, define the interaction stock $W_{i,1:2,t}$.

Finally, we consider the inclusion of a state dependence component to account for the possibility that purchasing decisions made in the previous period impact the purchasing decision in the current period, independently of advertising exposures. We leverage the same ad-stock framework to accommodate state dependence:

$$W_{iSt} = \sum_{j=0}^{t} \rho_{iS}^{j} \left(\mathbf{1}\left[Y_{i,t-1} > 0\right] + \varepsilon_{iSt} \right) \tag{4a}$$

$$= \rho_{iS} W_{iSt} + \mathbf{1}\left[Y_{i,t-1} > 0\right] + \varepsilon_{iSt}. \tag{4b}$$

where $\mathbf{1}\left[Y_{i,t-1} > 0\right]$ is an indicator for whether customer i purchased in the prior period. This form nests several extant models for state dependence, e.g., when $\rho_{iS} = 0$ in equation 4a) we get the classic single-period period "bump" (i.e., first order Markov process) for buying a given product in the previous period. We note that many empirical studies have shown support for state dependence (see Dubé, Hitsch and Rossi, 2010) and without controlling for state dependence the effect of advertising can be severely over-estimated (Seetharaman 2004). However, if state dependence is not supported in a particular data set, the estimate for the effect of W_{iSt} on the response variable will go to zero.

There is also a computational advantage to defining both the interactions and the state dependence using the same form as the ad stocks. Note that the entire vector of stocks consisting of the K multi-channel stocks, $\left(K^2 - K\right)/2$ interaction stocks and the state-dependence stock can be stacked as a vector

$$\mathbf{W}_{it} = \left[\underbrace{W_{i1t}, \ldots, W_{iKt}}, \underbrace{W_{i,1:2,t}, \ldots, W_{i,(K-1):K,t}}, \underbrace{W_{iSt}} \right]'$$

Multi-Channel + Pair-wise Interactions+ State Dependence.

Then we can derive a more compact representation of the dynamics in the model by simply writing:

$$\mathbf{W}_{it} = \rho_i \mathbf{W}_{i,t-1} + \mathbf{X}_{it} + \varepsilon_{it} \quad \varepsilon_{it} \sim N\left(0, \Sigma_{\varepsilon}\right) \tag{5}$$

where the bold fonts denote the vectorized components of the multi-channel ad-stock model over channels, interactions, state-dependence stocks and their exposures. $\Sigma_\varepsilon = diag(\sigma_{\varepsilon_1}, ..., \sigma_{\varepsilon_K}, ...)$ denotes a diagonal matrix of the error terms. Note that defining Σ_ε as a diagonal matrix avoids rotations leading to unidentifiability of the latent stocks (West and Harrison, 1997). Finally, as mentioned earlier, the vector of unobserved stock variables in formula (5) can be interpreted as the dynamic states equation in a state-space model which can be exploited to facilitate computation (see Appendix A for details).

Now, with the ad stock dynamics in hand, we relate the ad-stocks to the observed purchases, Y_{it}, through a latent variable, Y^*_{it}, for individual i at time t. We assume that the latent process, Y^*_{it}, is related to the the ad-stock variables, including interactions and state dependence, by sensitivity parameters β_{ik} which measure the instantaneous effect of the stock on latent Y^*_{it}.

$$Y^*_{it} = \mu_i + \sum_k \beta_{ik} W_{ikt} + \sum_{k'>k} \beta_{i,k:k'} W_{i,k:k',t} + \beta_{iS} W_{iSt} + \eta_{it} \quad \eta_{it} \sim N\left(0, \sigma_\eta\right) \tag{6}$$

Note also the presence of the intercept, μ_i, that can account for baseline differences between individuals in their purchasing behavior. This consumer-level intercept helps to avoid biases caused by the advertiser targeting customers who are more likely to buy, i.e., the estimated effect of advertising measured by β_{ik} is above and beyond that expected given the individual's estimated baseline propensity to buy.

We then specify a Tobit I selection process, accounting, as mentioned earlier, for sparsity in the purchases, by relating the latent values, Y^*_{it} to the observed outcomes, Y_{it}:

$$Y_{it} = \begin{cases} Y^*_{it} \text{ if } Y^*_{it} > 0 \\ 0 \text{ if } Y^*_{it} \leq 0 \end{cases} \tag{7}$$

To summarize, we develop an heterogeneous model of consumer purchasing behavior where the underlying processes are derived using Koyck stock-variables. The model allows for separate stock-variables for each channel and an interaction between them. The set of parameters determining the dynamics of the stock-variables can be easily interpreted as the simultaneous effectiveness and exponential decay rate of each advertising channel (or interaction or state dependence). To account for purchasing carry-over that is unrelated to advertising exposures but related to individuals' purchase behavior, we have introduced a structural state-dependence component using the same Koyck form for computational convenience. Finally to account for sparsity in the consumer-level purchase data we have specified a Tobit I selection process.

2.1 Priors and Computation

We employ a Bayesian approach to accommodate individual heterogeneity in both the instantaneous and carry-over effects (Rossi and Allenby 2003). We use conjugate Normal-Inverse Wishart distributions for all individual-level parameters. We define β_i as the stacked vector of μ_i,β_{ik},$\beta_{i,k:k'}$ and β_{iS}. Similarly, we define ρ_i as the stacked vector of ρ_{ik}, $\rho_{i,k:k'}$ and ρ_{iS}. We then define the distribution of β_i and ρ_i across the population of consumers as follows:

$$\beta_i \sim N_{\frac{K(K+1)}{2}+2}\left(\bar{\beta}, \Sigma_\beta\right) \quad \rho_i \sim N_{\frac{K(K+1)}{2}+1}\left(\bar{\rho}, \Sigma_\rho\right)$$

where N represents a multivariate normal distribution. We ensure the identification of the latent ad-stock processes via truncation and rejection sampling over the sets $0 \leq \rho_i < 1$.

We put weakly informative priors on the population-level parameters and the error terms.

$$\bar{\beta} \sim N_{\frac{K(K+1)}{2}+2}(0, 10I) \quad \Sigma_\beta \sim IW\left(I, \frac{K(K+1)}{2} + 2 + 2\right)$$

$$\bar{\rho} \sim N_{\frac{K(K+1)}{2}+1}(0, I) \quad \Sigma_\rho \sim IW\left(I, \frac{K(K+1)}{2} + 1 + 2\right)$$

$$\sigma^2_{\varepsilon_k} \sim IG(1,1) \quad \sigma^2_{\eta} \sim IG(1,1)$$

where IW is an inverse-Wishart distribution, IG is an inverse-Gamma distribution, and I is an identity matrix of appropriate dimension.

Given the state-space representation for the latent parameters, we employ standard Kalman, Dynamic Linear Model relationships (West and Harrison 1997). See also Naik and Raman (2003) and Ataman, van Heerde and Mela (2010) for similar derivations. The Tobit I selection induces a standard data augmentation step as described in Chib (1992). Appendix A also provides more details of the estimation algorithm.

2.2 Cumulative Impulse Response (CIR)

The key parameters of the model, β_{ik} and ρ_{ik}, define the response of individual i to advertising on channel k. However, because of the ad-stock formulation and the Tobit mechanism, these parameters do not directly relate to the economic value of delivering an additional advertisement to consumer i on channel k at time t. Instead, the cumulative impulse response, defined as the expected cumulative incremental effect on future purchases for one-impression impulse in X_{ikt}, is a more economically meaningful measure of the expected return from an additional exposure to consumer i on channel k. In this section, we derive the cumulative impulse response in closed-form, so that it is easy to compute from the estimated consumer-level parameters, and can therefore be used to categorize individuals into those who are expected to be more or less responsive to advertising. This provides a basis for targeting strategies that we illustrate in Section 3.

First, consider the change in the expected value of Y_{it} due to an increase in advertising exposures from channel k at time t. This instantaneous marginal effect can be written as

$$\frac{\partial E_t\left(Y_{it}\right)}{\partial X_{ikt}} = \underbrace{P\left(Y_{it}^* > 0\right) \cdot \frac{\partial E\left[Y_{it}|Y_{it}^* > 0\right]}{\partial X_{ikt}}}_{I} + \underbrace{\left(E\left[Y_{it}|Y_{it}^* > 0\right]\right) \frac{\partial P\left(Y_{it}^* > 0\right)}{\partial X_{ikt}}}_{II} \tag{8}$$

which is referred to as the McDonald and Moffitt (1981) decomposition. This allows one to see that a change in exposures on channel k affects the conditional mean of Y_{it}^* in the positive part of the distribution (I) and it affects the probability that the expected purchases will be non-zero (II).

By means of simple transformations dealing with the truncated Normal distribution due to the Tobit I structure (see Greene 2008), it follows that the marginal effect on the expected value for Y_{it} is given by:

$$\frac{\partial E\left(Y_{it}\right)}{\partial X_{ikt}} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\beta_{ik} \tag{9}$$

$$V_{it} = \sum_k \beta_{ik} W_{ikt} + \sum_{k'>k} \beta_{i,k:k'} W_{i,k:k',t} + \beta_{iS} W_{i,S,t}. \tag{10}$$

where Φ denotes the standard normal CDF. As can be seen in equation (10) the contemporaneous effect on expected purchases due to advertising depends on both β_{ik} and on the expected *activation* of the individual i at time t captured by the $\Phi\left(\right)$ function, which is time-varying due to the fact that each individual's ad-stock varies based on his or her prior advertising exposures and shocks. Thus, given two customers with identical ρ_{ik} and β_{ik} parameters, it is more beneficial to advertise to the one who has had more prior exposures and therefore a higher ad-stock, as the activation propensity will be greater.

While equation (10) gives the contemporaneous response to advertising, marketers are typically concerned with the total cumulative response that they can expect from an additional exposure to channel k at time t, i.e., the area under the impulse response curve. In the limit, by taking $T \to \infty$ and by the properties of the mean of a geometric series, the closed-form for the expected CIR is given by:

$$CIR_{ikt} = \sum_{j=0}^{\infty} \frac{\partial E\left(Y_{i,t+j}\right)}{\partial X_{ikt}} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\left(\frac{\beta_{ik}}{1-\rho_{ik}} + \phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\frac{\beta_{ik}\beta_{iS}}{1-\rho_{iS}}\right) \tag{11}$$

While the effect of an impression on any individual can be forward simulated from the model, equation (11) provides a computationally convenient way to identify customers who would be most affected by advertising at time t. We will illustrate the use of the CIR for customer scoring in our empirical applications.

It is easy to extend the reasoning to derive a closed-form expression for an $\varepsilon-$shock on both channels k and k' in the same period:

$$\begin{aligned} CIR_{i,k:k',t} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\cdot \\ \cdot\left(\frac{\beta_{i,k}}{1-\rho_{i,k}} + \frac{\beta_{i,k'}}{1-\rho_{i,k'}} + \frac{\beta_{i,k:k'}}{1-\rho_{i,k:k'}} + \phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\frac{\beta_{i,S}\left(\beta_{ik} + \beta_{i,k'} + \beta_{i,k:k'}\right)}{1-\rho_{i,S}}\right) \end{aligned} \tag{12}$$

which allows us to gauge the additional contribution of the interaction terms.

3 Empirical Analysis

In this section, we demonstrate the use of the model described in the previous section using three different datasets comprising six different brands. We next describe the data sets and how they fit into the "single-source" data types described earlier.

3.1 Data

Our focal data for this paper describes e-mail and catalog advertising from a company (who chooses to remain anonymous) that owns three independently-managed retail brands. Each brand sells their own product lines through a website and brick-and-mortar stores in the United States. Each brand sends out emails and catalogs[1] that are designed to drive purchases at the website or stores, typically by informing customers about newly-available products. Unlike some multi-brand retailers (e.g., Gap) this company avoids any cross-brand marketing and there is no coordination of marketing across the three retail brands. The brand websites do not reference each other and few customers are aware that the brands are even owned by the same company. None of the brands does substantial mass media advertising, so the email and catalog exposures we observe represent all the advertising that the customers see.

The company maintains a CRM system that tracks all customers who have expressed interest in the brand, either by purchasing something or signing up to receive emails or catalogs. The CRM system also records the emails and catalogs that are then sent to each customer. Customers typically pay with a credit card (either online or in-store), which allows the purchases to be matched to the advertising that the customer has received via their name and address. If the customer pays at the store in cash, clerks are trained to request an email address or phone number, which can also be used to track the purchase back. More than 80% of purchases are matched to existing customers in the CRM system. (Brand C has a match

[1]We treat catalogs as a form of advertising rather than a channel because the company uses catalogs as advertising that directs customers to online or physical stores. There is no way to purchase "through" the catalog.

rate approaching 95%.)

For each brand, we analyzed a different random sample of 300 customers from their CRM data, with all of their direct marketing exposures and purchases observed over two years. Table 3 summarizes these three data sets. Our goal in analyzing this data was to understand the differential effects of catalog versus email, to gauge which channel drives the most lift in purchasing and to compare advertising effectiveness across the three brands.

	Brand A	**Brand B**	**Brand C**
Time Periods	102 weeks	102 weeks	102 weeks
Average Exposure Incidence per Week	Catalog=11.4% Email=67.7%	Catalog=13.9% Email=64.1%	Catalog=15.5% Email=30.9%
Average Purchase Incidence Per Week	Purchases=8.8%	Purchases=7.8%	Purchases=4.2%

Table 3: Summary of linked direct marketing and purchase data for three independent speciality retailer brands.

Note that despite the fact that the company sends an email campaign nearly every week, there are substantial portions of the population that do not receive any emails, either because the customer has not provided her email address or because she has requested that she not receive them. Incidence of catalog exposure is much lower, both because the brands send catalogs less than once per week and because some customers have not provided a mailing address or have requested that they not receive catalogs. This variation in exposures across people is critical in identifying the effect of the advertising on purchases. As can be seen in Table 3, there is also a substantial amount of sparsity in the purchase incidence; for Brand C, customers on average purchase in only 4.2% of the weeks that they are observed. This level of sparsity, typical of marketing applications, motivates the Tobit I structure we use.

In addition to the data described in Table 3, we also analyzed advertising exposure and purchase data for two CPG brands (from a different firm who also chooses to remain anonymous) that were collected using a traditional single-source panel in a Western European country. As part of their panel, all households had devices installed on their televisions and computers to monitor their television and online display advertising exposures. For “Brand X”, we observe all TV and online display exposures, as well as all purchases which were collected through a home UPC scanner. “Brand Y”, did not do any online display advertising during our observation window, but it had a popular Facebook fan page, and the software installed on the panelist’s computer also tracked each exposure to brand mentions on Facebook. This would occur when the person saw a post from the brand in their Facebook feed because they or a friend were a fan of the brand. Note, there was no Facebook paid advertising in the country during the time that this data was collected. Our goal in analyzing this data was to understand the influence of TV advertising, which is generally accepted to be the most powerful advertising medium, as compared to the newer digital channels.

Table 4 summarizes the data for Brands X and Y. Note that the purchase incidence in this data is very low. As we have noted, when there is sparse response incidence or too little variation across users and time in the advertising exposure, then it is more difficult to identify the advertising effects and the posteriors become wider.

	Brand X	**Brand Y**
Time Periods	80 Weeks	106 Weeks
Average Exposure Incidence Per Week	TV=13.3% Online=10.3%	TV=19.9% Facebook=0.5%
Average Purchase Incidence Per Week	Purchases=3.13%	Purchases=1.0%

Table 4: Summary of data on TV, Online Display and Facebook exposures linked to purchases for two fast-moving consumer goods brands.

Finally, to illustrate the breadth of contexts in which we can apply this research, we also analyzed a data set on online display advertising for an US automotive brand. The data was collected through the

DART system (see Section 1.1) that tracks display advertising exposures and "conversions" at the target website. (Since it is impossible to purchase a car online in the US, this company uses key customer events, such as contacting a dealer or searching inventory, as conversion events and there can be multiple such "conversions" per customer.) While all of the advertising is online display, this company divides websites into two "channels": general media websites such as CNN.com and automotive websites like Edmunds.com. The advertising exposures are roughly equal between the two. In this application, we analyze the advertising response for each of these two "channels," with an eye toward re-balancing the advertising mix between the two to maximize conversions.

	Onl. Display
Time Periods	79 Days
Average Exposure Incidence Per Day	Media=2.1% Automotive=1.65%
Average Conversion Incidence Per Day	Conversions=1%

Table 5: Summary of online display advertising data for two different types of websites linked to conversions at a target website.

3.2 Nested Model Comparison

To understand the importance of the various features of the model, and provide evidence that the full specification we proposed in the previous section is necessary, we fit a series of nested models to all six data sets. The nested models included (1) a model with stocks for each advertising channel, (2) a model with stocks for each channel plus pair-wise interaction effects, (3) a model with stocks for each channel plus state dependence (4) the complete model as described in Section 2 including both state-dependence and interaction effects. To estimate each model we ran a Metropolis within Gibbs sampler for 100,000 draws with a burn-in of 20,000 draws. All chains were converged according to the Raftery and Lewis' (1996) test implemented in the CODA package.

As suggested in West and Harrison (1999, p. 393-394) we evaluated the model fit based on in-sample and out-of-sample posterior predictive checks[2] To assess in-sample fit, we use the root mean squared error between the predicted and actual purchases for each individual in each week as well as the average log-likelihood for each observation. To assess out-of-sample fit, we use the root mean square error of the one step ahead forecast. These are defined as

$$RMSE = \frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\frac{\sqrt{(\hat{Y}_{it}-Y_{it})^2}}{T}\right)}{N}$$

$$RFME = \frac{\sum_{i=1}^{N}\sqrt{(\hat{Y}^F_{i,T+1} - Y_{i,T+1})^2}}{N}$$

where the predicted values, $\hat{Y}_{i,t}$ are for the in-sample data and $\hat{Y}^F_{i,T+1}$ for out-of-sample are obtained for each individual from the posterior predictive distribution of the individual-level parameters, consistent with our Bayesian framework.

Table 6 reports the in- and out-of-sample model fits across all six data sets. We normalize the root mean square errors to facilitate comparison across data sets[3]. Across all the data sets there is support for both the interaction and the state-dependence components of the model. Combined, including these two effects, reduces the in-sample RMSE by 15% and the out-of-sample RMFE by 20% over model 1. Based on these analyses, we find evidence of both interaction and state-dependence in all six datasets. For this

[2]Alternative approaches to model selection are based on entropy measures such as the Deviance Information Criterion (DIC) of Gelfand and Ghosh (1998). However, when dealing with hierarchical models with several latent variables, as noted by Shirley et al. (2010), model selection based on indicators such as the DIC can lead to contrasting results. See also the discussion in Duan, McAlister and Sinha (2011) in the context of choice models with cross-brand pass-through effects. For this reason we rely on classic in- and out-of-sample procedures in the spirit of West and Harrison (1997).

[3]This normalization is often used in forecasting to compare the performance of the same models across different data sets (see Diebold, 2006).

reason we focus for the remainder of the paper on model 4 and report the population-level parameters for this model in Table 7 for the retail dataset and Table 10 for the other datasets.

		Retailer's Direct Marketing			Single-Source		Online
		Brand A	Brand B	Brand C	Brand X	Brand Y	Display
(1) No Interactions or State Dependence	$RMSE$	1.11	1.10	1.12	1.79	1.84	1.40
	LL	−0.50	−0.56	−0.62	−1.10	−1.21	−0.70
	$RMFE$	1.75	1.68	1.75	2.80	2.83	2.71
(2) Interactions Only	$RMSE$	1.01	1.02	0.99	1.57	1.61	1.33
	LL	−0.53	−0.56	−0.62	−1.10	−1.21	−0.62
	$RMFE$	1.48	1.52	1.48	2.40	2.46	2.14
(3) State Dependence Only	$RMSE$	0.95	0.92	0.94	1.49	1.47	1.29
	LL	−0.44	−0.56	−0.42	−0.91	−0.88	−0.60
	$RMFE$	1.46	1.38	1.43	2.23	2.29	2.05
(4) Interactions and State Dependence	$RMSE$	**0.94**	**0.88**	**0.89**	**1.39**	**1.43**	**1.26**
	LL	**−0.43**	**−0.44**	**−0.43**	**−0.86**	**−0.91**	**−0.58**
	$RMFE$	**1.36**	**1.31**	**1.30**	**2.12**	**2.27**	**1.97**

Table 6: Model comparison based on in- and out-of-sample goodness-of-fit measures across models with different components and datasets. RMSE represents the averaged root mean squared error, LL the average log-likelihood, RMFE the averaged root mean forecast error. The best-fitting model for each data set is highlighted in boldface.

3.3 Direct Marketing Case Study

To illustrate the interpretation and use of the estimated model, we focus in this section on the direct marketing data sets.

Table 7 shows the estimated parameters for all three retail brands. The estimated parameters related to instantaneous effectiveness of advertising, $\bar{\beta}_k$, suggest that email communications are instantaneously more effective than catalogs across all three brands. There is also a significant interaction between the two channels in all three cases. The different levels of instantaneous effectiveness in combination with a significant interaction effect may be explained (albeit it is somewhat speculative and an open area for field

	Brand A	**Brand B**	**Brand C**
$\bar{\mu}$	-18.82 $(-19.23, -18.41)$	-19.83 $(-19.99, -19.66)$	-19.89 $(-20.01, -19.76)$
$\bar{\beta}_{\text{CAT}}$	0.38 (0.11, 0.67)	0.33 (0.09, 0.58)	0.28 (0.06, 0.51)
$\bar{\beta}_{\text{EM}}$	0.87 (0.79, 0.95)	0.55 (0.47, 0.62)	0.61 (0.53, 0.69)
$\bar{\beta}_{k\,:\,k'}$	0.37 (0.10, 0.65)	0.43 (0.13, 0.72)	0.48 (0.14, 0.85)
$\bar{\beta}_S$	2.32 (2.07, 2.57)	3.35 (3.04, 3.68)	3.29 (3.04, 3.75)
$\bar{\rho}_{\text{CAT}}$	0.30 (0.10, 0.50)	0.51 (0.27, 0.77)	0.25 (0.07, 0.43)
$\bar{\rho}_{\text{EM}}$	0.09 (0.01, 0.17)	0.11 (0.02, 0.21)	0.08 (0.01, 0.18)
$\bar{\rho}_{k\,:\,k'}$	0.25 (0.07, 0.42)	0.52 (0.27, 0.77)	0.29 (0.09, 0.49)
$\bar{\rho}_S$	0.73 (0.67, 0.79)	0.87 (0.80, 0.94)	0.87 (0.80, 0.95)

Table 7: Population-level parameter estimates for catalog and e-mail exposures for Brands A, B & C. Control variables including dummies to account for seasonal effects are included but not reported.

experimentation) as the email advertising's ability to stimulate and initiate information search which is then completed using information presented in the catalog (see also Dinner, Van Heerde and Neslin 2011).

The carry-over coefficients, $\bar{\rho}_k$, also present an interesting pattern. Across all three brands, catalog has a significantly higher carry over than email. This suggests that while email is more effective at increasing same-week sales, a catalog exposure has a more long-lasting effect. These parameters indicate that 90% of the email advertising effects dissipate on average in two weeks while catalogs are more durable with their effects dissipating in about four weeks. Interestingly, Brand B tends to mail a catalog approximatively once per month which suggests the brand's marketers have an intuitive sense of this carry-over. Finally, the estimated state-dependence effects are substantially larger than advertising effects and are quite persistent, which is consistent with other studies such as Seetharaman (2004).

As the estimated parameters for the brands are similar, we next explore the estimated model for Brand

B in more detail, focusing on how including state dependence in the model affects the model predictions. Continuing with Brand B, we then show how the model can be used by practitioners to compute CIRs for each channel, how to compare the dynamics of the aggregate advertising response for different channels through impulse response curves, and how to target individual customers based on the heterogeneity across customers in advertising responsiveness.

3.3.1 State dependence

Table 8 shows the effect of the inclusion of the state-dependence component on the estimated population parameters. As can be seen from the estimated parameters, including the state-dependence stock leads to a substantial decrease in the population-level decay parameters for catalog, email and their interaction. This means that without including state-dependence one would potentially over-estimate the proportion of the response that should be attributed directly to advertising. Although, as we will show subsequently, ads can directly induce a purchase, they also produce additional future sales through state dependence. Since the advertiser still benefits from these additional sales that are attributed to state dependence, we include them in our calculation of the CIR, as can be seen equation (11).

Figure 1 graphically illustrates the role of state dependence in the model dynamics by comparing the predicted lift from an impulse of one catalog exposure combined with one email exposure in the same week for the full model, model 4, versus model 2 which does not include state-dependence. As Figure 1 shows, the shape of the response function is substantially different between the two models. The model with no state dependence predicts a response that peaks in the week of the impulse (week 0) and is monotonically decreasing in subsequent weeks. The model with state dependence predicts a peak response in the first week after the impulse due to the large state-dependence effects which begin in week 1. The better fit statistics in Table 8 reflect the fact that the response with state dependence may represent better the shape of the response observed in the data (albeit this is conditional on the exponentially weighted

	(2) Interactions Only	(4) Interactions and State-Dependence
$\bar{\mu}$	-19.78 $(-19.96, -19.52)$	-19.83 $(-19.99, -19.66)$
$\bar{\beta}_{\text{CAT}}$	0.51 $(0.10, 1.25)$	0.33 $(0.09, 0.58)$
$\bar{\beta}_{\text{EM}}$	0.43 $(0.03, 0.77)$	0.55 $(0.47, 0.62)$
$\bar{\beta}_{k\,:\,k'}$	0.95 $(0.21, 1.69)$	0.43 $(0.13, 0.72)$
$\bar{\beta}_S$		3.35 $(3.04, 3.68)$
$\bar{\rho}_{\text{CAT}}$	0.75 $(0.54, 0.96)$	0.51 $(0.27, 0.77)$
$\bar{\rho}_{\text{EM}}$	0.14 $(0.01, 0.28)$	0.11 $(0.02, 0.21)$
$\bar{\rho}_{k\,:\,k'}$	0.80 $(0.63, 0.97)$	0.52 $(0.27, 0.77)$
$\bar{\rho}_S$		0.87 $(0.80, 0.94)$

Table 8: Effect of state dependence on the estimation of the population parameters for Brand B.

decay reasonably fitting the advertising response.) Further testing on the different functional forms for advertising response, including response functions that allow for the advertising response to peak even later than the subsequent week would be a fruitful area for future research.

3.3.2 Comparing the Dynamic Response to Different Advertising Channels

To illustrate the model's dynamic predictions about how the response to advertising plays out over time, we plot predicted impulse response curves in Figure 2. The plot shows the impulse response for sending one email to all 300 customers versus one catalog to all 300 customers. Note that the full predicted impulse response includes both the direct contribution of the advertising and the indirect contribution of state dependence. The large estimated state dependence effect for Brand B induces a peak response in week 1 for both channels. As was suggested by the estimated population parameters, emails are instantaneously more effective, as can be seen by the predicted increase in sales in week zero, which is slightly higher for

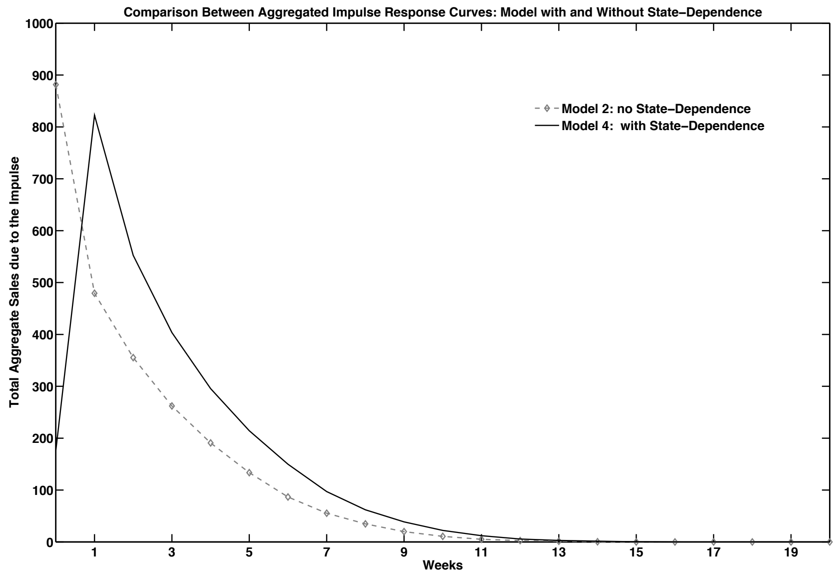

Figure 1: Comparison of predicted impulse response functions for models with and without state dependence. The plot depicts the predicted lift in sales for one additional email and catalog exposure for all 300 customers of Brand B.

the email impulse. However, the effect of catalogs is more long-lasting and larger overall.

As noted above, we find a strong interaction between catalog and email and Figure 3 shows the predicted impulse response curve for a simultaneous exposure to catalog and email for all 300 customers. The chart breaks down the overall response into that directly attributable to the catalog and email, and that driven by the interaction and the state dependence. In week zero, the same-week response is about $200, with a modest contribution from the catalog, a larger contribution from email and a large contribution due to the interaction between email and catalog. However, in week 1, the large effect of state dependence kicks in and we see that state dependence accounts for about 60% of the total response.

The impulse response functions we present here are illustrative of how advertisers can use the model to decompose their observed advertising response and attribute "credit" to each channel for the observed lift in sales. This is one of the major benefits to advertisers of using the model to analyze their past advertising and predict which channels exceed their costs and bring the largest return. We next turn to the other major potential use for the model: targeting advertising to individual customers.

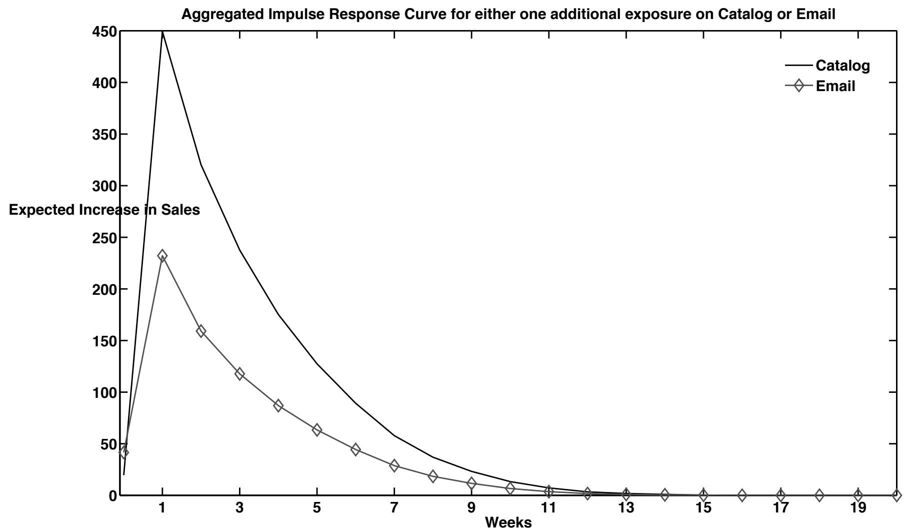

Figure 2: Comparison between aggregated cumulative impulse response curve for a one-impression shock to all customers on either email or catalog.

3.3.3 Targeting Individual Customers

While the population-level parameters give us a sense of the overall effectiveness of each channel, the model also gives us information about the responsiveness of each customer through the β_{ik} and ρ_{ik} parameters. Figure 4 plots posterior means of these parameters for the individuals in the Brand B data set. Somewhat surprisingly, we don't find a strong *a posteriori* relationship between β_{ik} and ρ_{ik} for either catalog or email (Panels A and B). That is, individuals with higher initial response to advertising don't seem to have systematically more or less carryover for either channel. However, we do see some negative correlation between the ρ_{ik} for email and the ρ_{ik} for catalog (Panel D) indicating that people who have a longer-lasting response to catalog, tend to have a less long-lasting response to email and vice-versa. Most importantly, we find that there is a great deal of heterogeneity between customers in their advertising response particularly for catalog, which opens up the opportunity to target catalog advertising (which is relatively expensive) to the most responsive customers.

In Table 9, we illustrate the economic advantages of targeting the most responsive customers. (We

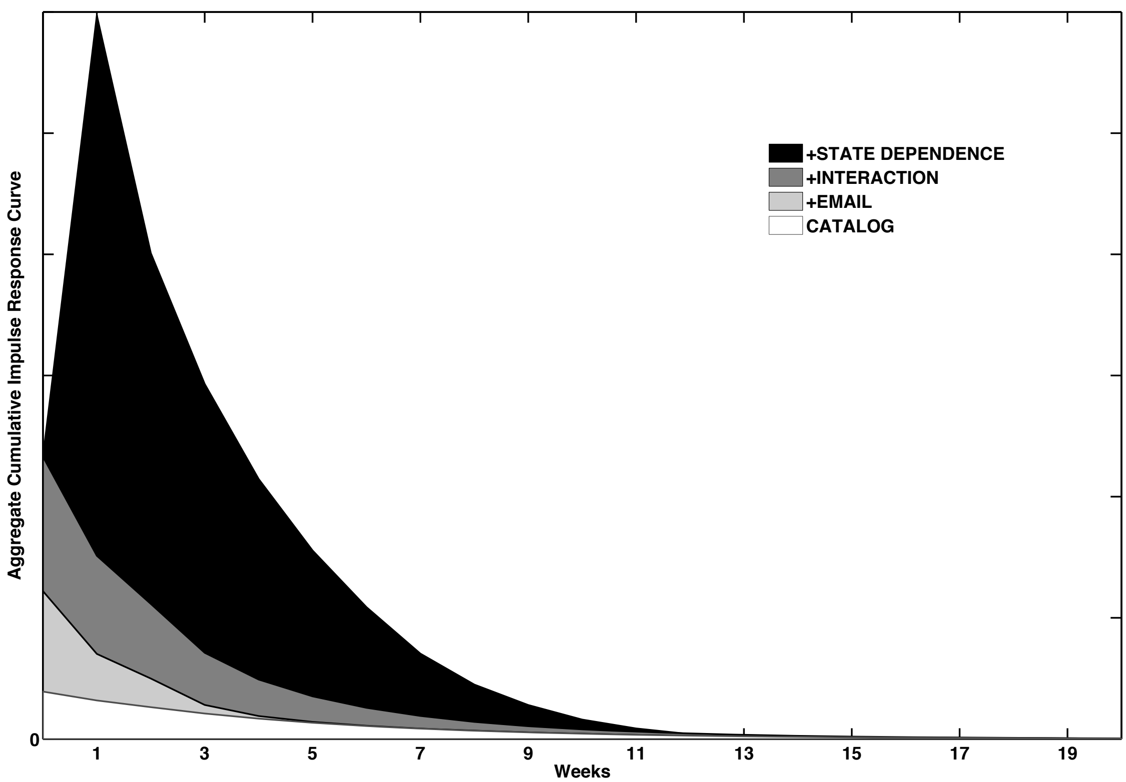

Figure 3: Comparison of the differential effect of each component in the model on the aggregated impulse response curve for an increase in exposures on both email and catalog for the full model.

report this for two different points in time, since the predicted CIR for a customer will vary over time depending on the customer's exposure to advertising and associated ad stock.) In the first row, we report the average model-predicted CIR across all customers; this represents the per-customer lift in sales that is obtained by exposing *every* individual to one additional email and catalog in the same week. The values in parentheses represent the variation across the population in the estimated CIR. As shown in the first row of Table 9, the expected response from sending a catalog and and email simultaneously is about \$0.04, but ranges from about \$0.01 to about \$0.35 per customer across customers.

The return on advertising can be increased substantially by targeting specific individuals based on their prior behavior. In the second row of Table 9 we show the results of targeting the top 10% of customers based on their *spending in the prior year*, a targeting strategy that is common in practice. This approach

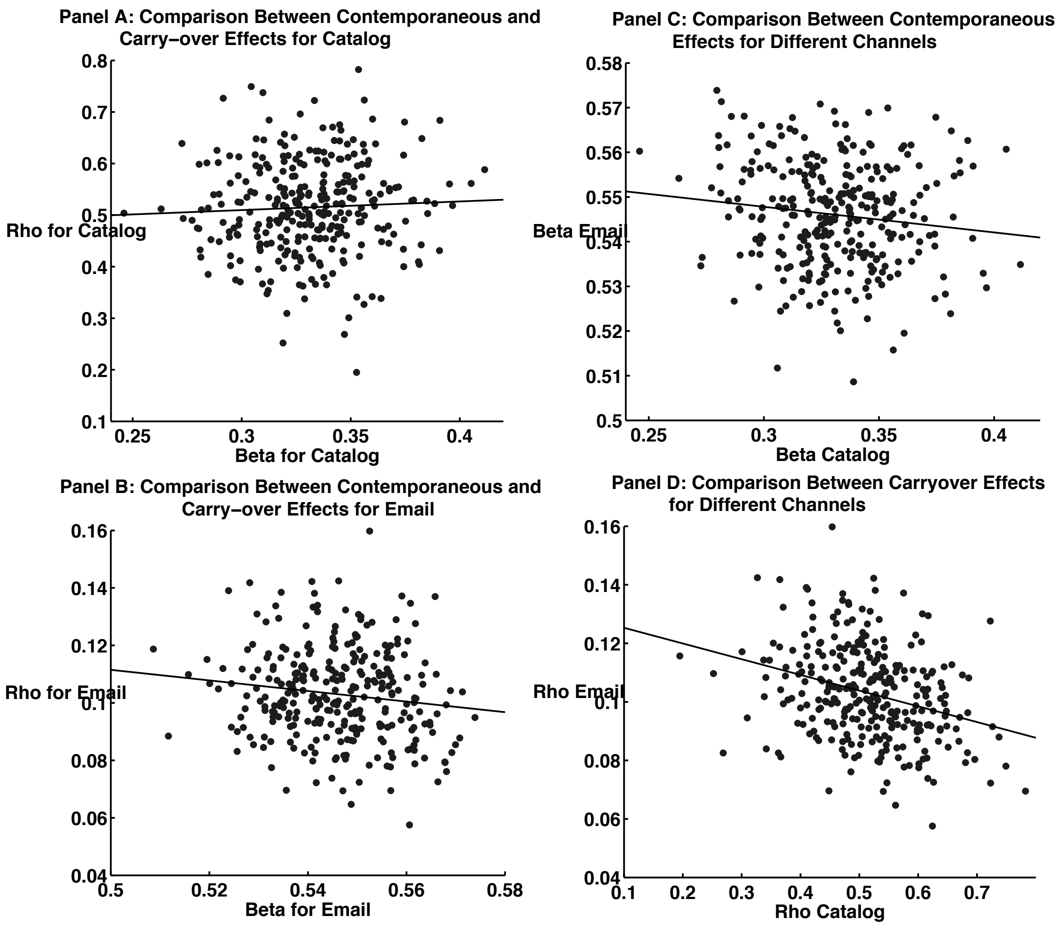

Figure 4: Comparison of posterior mean of the effectiveness and carry-over parameters across individuals for different channels.

substantially raises the average CIR for the targeted customers to about \$0.20 per customer. However, it is even more effective to target customers who have shown responsiveness to advertising in the past as measured by their model-estimated parameters and associated individual CIR. In the last row of Table 9, we show average per customer response that can be achieved when targeting the highest 10% *responsive* customers. Table 9 shows that the response nearly doubles to about \$0.37. Thus, there are clear advantages to using the proposed model to score customers for their responsiveness based on the CIR and then using that score to target advertising.

	at week 50	at week 101
No Targeting (all customers)	0.039 (0.010, 0.334)	0.042 (0.021, 0.373)
Targeting based on prior year's spending (top 10%)	0.188 (0.147, 0.402)	0.214 (0.196, 0.240)
Targeting based on predicted CIR (top 10%)	0.366 (0.296, 0.412)	0.382 (0.310, 0.482)

Table 9: Per customer cumulative impulse response for various customer targeting strategies. For each entry we report the average marginal response for the targeted group and the 2.5th and 97.5th quantiles across the targeted group.

3.4 Application to Other Datasets

As discussed above, we illustrate the broad applicability of the model by estimating it for three other data sets: two data sets collected in a single-source panel (advertising exposures and response for Brand X and Brand Y) and one data with with online display advertising. In Table 10, we present the population mean parameters for these other data sets, which give us some insight into the effectiveness of other types of advertising. For Brand X, we find that television and online advertising have roughly the same contemporaneous effect. For Brand Y, we find that television has a much more long-lasting effect than for Brand X. As in the direct marketing data sets, we find that the interaction effect and the state-dependence are substantial for both Brands X and Y, particularly the interaction between television and Facebook exposures for Brand Y.

Unlike the other five data sets, the online advertising data sets shows a much different pattern in the estimated parameters. Unsurprisingly, we find that the advertisements on automotive shopping sites are substantially more effective than those on general media sites. While we do find some synergy between the two channels, there is relatively little state dependence in the online data set, a finding that may relate to the well-studied stickiness or "frictionlessness" of online user behavior (Bucklin and Sismeiro, 2009) or simply that requesting a quote or searching inventory on a automotive website is something you only need to do once.

	Brand X	**Brand Y**	**Online Display**
$\bar{\mu}$	-9.19 $(-9.44, -8.95)$	-11.00 $(-11.38, -10.60)$	-8.88 $(-9.11, -8.63)$
$\bar{\beta}_{\text{TV}}$	0.18 (0.04, 0.32)	0.10 (0.01, 0.21)	
$\bar{\beta}_{\text{ONL}}$	0.19 (0.05, 0.34)		
$\bar{\beta}_{\text{FB}}$		0.20 (0.05, 0.34)	
$\bar{\beta}_{\text{GEN}}$			0.24 (0.08, 0.39)
$\bar{\beta}_{\text{SPE}}$			1.68 (1.52, 1.85)
$\bar{\beta}_{k\,:\,k'}$	0.27 (0.12, 0.42)	1.24 (1.08, 1.41)	0.35 (0.19, 0.52)
$\bar{\beta}_{S}$	2.03 (1.88, 2.18)	2.65 (2.48, 2.82)	0.44 (0.28, 0.61)
$\bar{\rho}_{\text{TV}}$	0.21 (0.08, 0.34)	0.49 (0.37, 0.62)	
$\bar{\rho}_{\text{ONL}}$	0.24 (0.10, 0.37)		
$\bar{\rho}_{\text{FB}}$		0.19 (0.06, 0.32)	
$\bar{\rho}_{\text{M}}$			0.22 (0.09, 0.34)
$\bar{\rho}_{\text{AS}}$			0.20 (0.08, 0.33)
$\bar{\rho}_{k\,:\,k'}$	0.23 (0.10, 0.36)	0.17 (0.05, 0.29)	0.23 (0.10, 0.36)
$\bar{\rho}_{S}$	0.17 (0.05, 0.30)	0.69 (0.56, 0.81)	0.47 (0.32, 0.61)

Table 10: Comparison of population parameter estimates for Brand X, Brand Y, and the Online Display data sets.

To further illustrate the value of the model across a wide variety of contexts, Table 11 presents the cumulative expected lift in response for the targeting strategies we discussed in the previous section. Across all five data sets, we see substantial increases in lift (on average an 80% increase across all five data sets) for targeting individuals based on advertising response versus targeting based on prior purchases.

In total, these results suggest that there can be remarkable differences in the advertising response for

	Brand A	**Brand C**	**Brand X**	**Brand Y**	**Onl. Display**
No Targeting (all customers)	0.015 (0.001, 0.134)	0.023 (0.001, 0.157)	0.043 (0.023, 0.080)	0.073 (0.042, 0.116)	0.057 (0.002, 0.257)
Targeting based on prior year's spending (top 10%)	0.061 (0.002, 0.462)	0.123 (0.058, 0.831)	0.130 (0.076, 0.233)	0.244 (0.154, 0.543)	0.207 (0.008, 0.46)
Targeting based on predicted CIR (top 10%)	0.179 (0.055, 0.697)	0.193 (0.070, 0.885)	0.171 (0.092, 0.274)	0.485 (0.213, 0.756)	0.343 (0.186, 0.64)

Table 11: Comparison of targeting strategies based on consumers predicted CIR for different brands. Note that a similar analysis for Brand B is presented in Table 9.

different channels. We find little evidence to suggest that broad generalizations can be made about online versus traditional mediums, about social media, etc. The only thing we find in common across all data sets is that there is consistent evidence for interactions between channels, suggesting that there is good reason for advertisers to use multi-channel advertising campaigns and that these campaigns should be coordinated. However, the magnitude of the main and interaction effects seems to be highly specific to each product (and probably varies over time depending on advertising copy), suggesting that advertisers should monitor advertising effectiveness for each product. With the rapid emergence of customer-level advertising response data, this goal is more attainable than ever and models like the one we propose here can be readily applied across a broad variety of contexts to monitor advertising effectiveness on an ongoing basis.

4 Model Extensions

In the previous sections, we proposed a model which can be applied broadly to different types of data to gauge consumers' response to different advertising channels and target individuals based on their responsiveness to advertising in a particular channel. In developing the model, we have focused on the features that are common across many data sets. In this section, we discuss several model extensions that can be used to tailor the approach to a particular context. These approaches include (a) incorporating saturation effects (b) addressing potential endogeneity biases (c) modeling competitive advertising effects and spillovers between brands and (d) incorporating alternative forms of state dependence. Our aim is to provide advice to those looking to implement our model, and so we discuss when these extensions might be most important in practice and propose several implementation choices that can mitigate the need for these extensions.

Saturation Effects. In our proposed model, consumers' response to advertising is a linear function of the number of advertising exposures. In aggregate models, researchers have found evidence of saturation and wear-in/out effects which are usually introduced using some non-linear function of the actual exposures, $f(X_{ikt})$(Bass et al. 2007, Dube, Hitsh and Manchanda 2005). It is certainly possible to extend this idea to the consumer-level, estimating saturation functions for each individual. For example, we could allow for a diminishing response to X_{ikt} with just one additional parameter as follows:

$$W_{ikt} = \rho_{it} W_{i,t1} + X_{ikt}^{\delta_{ik}} + \varepsilon_{ikt} \tag{13}$$

where $0 \leq \delta_{ik} \leq 1$. While conceptually simple, this would be computational burdensome requiring non-linear filtering techniques beyond the standard DLM filters employed in our estimation algorithm. Going well beyond this simple formulation, which can only account for within-period saturation, one might also consider employing a wear-out/restoration framework, like that of Braun and Moe (2013). However, we note that data sets that are very sparse in X_{ikt}, as is common, may result in extremely weak identification of saturation or wear-out effects.

Competitive Advertising. As discussed above, we have not considered competitive advertising effects because most of the readily-available consumer-level data do not include it. However, if data on competitive advertising is available, it could be readily incorporated into our model by adding a cross-effect term to the ad-stock equation to control for the competitors advertising:

$$W_{ikt} = \rho_i W_{it-1} + X_{ikt} + \gamma_k \tilde{X}_{ikt} + \varepsilon_{ikt} \tag{14}$$

where $\tilde{X}_{ikt}$ refers to the advertising exposures on the same channel for the competitive brand and γ_k represents the advertising spillover from another brand. A more comprehensive, yet computationally burdensome approach would be to simultaneously estimate stocks for each brand and model consumers' choice among brands (c.f., Bollinger, Cohen and Lai 2013). These extensions addressing competitive effects are left for future research, again not because they are conceptually difficult, but rather because these are specific to certain data structures which are not considered in this work.

Endogeneity Biases. In the proposed model, we assume that advertising and pricing are exogenous variables. As hinted in the introduction, since managers allocate advertising and make pricing decisions in an integrated fashion, it is reasonable to argue that both price and advertising could be treated as endogenous variables (Schweidel and Knox 2013) despite our use of granular individual-level data. Endogeneity of advertising can be addressed with instruments if they are available; unfortunately these are not common in advertising response studies (Shugan 2004). Alternatively, they can be addressed by a model of advertising targeting and using that in the empirical estimation (Manchanda, Rossi and Chintagunta 2004).

In this paper, we consider the case in which instruments or information about targeting are not necessarily available. Under this premise, we now argue that a stochastic form of goodwill as per the error term in equation (1b) can potentially moderate the degree of endogeneity bias, and hence provide a practical motivation for the use of stochastic stock transitions.

Consider a simple static advertising response model of the form

$$Y_{it} = \mu_i + \gamma_i X_{it} + \eta_{it} \tag{15}$$

If we have endogeneity, $Cov\,(X_i, \eta_i) \neq 0$, then it is well-known (see Greene, 2008) that the least squares estimate of γ_i, the contemporaneous effect, is biased where the bias is given by

$$Bias\,(\hat{\gamma}_i) = \frac{Cov\,(X_{it}, \eta_{it})}{Var\,(X_{it})} \tag{16}$$

In other words, in the classic static advertising response model, if exposures are positively (negatively) correlated with the purchasing shocks, possibly due to an omitted variable that is correlated with advertising, then we are (under-) over-estimating the contemporaneous effects of advertising.

We now extend the same reasoning to a dynamic ad-stock model of the form:

$$\begin{aligned} Y_{it}^{*} &= \mu_i + \beta_i W_{it} + \eta_{it} \\ W_{it} &= \rho_i W_{i,t-1} + X_{it} + \varepsilon_{it} \end{aligned} \tag{17}$$

where $\varepsilon_{it} \perp \eta_i$. Assume again that $Cov\,(X_i, \eta_i) \neq 0$, then in this case

$$Bias\left(\hat{\beta}_i\right) = \frac{Cov\,(W_{it}, \eta_{it})}{Var\,(X_{it})} \tag{18}$$

$$= \left(\frac{1-\rho_i^2}{1-\rho_i}\right) \frac{Cov\,(X_{it}, \eta_{it})}{\left(\sigma_\varepsilon^2 + Var\,(X_{it})\right)} \tag{19}$$

Formula (19) shows that there are two contrasting forces that can either increase or decrease the bias in the estimation of the contemporaneous advertising effect: the carry-over parameters increase the bias (although this increase is negligible when ρ_i is small) while the error term in the ad-stock equation (which is orthogonal to η_{it}) moderates the bias.

To compare the relative biases under the simple static advertising response versus a dynamic ad-stock model one can compute the "bias ratio":

$$\text{Bias Ratio}\left(\rho_i, \sigma_\varepsilon^2\right) = \frac{Bias\left(\hat{\beta}_i\right)}{Bias\left(\hat{\gamma}_i\right)} = \left(\frac{1-\rho_i^2}{1-\rho_i}\right) \frac{Var(X_{it})}{\left(\sigma_\varepsilon^2 + Var\left(X_{it}\right)\right)^2}. \tag{20}$$

This does not depend on the unobservable $Cov\left(X_{it}, \eta_{it}\right) \neq 0$ and allows us to tabulate summary statistics, individual by individual for each channel for all six datasets considered in this research. Assuming that the true data generating process is given in equation (17), which nests equation (15), we consider two scenarios: (a) Bias Ratio$\left(\rho_i, \sigma_\varepsilon^2 = 0\right)$ and (b) Bias Ratio$\left(\rho_i, \sigma_\varepsilon^2\right)$. That is, in Table 12 we compute the bias ratio under a deterministic ad-stock model, whereas in Table 13, we show the bias ratio under the unconstrained model. These numbers suggest a potential advantage to fitting the more general model, which includes carry-over and an error term in the ad-stock.

	Brand A		**Brand B**		**Brand C**
Catalog	129.83%	Catalog	150.74%	Catalog	124.60%
Email	129.46%	Email	151.22%	Email	124.70%
	Brand X		**Brand Y**		**Online Dis.**
TV	137.17%	TV	119.46%	Media	122.10%
FB	136.85%	Online	119.41%	Shopping	121.72%

Table 12: Average posterior Bias Ratio$\left(\rho_i, \sigma_\varepsilon^2 = 0\right)$ channel by channel for the six data sets.

	Brand A		**Brand B**		**Brand C**
Catalog	12.24%	Catalog	14.84%	Catalog	13.76%
Email	65.28%	Email	7.32%	Email	6.85%
	Brand X		**Brand Y**		**Online Dis.**
TV	24.58%	TV	37.86%	Media	12.30%
FB	3.08%	Online	2.94%	Shopping	31.91%

Table 13: Average posterior Bias Ratio$\left(\rho_i, \sigma_\varepsilon^2\right)$ channel by channel for the six data sets.

In a related “note”, there is also another feature of the model we propose that has the potential of moderating endogeneity biases. This deals with the problem of temporal aggregation. Consider aggregation of exposures and purchases with step h. To exemplify, consider the case in which we aggregate data from daily to weekly observations, so that $h = 7$. This implies that if $X_i \sim Poisson\left(\lambda_i\right)$ as in Manchanda, Rossi and Chintagunta (2004), then $X_i^h = \sum_{i=1}^{h} X_i \sim Poisson\left(h\lambda_i\right)$. Denote by β^h the effectiveness parameter in the model with h steps of aggregation, then it is easy to show that:

$$Bias\left(\hat{\beta}_i^h\right) = \left(\frac{1-\rho_i^{2h}}{1-\rho_i^h}\right)\frac{Cov\left(X_{it}, \eta_{it}\right)}{\left(h\sigma_\varepsilon^2 + Var(X_{it})\right)} \tag{21}$$

Thus temporal aggregation is another factor possibly moderating the endogeneity bias. The intuition is that the variance of a Poisson model grows linearly with the aggregation step, while the variance of the error in the stock equation grows quadratically. Also by property of the AR(1) process, the decay magnitudes decrease exponentially with the aggregation step. In summary, this suggests a trade-off between the benefits of analyzing the model at a finer time-scale and potential endogeneity biases driven by contemporaneous correlation between exposures and the purchasing shocks.

Alternative Forms of State-Dependence. In this work, we have adopted a form of state dependence that allows for prior periods' incidence of purchase ($\mathbf{1}\left[Y_{i,t-1} > 0\right]$) to affect future purchases with an exponential decay rate that is estimated. Our approach nests the common form of state dependence where only prior period purchase incidence can affect next period sales. We could certainly extend the framework to allow for other forms of state dependence such as "habit formation" where there is a carry-over effect of Y_{it}^* or a brand loyalty variable as in Guadagni and Little (1983). Since our focus is in measuring ad response, we leave the comparison of these different forms of state dependence to future research and, from a practical point of view, we have noted that different forms of state-dependence may not be observationally distinguishable in an individual-level model with sparse data.

5 Conclusion

In this paper, we have developed an integrated approach to measuring advertising effectiveness with consumer-level advertising response data. Our model accounts for (i) multiple channels and their interactions (ii) dynamic advertising effectiveness (iii) sparsity in the response variable (iv) heterogeneity across individuals and (v) state-dependence effects. The model is grounded in the traditional ad-stock literature, but we have employed a state-space formulation allowing for efficient and scalable computation of the latent stock variables. Our model addresses many key research questions that were originally raised by Little (1979) by showing how it is possible to construct from individual-level exposures, aggregate measures of advertising response that can be used to plan the marketing mix. We have demonstrated the broad applicability by applying the model to six different data sets.

Our intent was to develop a tool that would be useful to practitioners across many contexts, and so our specification is intentionally simple. We only rely on the advertiser having data on their own advertising linked at the customer level to their own purchases, which, as we have discussed, is rapidly becoming available for many different types of advertising. We have focused on using the data to identify how much lift is produced by each advertising channel and which consumers are most responsive to advertising. However, we recognize that the model could be extended in many different directions depending on what data is available and which decisions the advertiser wants to focus on. For example, if the data included information of the specific advertising copies, one could estimate the effects of individual advertising creatives (Braun and Moe 2013) or even the interaction between the copy and the advertising channel. One could develop models that allow advertisers to combine data that is observed at different scales, e.g. weekly direct mail exposures and daily online advertising. If there was more information about customers browsing behavior, one could develop more complex models of how advertising affects customers as they move through the purchase funnel (Abhishek, Fader and Hosanagar 2012). With sufficient variation in the advertising exposures, one could also allow for non-contemporaneous interactions between advertising channels (Bollinger, Cohen and Lai 2013). What we have attempted to show here is that whichever way data evolves, the framework developed in this work is a flexible start to a decision support tool to manage multi-channel advertising.

References

[1] K. H. Abhishek, K. Hosanagar, and P. Fader. Aggregation bias in sponsored search data: The curse and the cure. Working paper available at `http://ssrn.com/abstract=1490169`, CMU, 2012.

[2] B. Ataman, H. J. van Heerde, and C. F. Mela. The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research*, 47(5):866–882, 2010.

[3] F. Bass, N. Bruce, S. Majumdar, and B. Murthi. Wearout effects of different advertising themes: A dynamic bayesian model of the advertising-sales relationship. *Marketing Science*, 26(2):179–195, 2007.

[4] B. K. Bollinger, M. A. Cohen, and J. Lai. Measuring asymmetric persistence and interaction effects of media across platforms. Working paper available at `http://ssrn.com/abstract=2342349`, NYU, 2013.

[5] D. Bowman and H. Gatignon. *Market response and marketing mix models*. Number 14. Now Publishers, 2010.

[6] M. Braun and W. Moe. Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science*, 32(5):753–767, 2013.

[7] S. Broadbent. Modeling with ad-stock. *Journal of the Market Research Society*, 26:295–312, 1984.

[8] R. E. Bucklin and C. Sismeiro. Click here for internet insights: Advances in clickstream data analysis. *Journal of Interactive Marketing, Volume*, 23(1):35–48, 2009.

[9] S. Chib. Bayes inference in the tobit censored regression model. *Journal of Econometrics*, 51:79–99, 1992.

[10] P. Danaher, A. A. Bonfrer, and S. S. Dhar. The effect of competitive advertising interference on sales for packaged goods. *Journal of Marketing Research*, 45(2):211–25, 2008.

[11] P. J. Danaher and T. S. Dagger. Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. *Journal of Marketing Research*, 50(4):517–534, 2013.

[12] J. Deighton, C. M. Henderson, and S. A. Neslin. The effects of advertising on brand switching and repeat purchasing. *Journal of Marketing Research*, 31:28–43, 1994.

[13] M. G. Dekimpe and D. M. Hanssens. Time-series models in marketing: Past, present and future. *International Journal of Research in Marketing*, (17):183–193, 2000.

[14] F. X. Diebold. *Elements of Forecasting*. South-Western Publishing, 2006.

[15] I. M. Dinner, H. J. Van Heerde, and S. Neslin. Driving online and offline sales: The cross-channels effects of digital versus traditional advertising. Technical report, working paper, 2011.

[16] J. A. Duan, L. McAlister, and S. Sinha. Reexamining bayesian-model-comparison evidence of cross-brand pass-through. *Marketing Science*, 30(3):550–561, 2011.

[17] J. P. Dubé, G. Hitsch, and P. Manchanda. An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, 3(2):107, 2005.

[18] J. P. Dubé, G. Hitsch, and P. E. Rossi. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445, 2010.

[19] A. Gelfand and S. K. Ghosh. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11, 1998.

[20] W. Greene. *Econometric Analysis*. Pearson Prentice Hall, Upper Saddle River, 6th edition, 2008.

[21] P. M. Guadagni and J. D. C. Little. A logit model of brand choice calibrated on scanner data. *Marketing Science, Volume*, 2(3):203–238, 1983.

[22] J. J. Heckman. Heterogeneity and state dependence. *In "Studies in Labor Markets", University of Chicago Press*, pages 91–140, 1981.

[23] J. J. Heckman. Statistical models for discrete choice data. *In C. Manski, D. McFadden, eds, "Structural Analysis of Discrete Data with Applications", MIT Press, Cambridge, MA*, pages 179–195, 1981.

[24] J. Huang, M. Leng, and L. Liang. Recent developments in dynamic advertising research. *European Journal of Operations Research, Volume*, 220(3):591–609, August 2012.

[25] D. W. Jorgenson. Rational distributed lags. *Econometrica*, 32:135–149, 1966.

[26] R. J. Kent and D. A. Schweidel. Introducing the ad ecg: How the set–top box tracks the lifeline of television. *Journal of Advertising Research*, 51(4):586–593, 2011.

[27] L. M. Koyck. *Distributed lags and investment analysis*. North-Holland, Amsterdam, 1954.

[28] J. D. C. Little. Aggregate advertising model: The state of the art. *Operations Research*, 27(4):629, 1979.

[29] P. Manchanda, J. P. Dubé, K. Y. Goh, and P. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108, 2006.

[30] P. Manchanda, P. E. Rossi, and P. K. Chintagunta. Response modeling with non-random marketing mix variables. *Journal of Marketing Research*, 41(3):467–478, 2004.

[31] L. McAlister, R. Srivastava, J. Horowit, M. Jones, W. Kamakura, J. Kulchitsky, B. Ratchford, G. Russel, F. Sultan, and T. Yai. Incorporating choice dynamics in models of consumer behavior. *Marketing Letters*, 2(3):241–252, 1991.

[32] J. F. McDonald and R. A. Moffitt. The uses of tobit analysis. *The Review of Economics and Statistics*, 62(2):318–321, 1981.

[33] P. A. Naik and K. Peters. A hierarchical marketing communications model of online and offline media synergies. *Journal of Interactive Marketing*, 23(4):288–299, 2009.

[34] P. A. Naik and K. Raman. Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research*, 40(3):375–388, 2003.

[35] P. A. Naik, K. Raman, and R. S. Winer. Planning marketing-mix strategies in presence of interactions. *Marketing Science*, 24(1):25–34, 2005.

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

[37] J. H. Pedrick and F. S. Zufryden. Evaluating the impact of advertising media plans: A model of consumer purchase dynamics using single–source data. *Marketing Science, Volume*, 2(10), May 1991.

[38] A. E. Raftery and S. M. Lewis. Implementing markov chain monte carlo. In *Markov chain monte carlo in practice (W.R. Gilks, D.J. Spiegelhalter and S. Richardson, eds.)*, pages 115–130. Chapman and Hall, London, 1996.

[39] P. E. Rossi and G. M. Allenby. Bayesian statistics and marketing. *Marketing Science*, 22(3):304–328, 2003.

[40] P. E. Rossi, G. M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. John Wiley and Sons, New York, 2005.

[41] D. A. Schweidel and G. Knox. Incorporating direct marketing activity into latent attrition model. *Marketing Science*, 32(3):471–487, 2013.

[42] P. B. Seetharaman. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science*, 23(2):263–271, 2004.

[43] K. Shirley, D. Small, K. Lynch, S. Maisto, and D. Oslin. Hidden markov models for alcoholism treatment trial data. *Annals of Applied Statistics*, 4:366–395, 2010.

[44] S. M. Shugan. Endogeneity in marketing decision models. *Marketing Science*, 23(1):1–3, 2004.

[45] G. J. Tellis. *Advertising and Sales promotion strategy*. Reading, MA, 1998.

[46] G. J. Tellis. *Effective Advertising: Understanding when, how, and why advertising works*. Sage Publications, Thousand Oaks, 2004.

[47] G. J. Tellis and T. Ambler. *Handbook of Advertising*. Sage Publications, London, UK, 2007.

[48] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer–Verlag, NewYork, 2nd edition, 1999.

[49] K. C. Wilbur. A two-sided, empirical model of television advertising and viewing markets. *Marketing Science*, 27(3):356–378, 2008.

A Estimation Algorithm

We draw samples from the posterior of all parameters using a Gibbs sampler, which includes a combination of closed-form draws, Metropolis-Hastings draws and a DLM filtering procedure. After initialization, parameters are drawn in four blocks:

1. Sample W_{ikt} for each individual using DLM relationships
2. Sample Y_{it}^* for each individual from a truncated normal distribution based on the observed Y_{it} (Chib 1992)
3. Sample $\beta_i, \rho_i, \sigma_\eta^2, \sigma_\varepsilon^2$ for each individual using Metropolis Hastings updates
4. Sample $\bar{\beta}, \bar{\Sigma}_\beta$, $\bar{\rho}$, and $\bar{\Sigma}_\rho$ using standard conjugate draws (c.f. Rossi, Allenby and McCulloch 2003 pp 72-73)

We provide details of the initialization and steps 1 and 3 below.

A.1 Initialization

In order to initialize the relevant parameters in the model we estimate a pooled model, i.e., a model with common parameters for all individuals, where the latent stocks for each period are constructed explicitly based on the ten prior periods. Specifically, we estimate the model:

$$\begin{aligned} \mathbf{Y}_t^* &= \mu + \sum_{k=1}^{K} \beta_\mathbf{k} \sum_{j=0}^{9} \rho_k^{j-1} \mathbf{X}_{t-j,k} + \eta_\mathbf{t} \\ \mathbf{Y_t^*} &= \mathbf{Y_t} \text{ if } \mathbf{Y_t} > 0, \mathbf{Y_t}^* < 0 \text{ if } \mathbf{Y_t} = 0 \end{aligned} \tag{22}$$

where the bold fonts denote the vectorized observations "stacked" individual by individual. (For clarity, we suppress the notation for the interaction terms and state dependence.) We obtain estimates $\hat{\beta}_k$, $\hat{\rho}_k$ and and $\hat{\sigma}_\eta$ by constrained maximum likelihood. These MLE estimates together with the predicted $\hat{Y}_t^* < 0$ are used used to initialized the sampler.

The latent stocks W_{ik0} are initialized at the "steady-state" implied by the maximum likelihood estimates. Namely, for each individual denote by $\bar{X}_{ik}$ the in-sample mean of the number of exposures for channel k and individual i. Then we initialize $W_{ik0} = \bar{X}_{ik}/\,(1 - \hat{\rho}_k)$. Note that these starting values for the stocks are updated in each pass of the Gibbs sampler by means of smoothing densities derived using the DLM relationships in step 2 of the algorithm.

Finally, Σ_ε is initialized at the identity matrix.

A.2 Step 1. Update of the Latent Stocks

Recall the vectorized system of equations determining the dynamics of the stocks in the latent space:

$$\begin{aligned} Y_t^* &= \mu + \beta \mathbf{W}_t + \eta_t \\ \mathbf{W}_t &= \rho \mathbf{W}_{t-1} + \mathbf{X}_t + \varepsilon_t \end{aligned}$$

(We suppress the index for the individual, i, for clarity.) The DLM relationships determine a system of recursive densities based on a set of sufficient statistics for the predicted (and corrected) mean and variance of the stocks: $\mathbf{W}_t \sim N\left(\mu_{Wt}^t, V_{Wt}^t\right)$, where $\mu_{W_t}^t$, $V_{W_t}^t$ represent the set of sufficient statistics based on all the information available up to time t. We also denote with a subscript $t+1$, the predicted or estimated sufficient statistics as traditional in filtering studies (West and Harrison, 1997). These statistical summaries are derived in two steps commonly referred as the Forward Filtering algorithm (FF). For time $t = 1, ..., T-1$

- Forward Filtering:

$$\begin{aligned}
\mu_{W_t}^{t+1} &= \rho\mu_{W_t}^t + \mathbf{X}_t \\
V_{W_t}^{t+1} &= \rho V_{W_t}^t \rho' + \sigma_\varepsilon^2
\end{aligned}$$

$$\begin{aligned}
k_{t+1} &= V_{W_t}^{t+1}\beta'\left(\beta V_{W_t}^{t+1}\beta + \sigma_\eta^2\right)^{-1} \\
\mu_{W_{t+1}}^{t+1} &= \mu_{W_t}^{t+1} + k_{t+1}\left(Y_{t+1}^* - \mu - \beta\mu_{W_t}^{t+1}\right) \\
V_{W_{t+1}}^{t+1} &= V_{W_t}^{t+1} - k_{t+1}\left(\beta V_{W_t}^{t+1}\beta + \sigma_\eta^2\right)k_{t+1}'
\end{aligned}$$

Similarly, once time T is reached, it is possible to "smooth" the densities back in time. Importantly, this provides a way to estimate "the best" (in mean squared sense) prediction of the initial conditions of the stock equations: these updates are also characterized by a set of sufficient statistics denoted as $\mu_{W_{it}}^T$ and $V_{W_{it}}^T$ where the T superscript points out that the smoothing update is based on all the information in the sample and t in this case runs from $T-1$ to 0. Thus from time T consider the most recent set of sufficient statistics $\mu_{W_T}^T$, $V_{W_T}^T$ from the forward filtering step and move backwards in time to obtain:

- Backward Sampling (BS)

$$\begin{aligned}
g_t &= V_{W_t}^t\beta'\left(V_{W_t}^{t+1}\right)^{-1} \\
\mu_{W_t}^T &= \mu_{W_t}^t + g_t\left(\mu_{W_{t+1}}^T - \mu_{W_t}^t\right) \\
V_{W_t}^T &= V_{W_t}^t - g_t\left(V_{W_t}^{t+1} - V_{W_{t+1}}^T\right)g_t'
\end{aligned}$$

A detailed treatment of the derivations leading to the above can be found in West and Harrison (1997) and similarly in Bass et al. (2007). Armed with the set of sufficient statistics derived from the FF and BS steps, $\mu_{W_{it}}^T$, $V_{W_{it}}^T$ we can then sample the latent stocks for each individual:

$$\mathbf{W}_t \sim N\left(\mu_{W_t}^T, V_{W_t}^T\right).$$

A.3 Step 3. Update the Individual-level Parameters

We draw the individual-level parameters using a Metropolis algorithm which uses candidate sampling distributions that are customized to the unit-level likelihoods. We use a "fractional likelihood" approach as in Rossi, Allenby and McCulloch (2005 pp 135) to set the proposal density for each individual:

$$L_i^* = L_i \cdot \bar{L}^\alpha \tag{23}$$

where $\bar{L}^\alpha$ is the pooled likelihood described in the initialization step. The weight α is set to $T/(2 * N)$ so that it does not dominate the unit-level likelihood. The pooled likelihood has the purpose of regularizing the likelihood for the units that due to the high sparsity do not have a local maximum. We can use the maximum and Hessian of this likelihood to construct a proposal for each individual as follows: let $\hat{\beta}_i$ be the set of indivdiual-level parameters that maximizes the likelihood in equation (23) and let $\hat{V}_i = -\left.\frac{\partial^2 L_i^*}{\partial\beta\partial\beta'}\right|_{\beta=\hat{\beta}_i}$. These can then be combined with the priors presented in Section 2.1 to form a Metropolis proposal distribution. The update for the individual parameters then uses a standard M-H update with an additional rejection step to ensure that the ρ_i coefficients are sampled from the set $[0, 1)$.

Finally, we note that the individual level Tobit I likelihood is potentially invariant to sign transformations. Specifically, for any $(\beta, \mathbf{W})$ there is a sign transformation $(-\beta, -\mathbf{W})$ providing the same value of the likelihood. This will manifest itself while filtering W_{it} using the DLM relationship described above, by means of “reflected paths” at zero. In practice we have verified that the situation arises when the exposures X_{ikt} are sparse. To overcome the potential unidentifiability due to sign transformation, we suggest either post-processing the draws as in Rossi, Allenby and McCulloch (2005, Ch. 4) or (equivelently) restricting the βs over the positive real line for those individual having sparse exposures.