MANAGEMENT SCIENCE
Articles in Advance, pp. 1–23
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2016.2451

Measuring Multichannel Advertising Response

Daniel Zantedeschi
Department of Marketing and Logistics, Fisher College of Business, The Ohio State University, Columbus, Ohio 43210, zantedeschi.1@osu.edu

Eleanor McDonnell Feit
Drexel University, Philadelphia, Pennsylvania 19104, efeit@drexel.edu

Eric T. Bradlow
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, ebradlow@wharton.upenn.edu

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
History: Received July 7, 2015; accepted January 8, 2016, by Pradeep Chintagunta, marketing. Published online in *Articles in Advance*.

1. Introduction

There are an increasing number of platforms that enable advertising that can be targeted to individual customers, including direct marketing, online advertising, social media, and addressable TV. From a measurement perspective there are two key advantages of targeted advertising platforms. First, they collect consumer-level exposure data, and so 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 the advertising at each point in time. When these advertising exposure data are linked to data on individuals' purchases, advertisers can obtain consumer-level advertising response data at a very low cost. While it has long been recognized that linking individual consumers' advertising exposures to those same consumers' purchase behavior (or any other desired behavior) would be helpful in measuring advertising response (Little 1979, Tellis 2004), until recently those data were only available through relatively expensive "single-source" panels. Consequently, the literature on measuring advertising response has focused on methods that use aggregate advertising and sales data (see Dekimpe and Hanssens 2000).

A second key feature of modern advertising platforms is that they allow marketers to target individual customers. This capability allows advertisers to create control groups, sometimes called "holdout groups" where the marketer randomly assigns a fraction of the targeted customers to be "held out" from exposure to a particular advertisement. Because treatment is assigned randomly, a simple comparison of the purchases made subsequent to the ad exposure between the treated and the control groups gives an estimate of the causal effect of exposure to the ad for the targeted population. It is increasingly common for direct and online marketers to create randomized holdout groups consisting of a small fraction of the population that was targeted to receive each ad (e.g., Sahni et al. 2016). Our data are somewhat unique, however, in that we observe multiple catalog and email exposures targeted to the same customers over a two-year period, with different randomly selected holdout groups for each exposure, thus creating exogenous variation in the sequence of advertisements shown to each customer.

In Section 2, we illustrate how each experiment in the data can be analyzed independently to estimate the treatment effect. The analysis of individual experiments shows clear positive effects of catalog marketing on subsequent purchases and somewhat weaker evidence of positive effects of emails on sales. However, holdout groups are typically small (due to the opportunity cost of not advertising to some customers), and, as we will illustrate with our data, simple hypothesis tests on the effect of individual ad exposures are often underpowered because of the small size of the holdout group. In this paper, we move beyond simple hypothesis testing on individual ads or campaigns and employ a hierarchical modeling framework that allows us to pool information on customers' response over multiple ad exposures to obtain more precise estimates of the effect of ad exposures and how those effects decay after the exposure.

Pooling randomized holdout campaigns to increase power has also been proposed by Sahni et al. (2016), who pool data across repeated holdout experiments where each experiment includes a different set of customers using a meta-analysis approach. The panel structure of our data, with experiments that are repeatedly conducted with the same customers, affords the opportunity to model advertising response at the customer level. That is, our estimates of advertising response are informed by comparing subsequent purchases for customers who are exposed to an advertisement at a particular point in time versus those who were randomly selected to not be exposed. This allows us to estimate short- and long-term advertising response for each customer and for each channel, while avoiding the collinearity and endogeneity problems that often occur in aggregate observational data. Our model and focus on customer-level data across multiple marketing channels places us among recent academic papers that focus on the "multichannel advertising attribution" problem (Li and Kannan 2014, Bollinger et al. 2013). However, to our knowledge, we are the first to estimate a multichannel advertising response model using data with randomized control groups.

Our individual-level advertising response function follows in the tradition of the ad-stock literature (Nerlove and Arrow 1962, Jorgenson 1966) and is linked to purchases through a Tobit I structure, to accommodate the relative infrequency of purchases. (While we focus on purchases measured in dollars as the ultimate measure of advertising response, it is straightforward to extend the model to other measures of advertising response, such as quantity purchased, website visits, sign-ups, conversions, new subscriptions, etc.) The ad-stock structure allows us to relate ad exposures to subsequent sales with just two parameters for each channel (estimated for each individual): the first characterizes the contemporaneous effect of advertising and the second relates to the carryover of advertising response 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 further extend the ad-stock framework to allow for interactions among the different channels, i.e., "synergies" (Naik et al. 2005, Naik and Peters 2009). Using a hierarchical Bayesian framework, advertising response and interaction parameters are all estimated for each consumer, allowing us to score and target consumers based on their predicted cumulative response to an advertisement on a particular channel.

As we discuss, the exogenous variation produced by the randomized holdouts resolves some, but not all, of the potential endogeneity issues that can arise when modeling multichannel advertising response over repeated ad exposures. In particular, if all variation in exposures is due to randomized holdouts, as is the case in our data, then there is no possibility for endogeneity induced by systematic differences between the customers in the treated and the exposed groups, i.e., selection bias. Without randomized holdouts, there is a high potential for selection either due to the advertiser targeting high-propensity users (Manchanda et al. 2004) or because interest in the product is correlated with exposure, e.g., "activity bias" (Lewis et al. 2011). While the potential bias associated with exposure to advertising in periods of high demand (known to the manager but not observable by the researcher) is not directly addressed via randomized holdouts, we demonstrate that for our data the highly regular timing of ad campaigns in both channels mitigates this endogeneity threat as well.

Using a direct marketing data set describing the catalogs and emails sent by a retailer, we show that the total response to a catalog is greater than the response to an email and is spread out over a more extended period of time (approximately 30 days versus 15 days), which has important implications for whether and how often the advertiser should send catalogs or emails. Our estimates of advertising response by channel can be compared directly to costs to assess advertising ROI. Further, by using customer-level data, we are also able to estimate response parameters for each customer, and we show, by means of a counterfactual scenario, how targeting the most responsive customers can lead to increases in the returns to advertising of approximately 70% versus the common customer relationship management (CRM) practice of targeting customers based on past purchases (i.e., recency, frequency, and monetary value (RFM) targeting), irrespective of their prior advertising responsiveness.

To summarize, building on the traditional ad-stock framework, we develop a model that can be used to exploit consumer-level advertising response data with randomized holdouts to gauge advertising response by channel for each consumer. We apply the model to a direct marketing data set to estimate the effect of catalog and email mailings on subsequent purchases for each customer. These estimates can be compared with costs to determine the ROI of mailing on a particular channel or to a particular customer. In the conclusion section, we discuss the challenges of applying this approach in other advertising platforms, such as online or addressable television advertising, where potential customers exercise more influence over how many opportunities the advertiser has to serve an ad to them. We believe these challenges are surmountable and we hope that this paper motivates direct marketers and other advertisers to find ways to make advertising measurement and randomized holdouts a regular feature of their advertising campaigns.

1.1. Related Literature

Our work lies at the intersection of two distinct approaches to measuring advertising effects: field experiments typically analyzed by nonparametric methods and econometric modeling with observational data. Field experiments have recently become a popular tool for measuring advertising effects, primarily because they can be used to estimate the causal effect of advertising, often using relatively simple analysis methods. While several researchers have described geographic experiments that compare treated and control markets (Eastlack and Rao 1989, Blake et al. 2015, Kalyanam et al. 2015), our work is most closely related to those that use a randomly selected group of customers to form the control group. The idea of randomizing advertising exposures at the customer level goes back at least as far as Lodish et al. (1995a), who measure the effect of television advertising on purchases, using a complex split-cable system that allowed them to randomize treatment at the household level with purchases measured using an expensive scanner panel. More recently, as platforms that allow targeting have emerged, a number of researchers have utilized randomized control groups on a variety of newer digital channels. Goldfarb and Tucker (2011) randomize users' exposure to display advertising to gauge the effect on purchase intent (measured using a survey); Hoban and Bucklin (2014) randomize exposure to display advertising to measure the effect on website visits; Sahni (2015a, b) randomizes exposure to search advertising to measure the effect on seeking more detailed product information; and Sahni et al. (2016) randomize promotional offers in emails to measure the effect on purchases. Our work stands out from other field experiments because (1) we use field experiments conducted across multiple channels

Table 1 **Examples of Papers Using Alternative Paradigms for Measuring Advertising Response**

	Field experiments	Observational data
Market level		
Single channel	Eastlack and Rao (1989)	Little (1979)
	Blake et al. (2015)	Broadbent (1984)
	Kalyanam et al. (2015)	Danaher et al. (2008)
Multichannel		Naik and Raman (2003)
		Naik and Peters (2009)
		Ataman et al. (2010)
		Dinner et al. (2014)
Customer level		
Single channel	Lodish et al. (1995a)	Tellis (1998)
	Goldfarb and Tucker (2011)	Manchanda et al. (2006)
	Hoban and Bucklin (2014)	Braun and Moe (2013)
	Sahni et al. (2016)	
	Lewis and Rao (2015)	
	Barajas et al. (2016)	
	Johnson et al. (2016)	
Multichannel	This work	Danaher and Dagger (2013)
		Li and Kannan (2014)
		Bollinger et al. (2013)
		Xu et al. (2014)

Note. The table shows paradigms including field experiments versus econometric modeling with observational data, single channel versus multichannel, and market level versus customer level.

and (2) the panel structure of our data, where we observe advertising exposures and purchases for the same customers over time, allows us to measure not just immediate advertising effects, but the decay of advertising effects over time, while controlling for individual differences in baseline purchase propensity. Table 1 illustrates the position of our work within this literature.

Further, although many of the previously cited papers find significant positive effects of advertising, one of the ongoing issues with field experiments in advertising has been findings of nonsignificant effects of advertising on purchases (Lodish et al. 1995b, Lewis and Rao 2015), presumably because advertising's effect on purchases is small and the sample sizes used in reported experiments are too small to detect these effects. For instance, Lewis and Rao (2015) describe display advertising experiments with samples reaching into the millions, which are still too "small" to obtain reasonable statistical power. This suggests a gap in the advertising effectiveness literature between large-scale experiments and parametric modeling. Our parametric modeling approach attempts to "fill this gap" by pooling information across a series of randomized controlled experiments, conducted over two years using the same customers. By pooling across this two-year period, we find significant causal effects of advertising on purchases and are able to estimate the decay of advertising response over time.[1]

[1] As noted by one of the referees, Lewis and Rao (2015) find a very modest effect of display advertising intended to attract new

Our work also builds on the long tradition of using econometric models with observational data to measure advertising effects. We do not attempt to provide a comprehensive overview of the literature; see Tellis (2004) and Tellis and Ambler (2007) for comprehensive reviews.

Aggregate advertising response models are one of the earliest tools proposed in marketing science (Little 1979), and very shortly thereafter, researchers extended the models to account for advertising across multiple-channels, establishing marketing response modeling and marketing mix optimization as key tools 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 et al. 2010) and whether there are advertising synergies (i.e., interactions) between channels (Naik and Raman 2003, Danaher et al. 2008, Naik and Peters 2009). While these are just a few representative papers, there is a long literature on understanding multichannel 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 1980s, and this work tended to focus on just one channel: television (Pedrick and Zufryden 1991, Deighton et al. 1994, Tellis 1998). 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 advertising exposures and purchases are regularly tracked at the cookie level. So, most of the work on advertising response with individual-level data has focused on a single channel, largely due to data limitations.

More recently, as individual-level, multichannel advertising data have become available, research has started to emerge that focuses on using these data to measure ad response. Danaher and Dagger (2013) develop a clever measurement strategy for collecting multichannel 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 clearance sale at a retailer that lasted only 26 days, 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. Li and Kannan (2014) develop a complex model of online customers' "path to purchase," which allows them to not only gauge the effect of advertising on customer's likelihood to visit the advertiser's website and ultimately purchase services but also focus on contemporaneous advertising effects. More closely related to our work is that of Bollinger et al. (2013), who also develop a hierarchical ad-stock framework and estimate different decay parameters for each channel by extending the competitive advertising model of Dubé et al. (2005).

customers, while we find strong positive effects of catalog and email advertising that is designed to encourage existing customers to purchase more. Both the advertising context and our empirical strategy, which pools data over two years, may explain the differences in findings.

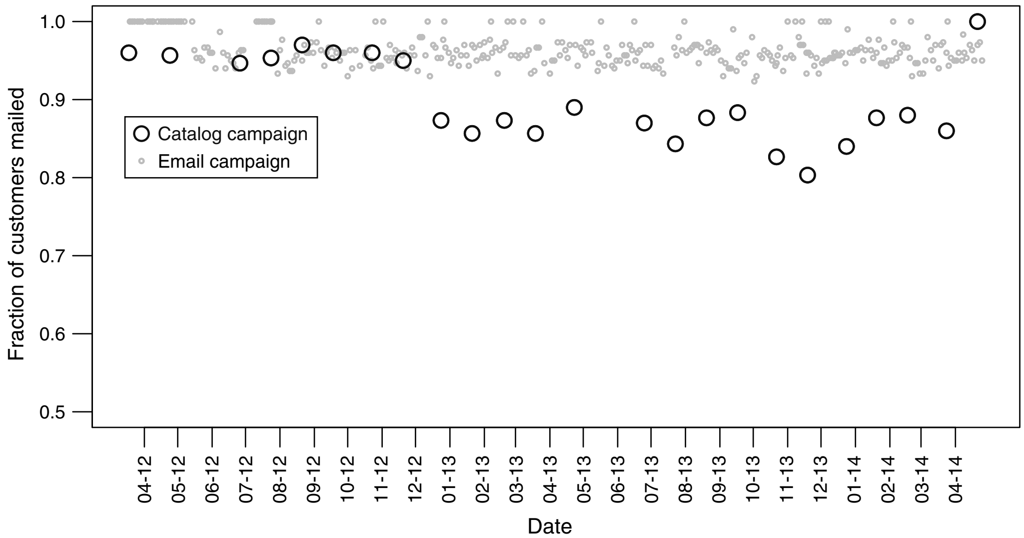
2. Consumer-Level Advertising Response Data with Randomized Holdouts

We focus on a data set from a multichannel specialty retailer that records all the catalogs and emails that were sent to individual customers along with those same customers' purchases (in dollars) across all channels (online and stores) over a 26-month period. Customers at this retailer nearly always use a credit card to transact, so that as soon as a customer makes her first purchase, the company has access to her name and address and can begin sending marketing communications. The retailer sends out emails and catalogs[2] that are designed to drive purchases at the website or stores, typically by informing customers about newly available products. The retailer did almost no mass media and little online advertising during this period, so the email and catalog exposures we observe represent nearly all the marketing communications that the customers see from the retailer. Our analysis focuses on a sample of 300 North American customers who had made at least one purchase in the year before the observation window and had not requested to be put on a "do not mail" list for email or catalog.

While the strategy we use to measure ad response can be used across a variety of contexts, there are several advantages to focusing on direct marketing data. Typically, direct marketers find it easier to link exposures and purchases; in our data set, the multichannel retailer uses name and address matching to track whether those who received catalogs and emails made purchases via any channel. But more importantly, the retailer is in complete control of whether or not each customer receives a catalog or email, which makes it easy for them to execute randomized holdouts. Like most direct marketers, the retailer sets a fixed schedule

[2] 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.

Figure 1 **Cadence (Timing) and Reach for Direct Marketing Campaigns**



for mass mailings of catalogs and emails throughout the year. Each occasion for mailing a catalog or email is referred to as a "campaign" in the direct marketing community; an email sent on Monday to a group of customers would be one campaign and an email sent on Wednesday would be a second campaign, even if the two emails have related messaging. (This usage of the term "campaign'' is different from paid media advertising, where "campaign" might refer to a series of media placements over several weeks.)

All 300 customers in our sample were initially targeted to receive all of the catalog and email campaigns that the retailer planned during the observation period (based on their past purchases). For each campaign, a fraction of customers were randomly selected to be part of the holdout group and did not receive that particular campaign. Because the retailer controls the direct marketing, the remaining customers were all sent one catalog on the planned campaign date; i.e., they were "treated" with exactly the same "dose" of advertising.[3] Holdout groups were selected independently for each campaign, so that across the two-year period, customers receive 10–24 catalogs and 311–337 emails, depending on how many holdout groups they are randomly assigned to. It is more difficult to control the timing of online display or search advertising in the same way, since customers may not be available to be targeted on a particular day. In the conclusion, we will discuss the potential for and challenges of using randomized holdouts in other settings.

2.1. Advertising Cadence

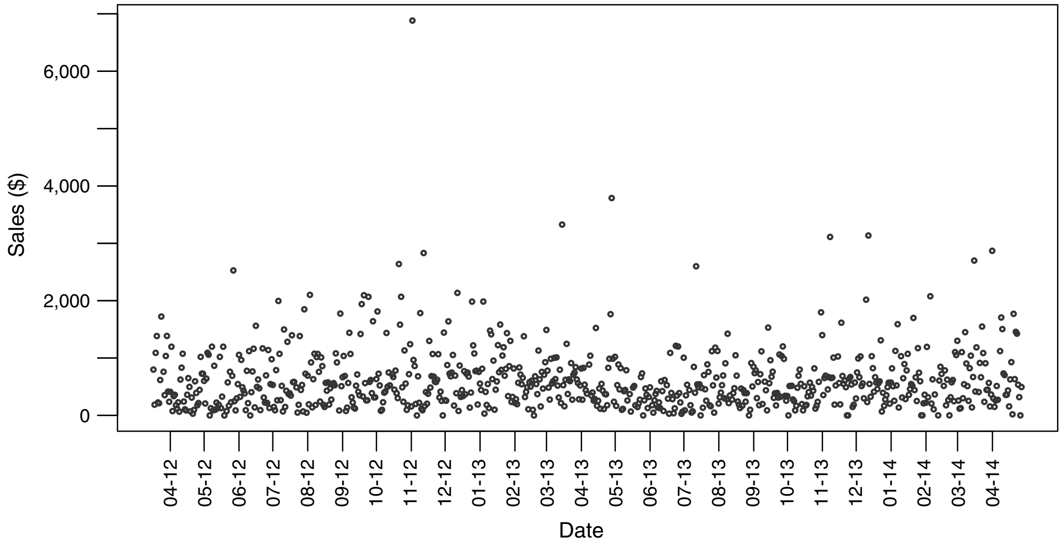
Figure 1 shows the timing of advertising campaigns over our observation period from March 31, 2012 to May 11, 2014. Direct marketers often refer to this schedule of planned campaigns as the "cadence." As shown in Figure 1, the timing of catalog and email mailings is extremely consistent; the company sends a catalog campaign during the first week of each month (except for June) and email campaigns nearly every Monday, Wednesday, and Friday. Notably, this retailer does not increase the frequency of advertising during the peak U.S. retail season from Thanksgiving to New Year's Day. As we will discuss further, this limits the possibility of simultaneity between advertising exposures and periods with large unexplained shocks in demand, which could induce a bias in our model-based estimates of advertising effects.

Figure 1 also shows the fraction of customers that were randomly selected to be contacted for each campaign (with the remaining customers making up the holdout group). All but a few email campaigns held out approximately 5% of the customers in our sample, and most catalog campaigns held out 5%–20% of the sample. The retailer changed the holdout fraction for catalogs over the course of our observation period as they learned more about how large a fraction they would need to analyze each campaign as an A/B test. There is nothing to suggest that the retailer systematically varied the holdout fraction as a function of the expected response to each campaign.[4]

[3] As is common in most CRM systems, the assignment to a holdout group is explicitly recorded in the company's database, which made it possible for us to confirm that all customers were initially targeted for all campaigns.

[4] As described in Sahni et al. (2016), if the company does adjust the holdout rate from one campaign to another in a way that is correlated with their expected response to the campaign, a bias in estimated advertising response can be produced.

Figure 2 Sales over the Observation Period



2.2. Sales

When a customer makes a purchase, the company matches the name and address associated with the credit card to the customer's record in the database, resulting in data on advertising exposures and purchases that are linked at the individual level, akin to "single-source" data. 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 match the purchase to a customer record in the database. This retailer finds that nearly 95% of customer transactions at stores (and 100% of online transactions) can be tracked back to an existing customer. In our sample, 263 of the 300 customers made at least one purchase in the two-year window, and the number of purchases for each customer ranges from 0 to 123, with total purchase amounts over the 26 months ranging from zero to thousands of dollars. Since these are regular customers of the retailer, our estimates of advertising response represent the increase in sales that the retailer can gain by further cultivating these "good" customers.

The aggregate sales for the sample of 300 customers is shown in Figure 2. Total sales are relatively constant over time, with a few spikes around major U.S. retail days like Black Friday and the days around Christmas. Comparing the aggregate sales to the campaign cadence in Figure 1, there is no obvious relationship between the aggregate advertising exposures and aggregate sales. The frequency of campaigns, the number of customers exposed to each campaign, and subsequent sales are all relatively constant over time, suggesting the aforementioned benefit of exploiting customer-level variation (as is done here).

2.3. Estimated Treatment Effects

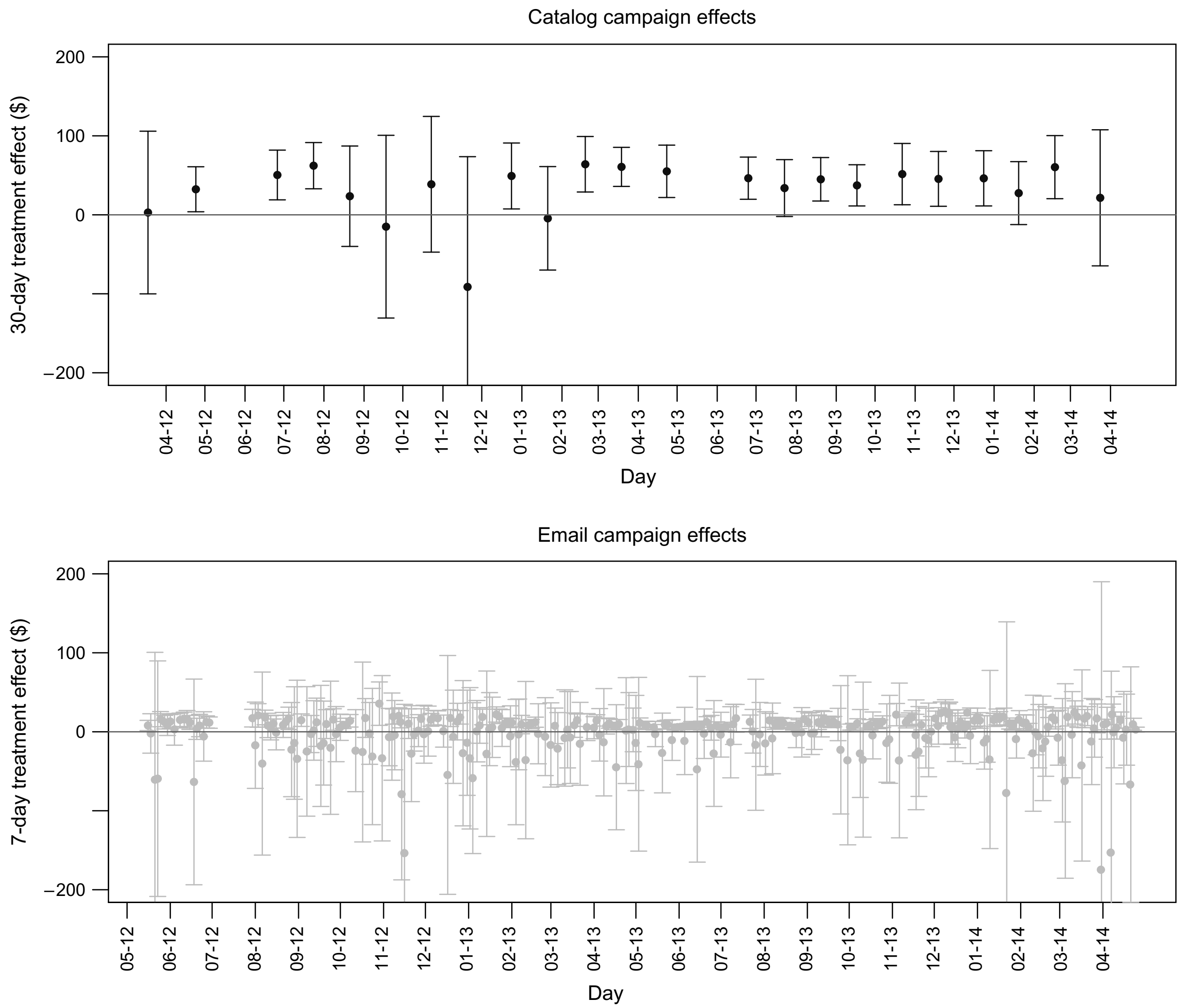
Each catalog or email campaign can be considered individually as a field experiment or A/B test by comparing the sales subsequent to a campaign for the treated and holdout customers. Table 2 shows an example of such an analysis, comparing the 30-day average spending per person for the treated and the control groups for the July 2012 monthly catalog campaign, which shows a significant lift in both the fraction of customers transacting and the average sales in the period 30 days after.[5] Although there were catalog and email campaigns before July 2012 and email campaigns were ongoing during the period when we measure sales response for this catalog campaign, since the holdout is randomized, the exposure to other campaigns is nearly balanced across the treated and holdout groups. We confirm that the holdout group was randomized by comparing the sales in the 30 days before the campaign and find no significant difference in precampaign sales between the two groups. This is further confirmation that the difference in average sales subsequent to exposure

Table 2 Example Campaign-Level A/B Test Analysis for July 2012 Catalog Campaign

	Customers (N)	Average 30-day sales ($)	30-day purchase incidence	Average prior 30-day sales ($)
Treated	284	62.34	0.25	58.82
Holdout	16	11.90	0.06	62.65
Difference		50.44	0.18	−3.84
95% CI		(18.95, 81.92)	(0.04, 1)	(−91.82, 84.13)

[5] Discussions with the retailer and the cadence that they utilized suggested the appropriateness of a 30-day window for measuring catalog sales response and a 7-day window for email for exploratory purposes. Sensitivity analysis with longer and shorter windows (available upon request) did not change the overall findings that catalog effects are usually positive and significant and that email effects are usually positive, but often not significant.

Figure 3 **Estimated Effects on Sales For All Catalog and Email Campaigns**



between the two groups can be treated as an estimate of the causal effect of the campaign.[6] However, when we analyze a single campaign, the estimated confidence intervals (CIs) for the treatment effects are quite wide, reflecting the limited amount of information we can get by comparing just 16 holdout customers (in this case) to the 284 customers in the treated group.

Our choice to focus on July 2012 as an example was arbitrary; the only difference between July 2012 and other months is that the firm does not send a catalog in June, which reduces the variation in prior ad exposure across customers. Repeating this single-campaign analysis for all 23 catalog campaigns, we find that the estimated 30-day treatment effect, i.e., the difference in per customer sales between the treated and holdout groups, ranges from −$91.26 to $64.03, with a median across campaigns of $44.99 and a mean of $32.33 (a 95% CI for the mean treatment effect across campaigns is $17.54–$47.12), suggesting that the catalog campaigns have positive effects on subsequent sales and that July 2012 is a fairly typical campaign. The estimated treatment effect for each catalog campaign is shown in Figure 3, which suggests clear exploratory evidence that receiving a catalog has a positive effect on sales in the subsequent 30 days.[7] We also find that, for any given campaign, the treatment effect is not precisely estimated and has fairly wide confidence intervals similar to those reported in Table 2. In fact,

[6] We also explored analyzing the campaigns using a difference-in-differences approach comparing the difference in sales for each customer in the 30 days before and after the campaign. Using this approach we find a similar average treatment effect of $54.28 with a wider 95% interval of (−$40.42, $148.98).

[7] The fact that we find substantial treatment effects when Lewis and Rao (2015) find much smaller effects should not be too surprising. Here we study catalogs intended to increase the purchase rate among existing customers; Lewis and Rao (2015) study display advertising intended to attract new customers and targeted at a broad audience.

Table 3 **Analysis of Combined Effect of July 2012 Catalog Campaign and Contemporaneous (+/− One Day) Email Campaigns**

	Customers (N)	Average 30-day sales	95% CI for sales ($)	30-day purchase incidence
Catalog + Email	263	67.02	(45.13, 88.90)	0.25
Catalog only	21	14.47	(−3.34, 32.29)	0.14
Email only	15	12.69	(−14.53, 39.92)	0.07
Neither	1	0.00	NA	0.00

we observe some catalog campaigns where the average 30-day sales are higher for the holdout group, typically due to one or two holdout customers making a large purchase; however, these single-campaign negative effect estimates never reach significance.

Similarly, we computed the 7-day effect for each of the 286 email campaigns that include holdouts and found estimated effects for each campaign that ranged from −$174.60 to $35.59 with a median of $6.08 and a mean of −$1.85. The individual-campaign treatment effects are plotted in the lower panel of Figure 3. The estimated effect of individual mail campaigns seldom reaches significance, but most are positive, suggesting that emails may have a positive, but more modest, effect (in a meta-analytic way) on sales.

2.3.1. Interactions Between Catalog and Email. To provide preliminary evidence of interaction effects between catalogs and emails, Table 3 analyzes the July 2012 catalog campaign and the contemporaneous email campaigns. Dividing customers into those who received email only, catalog only, both, or neither in three days around the catalog campaign, we see that the average 30-day sales for those who received both catalog and email is more than twice the sum of the average sales for the catalog-only and email-only groups. The estimated interaction effect for this campaign is $39.85 and the estimated main effect for catalogs is reduced to $14.47. (Note that the main effects reported in Table 3 that account for interactions between the two channels can be compared to our model-based estimates reported below.) However, you can see that the sample sizes for the holdout groups are quite small (only one customer was held out from both campaigns), and a simple regression of 30-day sales on the two treatments shows that the interaction effect is nonsignificant. Repeating this analysis across 16 catalog campaigns for which all four cells are populated, we find that the median interaction effect across campaigns is $21.79 but is nonsignificant for any individual campaign. This provides further evidence of the need for a model that can effectively pool across campaigns, given the common situation where the joint distribution of campaign exposures, for any given campaign, is likely to contain very sparse cells.

2.3.2. Response to Consecutive Campaigns. In most models of advertising, including the ad-stock framework we employ, customers who have received more advertising in the recent past should make more purchases than those that have been exposed to less advertising. To provide evidence that our data are consistent with this assumption, Table 4 compares customers who were sent catalogs in both July and August to those customers who received the August catalog only, the July catalog only, or neither the July nor August catalogs. Comparing the sales and purchase incidence in the 30 days after the August campaign, we find that those who were sent both catalogs purchase significantly more than any of the other groups. For example, the 95% confidence interval for the difference in mean 30-day sales between the group that was sent both catalogs and the group that was sent only the July catalog is ($8.05, $82.71). However, the differences between those groups that received only one catalog and those that received neither is nonsignificant (largely because of the small sample sizes in those groups.)

Repeating this analysis across 22 pairs of catalog campaigns, we find consistently that customers who receive two catalogs in a row have higher sales in the 30 days after the second campaign (median across campaigns of $65.05) than customers who received just the second catalog (median = $26.37), those who received just the first catalog (median = $18.55), and those who received neither (median = $0.00). The data, taken across multiple campaigns, suggest that the effect of a catalog seems to persist for more than 30 days and that the more recently a customer has received a catalog, the greater the effect.

Table 4 **Analysis of Combined Effect of July 2012 and August 2012 Catalog Campaigns**

		After August catalog		
	Customers (N)	Average 30-day sales ($)	95% CI for sales	30-day purchase incidence
Catalog in July and August	272	65.05	(43.85, 86.24)	0.25
Catalog in August only	14	13.60	(−15.78, 42.98)	0.07
Catalog in July only	12	19.67	(−12.51, 51.84)	0.17
No catalogs in July or August	2	0.00	NA	0.00

Similarly, looking at pairs of consecutive email campaigns, we see that those that receive two consecutive campaigns purchase more than those who receive just the second campaign (estimated effect ranges from −\$151.41 to \$40.86 across campaigns with a median of \$6.13) or those who receive just the first campaign (estimated 7-day effect ranges from −\$179.00 to \$37.09 across campaigns with a median of \$6.80). This suggests that email campaigns also have cumulative effects on sales that are consistent with a model that incorporates carryover.

2.4. Summary of Exploratory Data Analysis

In this section, we have provided an exploratory analysis of a sales and marketing data set from a specialty retailer that is linked at the customer level and includes randomized holdouts for each campaign. By analyzing each campaign as an independent experiment, we have shown evidence of a causal impact of both catalog mailings and emails on subsequent sales. We have also shown that the data are grossly consistent with a cumulative effect of advertising that decays over time.

There are two major limitations to analyzing each campaign as an independent test as we have here. First, each test individually has limited power, due to the small size of the holdout groups. While a 50/50 split between treatment and control would provide the most precise estimate of the treatment effect, companies are reluctant to do that, because of the potential lost revenue in the holdout group. (The design of holdout groups that optimally trade off this opportunity cost against the potential value of information is an interesting problem, but outside the scope of this paper.) A second, more subtle limitation of the analysis above is that we had to make an assumption about how long the treatment lasts. Our choice to analyze catalog response over 30 days and email response over 7 days was somewhat arbitrary (albeit sensitivity analysis shows that changing the window did not alter our conclusions.)

In the next section, we propose a model that can be used to analyze repeated randomized holdouts by pooling information across campaigns. This not only increases the precision of our estimates by controlling for prior advertising exposures, it also allows us to estimate the duration of advertising response separately for catalogs and emails, as well as any interaction effect. This model also gains power by comparing customer behavior on days when they have not been exposed to ads recently to days when they have been exposed to ads.

3. 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:

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

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

where ρ_{ik} is a consumer- and channel-specific decay parameter constrained to $[0, 1)$ and ε_{ikt} is a series of demand shocks that are also consumer- and channel-specific.

The formula in (1a) presents an exponentially decaying stock variable that has a long tradition in marketing (see the survey by Huang et al. 2012). However, there are practical difficulties in 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). We should also note that the inclusion of the error term in Equation (1b) is of significance, as there is some disagreement in the literature about whether the shocks should be included. Dubé et al. (2010) justify the error term on the basis that the data (in their case aggregate exposures and in our case individual-level exposures) do 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. The error term is also important in the computation of the stocks since it allows us to represent the ad-stock model as a state-space model, as we describe below. While in the original Koyck framework the shocks, ε_{ikt}, are assumed to be independent and identically distributed, we generalize this to allow for (potential) serial correlation in the errors.

We extend the single latent stock formulation given in Equation (1b) to a multichannel environment by introducing K stock variables, one for each channel, as in (2a):

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

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

$$\cdots$$

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

We also include additional stocks for the interactions among the different channels, a topic of significant empirical interest (see Wilbur 2008, Naik et al. 2005, Bollinger et al. 2013). 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:

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

The specification in Equation (3) allows for the possibility that exposures on channels 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 carryover 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 error terms $\varepsilon_{i,1:2,t-j}$, define the interaction stock $W_{i,1:2,t}$. (For a similar treatment of interactions, see Bass et al. 2007, although at the aggregate level.)

We relate the ad stocks to the observed purchases, Y_{it} (measured in dollars in our application), through a latent variable, Y_{it}^*, for individual i at time t. We assume that the latent process, Y_{it}^*, is related to the ad-stock variables, including interactions, by sensitivity parameters β_{ik} that measure the effect of the stock on latent Y_{it}^*.

$$Y_{it}^* = \mu_i + \sum_k \beta_{ik} W_{ikt} + \sum_{k'>k} \beta_{i,k:k'} W_{i,k:k',t} + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_{\eta_i}). \tag{4}$$

Note also the presence of the intercept, μ_i, in Equation (4) that can account for baseline differences between individuals in their purchasing behavior. This consumer-level intercept also helps to avoid biases (albeit not applicable in our case because of randomized holdouts) caused by the advertiser targeting customers who are more likely to buy, because the estimated effect of advertising measured by β_{ik} is above and beyond that expected, given the individual's estimated baseline propensity.

There are two sources of variation that identify the advertising effect β_{ik}: the variation across customers in advertising exposures in a given time period and the variation over time in the ad stock. That is, some of the identification comes from comparing the treatment and the holdout group (as we did in the exploratory data analysis), and some of the identification comes from comparing a customer's behavior on the days just after she received a marketing communication to days where she has not recently received marketing. Additional power to estimate the advertising effects, β_{ik}, is provided by the individual-level intercepts, μ_i, which control for the overall purchase propensity of each customer and the fact that the model controls for prior ad exposure at the individual level, through the ad-stock W_{ikt}.

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}:

$$Y_{it} = \begin{cases} Y_{it}^* & \text{if } Y_{it}^* > 0, \\ 0 & \text{if } Y_{it}^* \le 0. \end{cases} \tag{5}$$

Note that, while the latent variable in the Tobit I structure responds linearly to advertising exposures, this does not imply that advertising response is linear. As we will discuss in Section 5.2, the Tobit I structure implies that this response has increasing returns as the customer's ad-stock approaches the threshold, consistent with wear-in effects.

Finally, as mentioned above, we allow for the possibility of a first-order serial correlation in the errors. We implement this by creating an additional "stock" for serial correlation,

$$W_{iCt} = \phi_i W_{i,C,t-1} + \varepsilon_{iCt}, \tag{6}$$

that is equivalent to allowing for a first-order autocorrelation in ε_{ikt}. With the additional "stock" that captures serial autocorrelation, the resulting full-model equation for Y_{it}^* is

$$\begin{aligned} Y_{it}^* &= \mu_i + \sum_k \beta_{ik} W_{ikt} + \sum_{k'>k} \beta_{i,k:k'} W_{i,k:k',t} \\ &\quad + W_{iCt} + \gamma_i Z_{it} + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_{\eta_i}), \end{aligned} \tag{7}$$

where Z_{it} is a vector of dummy variables for seasonal effects. As we will discuss below, including seasonal effects to account for periods of peak demand is valuable in preventing simultaneity bias in the estimates of advertising effectiveness, which can occur if advertising is timed to periods of peak demand.

3.1. State-Space Formulation

As noted, there is a substantial computational advantage to defining the interactions and the serial correlation using the same form as the ad stocks since the entire vector of stocks consisting of the K multichannel stocks, $(K^2 - K)/2$ interaction stocks, and the serial correlation stock can be stacked as a vector

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

$$\text{Channels} + \text{Interactions} + \text{Serial Correlation}.$$

As a result, we derive a more compact representation of the dynamics in the model by simply writing

$$\mathbf{W}_{it} = \boldsymbol{\rho}_i \mathbf{W}_{i,t-1} + \mathbf{X}_{it} + \boldsymbol{\varepsilon}_{it}, \quad \boldsymbol{\varepsilon}_{it} \sim N_{K(K-1)/2+1}(0, \Sigma_\varepsilon), \tag{8}$$

where the bold fonts denote the vectorized components of the multichannel ad-stock model over channels, interactions, and serial correlation stocks and their exposures. As we show in Appendix A, this representation allows for an exact likelihood evaluation for each individual at each time period by virtue of an efficient forward-filtering and backward-smoothing algorithm. This substantially improves the speed of the estimation algorithm.

To summarize, we propose a heterogeneous model of consumer purchasing behavior where the advertising response follows a Koyck structure. 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 sparsity in the consumer-level purchase data we have specified a Tobit I process for purchasing as a function of advertising stocks.

3.2. Priors and Computation

We employ a Bayesian approach to accommodate individual heterogeneity in both the instantaneous and carryover effects (Rossi and Allenby 2003). We use conjugate normal-inverse-Wishart distributions for all individual-level parameters. We define $\boldsymbol{\beta}_i$ as the stacked vector of μ_i,β_{ik} and $\beta_{i,k:k'}$. Similarly, we define $\boldsymbol{\rho}_i$ as the stacked vector of ρ_{ik}, $\rho_{i,k:k'}$, and ϕ_i. We then define the distribution of $\boldsymbol{\beta}_i$ and $\boldsymbol{\rho}_i$ across the population of consumers as follows:

$$\boldsymbol{\beta}_i \sim N_{K(K+1)/2+1}(\bar{\boldsymbol{\beta}}, \Sigma_\beta) \quad \boldsymbol{\rho}_i \sim N_{K(K+1)/2+1}(\bar{\rho}, \Sigma_\rho)$$

where N_d represents a multivariate normal distribution of dimension d. To ensure stationarity in the latent ad-stock processes, we additionally constrain each $0 \le \rho_{ik} < 1$ and $-1 \le \phi_i < 1$ via truncation and rejection sampling.[8] We put weakly informative priors on the population-level parameters and the error terms.

$$\bar{\boldsymbol{\beta}} \sim N_{K(K+1)/2+1}(0, 5I), \quad \boldsymbol{\Sigma}_\beta \sim IW\left(I, \frac{K(K+1)}{2}+4\right),$$
$$\bar{\boldsymbol{\rho}} \sim N_{K(K+1)/2+1}(0, I), \quad \boldsymbol{\Sigma}_\rho \sim IW\left(I, \frac{K(K+1)}{2}+4\right),$$
$$\sigma^2_{\varepsilon_k} \sim IG(1,1), \quad \Sigma_\eta \sim IW\left(I, \frac{K(K+1)}{2}+4\right).$$

where IW is an inverse-Wishart distribution, IG is an inverse-Gamma distribution, and I is an identity matrix of appropriate dimension. We also explored several prior specifications where we decreased the expected precision of Σ_β and Σ_ρ under alternative scenarios and found the resulting posteriors largely unchanged, indicating that the population-level priors are not driving the reported effects. Given the state-space representation for the latent stocks, we employ Kalman dynamic linear model relationships (West and Harrison 1999 with the modifications provided in Harvey 1991, Section 3.7.2; and Liao et al. 2014 to account for the Tobit I structure. See also Naik and Raman 2003 and Ataman et al. 2010). Appendix A provides additional details of the estimation algorithm.

[8] Alternatively, one could use a logit transformation to restrict the carryover parameters.

4. Endogeneity in Advertising Response Models

It is well known that there is potential for endogeneity in advertising response models, although there may be some confusion as to the sources of this endogeneity. An endogeneity bias occurs when one of the regressors, which in our case are the goodwill stocks, W_{ikt}, is correlated with the error term, η_{it}.

There are two key ways that this correlation might arise in customer-level advertising response models: (1) a selection bias, where the customers who are exposed to advertising are systematically different that those who are not exposed (see Manchanda et al. 2004), and (2) a temporal simultaneity between advertising and demand shocks, which is typically produced by the advertiser timing advertising to periods of high-demand shocks. Either case produces an unconditional correlation between W_{ikt} and η_{it}, typically resulting in an upward bias in the estimated advertising effects. We discuss these two sources of endogeneity in the following subsections.

4.1. Endogeneity due to Customer Selection

Recent literature on advertising response has focused on the endogeneity that occurs when the group who is exposed to the advertising is systematically different than the nonexposed customers. In our notation, this happens when η_{it} is correlated with X_{ikt} and therefore W_{ikt} across customers i for a given period t. Lewis et al. (2011) describe how this can occur in online advertising because customers who are more active online are more likely both to get exposed to display ads and to make purchases online. Blake et al. (2015) point out that this problem can be extreme in the case of search advertising, where searching for a related keyword makes the customer more likely to be exposed to a search ad and more likely to purchase. In a direct marketing context like ours, this type of selection happens when the marketer knows something (unknown to the researcher) about each customer's potential to buy or their responsiveness to advertising, and uses that information to target customers. Manchanda et al. (2004) and Dong et al. (2009) describe this type of targeting in the context of pharmaceutical detailing. When high-propensity customers are targeted, the resulting correlation between W_{ikt} and η_{it} can produce an upward bias in the estimate of the advertising effect.

Randomized holdouts, like the ones in our data, prevent endogeneity due to targeting or selection (Lewis and Rao 2015, Hoban and Bucklin 2014, Sahni 2015b). With our randomized holdouts, all the variation in exposures at a given point in time is due to random assignment, so targeting or selection is not a concern. While not applicable to our data, we note that targeting can be admissibly ignored, without biasing the posterior, if behavior is modeled at the consumer level and the targeting is based on past observables, such as prior purchases. (See also Liu et al. 2007, who discuss this in the setting of adaptive conjoint surveys.)

4.2. Temporal Simultaneity Between Advertising and Demand Shocks

While selection of customers has been a concern in the recent literature, another source of endogeneity in advertising response models occurs when the ad stock in period t is correlated with demand shocks in the same period. The classic example of this is when there is a demand driver that varies over time (for instance, temperature) that is observed by the advertiser and unobserved by the researcher. If the advertiser chooses to time the advertising to occur during or just before periods of peak demand, then a correlation is induced between η_{it} and W_{ikt} over time, which can produce an upward bias in the estimated response to advertising. In the direct marketing context we study, the advertiser can't simply increase budgets, as with a media campaign. The main way direct marketers can "heavy up" on advertising is by increasing the frequency of campaigns. If they increase the frequency of campaigns during periods of high demand, then a bias would result (even with the use of randomized holdouts.) Similarly, if the firm decreases the holdout fraction during periods of peak demand, an upward bias in the estimated advertising effect would result.

If these periods of high demand are accounted for by including the demand drivers that the advertiser is reacting to as variables in the model, then the error term η_{it} will not be correlated W_{itk} and there is no endogeneity of this type. Thus the "first line of defense" against temporal simultaneity in advertising response models is to include appropriate controls (Rossi 2014), which we attempt to do in our application with day-of-week and holiday effects. If the seasonal demand drivers are not known, then the researcher might estimate monthly or weekly random effects to attempt to control for seasonality.

Additionally, in our application, the advertiser has maintained a steady cadence of advertising (see Figure 1) and has not increased marketing communications during holidays and other periods of peak observed demand. The catalog campaigns occur on the first Tuesday or Wednesday of each month and the email campaigns are nearly every Monday, Wednesday, and Friday; it seems highly improbable that there is a demand shock that occurs regularly on those days. (We also discussed this directly with the advertiser and they explained that their decisions about when to run campaigns were driven primarily by their capacity to produce the advertisements.) Similarly, the holdout fraction is not substantially varied over time. Considering this regular cadence, it seems extremely unlikely that the advertiser is sending marketing more frequently during periods with high η_{it} If the campaigns are not being timed to periods of peak demand, then resulting variation in individual's goodwill stock, W_{ikt}, is either due to the cadence (which is not correlated with η_{it}) or the randomized holdout (which is also not correlated with η_{it}).

In the ad-stock model, a more subtle potential source of simultaneity is a correlation between the demand shock, η_{it}, and the ad-stock shock, ε_{ikt}. This might occur if there was some temporal shock that made a particular customer more likely to be persuaded by the catalog or email (resulting in a shock to the ad stock) and that simultaneously made the customer more likely to purchase from the retailer in that same time period (resulting in a shock in the demand equation). While plausible, we do not believe this source of endogeneity is likely to occur in our data.

5. Empirical Analysis

In this section, we apply the model to the two years of advertising exposure and sales data for the 300 customers in our data set, which allows us to estimate the response and decay for catalog and email exposures, the decay for each channel, and the interaction between them. In applying the model to these data, we include three (fixed effects) dummy variables to account for seasonal events. These controls (Z_{it} in Equation (7)) are a dummy for weekend days (from Friday to Sunday), a dummy for days around Thanksgiving (from Wednesday to Monday), and a dummy for days in the Christmas week (from December 23 to January 2). 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 test proposed in Raftery and Lewis (1996) implemented in the CODA package.

5.1. Nested Model Comparison

To understand the importance of the various features of the model presented in Section 3 and provide evidence that the full specification we propose is necessary, we fit a series of nested models to the data. In Table 5, we report the harmonic estimator of the marginal likelihood (Gelman et al. 2013), as well as the log-likelihood averaged over each Markov chain Monte Carlo (MCMC) draw for each model. As suggested in

Table 5 **Model Comparison Based on Marginal Likelihood and In-Sample and Out-of-Sample Goodness-of-Fit Measures Across Models with Different Components**

Model	Marginal likelihood	Log-likelihood	RMSE	RMFE
Model 1				
Common decay	12,665	−126,145	1.89	1.93
Model 2				
Proposed model without serial correlation or interaction	12,880	−123,449	1.69	1.80
Model 3				
Proposed model without serial correlation	13,040	−119,967	1.45	1.61
Model 4				
Proposed model without interactions	13,671	−120,196	1.41	1.56
Model 5				
Proposed model	16,475	−116,810	1.35	1.49

Notes. Marginal likelihood is the harmonic mean estimator of the integrated likelihood. Log-likelihood is the average over the draws of the likelihood across the chains of the algorithm. RMSE is the averaged root mean square error. RMFE is the averaged root mean square forecast error. Our proposed model provides improvement along all the metrics reported, suggesting that all model components are important in fitting the data well.

West and Harrison (1999, pp. 393–394), we also evaluated the model fit based on a combination of in-sample and out-of-sample posterior predictive checks.[9] To assess in-sample fit, we use the root mean square error (RMSE) between the predicted and actual purchases for each individual in each day. To assess out-of-sample fit, we use the root mean square error of one-period-ahead forecasts (RMFE), which condition on the past amount of advertising exposure. These are defined as

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

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

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

In addition to estimating the full model proposed in Section 3 (Model 5), we estimate a model with a single common latent stock for the advertising channels and the interaction (Model 1), which is the approach that has been taken in past work on multichannel advertising (Bass et al. 2007), at the aggregate level. See Dinner et al. (2014) and Bollinger et al. (2013) for similar considerations on the importance of different decays per channel. As Table 5 shows, the full model reduces the in-sample RMSE by 25% and the out-of-sample RMFE by 30% over Model 1, suggesting that separate decays are empirically supported in these data.

We also estimate models that include separate decay for each channel but which do not include serial correlation (Model 3), interactions (Model 4), or either of the two (Model 2). Table 5 shows that these nested models all fit substantially worse than our proposed model (Model 5), suggesting that an interaction between channels and serial correlation in the error terms is empirically supported in this data set.

Table 6 shows the estimated parameters for the five models. Focusing on the full proposed model, the estimated parameters related to instantaneous effectiveness of advertising, $\bar{\beta}_k$, suggest that email communications are more effective on the day of exposure than catalogs, 0.13 versus 0.07 on average at the population level. The estimate of the interaction effect between the two channels is positive but nonsignificant. We do not find any significant effects for the seasonal controls.

The carryover coefficients, $\bar{\rho}_k$, also present an interesting pattern. In Model 5, catalog has a moderately higher carryover than email. This suggests that, while email is more effective at increasing same-day sales, a catalog exposure has a more long-lasting effect. These parameters indicate that 90% of the email advertising effects dissipate on average in one week while catalogs are more long lasting, with their effects dissipating in approximately two to three weeks. Interestingly, on average the company mails catalogs approximately once per month, which suggests that the brand's marketers have an intuitive sense of the duration of catalog response.

[9] 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 et al. (2011) on the context of choice models with cross-brand pass-through effects. For this reason, we rely on classic in-sample and out-of-sample procedures in the spirit of West and Harrison (1999).

Table 6 Population-Level Parameter Estimates for Our Proposed Model Along with Four Nested Models with 2.5th and 97.5th Percentiles of the Posterior Shown in Parentheses

Parameter	Model 1 Common decay	Model 2 Without serial correlation or interaction	Model 3 Without serial correlation	Model 4 Without interaction	Model 5 Complete model
$\bar{\mu}$	−1.24 (−2.45, −0.03)	−1.61 (−2.93, −0.48)	−1.30 (−2.52, −0.07)	−1.11 (−2.05, −0.06)	−1.39 (−2.37, −0.40)
$\bar{\beta}_{CAT}$	0.10 (0.03, 0.17)	0.12 (0.05, 0.20)	0.08 (0.05, 0.11)	0.09 (0.06, 0.12)	0.07 (0.02, 0.11)
$\bar{\beta}_{EM}$	0.12 (0.04, 0.20)	0.15 (0.04, 0.26)	0.18 (0.08, 0.28)	0.21 (0.14, 0.28)	0.13 (0.04, 0.22)
$\bar{\beta}_{k:k'}$	0.04 (−0.03, 0.11)	NA	0.04 (−0.12, 0.21)	NA	0.03 (−0.12, 0.17)
$\bar{\rho}_{CAT}$	0.71 (0.41, 0.95)	0.84 (0.47, 0.98)	0.82 (0.53, 0.97)	0.83 (0.52, 0.96)	0.81 (0.49, 0.92)
$\bar{\rho}_{EM}$	0.71 (0.41, 0.95)	0.81 (0.53, 0.98)	0.72 (0.41, 0.96)	0.73 (0.41, 0.96)	0.71 (0.39, 0.94)
$\bar{\rho}_{k:k'}$	0.71 (0.41, 0.95)	NA	0.76 (0.47, 0.98)	NA	0.73 (0.42, 0.95)
$\bar{\phi}_C$	−0.55 (−0.93, 0.15)	NA	NA	−0.38 (−0.76, 0.21)	−0.53 (−0.90, 0.06)
γ_W	0.12 (−0.20, 0.33)	0.10 (−0.23, 0.30)	0.09 (−0.22, 0.29)	0.09 (−0.23, 0.28)	0.09 (−0.23, 0.29)
γ_{TG}	0.14 (−0.16, 0.30)	0.13 (−0.18, 0.28)	0.12 (−0.13, 0.27)	0.11 (−0.15, 0.25)	0.10 (−0.16, 0.24)
γ_{CH}	0.11 (−0.18, 0.29)	0.10 (−0.21, 0.32)	0.08 (−0.21, 0.26)	0.07 (−0.19, 0.25)	0.07 (−0.18, 0.25)

We also find that there is a negative estimated serial correlation (although not significant). Although we do not report it in Table 6, we also find that there is substantial heterogeneity in the serial correlation coefficients. Across individuals, we find that the posterior means for the ϕ range from −0.95 to 0.15, indicating that the individual-level dynamics vary. We believe this range of coefficient to be realistic as a shock leading to purchase on day t will lead to less purchasing on day $t+1$, for most customers.

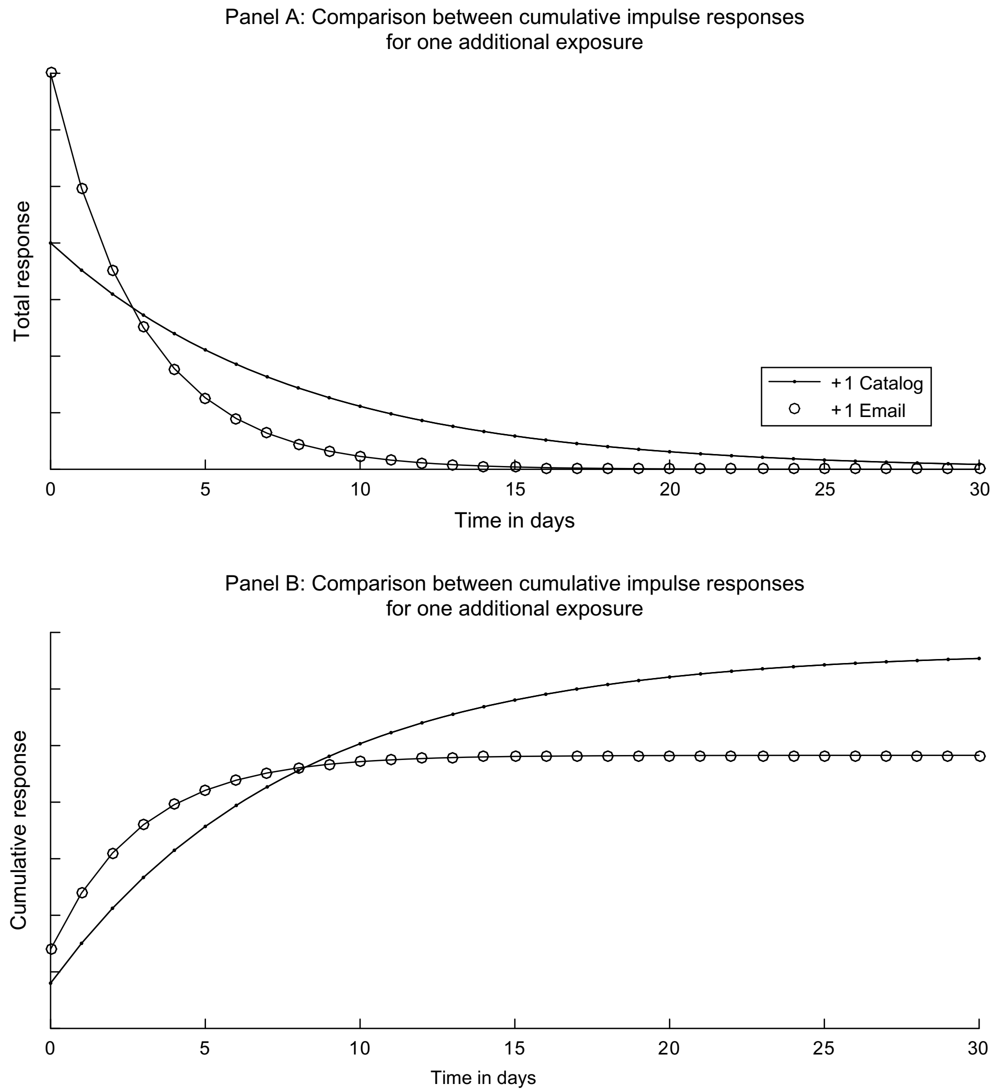
5.1.1. Comparing the Dynamic Response to Advertising for Different Channels. To illustrate the model's dynamic predictions about how the response to advertising plays out over time, we plot predicted impulse response curves in Figure 4. The plot shows the impulse response for sending one email to all 300 customers versus one catalog to the same 300. 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 much higher for the email impulse. However, a catalog becomes cumulatively more effective after approximately 8 days as shown in panel B of Figure 4.

As noted above, we find a positive but nonsignificant interaction between catalog and email, and Figure 5 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. As Figure 5 shows, the model predicts a small contribution to total purchase response due to the interaction between email and catalog.

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 past advertising and predict which channels exceed their costs and bring the largest return on marketing. We next turn to the other major potential use for the model: targeting advertising to individual customers.

5.1.2. 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 6 plots posterior means of these parameters for the individuals in our data set. Panel A suggests that there is a positive correlation across customers between the instantaneous effectiveness parameters of catalogs and email; that is, customers who are more responsive to email are also more responsive to catalog. In panels B and C, we find a positive a posteriori relationship between β_{ik} and ρ_{ik} for both catalog and email. That is,

Figure 4 Comparison Between Aggregate CIR Curve for a One-Impression Shock to All Customers on Either Email or Catalog



Note. The scale of the y axis has been omitted per request of the retailer providing the data.

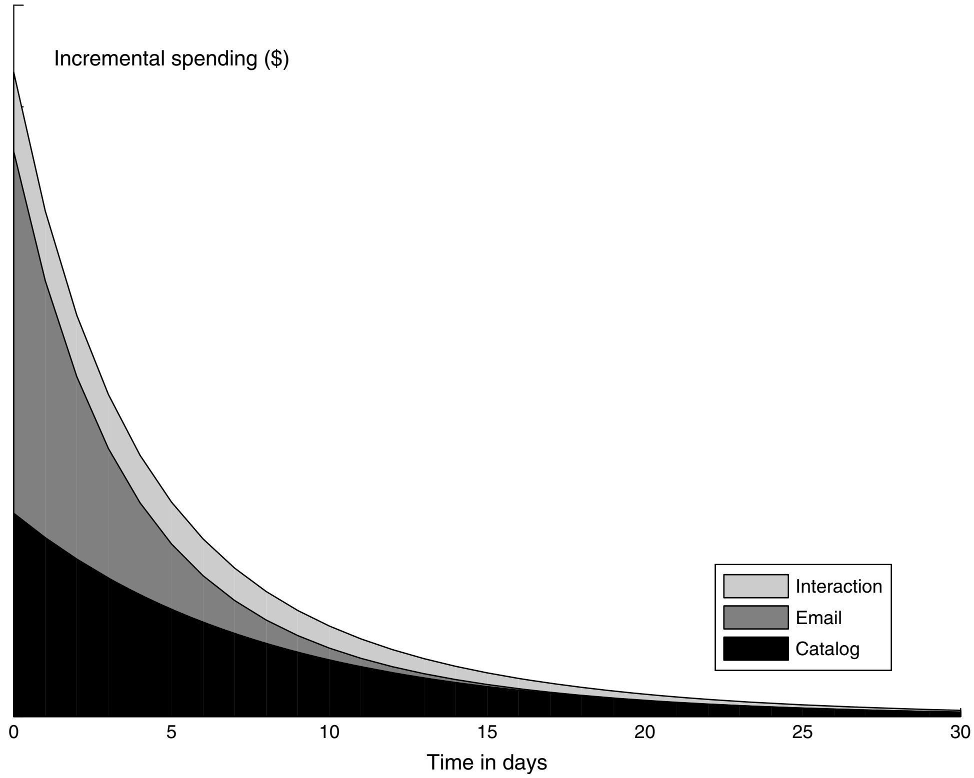
individuals with higher initial response to advertising on one channel seem to have systematically more carryover on that same channel. However, there is not a significant correlation between ρ_{ik} for email and the ρ_{ik} for catalog (panel D). Most importantly, we find that there is a great deal of 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.

5.2. 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 stochastic 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 a one-impression "impulse" in W_{ikt}, is a more economically meaningful measure of the expected return from an additional exposure to consumer i on channel k, which can be directly compared to the cost of the advertisement. In this section, we derive the cumulative impulse response for an individual in closed form, demonstrating that it is easy to compute from the estimated consumer-level parameters and therefore can be used to categorize individuals into those who are expected to be more or less responsive to advertising.

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

Figure 5 **Comparison of the Differential Effect of Each Component in the Model on the Aggregate Impulse Response Curve for an Increase in Exposures on Both Email and Catalog for the Full Model**



Note. The scale of the y axis has been omitted per request of the retailer providing the data.

can be written as

$$\frac{\partial E_t(Y_{it})}{\partial W_{ikt}} = \underbrace{P(Y_{it}^* > 0) \cdot \frac{\partial E(Y_{it} \mid Y_{it}^* > 0)}{\partial W_{ikt}}}_{I} + \underbrace{(E(Y_{it} \mid Y_{it}^* > 0)) \frac{\partial P(Y_{it}^* > 0)}{\partial W_{ikt}}}_{II}, \quad (9)$$

which is referred to as the McDonald and Moffitt (1981) decomposition for the Tobit model. 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 nonzero (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

$$\frac{\partial E(Y_{it})}{\partial W_{ikt}} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_\eta}\right)\beta_{ik}, \quad (10)$$

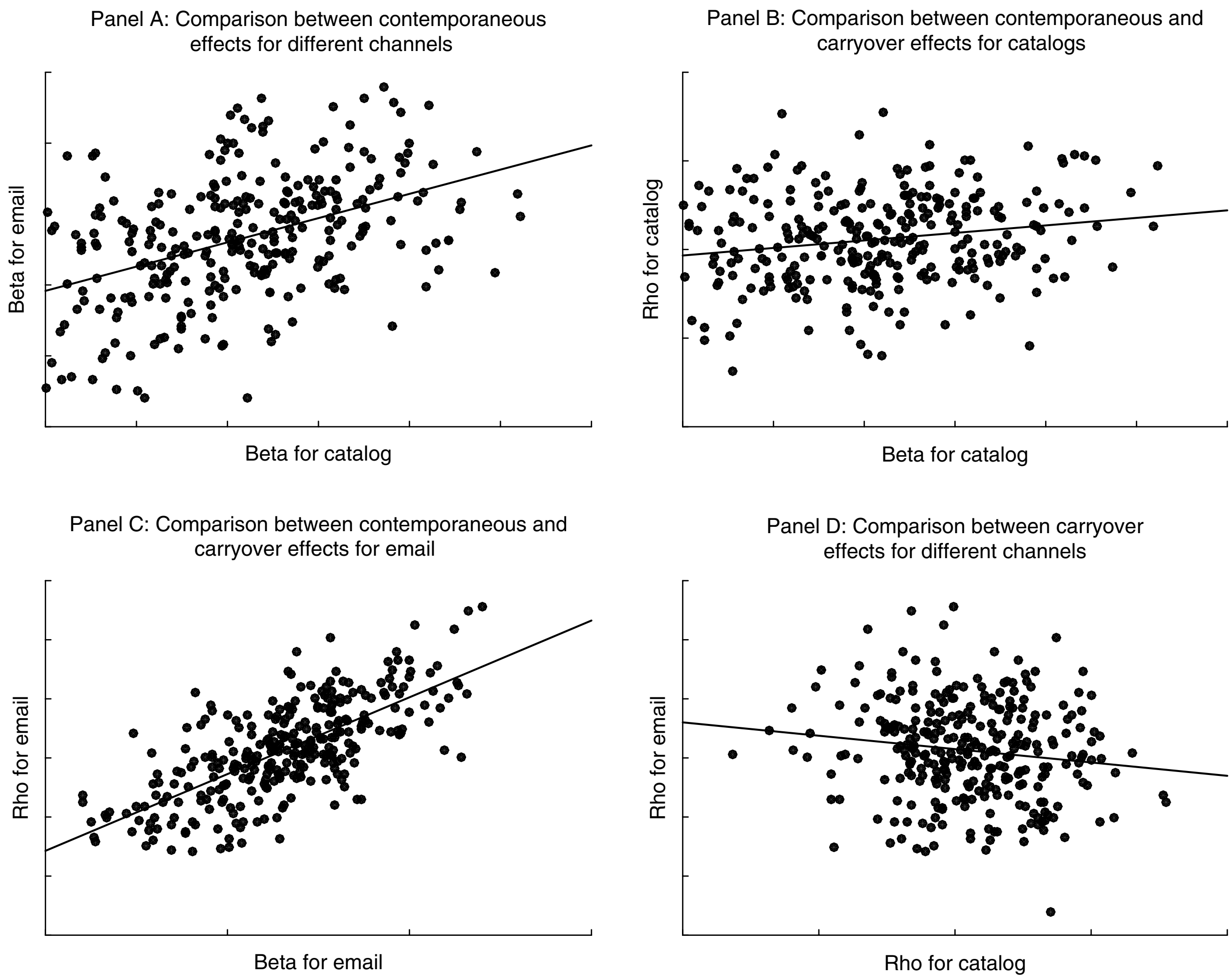
$$V_{it} = \textstyle\sum_k \beta_{ik} W_{ikt} + \sum_{k'>k} \beta_{i,\,k:k'} W_{i,\,k:k',\,t} + W_{i,\,Sc,\,t} + \gamma_i Z_{it}, \quad (11)$$

where Φ denotes the standard normal cumulative distribution function (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 $\Phi()$ 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. Equation (10) implies that, for individuals who are just below the threshold, the purchase response will be increasing in the number of ad exposures, while for customers who have had many exposures, the response will be decreasing. This is consistent with the theoretical S-shaped advertising response. Thus, under the model assumptions, given two customers with identical ρ_{ik} and β_{ik} parameters, it is more beneficial to advertise to the ones whose combination of μ_i and prior exposures puts them near the Tobit threshold, 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 (CIR) 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

$$CIR_{ikt} = \sum_{j=0}^{\infty} \frac{\partial E(Y_{i,\,t+j})}{\partial W_{ikt}} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_{\eta_i}}\right)\left(\frac{\beta_{ik}}{1-\rho_{ik}}\right). \quad (12)$$

Figure 6 **Comparison of Posterior Mean of the Effectiveness and Carryover Parameters Across Individuals for Different Channels**



While the effect of an impression on any individual can be forward simulated from the model, Equation (12) provides a computationally convenient way to identify customers who would be most affected by advertising at time t.

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

$$CIR_{i,\,k:k',\,t} = \sum_{j=0}^{\infty} \frac{\partial^3 E(Y_{i,\,t+j})}{\partial W_{ikt} \partial W_{ik't} \partial W_{ik:k't}} = \Phi\left(\frac{\mu_i + V_{it}}{\sigma_{\eta_i}}\right) \cdot \left(\frac{\beta_{i,\,k}}{1-\rho_{i,\,k}} + \frac{\beta_{i,\,k'}}{1-\rho_{i,\,k'}} + \frac{\beta_{i,\,k:k'}}{1-\rho_{i,\,k:k'}}\right), \quad (13)$$

which allows us to gauge the contribution of both stocks with the inclusion of the interaction term.

In Table 7, we illustrate the economic advantages of targeting the most responsive customers, based on their individual predicted cumulative impulse response. Note that the CIR depends on the individual's ad stock at time t, namely, the perturbed contribution of the state-space equation in (8). We report the results of targeting customers at two different points in time: June 6, 2012, and June 6, 2013. We chose these two days because the retailer did not send catalogs in June, and so our estimates represent the potential revenue that could be gained by sending a catalog in that month. (Sending more than one catalog in a month would likely produce even greater returns as more customers who have already received a catalog would have a higher stock and would be more likely to exceed the Tobit threshold.)

In the first row of Table 7, we report the average model-predicted increase in sales in a randomly picked group; this represents the average per customer lift in sales that is obtained by exposing every individual in the group to either one additional email or one additional catalog. As shown in the first row of Table 7, the expected response from sending an email is approximately \$2.25 and the expected response for a catalog is approximately \$5. (These estimates can be compared

Table 7 Per Customer Predicted Increase in Sales for Various Targeting Strategies

	June 6, 2012	June 6, 2013
No targeting (10% of customers picked randomly)	+1 Email 2.24 +1 Catalog 4.99	+1 Email 2.36 +1 Catalog 5.21
Targeting based on prior year's recency, *R* (top 10%)	+1 Email 4.70 +1 Catalog 9.08	+1 Email 5.58 +1 Catalog 9.87
Targeting based on prior year's number of purchases, *F* (top 10%)	+1 Email 3.67 +1 Catalog 5.21	+1 Email 3.80 +1 Catalog 5.40
Targeting based on prior year's spending, *M* (top 10%)	+1 Email 4.01 +1 Catalog 7.92	+1 Email 4.73 +1 Catalog 8.39
Targeting based on heuristic scoring using RFM metrics (top 10%)	+1 Email 6.21 +1 Catalog 10.77	+1 Email 6.98 +1 Catalog 11.03
Targeting based on predicted CIR (top 10%)	+1 Email 10.67 +1 Catalog 16.72	+1 Email 11.26 +1 Catalog 17.78

Notes. For each entry we report the average in the targeted group of the posterior median lift in sales for each individual. We compare our CIR-based strategy with traditional recency, frequency, and monetary value (RFM) targeting. See Baier (1996) for a description of the heuristic coding system we used to combine the *R*, *F*, and *M* measures.

to those reported in Table 3 in the exploratory analysis in Section 2.) In the second, third, and fourth rows of Table 7, we show the results of targeting the top 10% of customers based on their recency (i.e., the time of the last purchase), their frequency (i.e., the number of purchases in the prior 12 months), and their monetary value (i.e., their spending in the prior 12 months). Table 7 shows that these popular targeting strategies all increase response over targeting customers at random, in some cases doubling the response. We also find that recency targeting performs better than frequency or monetary value targeting, consistent with conventional wisdom in direct marketing. In the fifth row of Table 7, we report the cumulative response to a strategy that combines recency, frequency, and monetary value, following the RFM targeting approach based on a heuristic coding system proposed by Baier (1996), which is typical of the RFM targeting done in practice. This approach substantially raises the average response for the targeted customers to \$6–\$7 for email and \$10–\$11 for catalog.

However, when we target customers based on their predicted CIR, we can improve predicted sales even further. In the last row of Table 7, we show average per customer response that can be achieved when targeting the most responsive 10% of customers based on their model-predicted CIR. Table 7 shows that the response increases to approximately \$11 for email and \$17 for catalog. Thus, there are clear economic advantages to using the proposed model to score customers for their responsiveness based on the CIR, rather than traditional RFM targeting. The model-based targeting derives this advantage by selecting customers who have been most responsive to marketing in the past. For example, the average posterior mean of the $\beta_{i,CAT}$ for those who are selected based on the CIR is approximately 20% higher than for those who are selected based on RFM. The situation is analogous with email with a positive difference of 15% in favor of CIR for $\beta_{i,EM}$. In contrast, the RFM targeting tends to select customers who have a higher intercept; the average of the posterior means of μ_i is approximately 17% higher for the RFM-targeted customers with respect to the CIR-targeted customers. While the industry-standard RFM targeting does improve marketing response, targeting customers who have shown a strong advertising response in the past can raise marketing ROI substantially. (We should note, ironically, that there is an associated disadvantage. Targeting of any sort raises a potential endogeneity bias were the model to be estimated from data where customers had been targeted. This could prevent the firm from using the resulting data to reestimate models in the future.)

We can also use the results in the first row of Table 7 to compute long-term advertising elasticities based on the CIR estimates in Table 7.[10] With the revised data, our model-based estimate of the long-term, cumulative effect of an additional catalog among these customers is approximately \$5, which corresponds to an elasticity of approximately $5.1/750/1/11 = 0.075$, where \$750 is the approximate annual purchase amount for these customers and 11 is the number of catalog mailings in a year. Similarly, for emails the elasticity is approximately $2.2/750/1/125 = 0.367$. These approximate elasticities appear to be consistent with results reported in both the experimental and modeling literature on advertising response. For instance, a meta-analysis of split-cable television experiments (Lodish et al. 1995a) reports elasticity of 0.13 for established products with a standard deviation of 0.40 across campaigns. The observational study reported by Danaher and Dagger (2013) reports an elasticity of 0.05 for email and 0.10 for catalogs, albeit in a radically different context than ours. A meta-analysis of observational econometric studies, potentially subject to endogeneity bias, reports elasticities between −0.35 and 1.80 (Sethuraman et al. 2011). More recently Danaher and van Heerde (2015), in the context of a multichannel retailer, find elasticities across multiple brands of approximately 0.41 for email and 0.18 for catalog. Thus, the model-based elasticities we report are well within the range previously reported in the literature.

6. Other Applications

In the previous sections, we have presented a model that can be applied broadly to different types of data

[10] We should note that elasticity of advertising with respect to sales can't be computed directly from the β parameters, due to the Tobit structure of the model. Since the latent Y_{it}^* is negative for most customers in most time periods, the increase in Y_{it}^* due to β will often not translate into sales.

to gauge consumers' response to different advertising channels and to target individuals based on their responsiveness to advertising in a particular channel. In determining our empirical strategy and developing the model, we have focused on the features that are common across many data sets and, in particular, the direct marketing data set described in the previous section. In this section, we discuss several model extensions that can be used to tailor both the data collection and the modeling to other contexts.

6.1. Additional Modeling Features

There are a number of modeling extensions that one might consider, depending on the empirical setting and the richness of the data. For example, in our proposed model, the latent variable responds linearly to advertising, but others, including Hoban and Bucklin (2014), have shown experimental evidence of saturation effects at the individual level. Given data where advertising exposure varies substantially across customers, diminishing response to advertising could be modeled; see Bass et al. (2007), Dubé et al. (2005) for examples in the aggregate modeling literature. While conceptually simple, this would be computationally burdensome, requiring nonlinear techniques beyond the standard dynamic linear model (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 in data sets that are sparse in X_{ikt}, as is common in most advertising contexts, this may result in extremely weak identification of saturation or wear-out effects.

Similarly, competitive advertising effects could be incorporated in the model, were competitive advertising data available. (In our data set, there is no practical way the retailer could gain access to data on which competitor's emails or catalogs each customer has received.) If data on competitive advertising are available, as is the case with single-source panel data, they could be readily incorporated into our model by adding a cross effect term to the ad-stock equation to control for the competitors' advertising. A more comprehensive, yet computationally burdensome approach would be to simultaneously estimate stocks for each brand and model consumers' choice among brands (see Bollinger et al. 2013).

Finally, one could add controls for state dependence to the model. We have not done that here, noting that, with daily data, this was difficult to separate from serial correlation. However, a form of state dependence that allows for prior periods' incidence of purchase ($\mathbf{1}[Y_{i,\,t-1} > 0]$ or other similar parametric forms) to affect future purchases with an exponential decay rate could be easily stacked in the vector of latent stocks as shown in Section 3.

6.2. Application to Other Targeted Advertising Platforms

While we have applied our proposed method in a direct marketing context, the model that we propose could be applied to data where (1) advertising exposures are recorded at the individual level and can be linked to purchases and (2) there is a way to withhold advertising from some customers to execute randomized holdouts. All of the major digital advertising platforms—display, online video, social, mobile—allow advertisers to track users' ad exposures and sales response through a variety of mechanisms and these data are increasingly available, spawning the "multitouch attribution" industry. Hypothetically, it should also be very easy for online advertisers to execute randomized holdouts, since advertisements are served at the cookie or device level; however, the practice is not common due to several challenges in executing randomized holdouts. First, as we discussed earlier, online advertisers are not completely in control of when and (especially) how much each user is exposed to an ad. Users visit websites when they want to visit and so the opportunity to advertise to a particular cookie is under control of the user. New cookies can show up at any time, so a mechanism for executing randomized holdouts must assign cookies to holdout or treatment on the fly and address whether a user who has been exposed once should be exposed again if another opportunity arises. If the number of exposures is not controlled, then there is the potential to reintroduce a "activity bias" among those who are treated, which could induce a bias in the model estimates.

Second, most digital advertising platforms currently choose which ads to show each cookie based on how that cookie is likely to respond, in a way that will maximize revenue for the advertisng platform. It will be challenging to persuade advertising platforms to randomly select which users receive a particular campaign since this goal is in conflict with their profit motives.

Finally, even if the advertising platform will assign treatment and control groups randomly, the control group also needs to be served some ad, and so in paid media, unlike direct marketing, one has to consider what that ad is. If the control group ends up being exposed to more competitor ads, we may see a difference between treatment and control that is caused by the competitor ads and not the target ads. The standard solution to this is to pay for and serve a noncompetitive ad (e.g., a public service announcement as described in Lewis and Rao 2015 and Hoban and Bucklin 2014), yet there seems to be little appetite for this among advertisers due to the added expense of paying for the control group exposures. Add these three challenges to the overall complexity of the digital advertising ecosystem—with publishers, major ad platforms, ad resellers, ad agencies, and advertisers who would all have to coordinate—and it becomes very difficult to execute randomized holdouts on digital platforms. All of these same complexities will apply as addressable television systems are rolled out; yet, the opportunity to accurately measure response to television, arguably the most potent medium, remains enticing.

7. Conclusion

In this paper, we estimate the effects of advertising on sales using a unique panel data set where customers were randomly exposed or not exposed to each individual catalog and email mailing, within a direct marketing cadence. We first analyze these randomized controlled experiments individually, as would be common in the experimental literature, and find substantial exploratory evidence of positive advertising response. We then pooled information across campaigns using a heterogeneous consumer-level ad-stock model, which allows us to estimate the advertising response and decay for catalogs and emails and the interaction between them. 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. We then show how the model estimates, which include estimates of the effects for individual customers, can be used to improve targeting strategies, leading to substantial gains over traditional RFM targeting, common in the direct marketing industry.

Our intent was to develop an approach to measuring the short- and long-term effects of advertising across multiple channels that would be useful to practitioners across many contexts, and so the data requirements and model specification are intentionally simple. As we discuss, a critical feature of the approach is to induce exogenous variation in ad exposures, using randomized holdouts. This, as we have discussed, is rapidly becoming a possibility 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 are available and which decisions the advertiser wants to focus on. For example, if the data included information about the specific advertising copy, 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 are observed at different time 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 (Li and Kannan 2014, Abhishek et al. 2015). With sufficient variation in the advertising exposures, one could also allow for noncontemporaneous interactions between advertising channels. Bollinger et al. (2013), for instance, introduce a novel way to parameterize interactions symmetrically across stocks and exposures. Moreover, with detailed single-source panels comprising different brands and competitors, one could extend the model presented in this work in the spirit of Danaher et al. (2008) to provide individual-level estimates of cross competitor (brands) elasticities. What we have attempted to show here is that, whichever way data evolve, the framework developed in this work, which combines randomized field experiments with parametric modeling, is a flexible basis for a decision support tool to manage multichannel advertising.

Acknowledgments
The authors thank the Wharton Customer Analytics Initiative (WCAI) and one of its corporate partners that provided data. WCAI staff Melissa Hartz, Ben Adams, and Ankur Roy all assisted with data preparation.

Appendix 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 three blocks:

1. Sample W_{ikt} for each individual using modified second-order efficient DLM relationships accounting for the truncation due to the Tobit model.
2. Sample β_i, ρ_i, γ_i, ϕ_i, σ^2_η, σ^2_ε for each individual using Metropolis–Hastings updates.
3. Sample $\bar{\beta}$, $\bar{\Sigma}_\beta$, $\bar{\gamma}$, $\bar{\Sigma}_\gamma$, $\bar{\rho}$, and $\bar{\Sigma}_\rho$ using standard conjugate draws (Rossi et al. 2005, pp. 72–73).

We provide details of the initialization and steps 1 and 2 below.

Initialization
To initialize the relevant parameters in the model, we estimate an auxiliary pooled model by Rossi et al. (2005, Chaps. 4 and 5), i.e., a model with common parameters for all individuals, where the latent stocks for each period are constructed explicitly based on 21 periods (three weeks). Specifically, we estimate the model

$$\begin{gathered}\mathbf{Y}^*_t = \boldsymbol{\mu} + \sum_{k=1}^{K} \boldsymbol{\beta}_k \sum_{j=0}^{20} \rho_k^{j-1} \mathbf{X}_{t-j,\,k} + \gamma \mathbf{Z}_t + \boldsymbol{\eta}_t, \\ \mathbf{Y}^*_t = \mathbf{Y}_t \quad \text{if } \mathbf{Y}_t > 0,\, \mathbf{Y}^*_t < 0 \text{ if } \mathbf{Y}_t = 0,\end{gathered} \tag{A1}$$

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

The latent stocks W_{ik0} are initialized at the "steady state" implied by the maximum likelihood estimates. (See Smith 2007 for an accessible introduction to the long-run and stationary properties of state-space systems.) 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.

Step 1. Update of the Latent Stocks
Recall the vectorized system of equations determining the dynamics of the stocks in the latent space:

$$Y_t^* = \mu + \beta \mathbf{W}_t + \gamma \mathbf{Z}_t + \eta_t,$$
$$\mathbf{W}_t = \rho \mathbf{W}_{t-1} + \mathbf{X}_t + \varepsilon_t.$$

(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 $\mathbf{W}_t \sim N(\mu_{Wt}^t, V_{Wt}^t)$, 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 1999). These statistical summaries are derived in two steps commonly referred as the forward filtering algorithm (FF). For time $t = 1, \ldots, T-1$,

- Forward filtering when Y_{t+1} is positive (nothing changes with respect to classic Kalman DLM relationships):

$$\mu_{W_t}^{t+1} = \rho \mu_{W_t}^t + \mathbf{X}_t,$$
$$V_{W_t}^{t+1} = \rho V_{W_t}^t \rho' + \sigma_\varepsilon^2,$$
$$k_{t+1} = V_{W_t}^{t+1} \beta' (\beta V_{W_t}^{t+1} \beta + \sigma_\eta^2)^{-1},$$
$$\mu_{W_{t+1}}^{t+1} = \mu_{W_t}^{t+1} + k_{t+1}(Y_{t+1}^* - \mu - \beta \mu_{W_t}^{t+1} - \gamma \mathbf{Z}_t),$$
$$V_{W_{t+1}}^{t+1} = V_{W_t}^{t+1} - k_{t+1}(\beta V_{W_t}^{t+1} \beta + \sigma_\eta^2) k_{t+1}'.$$

- Forward filtering when Y_{t+1} is zero (using second-order sufficient statistics):

$$\mu_{W_t}^{t+1} = \rho \mu_{W_t}^t + \mathbf{X}_t,$$
$$V_{W_t}^{t+1} = \rho V_{W_t}^t \rho' + \sigma_\varepsilon^2,$$
$$k_{t+1} = V_{W_t}^{t+1} \beta' (\beta V_{W_t}^{t+1} \beta + \sigma_\eta^2)^{-1},$$
$$\mu_{W_{t+1}}^{t+1} = \mu_{W_t}^{t+1} + k_{t+1}(E_t(Y_{t+1}^* \mid Y_{t+1}^* < 0) - \mu - \beta \mu_{W_t}^{t+1} - \gamma \mathbf{Z}_t),$$
$$V_{W_{t+1}}^{t+1} = V_{W_t}^{t+1} - k_{t+1}(\beta V_{W_t}^{t+1} \beta + \sigma_\eta^2 - \mathrm{Var}_t(Y_{t+1}^* \mid Y_{t+1}^* < 0)) k_{t+1}'.$$

Note that the corresponding conditional expectations under truncation can be easily computed using the formulas in Harvey (1991, Section 3.7.2) and Shepard (1994) and, more recently, by Liao et al. (2014), who present a multivariate extension that can be used to estimate dynamic higher-order Tobit/Probit models.

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-square-error 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 backward in time to obtain backward sampling (BS):

$$g_t = V_{W_t}^t \beta' (V_{W_t}^{t+1})^{-1},$$
$$\mu_{W_t}^T = \mu_{W_t}^t + g_t(\mu_{W_{t+1}}^T - \mu_{W_t}^t),$$
$$V_{W_t}^T = V_{W_t}^t - g_t(V_{W_t}^{t+1} - V_{W_{t+1}}^T) g_t'.$$

A detailed treatment of the derivations leading to the above can be found in West and Harrison (1999) 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:

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

These can be used for prediction or evaluation of counterfactual scenarios.

Step 2. Update the Individual-Level Parameters
We draw the individual-level parameters using a Metropolis algorithm that uses candidate sampling distributions that are customized to the unit-level likelihoods. We use a "fractional likelihood" approach as in Rossi et al. (2005, p. 135) to set the proposal density for each individual:

$$L_i^* = L_i \cdot \bar{L}^\alpha, \tag{A2}$$

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 individual-level parameters that maximizes the likelihood in Equation (A2) and let $\hat{V}_i = -\partial^2 L_i^* / \partial\beta\partial\beta' |_{\beta=\hat{\beta}_i}$. These can then be combined with the priors presented in Section 3.2 to form a Metropolis proposal distribution. The update for the individual parameters then uses a standard Metropolis–Hastings update with an additional rejection step to ensure that the ρ_i and ϕ_i coefficients are sampled from the stationary region as in Chib (1992).

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 postprocessing the draws as in Rossi et al. (2005, Chap. 4) or (equivalently) restricting the βs over the positive real line for those individuals having sparse exposures.

Appendix B. Empirical Identification: Parameter Recovery and Initial Conditions

Table B.1 illustrates the recovery of the parameters at the population level for a simulated example in which serial correlation has a positive carryover coefficient. This scenario can be considered as pessimistic, as, in principle, positive serial correlation should render the recovery of the parameters more difficult. In detail, we generate two years of daily observations for 300 customers according to the population parameters presented in the second column of Table B.1. The fraction of time periods in which the customer does not buy is similar to the data presented in the empirical analysis.

We then repeat the same analysis without the portion of the algorithm (imputation of the initial condition by backward smoothing step) responsible for the estimation of the initial condition. In this way we can gauge the robustness of the estimation algorithm to the initial conditions known to be problematic in panels with unobserved heterogeneity (Wooldridge 2005).

We note that the results in Table B.2 are not necessarily "biased," but are moderately more dispersed around the true values. The practical recommendation that we offer is to use the imputation step for the initial condition to reduce variability of the estimates even in presence of a long time horizon consistent with the findings presented in Naik and Raman (2003).

Table B.1 Parameter Recovery for the Population Parameters in the Complete Model Specification

Population parameter	True	Posterior mean	Post q2.5	Post q97.5
$\bar{\mu}$	−2.00	−1.99	−3.30	−0.58
$\bar{\beta}_{CAT}$	1.00	1.02	0.61	1.45
$\bar{\beta}_{EM}$	2.00	1.95	1.79	2.08
$\bar{\beta}_{k:k'}$	0.5	0.52	0.44	0.60
$\bar{\rho}_{CAT}$	0.66	0.65	0.51	0.79
$\bar{\rho}_{EM}$	0.40	0.43	0.32	0.55
$\bar{\rho}_{k:k'}$	0.20	0.20	0.04	0.39
$\bar{\phi}_C$	0.80	0.76	0.60	0.93

Table B.2 Parameter Recovery for the Population Parameters for the Algorithm Without the Imputation of the Initial Conditions

Population parameter	True	Posterior mean	Post q2.5	Post q97.5
$\bar{\mu}$	−2.00	−1.98	−3.40	−0.45
$\bar{\beta}_{CAT}$	1.00	1.05	0.50	1.58
$\bar{\beta}_{EM}$	2.00	2.04	1.74	2.33
$\bar{\beta}_{k:k'}$	0.5	0.48	0.25	0.73
$\bar{\rho}_{CAT}$	0.66	0.63	0.39	0.85
$\bar{\rho}_{EM}$	0.40	0.47	0.22	0.70
$\bar{\rho}_{k:k'}$	0.20	0.24	0.02	0.42
$\bar{\phi}_C$	0.80	0.76	0.50	0.97

Note. All latent stocks are initialized with a $\mu_{W_0} = 0$ and $V_{W_0} = I$.

References

Abhishek KH, Hosanagar K, Fader P (2015) Media exposure through the funnel: A model of6 multistage attribution. Working paper, Carnegie Mellon University, Pittsburgh.

Ataman B, van Heerde HJ, Mela CF (2010) The long-term effect of marketing strategy on brand sales. *J. Marketing Res.* 47(5):866–882.

Baier M (1996) *How to Find and Cultivate Customers Through Direct Marketing* (NTC Business Books, Lincolnwood, IL).

Barajas J, Akella R, Flores A, Holtan M (2016) Experimental designs and estimation for online display advertising attribution in marketplaces. *Marketing Sci.*, ePub ahead of print May 3, http://dx.doi.org/10.1287/mksc.2016.0982.

Bass F, Bruce N, Majumdar S, Murthi BPS (2007) Wearout effects of different advertising themes: A dynamic Bayesian model of the advertising–sales relationship. *Marketing Sci.* 26(2):179–195.

Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica.* 83(1):155–174.

Bollinger BK, Cohen MA, Lai J (2013) Measuring asymmetric persistence and interaction effects of media across platforms. Working paper, Duke University, Durham, NC.

Bowman D, Gatignon H (2010) *Market Response and Marketing Mix Models: Trends and Research Opportunities, Foundations and Trends in Marketing* (Now Publishers, Delft, Netherlands).

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

Broadbent S (1984) Modeling with ad-stock. *J. Market Res. Soc.* 26(4):295–312.

Chib S (1992) Bayes inference in the Tobit censored regression model. *J. Econometrics* 51(1–2):79–99.

Danaher PJ, Bonfrer A, Dhar SS (2008) The effect of competitive advertising interference on sales for packaged goods. *J. Marketing Res.* 45(2):211–225.

Danaher PJ, Dagger TS (2013) Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. *J. Marketing Res.* 50(4):517–534.

Danaher PJ, van Heerde HJ (2015) Delusion in attribution: Caveats in attributing purchases to multimedia exposures. Working paper, Monash University, Melbourne, Australia.

Deighton J, Henderson CM, Neslin SA (1994) The effects of advertising on brand switching and repeat purchasing. *J. Marketing Res.* 31(1):28–43.

Dekimpe MG, Hanssens DM (2000) Time-series models in marketing: Past, present and future. *Internat. J. Res. Marketing* (17):183–193.

Dinner IM, Van Heerde HJ, Neslin S (2014) Driving online and offline sales: The cross-channels effects of digital versus traditional advertising. *J. Marketing Res.* 51(5):527–545.

Dong X, Manchanda P, Chintagunta PK (2009) Quantifying the benefits of individual-level targeting in the presence of firm strategic behavior. *J. Marketing Res.* 46(2):207–221.

Duan JA, McAlister L, Sinha S (2011) Reexamining Bayesian-model-comparison evidence of cross-brand pass-through. *Marketing Sci.* 30(3):550–561.

Dubé JP, Hitsch G, Manchanda P (2005) An empirical model of advertising dynamics. *Quant. Marketing Econom.* 3(2):107–144.

Dubé JP, Hitsch G, Rossi PE (2010) State dependence and alternative explanations for consumer inertia. *RAND J. Econom.* 41(3): 417–445.

Eastlack JO Jr, Rao AG (1989) Advertising experiments at the Campbell Soup Company. *Marketing Sci.* 8(1):57–71.

Gelfand A, Ghosh SK (1998) Model choice: A minimum posterior predictive loss approach. *Biometrika* 85(1):1–11.

Gelman A, Carlin JB, Stern HS, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis*, 3rd ed. (Chapman and Hall/CRC, Boca Raton, FL).

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

Greene W (2008) *Econometric Analysis*, 6th ed. (Pearson Prentice Hall, Upper Saddle River, NJ).

Harvey AC (1991) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge, UK).

Hoban PR, Bucklin RE (2014) Effects of Internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *J. Marketing Res.* 52(3):375–393.

Huang J, Leng M, Liang L (2012) Recent developments in dynamic advertising research. *Eur. J. Oper. Res.* 220(3):591–609.

Johnson GA, Lewis RA, Reiley DH (2016) Location, location, location: Repetition and proximity increases advertising effectiveness. Working paper, University of Rochester, Rochester, NY.

Jorgenson DW (1966) Rational distributed lags. *Econometrica* 32(1): 135–149.

Kalyanam K, McAteer J, Marek J, Hodges J, Lin L (2015) Cross channel effects of search engine advertising on brick and mortar retail sales: Insights from multiple large scale field experiments on Google.com. Working paper, Santa Clara University, Santa Clara, CA.

Koyck LM (1954) *Distributed Lags and Investment Analysis* (North-Holland, Amsterdam).

Lewis RA, Rao JM (2015) On the unfavorable economics of measuring the returns to advertising. *Quart. J. Econom.* 130(4):1941–1973.

Lewis RA, Rao JM, Reiley DH (2011) Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. *Proc. 20th Internat. Conf. World Wide Web* (ACM, New York), 157–166.

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

Liao Y, Anderson HM, Vahid F (2014) Multivariate dynamic Tobit model. Working paper, Queensland University of Technology, Brisbane, Australia.

Little JDC (1979) Aggregate advertising model: The state of the art. *Oper. Res.* 27(4):629–667.

Liu Q, Otter T, Allenby GM (2007) Investigating endogeneity bias in marketing. *Marketing Sci.* 26(5):642–650.

Lodish LM, Abraham MM, Livelsberger J, Lubetkin B, Richardson B, Stevens ME (1995a) A summary of fifty-five in-market experimental estimates of the long-term effect of TV advertising. *Marketing Sci.* 14(3):G133–G140.

Lodish LM, Abraham M, Kalmenson S, Livelsberger J, Lubetkin B, Richardson B, Stevens ME (1995b) How T.V. advertising works: A meta-analysis of 389 real world split cable T.V. advertising experiments. *J. Marketing Res.* 32(2):125–139.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with non-random marketing mix variables. *J. Marketing Res.* 41(3):467–478.

Manchanda P, Dubé JP, Goh KY, Chintagunta P (2006) The effect of banner advertising on internet purchasing. *J. Marketing Res.* 43(1):98–108.

McDonald JF, Moffitt RA (1981) The uses of Tobit analysis. *Rev. Econom. Statist.* 62(2):318–321.

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

Naik PA, Raman K (2003) Understanding the impact of synergy in multimedia communications. *J. Marketing Res.* 40(3): 375–388.

Naik PA, Raman K, Winer RS (2005) Planning marketing-mix strategies in presence of interactions. *Marketing Sci.* 24(1):25–34.

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

Pedrick JH, Zufryden FS (1991) Evaluating the impact of advertising media plans: A model of consumer purchase dynamics using single-source data. *Marketing Sci.* 10(2):111–130.

Raftery AE, Lewis SM (1996) Implementing Markov chain Monte carlo. Gilks WR, Spiegelhalter DJ, Richardson S, eds. *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London), 115–130.

Rossi PE (2014) Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Sci.* 5(33):655–672.

Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.

Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, New York).

Sahni NS (2015a) Advertising spillovers: Field experimental evidence and implications for the advertising sales–response curve. Working paper, Stanford University, Stanford, CA.

Sahni NS (2015b) Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quant. Marketing Econom.* 13(3):203–247.

Sahni NS, Zou D, Chintagunta PK (2016) Do targeted discount offers serve as advertising? Evidence from 70 field experiments. *Management Sci.* Forthcoming.

Sethuraman R, Tellis GJ, Briesch RA (2011) How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *J. Marketing Res.* 48(3):457–471.

Shepard N (1994) Partial non-Gaussian state space models. *Biometrika* 1(81):115–131.

Shirley K, Small D, Lynch K, Maisto S, Oslin D (2010) Hidden Markov models for alcoholism treatment trial data. *Ann. Appl. Statist.* 4(1):366–395.

Smith JO III (2007) *Introduction to Digital Filters with Audio Applications* (W3K Publishing), http://www.w3k.org/books/.

Tellis GJ (1998) *Advertising and Sales Promotion Strategy* (Reading, MA).

Tellis GJ (2004) *Effective Advertising: Understanding When, How, and Why Advertising Works* (Sage Publications, Thousand Oaks, CA).

Tellis GJ, Ambler T (2007) *Handbook of Advertising* (Sage Publications, London).

West M, Harrison J (1999) *Bayesian Forecasting and Dynamic Models*, 2nd ed. (Springer, New York).

Wilbur KC (2008) A two-sided, empirical model of television advertising and viewing markets. *Marketing Sci.* 27(3):356–378.

Wooldridge JM (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *J. Appl. Econometrics* 20(1):39–54.

Xu L, Duan J, Whinston A (2014) Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Sci.* 60(6):1392–1412.