A Dynamic Model for Digital Advertising:

The Effects of Creative Formats, Message Content and Targeting on Engagement

Norris I. Bruce
B.P.S. Murthi
Ram C. Rao

April 10, 2016

*** Please do not cite, quote or distribute without permission ***

Norris Bruce (972.883.6293, nxb018100@utdallas.edu) is an Associate Professor, B.P.S. Murthi (992-883-6355,murthi@utdallas.edu) is a Professor, and Ram C. Rao (972-883-2580, rrao@utdallas.edu) is the Founders Professor in marketing at Naveen Jindal School of Management at The University of Texas at Dallas, P.O. Box 830688, Richardson, TX 75083.

A Dynamic Model for Digital Advertising:

The Effects of Creative Formats, Message Content and Targeting on Engagement

ABSTRACT

The authors study the joint effects of creative format, message content, and targeting on the performance of digital ads over time. Specifically, they present a dynamic model to measure the effects of different sized static (GIF) and animated (Flash) display ad formats; and consider whether different ad contents, related to brand or a price offer, are more or less effective for different ad formats and targeted or re-targeted consumers. To this end, they obtain daily impressions, clicks, targeting, and ad creative content data for a six-month period, from a major US retailer; and develop a dynamic zero-inflated (DZI) count model. Given the sparse, non-linear, and non-Gaussian nature of the data, the study designs a particle filter/MCMC scheme for estimation. Results show that carryover rates for dynamic formats are greater than for static formats; yet static format can still be effective for price ads and re-targeting. Most interesting, results also show that re-targeted ads are effective only if they offer price incentives. The study then considers the import of these results for the retailer's media schedules.

Keywords: Online Advertising, Ad Formats (static versus animated), Ad content, Media Planner, Dynamic Zero-Inflated Poisson (DZIP), Particle Filtering/Smoothing, Sequential Monte Carlo (SMC), Markov Chain Monte Carlo.

INTRODUCTION

Advertisers often use multiple creative formats in their digital campaigns to target, and re-target, consumers with product-based messages and price incentives. These include static formats (GIF/JPG) that offer neither animation nor interactivity; simple flash formats (*.swf* format) that offer animation but no interactivity; and rich-media formats that offer both interactivity and animation, with elements such as sound, video, floating images, and takeovers, and so on. As a result, advertisers have the non-trivial task of jointly assessing the effects over time of design elements available in the large number of such formats as they decide on budgets, message objectives, and consumer targeting. There is however some evidence from industry studies that ad format size, location, and creative elements such as color, interactivity, and animation may all independently influence engagement (e.g., DoubleClick 2009, Cole et al. 2009). Yet this raises the difficult questions. For example, a retailer may still ask which message content, product-based or price incentives, is more suitable for animated and static ads; or which ad formats and message are more effective for re-targeting, the canonical tactic of tracking visitors to the firm's site, and then serving the firm's ads to them once they visit other sites (Lambrecht and Tucker 2013)?

The retailer may also be interested in the temporal effects of online ads but extant work has been largely cross sectional and so cannot help to formulate dynamic advertising strategies (Breuer et al 2011). Internet ad exposure-models in marketing (e.g., Danaher 2007; Danaher et al. 2010) have however explored the performance of ad formats over time. Thus, it is necessary to consider not only when formats work, but also how long these effects persist, so one may better match formats to ad messages and targeted consumers (e.g., Tellis et al. 2000). For instance, a large body of work on offline ads suggests that ads have instantaneous and long-term

or carryover effects (e.g., Sethuraman et al. 2011). Yet studies of digital ads have largely ignored carryover, attributing consumer engagement to recent impressions. Braun and Moe (2013) model carryover effects but treat it as homogenous. Given the evidence that carryover may differ, for example, across email and online channels (Breuer et al. 2011) a useful direction to explore is to model heterogeneous carryover effects. Furthermore, we know that the effects of ad messages may vary across media and markets (MacInnis et al. 2002; Deighton et al. 1994; D'Souza and Rao 1995), and so we may want to consider how such effects differ across online re-targeted consumers. Knowing these features of digital ads, the effects of carryover, format, target, and message could help managers improve ad engagement. This, in turn, could ultimately help firms better allocate ad resources throughout their digital advertising campaigns (Rust and Leon, 1984).

This study attempts to fill some of the central gaps in extant work by developing a dynamic response model to study the joint effects of creative format, message content, and targeting/re-targeting on the performance of digital ads over time. Specifically, the model examines the dynamic effects of thematic ads, price or product; presented in multiple creative formats, animated and static ads of varied sizes; contrasting the clicking behaviors of targeted and retargeted consumers. The substantive questions addressed are: How do carryover effects vary across animated and static ads, and their targeted consumers? What is the effect of format (size/position) on consumer clicking behavior? In addition, prior research posits that ad format effectiveness can vary with ad copy elements (Naik et al. 1998; Grass and Wallace 1969). Thus, what are the effects of price incentives and product ads within different digital formats? Most interesting, which ad formats and copy (price vs. product) are more effective for re-targeting, the industry standard tactic just described.

We also innovate methodologically to be able to extend econometric studies of advertising's dynamic and content effects (e.g., Clarke 1976; MacInnis et al. 2002; Tellis 2004; Chandy et al. 2001; Bass et al. 2007) to the domain of of digital advertising. Digital ad response data, consisting of clicks in this study, are time series of counts, which contains a high frequency of zeros due to non-response that result in "zero-inflation". This presents a challenge. Failure to account for zero-inflation and/or dynamics may result in misleading inference and the detection of spurious associations. To address these two concerns, and the above substantive questions, we propose a dynamic, state space, zero-inflated count model (e.g., Poisson, Negative-Binomial). The resulting response model is both dynamic and nonlinear, and therefore we estimate it using a combination of Particle Filtering and MCMC procedures (Liu and Chen 1998, Doucet et al. 2001, Ristic et al. 2004,). Particle Filtering in its many variations is widely applied in statistics; it is a flexible Bayesian inferential method used to estimate nonlinear/non-Normal dynamic systems. In these systems, the posterior distributions of the state space parameters are analytically intractable, and hence the filter operates by drawing weighted samples from a time-varying proposal distribution (i.e., an importance function). The analytic expression for the optimal form of this importance distribution, optimal in terms of computational efficiency, is available only in special cases (e.g., Doucet et al. 2001). It is possible, however, to obtain a linear/normal approximation of this function at its mode, where the mode arises from an iterative Newton-Raphson step embedded within the particle filter. The resulting algorithm provides an approach to estimate any state space model within the exponential family (Doucet 2000), and is more general than Gaussian filters such as the extended and the unscented Kalman Filters (Ristic et al. 2004, pp. 32).

The paper, therefore, contributes to an emerging literature on digital ad response models in the following ways. First, we find that animated ads have significantly higher carryover effects and impact consumer engagement over a longer duration than static ads, in all ad formats and among both targeted and retargeted consumers. Second, within the animated formats, price ads are more effective than product ads. Third, re-targeted ads are effective only when they offer price incentives, a finding consistent with Lambrecht and Tucker 2013 who found them effective only when consumers had strong preferences such that they have incentives to buy. Fourth, all ads (i.e., by formats and messages) targeted to the female segment were effective; this suggests that, in our example, female shoppers are more willing to engage, perhaps confirming the brick mortar studies that "women shop, men buy.[1]" Fifth, to answer our questions we had to introduce new Bayesian methods that respect features of digital ad data, and the underlying non-linear dynamics process that generates it. We believe the approach, our main contribution, complements other methods, including bandit problems (sequential experiments) and the Bayesian algorithms (e.g., Thompson Sampling) used to study them (Scott 2010); for modeling non-linear dynamics, data sparsity (e.g., non-response), and the effects of multiple exposures (i.e., ad repetition) present challenges to such algorithms (e.g., Agarwal 2010; Schwartz et al. 2013). Admittedly, hierarchical modeling (i.e., Bayesian) can help obviate the sparsity problem when using Thompson Sampling, but dynamics and multiple exposures are not straightforward extensions. Finally, we conduct simulations to show the import of our findings; these should be of interest to online retailers and digital media planners.

[1] 'Men Buy, Women Shop': The Sexes Have Different Priorities When Walking Down the Aisles, Knowledge@Wharton, Nov 28, 2007.

To address the questions in the study, we obtained panel data from a major US retailer, in an industry that provides products and services for the home. The data offers a selection of daily ad impressions and their associated clicks, with both clicks and impressions disaggregated by consumer targets, ad format and message content; ad networks commonly release such data to their clients. Specifically, the data contains a panel of click counts for 154 days, across six creative formats, and four targeted segments: formats Flash and GIF[2] by size and orientation, 160x600, 300x250, and 728x90; and segments classified as re-targeted, male, female, and age. A unique feature of the data is that daily impressions (within format and target) cluster into price, product, and control impressions, where price impressions are price promotion ads; product impressions are ads that stress brand benefits excluding price; and control impressions are blank impressions used to exclude non-US consumers from viewing specific ads. We (and the retailer) note that these blank ads often artificially inflate engagement because viewers click on them, largely from curiosity (e.g., white objects become visually salient) but also in error (McConnell 2012). Finally, we model impressions as potentially endogenous (Lee et al. 2015) because they may depend on omitted factors such as website content, format type, or clicking history.

The remainder of the paper is organized as follows. The next section provides a brief review of the relevant streams in the advertising content and dynamic effects literatures. Subsequent sections develop the empirical model, and describe its estimation and the data we employ, in that order. The last two sections describe our estimation results and conduct simulations to summarize their impact on a hypothetical media schedule. The paper concludes with an overview of the findings, and the limitations of the study.

[2] Simple Flash: Ads with animation frames, multiple click-through buttons, but without interactive elements. GIF: Ads with no animation frames, a single click-through button, but without interactive elements.

LITERATURE

We provide a brief review of the academic literature relevant to the effects of ad formats (size and animation), content, targeting, and carryover rates, all on response metrics such as click-through rates (CTR), attention and recall. Admittedly, we know much about the effectiveness of traditional ads, but our understanding of the effectiveness of digital ads is rapidly evolving. This review reflects that notion.

The Impact of Ad size

While one would expect larger banner ads to be more effective than smaller ads (*ceteris paribus*), the evidence seems inconclusive. Larger ads seemingly could improve memory for products, and are more likely to be seen and remembered relative to smaller ads (Cho 1999, Chandon et al. 2003). They have also been associated with greater attention and response (Baltas 2003); greater intention to spread positive word-of-mouth (Chtourou and Chandon 2000); higher recall (Chatterjee et al 2008); and higher click through rates (Rettie et al. 2004, Robinson et al. 2007) [3]. Yet Dreze and Hussherr (2003) and Cho (2003) found no significant effect of ad size on engagement. They suggested that users learn to avoid looking at ads, even though the ads may affect them through their peripheral vision. These differing results suggest perhaps a tension between the ability of large ads to attract attention and their more intrusive nature that leads to avoidance. Thus, the problem needs more study, with a focus on both the research methods and the ad context (e.g. type of products/websites).

The Impact of animation

Experiments on the other hand confirm that animation in banner ads can attract user's attention and increase engagement. For example, Li and Bukovac (1999) found that users were

[3] Given its interest in ensuring continued growth of digital advertising, not surprisingly DoubleClick (2009) also reported that click-through rates for large ads (300x600 and 240x400) were three times greater than that for smaller ad formats.

able to quickly identify and better recall animated banner ads than static banner ads. Cho et al. (2001) showed that a higher degree of forced exposure to animated banners ads yielded higher click-through rates and favorable attitudes among users. Animation has also been associated with greater clicking behavior in econometric studies (Lohtia et al. 2003; Hong et al. 2007 and Tsang and Tse 2007). These proposed that when consumers have not decided on the items they want, they are more likely to click on animated ads as they may attribute a higher quality to the advertised products or pay greater attention. Similarly, other works suggested that the animation is more likely to be effective when user experience and brand familiarity are low (Dahlen 2001); or when users are searching for fun rather than for specific information (Tuten et al. 2000).

Impact of Ad Content

There are several major studies on the effects of ad creative or content in off-line advertising; but there are, as of yet, no major studies considering these effects in digital ads. For example, MacInnis et al. (2002) found in a study of TV commercials that emotional content was more likely to increase sales; and ads that used rational appeal were less likely to produce increases. Chandy et al. (2001) studied the effects of advertising on sales across multiple creatives. While many creative executions were ineffective in increasing sales, they confirmed that emotional ads were more effective in mature markets, and argument-based appeals more effective in newer markets. Similarly, Bass et al. (2007) found that rational ads wear out faster than emotional ads: for example, price advertising had the highest wear-out among all appeals. Still, in the digital space evidence about the role of ad content is still emerging. Chtourou et al. (2001) suggested that banner ads with promotional incentives have higher click-through rates than those that lacked incentive offers. Xie et al. (2004) also found evidence that incentive offers improved click-through rates, but the effect varied by the type of appeal, rational vs. emotional.

Thus, for example, banners with positive emotional appeals and incentive offers generated higher click-through than those with positive appeals and no incentives. Similarly, Maureen and Grey (2005) showed that banner ads that offer a free sample achieved higher click-through than banner ads with information only. Braun and Moe (2013) also found that the effects of creative content in banner ads could differ, even though their data did not ascribe substantive meaning to these contents. Nevertheless, taken together, these studies indicate the importance of ad content for digital media.

Impacts of Ad targeting

The marketing literature has shown that more precise targeting can increase click through rates of banner ads (Briggs & Hollis 1997; Sherman & Deighton 2001; Chandon et al. 2003; Chatterjee et al. 2003). For example, with regards to retail shopping, there is some evidence that women are more invested in the experience, and thus more likely to spend more time browsing online; in contrast, men are more goal oriented (Passyn et al. 2011); and for many product categories, women are the primary purchasers. Moreover, given current and exact technology, once consumers browse a firm's website, an ad network can use their browsing histories to serve that firm's banner ads to them when they the visit other sites. Research suggests that such *re-targeted ads* are on average surprisingly ineffective; unless the consumers' preferences for products earlier viewed are well defined; i.e., "they have a detailed view of what product they wish to purchase" (Lambrecht and Tucker 2013, pp. 2). This suggests that re-targeted ads that offer the consumer incentives to buy should, on average, be more effective than ads that merely provide non-price information.

Carryover Rates for Different Media and Target

Braun and Moe (2013) evaluated the carryover effects of banner ads in a model designed to study effectiveness of creative content; where carryover is the extent to which past impressions, affect the contemporaneous effects of banner ads on response behavior (Bass et al. 2007). The study uses data across individuals to obtain a homogenous estimate of carryover. It is however well documented that the effects of advertising and hence carryover could differ across channel, target markets and media. For example, Sethuraman et al. (2011) reported, from a meta-analysis of 56 studies, that television advertising has higher short-term elasticity but lower long-term elasticity as compared to print advertising. Berkowitz et al. (2001) modeled weekly data from three stores of a large national retailer, and found that the carryover effect of radio was higher than that of billboards. Similarly, Naik and Raman (2003) in a study that considered the media synergy found that carryover for TV was approximately 2.5 times that of print. With regard to markets, the literature (Deighton et al.1994; D'Souza and Rao 1995) reported that advertising is more effective among consumers who are more loyal, and that advertising is more effective for experience than search goods (Hoch and Ha 1986). Finally, Breuer et al. 2011 found in a study of online channels that email advertising had the longer effect than banner advertising. This raise questions of how carryover effects may vary across different online formats, animated and static, and for different targets. Our study provides some answers to these questions.

DYNAMIC MODEL OF DIGITAL ADVERTISING

We now present a non-linear, state space model to track the effectiveness of online display ads over time, across digital formats and targets. The model adopts an observation equation in which daily clicks follow an event-count distribution (e.g., Poisson), extended

however to allow for different forms of non-response (e.g., zero clicks), for it can be shown that the presence of zeros in count data may lead to over-dispersion, where the variance of the count distribution exceeds its mean (Green 1994). This and other forms of dispersion violate basic assumptions in the standard event-count models. The state equation, on the other hand, assumes a model of advertising goodwill (Nerlove-Arrow, 1962) in which goodwill evolves over time as a function of banner size, animation, ad targeting, and different thematic impressions (price and product). The natural *thematic variation* in the data allows for the identification of the effects of price and product ads. Moreover, we control for the potential endogeneity of targeted impressions because such impressions may co-vary with unobservable site content, format type or clicking history.

First, in a dynamic model of display advertising, one has to account for presence of excess zeros (zero-inflation) because of the typical low level of response to digital ads. Here one may observe zero-clicks because online consumers are unaware of advertising impressions; or they are aware, but choose not to respond for numerous factors, many unobservable to the researcher. In the former case, we observe what is often termed structural zeros, which are inevitable; in the latter case, we observe *sampling* zeros, which occur at random; both events emerging from potentially distinct data generating processes (e.g., Greene, 1994; Lambert 1992).

Let $f\left(y_{ijkt} \middle| \lambda_{ijkt}\right)$ be the distribution for the random number of clicks y_{ijkt} during period *t*, for format i (flash or GIF) and size j (728x90, 300x250, 160x600), and target *k* with mean $E(Y_{ijkt}) = \lambda_{ijkt}$. For now, we adopt the familiar Poisson count distribution (Poisson-*P*) and in a later section consider a linear model, and other count models, including the negative binomial and the zero hurdle models:

1) $$f^{P}\left(y_{ijkt}\middle|\lambda_{ijkt}\right)=\frac{\exp\left(-\lambda_{ijkt}\right)\lambda_{ijt}^{\;y_{ijkt}}}{y_{ijkt}!}$$

As just described, we may observe zero clicks under distinct data generating processes: we may observe structural zeros when (say) impressions go unnoticed; or randomly as a count event from equation (1). Suppose p_{ijkt} is the probability of observing structural zero clicks for format i, size j, and target k; and conversely, $1-p_{ijkt}$ the probability that click-through occurs at some rate λ_{ijkt}, with $I(Y_{ijkt}=0)$ an indicator function; then the distribution of clicks is the following:

2) $$\pi\left(Y_{ijkt}=y_{ijkt}\right)=\begin{cases} p_{ijkt}+\left(1-p_{ijkt}\right)f\left(0\middle|\lambda_{ijkt}\right) \text{ if } y_{ijkt}=0 \\ \left(1-p_{ijkt}\right)f\left(y_{ijkt}\middle|\lambda_{ijkt}\right) \qquad y_{ijkt}>0 \end{cases}$$

or equivalently,

3) $$\pi\left(Y_{ijkt}=y_{ijkt}\right)=p_{ijkt}I\left(y_{ijkt}=0\right)+\left(1-p_{ijkt}\right)f\left(y_{ijkt}\middle|\lambda_{ijkt}\right)$$

$$\text{for } i=1,2,..,I;\ j=1,2,..,J;\ k=1,2,,,K;\ t=1,2,..,T.$$

Note that equations (2) and (3) can be viewed as a two-component mixture of an ordinary count distribution $f\left(y_{ijkt}\middle|\lambda_{ijkt}\right)$ and a degenerate distribution having a point mass at zero; that is, the probability of no-response $\pi\left(Y_{ijkt}=0\right)$ is a weighted average of both outcomes described above. Note here the standard Poisson count model (obtained when $p_{ijk}=0$) is fully embedded in equations (2) and (3). Finally, let c_{ijkt} be a dichotomous variable which indicates whether the observed response (i.e., non-response) comes from the degenerate ($c_{ijkt}=1$) or randomly from the ordinary even-count component ($c_{ijkt}=0$). We propose a simple hierarchical model for c_{ijkt}, where $c_{ijkt}\sim Bernoulli(p_{ijk})$, and the probability p_{ijk} of the degenerate event has the logistic transformation, $p_{ijk}=\left\{1+\exp\left(-\gamma_{ijk}\right)\right\}^{-1}$. Therefore, equations (1-3) constitute a zero-inflated Poisson (ZIP) (Green 1994; Lambert 1992); it is the most widely applied technique for

addressing over-dispersion in count data. Finally, the log-likelihood contribution from a single format and target in the Poisson case is therefore:

4) $$LL_{ijk} = \sum_{t=1}^{T} \log\left\{p_{ijk} I\left(y_{ijkt}=0\right) + \left(1 - p_{ijk}\right)\exp\left(-\lambda_{ijkt}\right)\lambda_{ijkt}^{\,y_{ijkt}} / y_{ijkt}!\right\}$$

Given the above familiar framework, one can now develop a model to study the dynamic effects of advertising throughout a digital campaign. For instance, one may ask whether online ads like offline ads exhibit carryover; and if they do, then does carryover vary by digital media formats and targets? That is, whether a consumer's decision to click on ad at time *t* depends not only on the current, but also on past or carryover impressions. One could assess, too, whether some formats are better for different thematic impressions; or specifically whether incentive based message are better for re-targeting; and assess the impact of size and animation on digital ad effectiveness. To do this, we employ a flexible state space model, where unobservable mean clicks λ_{ijkt} in equations (1-4) evolve over time in the following multiplicative way:

5) $$\lambda_{ijkt} = \lambda_{ijkt-1}^{\delta_{ijk}} \exp\left(\alpha_{ik} + \sigma_{jk} + \sum_{l=1}^{L} \beta_{ijkl} \bar{f}(a_{ijklt}) + v_{ijkt}^{g}\right), v_{ijkt}^{g} \sim N(0, \omega_{ijk}^{2})$$

where goodwill is log of the latent mean clicks, $g_{ijkt} = \log\left(\lambda_{ijkt}\right)$ and

g_{ijkt} = goodwill of ads in format *i,* size *j,* target *k* at time, *t*.

$\bar{f}(a_{ijklt})$ = a function of ad impression a_{ijklt} in format *i,* size *j,* target *k*, theme *l* at time, *t*.[4]

β_{ijkl} = effectiveness of impression in format *i,* size *j,* target *k,* theme *l*.

δ_{ijk} = carryover rate in (Flash or GIF) format *i,* size *j,* target *k.*

α_{ik} = fixed effect of animation in flash ads in target *k*.

[4] The estimation uses a semi-log transformation: $\bar{f}(a_{ijklt}) = \ln(1 + a_{ijklt})$. See Bass et al. 2007 for the justifications for these functional forms.

σ_{jk} = fixed effect of size across flash and GIF ads in target *k*.

v^{g}_{ijkt} = mean zero, normal error for in format *i,* size *j,* target *k*.

Thus, with log-link $g_{ijkt} = \log(\lambda_{ijkt})$, equation (5) is the familiar discrete-time goodwill model due to Nerlove-Arrow. That is, one assumes that goodwill g_{ijkt} decays in proportion to prior goodwill g_{ijkt-1}, and sustained here by an additive functions of advertising exposures $\bar{f}(a_{ijklt})$. Also, although in the digital arena consumers can only click-through if they see an ad, the decision to click is attributed to both the current and the cumulative effects, or goodwill of past ad impressions. Without carryover, our model would attribute ad response only to current impressions at time *t*. The fixed-effects parameters (α_{ik} and σ_{jk}) control for the possibility that digital media characteristics, size and animation, influence response; that is, online consumers may respond differently to messages in different ad formats and sizes. To separately identify both the effects of size and animation, though, we set $\alpha_{ik} = 0$ for the GIF formats. Apart from their simple interpretations in terms of our substantive questions, there are other desirable features of these fixed effect parameters: first, they help impose correlation among multivariate count data; second, they explicitly account for one source of *endogeneity* that could arise were they omitted, because indeed format and size effects could be correlated with ad impressions.

Endogenous Impressions

Yet more likely sources of endogeneity in ad impressions $\text{cov}(a_{ijklt}, v^{g}_{ijkt}) \neq 0$ may be due to the context consisting of unobservable site features such as the information content of the site. Ad networks are likely to serve more impressions to sites whose context matches the advertised product. Given these reliability concerns, we follow Naik and Tsai 2000 and Sonnier et al. 2012 to account for endogeneity:

6) $$\mathbf{m}_{kt} = \boldsymbol{\theta}_{kt} + \boldsymbol{\eta}_k \mathbf{Z}_{kt} + \mathbf{v}_{kt}^m$$

7) $$\boldsymbol{\theta}_{kt} = \mathbf{B}_k \boldsymbol{\theta}_{kt-1} + \mathbf{v}_{kt}^\theta$$

where $\mathbf{m}_{kt} = \left[\bar{f}(a_{11k1t}),.,\bar{f}(a_{23k1t}),\bar{f}(a_{11k2t}),..,\bar{f}(a_{23k2t})\right]'$, $\mathbf{v}_{kt}^m = \left[v_{11k1t}^m,.,v_{23k1t}^m, v_{11k2t}^m,..,v_{23k2t}^m\right]'$,

$\mathbf{v}_{kt}^\theta = \left[v_{11k1t}^\theta,.,v_{23k1t}^\theta, v_{11k2t}^\theta,..,v_{23k2t}^\theta\right]'$, $\mathbf{v}_{ijkt}^\theta \sim \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma}_{ijk})$, $\left[v_{ijkt}^g, \mathbf{v}_{ijkt}^m\right] \sim \mathbf{N}(\mathbf{0},\mathbf{H}_{ijk})$ and $\mathbf{H}_{ijk} = \begin{bmatrix} \omega_{ijk}^2 & \mathbf{S}_{ijk}' \\ \mathbf{S}_{ijk} & \boldsymbol{\Omega}_{ijk} \end{bmatrix}$.

That is, equations (6-7) model impressions ($\mathbf{m}_{kt}$) across format, message, and target as functions of: 1) covariates $\mathbf{Z}_{kt}$, dummy variables for the category of the product embedded in these impressions; 2) a random measurement noise $\mathbf{v}_{kt}^\theta$; and 3) a latent time varying component $\boldsymbol{\theta}_{kt}$, that is governed by an AR(1) process (Naik and Tsai 2000). The category dummies are proxy measures for the web context, aiming to capture endogeneity in impressions due to the matching of website with the retailers advertised product; the latent measure $\boldsymbol{\theta}_{kt}$ captures variation in impressions due to other unobserved factors. Recall, that our data captures the product related promotions of a (multi-category) retailer. As such, the objectives of these ads are to create awareness and engagement among targeted consumers for the retailer's products, using targeted price and product related messages. Now, many of these products have specific uses, and thus are often advertised on sites where they are related to the content of the sites; for the retailer's targets are likely to visit these locations; and when they visit contextually matched sites, they are more likely to engage. This type of matching suggests that our product category dummies are potentially valid instruments, related to website content and traffic[5]. Note also that our instrumental variable model is a state space model, with a time-varying intercept $\boldsymbol{\theta}_{kt}$ in the equation (6). This helps control for other time-varying un-observables that could co-vary with ad

[5] The population R-squares from the regression of log (1+ad impression) against these instruments range from 0.63-.73 across the four consumer targets (see Web Appendix A for further assessment of our instruments).

impressions. Formally, to control for potential endogeneity (which becomes relevant when elements of $\text{cov}\left(v_{ijkt}^{g}\ \mathbf{v}_{ijkt}^{m}\right) = \mathbf{S}_{ijk} \neq 0$), we condition the analysis of equations 1-5 on $\mathbf{v}_{ijkt}^{m}$ (see e.g., Rossi et al. 2005).

In summary, we proposed a model to investigate the effects of digital ads served across multiple formats, messages and targeted consumers over time. The model has three major components: a non-linear model of ad response that accounts for the presence of zeros in event counts data; a model of ad dynamics that links impressions and targeting decisions to ad response; and a linear measurement model that controls for endogeneity in ad impressions. The model can address several questions about the duration of advertising across digital formats; whether some formats and re-targeting strategies are more effective with price-based incentives; and the impact of size and animation on digital ad effectiveness. First, however, we will have to develop an estimation scheme, whose primary challenge will be to recover time-varying vectors that include both linear and non-linear components.

ESTIMATION AND INFERENCE

We adopt a Bayesian approach to estimation given its versatility and our need to evaluate non-linear, non-normal state-space parameters. With few exceptions (e.g., Lopes et al. 2010), the Bayesian approach to such problems relies upon conditional independence to iteratively sample a sequence of conditional posteriors (for the fixed and time-varying parameters) rather than sample directly from their intractable joint (Doucet et al. 2001)). How, then, does conditional independence help resolve our estimation problem defined by equations 1-7? First consider our essential task: it is to recover a joint, but intractable posterior $p(\boldsymbol{\theta}_t, \mathbf{g}_t | \mathbf{y}_t, \mathbf{m}_t, \zeta)$ (intractable because $\mathbf{y}_t$ is non-linear/non-Gaussian), where $\mathbf{g}_t = \{g_{11t}, g_{12t}, .., g_{IJt}\}$ ($\mathbf{g}_t = \log(\boldsymbol{\lambda}_t)$) and $\boldsymbol{\theta}_t = \{\theta_{11t}, \theta_{12t}, .., \theta_{IJt}\}$ are the vectors of goodwill and measurement state variables just described; ζ is a collection of

all the static parameters; and $(\mathbf{y}_t, \mathbf{m}_t)$ are click and impressions (respectively) at time *t*, (here we suppress the target subscript, *k*) (See equations 1-5). Thus, in our case, conditional on $\mathbf{g}_t$, the clicks $\mathbf{y}_t$ provide no further information for estimating the measurement state variable $\boldsymbol{\theta}_t$. In other words, $\mathbf{g}_t$ becomes a sufficient statistic for estimating $\boldsymbol{\theta}_t$; and thus equations (5-6) become the linear observation equations and equation 7 becomes the linear the system equation for the state $\boldsymbol{\theta}_t$. We can therefore apply the basic Kalman Filter/Smoother Algorithm to estimate $p(\boldsymbol{\theta}_t|, \mathbf{g}, \mathbf{m}_t, \zeta)$ and its related fixed parameters in ζ (See e.g., Carter and Kohn 1994; Fruhwirth-Schnatter, 1994; Bass et al. 2007).

In contrast, the conditional posterior $p(\mathbf{g}_t|, \mathbf{y}_t, \zeta)$ of goodwill is non-linear and non-Gaussian (because its observation equation (2) is ZIP-Poisson), and so there is no general (or closed form) expression for its pdf; we thus approximate it using the particle filter (Bruce 2008; Doucet et al. 2001; Liu and Chen 1998). Particle filtering (PF) belongs to a class of Sequential Monte Carlo (SMC) integration methods based on Bayesian inference. It is more flexible than the extended and unscented Kalman filters, methods that work with Gaussian approximations for posterior densities, which make them simpler to implement and faster to execute, but preclude them from modeling the higher order moments of truly non-Gaussian distributions (Ristic et al. 2004). PF involves the use of particles, samples drawn from an importance function, and their associated weights to approximate the pdf. The procedure, based on *Importance Sampling* (e.g., Geweke, 1989), provides a discrete approximation to the posterior density of the states through a set of support N_s points (or particles) $\{\mathbf{g}_{0:t}^n\}_{n=1}^{N_s}$, and their respective weights, $\{w_{0:t}^n\}_{n=1}^{N_s}$, where $w_t^n > 0$ and $\sum_{n=1}^{N_s} w_t^n = 1$. These particles are drawn from an importance function; the choice of this density is one of the most important decisions in constructing PF algorithms.

Choosing an Importance Function

In many applications of the particle filter, the chosen importance function is the transition density (equation 5) because it is simple and readily available from the model. Yet we know that particle filter algorithms that use this (prior) importance function often suffer from the degeneracy problem; that is to say, the variance of the importance weights increase over time. Intuitively, if the data is very informative (i.e., the variance of its distribution is very small), the algorithm would waste many samples and time by exploring regions of low importance. To make the method more effective, Doucet et al. (2001) and Liu and Chen (1998) suggest importance functions of the form $p\left(\mathbf{g}_t \middle| \mathbf{g}_{t-1,} \mathbf{y}_t\right)$), i.e., ones which incorporate both the system and observation processes. Indeed, Doucet et al. (2001) show that this importance function $p\left(\mathbf{g}_t \middle| \mathbf{g}_{t-1,} \mathbf{y}_t\right)$ addresses the degeneracy problem by minimizing the variance of the (un-normalized) importance weight $\left\{w_{0:t}^n\right\}_{n=1}^{N_s}$. Nevertheless, it is very difficult to derive optimal importance functions of the form $p\left(\mathbf{g}_t \middle| \mathbf{g}_{t-1,} \mathbf{y}_t\right)$ analytically, outside of a few special cases (e.g., Bruce 2008), and certainly not in our case where the observations are nonlinear and non-Gaussian. Yet we can derive a linear-normal approximation of this optimal function (See Web Appendix B):

(8) $$\tilde{p}(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t) = N\left(\mathbf{g}_t^*, -\left[l''(\mathbf{g}_t^*)\right]^{-1}\right) \text{ where,}$$

$l(\mathbf{g}_t) = \ln p(\mathbf{y}_t | \mathbf{g}_t) p(\mathbf{g}_t | \mathbf{g}_{t-1})$, with derivatives of this log distribution, l' and l'' evaluated at its mode,

$$l'(\mathbf{g}_t^*) = \left.\frac{\partial l(\mathbf{g}_t)}{\partial \mathbf{g}_t}\right|_{\mathbf{g}_t = \mathbf{g}_t^*} \text{ and } l''(\mathbf{g}_t^*) = -\left.\frac{\partial^2 l(\mathbf{g}_t)}{\partial \mathbf{g}_t \partial \mathbf{g}_t'}\right|_{\mathbf{g}_t = \mathbf{g}_t^*}.$$

We can obtain the mode $\mathbf{g}_t^*$ of $l(\mathbf{g}_t)$ by applying an iterative Newton-Raphson procedure, initialized with $\mathbf{g}_t^0 = \mathbf{g}_{t-1}$ at each step of the filter:

(9) $$\mathbf{g}_t^{k+1} = \mathbf{g}_t^k - \left[l''\left(\mathbf{g}_t^k\right)\right]^{-1}\left[l'\left(\mathbf{g}_t^k\right)\right]$$

DATA AND IDENTIFICATION

Recall our substantive aim is to explore how central features of digital ads affect consumer engagement over time. Thus, to identify the cross-sectional and temporal features of the problem, we acquired panel data from a major US retailer in an industry that provides products and services for the home. The data contains daily ad impressions served via an ad network and the resulting clicks, both disaggregated by target, format and message; for a period of *T=154* days from February 14, 2011 to July 17, 2011. In this campaign, the retailer targeted four broad segments, one behavioral (retargeted) and three demographic (male, female, age); and employed two ad formats, flash (animated) and GIF (static). Flash ads appear as a sequence of (~ 4-8) time delayed images, with the last identical to the static GIF image. Flash ads not only include colorful, attractive, animation but also deliver a longer message than GIF. There are as well three standard sizes-orientations for ads, 728x90 (leaderboard), 160x600 (skyscraper), and 300x250 (square box) (See Figure 1). The retailer deems ads as price messages if they mention price or price discounts; and products messages if they convey product attributes without reference to price. Lastly, the ad network serves blank impressions (white spaces) to non-US consumers to preclude them from viewing the ads; and serves product offer and control impressions exclusively in flash and GIF formats, respectively.

Model identification thus draws upon a balanced panel of 24 time series (24 = 4 targets x 2 (Flash, GIF) x 3 sizes) each of length T=154. How does this panel help identify our substantive parameters, the effect of ads across format, message and targets? First, we have day-to-day variations in clicks and their associated impressions within each of the 24 target-format time series, with the impressions in each series further classified into those attributed to price and

product for flash ads, and those to price and control for GIF. Moreover, the correlation between impression pairs (price and product, and price and control) in our sample is low (median = 0.0378); that is, there is daily variations in the number impressions served for each theme such that co-movement is negligible. This natural *thematic variation* (Schumann and Clemons 1989) allows one to recover the separate effects of price, product and control impression on click response in each of the 24 target-format (time series) combinations.

-- Insert Figure 1 here --

Table 1 provides summary statistics for the data. We see that the average total number of clicks for flash format ads is about 100 times more than clicks for GIF ads across all ad sizes and targets. Within the flash format ads, the leaderboard ads generate the highest average number of clicks while the box ads have the lowest average. The average total number of clicks within the age segment is considerably higher than clicks within the other three segments. Quick calculations show that the firm serves about 40% of product ad impressions and 60% of price ad impressions. Nearly 59% of the retailers ads are served to targets age segment, the remaining impressions are served to the retargeted (17%), male (8%) and female (16%) segments. Figure 2 plots time-series of clicks for T=154 across the formats. Note the spikes in the numbers of clicks around period 10; these occurred during the early spring, when consumers are interested in home improvement projects as winter is ending. Ad messages here offer specific promotions that take advantage of this interest. Figure 3 summarizes the data across format and targets.

-- Insert Table 1 here –

-- Insert Figure 2 and 3 here --

Table 2 compares the click through rates for different formats and ad sizes, across consumer targets. Flash ads presented to remarketing clickers in the GIF (300x250) format have

the highest average CTR of .14%, suggesting GIF ads can be effective in some context. Nevertheless, CTRs are expectedly very low. For example, the flash format ads have an average total CTR that ranges from 0.05% to 0.057%; similar CTR values for GIF ads range from 0.013% to 0.059%, if one ignores the large percentage of blank impressions served to non-US visitors, and their associated clicks (McConnell 2012).

-- Insert Table 2 here --

ESTIMATION RESULTS

Tables 3-7 display the results of our empirical analysis. Tables 3-5 report findings related to robustness checks and to the potential endogeneity of ad impressions. Tables 6 and 7 report estimates of the main parameters of the proposed model (DZIP). Significance estimates in bold fonts are estimates whose 95% HPDI interval excludes zeros. What follows are first reviews of the robustness and endogeneity results; then reviews in turn of results (Tables 6-7) related to the effects of ad format, carryover, and message content across the four consumer targets in the study; and then lastly a brief summary of the main conclusions.

Model Selection

Table 3 compares the proposed (DZIP) to seven alternate models, including the normal, dynamic linear model (NDLM) and alternative count models, specifically variants of the zero hurdle (*H*) and negative binomial (*NB*) models:

9) $$\pi^{H}\left(Y_{ijkt}=y_{ijkt}\right)=\begin{cases} p_{ijk} & \text{if } y_{ijkt}=0 \\ (1-p_{ijk})\,f\left(y_{ijkt}\middle|\lambda_{ijkt},y_{ijkt}>0\right) & \text{if } y_{ijkt}>0 \end{cases} \qquad (H)$$

10) $$f^{NB}\left(Y_{ijkt}=y_{ijkt}\middle|\lambda_{ijkt}\right)=\frac{\Gamma\left(k_{ijk}+y_{ijkt}\right)}{\Gamma\left(k_{ijk}\right)y_{ijkt}!}\left(\frac{k_{ijk}}{k_{ijk}+\lambda_{ijkt}}\right)^{k_{ijk}}\left(\frac{\lambda_{ijkt}}{k_{ijk}+\lambda_{ijkt}}\right)^{y_{ijkt}} \qquad (NB)$$

Note the proposed dynamic zero-inflated Poisson (DZIP) model dominates all alternatives, as indicated by its DIC (deviance information criterion) value. The DIC and similar

model selection methods (Akaike information criterion, AIC; and Bayesian information criterion, BIC) include penalty terms to offset gains in model fit due solely to added complexity, since more complex models, with more parameters generally provide better fit. For Bayesian hierarchical models, however, the number of parameters is less clear. Spiegelhalter et al. 2002 proposed the DIC to address this uncertain complexity in Bayesian hierarchical models (e.g., equation 5). With the DIC, then, the worst model (model 8) is the normal dynamic linear model (NDLM); this confirms that here a normal approximation to the distribution of the data (clicks) is inappropriate. The negative binomial models also perform poorly relative to Poisson models; and among the latter, the DZIP model dominates. Hence, we show that it is important to consider the dynamic effects of digital ads as well as to control for over-dispersion. Finally, figure 4 shows the fit of the DZIP model by plotting its posterior mean (λ_{ijkt}) against the actual number of daily clicks for different formats and targets. The proposed model fits the data quite well.

-- Insert Table 3 and Figure 4 here –

Endogeneity of Impressions

Tables 4 and 5 show the results of an analysis into the potential *endogeneity* of ad impressions. In implementation, this means controlling for the potential co-movement of the measurement ($\mathbf{v}_{ijkt}^{m}$) and goodwill noises (v_{ijkt}^{g}) (Rossi et al. 2005; Naik and Tsai 2000), after accounting for unobserved information context of the publishers site using product category dummies. Table 4 reports the effects of these dummies on the volume of impressions. The significant parameters show that some category dummies predict product and price ad impressions; and thus could account for the unobservable matching features of the publisher sites; and therefore if the correlations between ($\mathbf{v}_{ijkt}^{m}$) and (v_{ijkt}^{g}) are not significantly different from zeros, then one can conclude that measurement noise is inconsequential in our sample.

Table 5, however, shows that five to ten out of the twelve correlations in each target are still significant. This suggests controlling for endogeneity seems essential for determining the effectiveness of digitals ads (e.g., Lee et al. 2015).

-- Insert Tables 4-5 here --

Animated vs. Statics Display Ads

Tables 6 and 7 report estimates from the proposed DZIP model. First, table 6 shows that flash ads have significantly higher average clicks than GIF ads, as seen by the fixed effect of flash ads, supporting the notion that animation can foster engagement (Li and Bukovac 1999). These results are consistent across all consumer targets. Recall that the dependent parameter in equation (5) is the log-link, $log(\lambda_{ijkt})$; thus the effectiveness of flash ads across re-targeted, male, female, and age segments are 11.8, 10.4, 16.9 and 12.6 times (respectively) that of similar GIF ads, ceteris paribus; reflecting the much greater sparsity of click response to GIF ads (Table 1). The effects of orientation-sizes ($\exp(\sigma_{ik})$) across segments are significant too; but their relative effects on engagement are mixed as predicted (e.g., Chandon et al. 2003; Cho 2003). For example, box ads are most effective in the age segment; but leaderboard are most effective among re-targeted consumers; and all are equally effective among females. The latter result seems to support the prediction that female retail shoppers are more likely to browse (e.g., Passyn et al. 2011).

Carryover Effects

Carryover rates for flash (gif) format ads are significant across the four segments (Table 6), with values ranging from 0.52 to 0.75 (0.09 to 0.27). Thus, animated banner ads have significantly higher carryover rates than GIF ads, across consumer segments and size-orientations. The increase in carryover rates is roughly 3 to 5 times greater when one uses

animated ads rather than static ads across target and format. These results seem consistent with Naik and Raman (2003) who found that carryover for TV (animated) was approximately 2.5 times that of static print. Hence, here animated ads have the potential to engage consumers for longer periods. To make this result more concrete, we computed the 90% duration for each format and target (D_{90} days); that is, the number of days it takes for the effectiveness of an ad to lose 90% of its effect. Thus, in table 7 the average D_{90} across the four segments ranges from 5.50 to 7.6 days for flash ads, while it is about 2.6-2.8 days for GIF ads. Similarly, in Table 7 the mean ad elasticity (calculated using the posterior draws) for flash ads range from 0.2437-0.3708 for flash ads, while those for GIF range from 0.0682-0.1595.

-- Insert Tables 6 and 7 here --

Price vs. Product Based Messages

Consider now the effects of product and price incentive-based messages; specifically how these effects vary across creative formats and targeted consumers. Table 6 reports the immediate or short-term effects β_{ijk} of ads by themes, across formats, sizes, and among differing consumers. From these results, price ads are more effective than product ads within the flash format, in all sizes and target markets; building upon evidence that price incentives can motivate engagement (Chtourou et al. 2001; Xie et al. 2004; Maureen and Grey 2005). Product ads, nevertheless, are still effective in the male, female and age segments across all size-orientations (with one exception, leaderboard among males); and though these effects differ marginally, they are on average highest among targeted females (0.0325, 0.0335, 0.0372), who retailing studies predict are likely more engaged shoppers. Yet product ads are ineffective among re-targeted consumers, while, in contrast, price ads in similar Flash are effective in all segments, even among re-targeted consumers. Now recall, evidence suggests that retargeted ads are ineffective unless served to

consumers who have well defined preference such that they are willing to purchase (Lambrecht and Tucker 2013). Thus, our finding suggests that when re-targeting consumers, one should also recognize that price-incentives can be useful in making ads more effective by addressing consumer willingness to pay.

The discussion, hitherto, has reviewed the effects of flash formatted ads; and earlier reported that ceteris paribus flash ads garner more engagement than GIF. Table 6 however shows that static GIF ads with price offer messages can be effective among re-targeted and female shoppers. Furthermore, while price ads are more effective in generating engagement in the flash than in the GIF format for the male and age segments (Li and Bukovac 1999), price ads are equally effective for GIF and Flash among females and the re-targeted. Finally, we note the parameters for international GIF ads (e.g. β_{42}). Recall that because this campaign targets US consumers, the ad server sends blank impressions to non-US consumers. Nevertheless, these consumers may still click on blank images, usually from curiosity (e.g., when blank ads are visually salient, Wedel and Pieters 2008) but also in error (McConnell 2012); and as a result, the parameters that capture the effects of these clicks are large and significant. While these measures have no managerial interpretation in terms of ad content, they do show how the tactic of serving blanks can distort naive measures of campaign effectiveness (e.g., CTR).

In summary, tables 6-7 help reveal the workings of digital ads. For instance, in our sample, animated ads are more effective than static ads, and have longer duration. There is heterogeneity too in the performance of banner ads across creative formats, messages and targeted audiences. For example, within the flash format, price ads are more effective in generating engagement than product ads, and are effective in all three size-orientations and four target markets defined in this study. Product ads in contrast are ineffective among re-targeted

consumers. Thus, retargeted a consumers are less likely to engage when ads exclude price incentives. Finally, although flash ads engage more consumers than GIF ads, they are still effective for engaging retargeted and female consumers; the latter consumers seemingly more willing to engage with ads of all formats, and messages.

Robustness Check of Results

As a final step, we investigate the robustness of the above findings by comparing them to results from five (simpler) variations of the proposed (See Web Appendix C):

a. A linear state space model (NDLM) - click data on original scale
b. A log-linear state space model - click data log-transformed
c. A Dynamic Poisson model - no endogeneity or zero inflation
d. A Static Poisson Model - no endogeneity or zero inflation
e. A Dynamic negative binomial- no endogeneity or zero inflation

Notably, results from the (generalized linear models) GLMs (c, d, and e) are more consistent with the results from the proposed model (Table 3 Appendix; Tables 1R-5R, Web Appendix C). The NDLM by contrast reported mixed findings for the Flash and size effects. That is, in some cases, flash ads on average are no more effective at generating clicks than GIF ads, ceteris paribus. Similarly, the effects of some ad sizes are not significant. Although the log-linear model (b) had many findings similar to the GLM findings, it too reported mixed results for the fixed effects of size and format. (Note, we also estimated, but did not report a square root transformed data model, and found conflicting evidence). In general, we know that GLMs are better suited for count data, more so when we they include zero observations; and log transformations are more effective when mean counts are large and over-dispersions is small (e.g., O'Hara and Kotze 2010).

REALLOCATION ANALYSIS

The final task of this study is to conduct a simulation that summarizes the import of the previous results. One approach is to see how the above results influence the reallocation of ad impressions across the duration of the campaign. That is, given hyper-parameters ζ, we solve a problem that reallocates the total ad impressions (b_t) in each period, across ad format (GIF, Flash), sizes, themes and targets to maximize the total expected clicks $(\mathbf{y}_t)$ over $T = 154$ days. That is, with estimates of the state vectors from the particle filter $\{\mathbf{g}^n_{0:T}, \mathbf{w}^n_{0:T}\}_{n=1}^{N_s}$, we solve the problem (P1):

$$\max_{\mathbf{a}_{111}\ldots\mathbf{a}_{IJT}} \sum_{t=1}^{T}\sum_{n=1}^{N} \mathbf{w}^h_{ijt-1} E\left(\mathbf{y}_{ijt} \middle| \mathbf{g}^n_{ijt-1}\right)$$

$$st \sum_{k=1}^{K}\sum_{i}^{I}\sum_{j}^{J}\sum_{l}^{2} a_{ijklt} \le b_t, a_{ijklt} \ge 0, t = 1,..,T$$

where $\mathbf{a}_{ijt} = \{\{a_{ijklt}\}_{l=1}^{2}\}_{k=1}^{K}$, and $\{\mathbf{g}^n_{ijt}, \mathbf{w}^n_{ijt}\} = \{g^n_{ijkt}, w^n_{ijkt}\}_{k=1}^{K}$ are impressions, goodwill, and particle weights across format, message, and target. Also, $E(\mathbf{y}_{ijt} | \mathbf{g}^n_{ijt-1})$ is the 1-step ahead forecast vector at the particle $\{\mathbf{g}^n_{ijt-1}\}$, and a_{ijklt} the impression for ad theme (l) in format i, size j and target k at period t.[6] We select solutions to P1 that give allocations that represent improvements in ad efficiency. To do this, we exploit an advantage of the state space approach to optimize successively for one period (t) given the data at t-1; this is a more appealing solution, because impressions are bought real-time. (We also optimized similarly for two and four periods at a time and report all results in Web Appendix D)[7]. Lastly, we keep the total number of ad impressions for a given optimization interval to be the same as the actual number of impressions but allow a reallocation across formats, sizes, targets and content.

---

[6] Solved in Tomlab/SNOPT

[7] We thank the Area Editor for this suggestion.

-- Insert Tables 8 here --

Table 8 shows the solution to P1 with all hyper-parameters ζ at their mean values. It reports for target, digital format and message, the *actual* number of impressions, and the *model-based* prediction of the number of impressions needed to generate a higher number of clicks. Here the model-based allocations generated approximately 17% more clicks than the current allocation. The results in Table 8 are largely consistent with findings discussed above. Thus, overall the model suggests a 19% decrease in number of impressions of product ads and a 13% increase in impressions of price offer ads. Much of this increase is attributable to the shift from product to price ads in the flash formats. Similarly, we observe higher impressions in for the retargeted (21%) and female (5%) consumer segments. Consequently, the model recommends increases in static GIF (price) ad impressions, given these were effectiveness for ads retargeted and female consumers (Table 3). Finally, as a robustness check, we also solved P1 over 1000 random draws from the posterior, using a shorter period, given the computational complexity of solving P1. The results reported in the Web Appendix D are consistent with results in Table 8.

CONCLUSION

This study explored how the performance of digital ads is influenced by the joint effects of creative format, message content, and targeting as well as retargeting. Its goal was to reveal how central features of a digital campaign affect consumer engagement over time. The study accomplished this by constructing a dynamic model and estimating it using a panel dataset obtained from a major US retailer. Formally, it proposed a dynamic (state space) zero-inflated count model (Poisson), given the potential for zero-inflation and temporal correlation in count series. The resulting model was both dynamic and nonlinear; and therefore we estimated it using a combination of Particle Filtering and MCMC. The resulting algorithm provides one approach

to estimating any state space model within the exponential family, and is more flexible than Gaussian filters such as the extended and the unscented Kalman Filters. The estimation also allowed for endogeneity in ad impressions, possibly due to unobserved context of the publisher's site.

The study found a number of substantive results. First, animated ads had significantly higher carryover effects; and thus affected engagement over a longer duration than static ads, in all ad creative formats, among targeted and retargeted consumers. Second, among animated formats, price ads were more effective than product ads. Third, re-targeted ads were effective only when they offered price incentives; a finding consistent with Lambrecht and Tucker (2013) who found them to be effective only when consumers had strong preferences such that they had incentives to buy. Ours is a useful result because it suggests that price sensitivity (perhaps more observable than consumer preferences) could help select re-targeted consumers for engagement. Third, although flash ads were more effective at engaging consumers, simpler static GIF ads could also be effective; in our case, for price ads served to retargeted and female consumers. Lastly, note that all the retailer's ads (i.e., all formats and messages) targeted to the female segment were effective; this suggests that female shoppers were largely more willing to engage, perhaps confirming the brick mortar studies that "women shop, men buy."

Limitations and Extensions

Yet the work has a few potential limitations that could be addressed in the future. First, some of our findings may not generalize to other contexts. For example, the gender effects noted above may have arisen because of the match between the retailer's product category and gender. Similarly, our data comes from a single (albeit major) retailer, and so given recent findings (e.g., Li and Kannan, 2014), we would be reluctant to generalize these results to other industries or

smaller firms. Second, we estimated the model with daily but aggregate data, at the level of target, message, and format; and our method lacks features such as those of exploration and exploitation embedded in sequential experiments (e.g., Thompson sampling). Yet ad networks commonly release such data to their clients, and it is in these cases that our method is most applicable. Individual data nevertheless could obviate some of endogeneity issues we addressed statistically; but there are challenges to estimating dynamics at the individual/cookie level. For example, one would be required to estimate a large numbers of parameters from sparse data; for although the data will contain many individuals, many of these will be unique or one-time visitors. To address this sparseness problem, one could perhaps build a hierarchical dynamic model (e.g., Gamerman et al. 2002) using demographics and retargeting data (Agarwal 2010) to define segment level distributions from which individual behavior could arise. Notably, in this case, the substantive parameters would again be at the segment level. Finally, although our model fits the data satisfactorily, another potential criticism, given the full Bayesian approach, is that we adopted standard parametric assumptions for all model components: e.g., normal random noise in the state equation (equation 5). To mitigate this criticism, one could model errors as Gaussian mixtures; or take a fully Bayesian nonparametric approach in which the distributions of the errors are themselves unknown and treated as objects to be estimated (Phadia 2013 and Hjort et al. 2013). Still, again, a non-parametric approach could be more feasible at the segment level given data sparseness at the cookie level.

REFERENCES

Agarwal, Deepak (2010), "A modern Bayesian look at the multi-armed bandit by Steven Scott: Discussion," *Applied Stochastic Models Business and Industry*, 26(6), 639–658.

Baltas, George (2003), "Determinants of Internet Advertising Effectiveness: An Empirical Study," *International Journal of Market Research*, 45 (4), 505-513.

Bass, Frank, Norris Bruce, Sumit Majumdar and B.P.S. Murthi (2007), "Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of Advertising-Sales Relationship," *Marketing Science*, 26 (2), 179-195.

Berkowitz, David, Arthur Allaway and Giles D'Souza (2001), "Estimating Differential Lag Effects for Multiple Media across Multiple Stores," *Journal of Advertising*, 30 (4) 59-65.

Braun, Michael and Wendy Moe (2013), "Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories," *Marketing Science*, 32(5), 753-767.

Breuer, R., Brettel, Malte ; Engelen, Andreas (2011) "Incorporating Long-term Effects in Determining the Effectiveness of Different Types of Online Advertising," *Marketing Letters*, 22(4), 327-340.

Briggs, R., and N. Hollis (1997), "Advertising on the Web: Is There Any Response before clickthrough?," *Journal of Advertising Research*, 37(2), 33–46.

Bruce, Norris (2008), "Pooling and Dynamic Forgetting Effects in Multi-Theme Advertising: Tracking the Ad Sales Relationship with Particle Filters," *Marketing Science*, 27 (4), 659-673.

Carter, C., and R. Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541-553.

Chatterjee, Patrali (2008), "Are Unclicked Ads Wasted? Enduring Effects of Banner and Pop-up Ad Exposures on Brand Memory and Attitudes," *Journal of Electronic Commerce Research*, 9 (1), 51-61.

Chatterjee, P., Hoffman, D.L. and Novak, T.P. (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science*, 22(4), 520–541.

Chandon, J.L., Chtourou, M.S., & Fortin, D.R. (2003), "Effects of Configuration and Exposure Levels on Responses to Web Advertisements," *Journal of Advertising Research,* 34(3), 217-229.

Chandy, Rajesh K., Gerard J. Tellis, Deborah J. Macinnis, and Pattana Thaivanich (2001), "What to Say When: Advertising Appeals in Evolving Markets," *Journal of Marketing Research*, 38 (4), 399-414.

Cho, C. (1999), "How Advertising Works on the WWW: Modified Elaboration Likelihood Model," *Journal of Current Research in Advertising*, 27(1), 33–50.

Cho, C.H., J. G. Lee and M. Tharp (2001), "Different Forced Exposure Levels to Banner Advertisements," *Journal of Advertising Research*, 41(4), 45-56.

Cho, Chang-Hoan (2003), "The Effectiveness of Banner Advertisements: Involvement and Click-through," *Journalism and Mass Communication Quarterly*, 80 (3), 623-645.

Chtourou, M. S., and Chandon, J. L. (2000), "Impact of motion, picture and size on recall and word of mouth for Internet banners." INFORMS Internet and Marketing Science Conference, University of Southern California, Los Angeles, CA, USA, May.

Chtourou, M.S., J.L. Chandon, and M. Zollinger (2002), "Effect of Price Information and Promotion on Click-through Rates for Internet Banners," *Journal of Euromarketing*, 11(2), 23-40.

Clarke, Darell G. (1976), "Econometric Measurement of the Duration of Advertising Effect on Sales," *Journal of Marketing Research*, 13(4), 345–357.

comScore (2012), "comScore Introduces Validated Campaign Essentials™ (vCE): A Holistic Measurement Solution That Validates Advertising Impressions and Audiences Reached with Digital Advertising Campaigns," (accessed April 16, 2013), [available at http://www.comscore.com/Insights/Press_Releases/2012/1/comScore_Introduces_Validated_Campaign_Essentials].

Cole, Sally G., Leah Spalding, and Amy Fayer (2009), "The Brand Value of Rich Media and Video Ads," (accessed April 16, 2013), [available at http://static.googleusercontent.com/media/www.google.com/en/us/doubleclick/pdfs/DoubleClick-06-2009-The-Brand-Value-of-Rich-Media-and-Video-Ads.pdf].

Dahlen, Michael (2001), "Banner Advertisements Through a New Lens," *Journal of Advertising Research*, 41 (4), 23-30

Danaher (2007), "Modeling Page Views Across Multiple Websites With An Application to Internet Reach and Frequency Prediction," *Marketing Science*, 26(3), 422-437.

Danaher, P.J., Janghyuk Lee and Laoucine Kerbache (2010), "Optimal Internet Media Selection," *Marketing Science*, 29 (2), 336-347.

Deighton, John, Caroline Henderson, and Scott A. Neslin (1994), "The Effects of Advertising on Brand Switching and Repeat Purchasing," *Journal of Marketing Research*, 31 (1), 28-43.

DoubleClick (2009), "2009 Year-in-Review Benchmarks," (accessed April 16, 2013), [available at https://static.googleusercontent.com/media/www.google.com/en//doubleclick/pdfs/DoubleClick-07-2010-DoubleClick-Benchmarks-Report-2009-Year-in-Review-US.pdf].

Doucet, Arnaud, Simon Godsill , Christophe Andrieu (2000), "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, 10(3), 197-208.

Doucet, Arnaud, Nando de Freitas, and Neil Gordon, Editors (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer.

Dreze, X and F. X. Hussherr (2003), "Internet Advertising: Is Anybody Watching?" *Journal of*

Interactive Advertising, 17(4), 8-23.

D'Souza, Giles and Ram C. Rao (1995), "Can Repeating an Advertisement More Frequently than the Competition Affect Brand Preference in a Mature Market?" *Journal of Marketing*, 59 (2), 32-42.

Fruhwirth-Schnatter, S (1994), "Data Augmentation and Dynamic Linear Models," *Time Series Analysis*, 15(2), 183-202

Gamerman, Dani (1997), "*Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*," London: Chapman and Hall, 124-132.

Geweke, John (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration" *Econometrica*, 57(6), 1317-1339

Godsill, Simon, Arnaud Doucet, and Mike West (2004), "Monte Carlo Smoothing for Nonlinear Time Series," *Journal of Statistical Association,* 99(465),156-168.

Grass, Robert, W. H. Wallace (1969), "Satiation effects of TV Commercials," *Journal of Advertising Research*, 9 (3), 3-8.

Greene, William H. (1994), "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," New York University Department of Economics Working Paper EC-94-10.

Hoch, Stephen J. and Young-Won Ha (1986), "Consumer Learning: Advertising and the Ambiguity of Product Experience," *Journal of Consumer Research*, 13 (2), 221-33.

Hong, W., J. Y. L. Thong, and K. Y. Tam (2007), "How Do Web users Respond to Non-banner-ads Animation? The Effects of Task Type and User Experience," *Journal of the American Society for Information Science and Technology*, 58(10), 1467–1482.

Hongshuang (Alice) Li and P.K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51(1), 40-56.

Hjort, Nils L.,Chris Holmes, Peter Muller, and Stephen E. Walker, Editors (2010), *Bayesian Nonparametrics*. Cambridge University Press.

Kim, Chang-Jin (2008), "Dealing with Endogeneity in Regression Models with Dynamic Coefficients," *Foundations and Trends in Econometrics*, 3(3), 165-266.

Lambert, Diane (1992), "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing," *Technometrics* 34,1-14.

Lambrecht, Anja and Catherine Tucker. (2013), "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 50(5), 561-576.

Lee, Dokyun, Kartik Hosanagar, and Harikesh S. Nair (2015), "Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook," Working Paper, Wharton School of Business, University of Pennsylvania.

Li, Hairong and Janice L. Bukovac (1999), "Cognitive Impact of Banner Ad Characteristics: An Experimental Study," *Journalism & Mass Communication Quarterly*, 76 (Summer), 341-353.

Liu, Jun S., and Rong Chen (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93(443) 1032–1043.

Lohtia, Ritu, Naveen Donthu and Edmund K. Hershberger (2003), "The Impact of Content and Design Elements on Banner Advertising Click-through Rates,"*Journal of Advertising Research*, 43, 410-418.

Lopes, H. F., Carvalho, C. M., Johannes, M. S. and Polson (2010) *Particle Learning for Sequential Bayesian Computation*. In Bayesian Statistic 9, Edited by J. Bernardo, M. Bayarr1, J. Berger, A. Dawid, D Heckerman, A Smith, M. West, Oxford University Press, New York.

MacInnis, Deborah J, Ambar Rao, Allen Weiss (2002), "Assessing When Increased Media Weight of Real-world Advertisements Helps Sales," *Journal of Marketing Research*, 39 (4), 391- 407.

Maureen, H. and A. Grey (2005). "Getting Something for Nothing: The Impact of a Sample Offer and User Mode on Banner Ad Response," *Journal of Interactive Advertising*, 6(1), 105-117.

McConnell, Ted (2012), "How Blank Display Ads Managed to Tot Up Some Impressive Numbers." (accessed April 16, 2013), [available at http://adage.com/article/digital/incredible-click-rate/236233].

Naik, Prasad and Kalyan Raman (2003), "Understanding the Impact of Synergy in Multimedia Communications," *Journal of Marketing Research*, 40 (4), 375-388.

Naik, Prasad A. and Chih-Ling Tsai (2000), "Controlling Measurement Errors in Models of Advertising Competition," *Journal of Marketing Research*, 37 (1), 113-124.

Naik, Prasad, Murali Mantrala, and Alan Sawyer (1998) "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17(3), 214–235.

Nerlove, Marc, and Kenneth Arrow (1962), "Optimal Advertising Policy Under Dynamic Conditions," *Economica* 29 (May) 129-142.

O'Hara, R. and J. Kotze (2010), "Do not log transform Count Data," *Methods in Ecology*, 1,118-122.

Phadia, Eswar G., (2013), *Prior Processes and Their Applications*: *Nonparametric Bayesian Inference*. New York: Springer.

Passyn, Kirsten A., Memo Diriker, Robert B. Settle (2011), "Images Of Online Versus Store Shopping: Have The Attitudes Of Men And Women, Young And Old Really Changed?," *Journal of Business & Economics Research*, 9(1), 99-110.

Rettie, R., Grandcolas, U. & McNeil, C. (2004) *Post Impressions: Internet Advertising without Click-through*. Kingston Business School, Kingston University.

Ristic, B., S. Arulampalam, and N. Gordon (2004), *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House Publishers, Boston.

Robinson, Helen, Anna Wusocka and Chris Hand (2007), "Internet Advertising Effectiveness: The effect of design on click-through rates for banner ads," *International Journal of Advertising*, 26(4), 527-541.

Rossi, Peter, Greg Allenby, and Rob McCullogh (2005), *Bayesian Statistics and Marketing.* Hoboken, N.J.: Wiley Series in Probability and Statistics.

Rust, Roland T. and Robert P. Leone (1984), "The Mixed Media Dirichlet Multinomial Distribution: A Model for Evaluating Television-Magazine Advertising Schedules," *Journal of Marketing Research*, 21 (1), 89-99.

Schumann, D., D. Clemons. (1989) "The Repetition/Variation Hypothesis: Conceptual and Methodological Issues," *Adv. Consumer Res*., 16, 529–534.

Scott, Steven (2010), "A Modern Bayesian Look at the Multi-armed Bandit," *Applied Stochastic Models Business and Industry*, 26(6),639–658.

Sethuraman, Raj, Gerard J. Tellis, and Richard Briesch (2011), "How Well Does Advertising Work? Generalizations from a Meta-Analysis of Brand Advertising Elasticity," *Journal of Marketing Research*, 48, (3), 457-471.

Sherman, L. & Deighton, J. (2001), "Banner Advertising: Measuring Effectiveness and Optimizing Placement," *Journal of Interactive Marketing*, 15(2), 60–64.

Sonnier Garret P., Oliver Rutz and Leigh McAlister (2011), "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Marketing Science*, 30(4), 702-716.

Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2013), "Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments," Working paper, University of Michigan.

Spiegelhalter, David, N.G. Best, B.P. Carlin A. V. Linde (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society: Series B*, 64(4) 583-639.

Tellis, Gerard J., Rajesh Chandy and Pattana Thaivanich (2000), "Decomposing the Effects of Direct Advertising: Which Brand Works, When, Where, and How Long?" *Journal of Marketing Research*, 37 (1), 32-46.

Tellis, Gerard J., Rajesh Chandy, Deborah MacInnis, and Pattana Thaivanich (2005), "Modeling the Micro Effects of Television Advertising: Which Ad Works, When, Where, Why, and For How Long?" *Marketing Science*, 24(3), 359-366.

Tsang, P.M., and S. Tse (2005), "A Hedonic Model for Effective Web Marketing: An empirical Examination," *Industrial Management and Data System*, 105(8), 1039- 1052.

Tuten, Tracy L., Michael Bosnjak, and Wolfgang Bandilla (2000), "Banner-Advertised Web Surveys," *Marketing Research*, Spring, 17-21.

Wedel, Michel and Rik Pieters (2008), "Informativeness of Eye Movements for Visual Attention: Six Corner Stones," in *Visual Marketing: From Attention Action*, eds. Wedel, Michel and Rik Pieters, New York, Lawrence Erlbaum Associates, 43-71.

Xie, Tian (Frank), Naveen Donthu, Ritu Lohtia, Talai Osmonbekov (2004), "Emotional Appeal and Incentive Offering in Banner Advertisements," *Journal of Interactive Advertising*, 4 (2), 30-37.

APPENDIX A

This appendix provides an overview of the MCMC algorithm we employ to recover both time-varying $(\mathbf{g}_t, \boldsymbol{\theta}_t)$ and fixed parameters ζ. Recall, our main task is to estimate a joint conditional posterior $p(\boldsymbol{\theta}_t, \mathbf{g}_t | \mathbf{y}_t, \mathbf{m}_t, \zeta)$ that includes both linear and non-linear, time-varying components. It is easier to obtain this posterior from conditionals $p(\boldsymbol{\theta}_t |, \mathbf{g}, \mathbf{m}_t, \zeta)$ and $p(\mathbf{g}_t |, \mathbf{y}_t, \zeta)$. That is, we sample the first conditional with the Kalman Filter because its state vectors, the measurement parameters, $\boldsymbol{\theta}_{0:T} = \{\boldsymbol{\theta}_t\}_{t=1}^{T}$ are the linear; and the second with the Particle Filter because its state vectors are the non-linear, goodwill vectors, $\mathbf{g}_{0:T} = \{\mathbf{g}_t\}_{t=1}^{T}$. (Note for simplicity we will suppress the target index, *k*).

1.0 Sampling $p(\boldsymbol{\theta}_t | \mathbf{g}_t, \mathbf{m}_t, \varsigma)$

First, conditional on $\mathbf{g}_{1:T}$ it is straightforward to sample the posterior $p(\boldsymbol{\theta}_t | \mathbf{g}_t, \mathbf{m}_t, \xi)$. To see this, note that for any $\mathbf{g}_t = \{g_{11t}, g_{12t}, ..., g_{IJt}\}$, we can re-write equations (5-7) as the following system:

1A) $$\hat{\mathbf{m}}_{ijt} = \mathbf{F}\boldsymbol{\theta}_{ijt} + \mathbf{L}\mathbf{v}_{ijt}$$

where $\mathbf{F} = \begin{bmatrix} \beta_{ij1} & \beta_{ij2} & \cdots & \beta_{ij.} \\ 1 & 0 & \cdots & 0 \\ . \; 0 & \ddots & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$ $\mathbf{L} = \begin{bmatrix} 1 & \beta_{ij1} & \beta_{ij2} & \cdots & \beta_{ij.} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}$

$\hat{\mathbf{m}}_{ijt} = \left[g_{ijt} - \delta_{ij} g_{ijt-1} - \alpha_i - \sigma_j - \boldsymbol{\beta}'_{ij}\boldsymbol{\eta}_{ij}\mathbf{Z}_{ijt}, \mathbf{m}_{ijt} - \boldsymbol{\eta}_{ij}\mathbf{Z}_{ijt}\right]'$ and $\mathbf{v}_{ijt} = \left[v^g_{ijt}, \mathbf{v}^m_{ijt}\right]'$.

2A) $$\boldsymbol{\theta}_{ijt} = \mathbf{B}_{ij}\boldsymbol{\theta}_{ijt-1} + \mathbf{v}^{\theta}_{ijt}, \text{ with}$$

$\mathbf{v}^{\theta}_{ijt} \sim N(0, \boldsymbol{\Sigma}_{ij})$, $\mathbf{v}_{ijt} = \left[v^g_{ijt}, \mathbf{v}^m_{ijt}\right]'$, $\mathbf{v}_{ijt} \sim \mathbf{N}(\mathbf{0}, \mathbf{H}_{ij})$, and $\mathbf{H}_{ij} = \begin{bmatrix} \omega^2_{ij} & \mathbf{S}'_{ij} \\ \mathbf{S}_{ij} & \boldsymbol{\Omega}_{ij} \end{bmatrix}$.

Thus, the equations (1A-2A) constitute a linear state-space model with respect to $\boldsymbol{\theta}_{ijt}$ and so we sample $p(\boldsymbol{\theta}_t |, \mathbf{g}, \mathbf{m}_t, \zeta)$ using Kalman Filter/MCMC ideas (See Bass et al. 2007 for details).

2.0 Sampling $p(\mathbf{g}_t|,\mathbf{y}_t,\zeta)$

The posterior $p(\mathbf{g}_t|,\mathbf{y}_t,\zeta)$ is however nonlinear in the goodwill state vector $\mathbf{g}_t$ because the observation equation (1) is non-linear/non-Gaussian, and thus sampled using particle filters. The basic procedure, based on sequential *Importance Sampling*, a Monte Carlo Integration method (e.g., See Geweke, 1989; Bruce 2008), provides a discrete approximation to the posterior density of the states through a set of support N_s points (or particles) $\{\mathbf{g}_{0:t}^n\}_{n=1}^{N_s}$, and their respective weights, $\{w_{0:t}^n\}_{n=1}^{N_s}$, where $w_t^n > 0$ and $\sum_{n=1}^{N_s} w_t^n = 1$. Unfortunately, the optimal importance function $p(g_t|g_{t-1}, y_t)$ is unavailable here because of the ZIP-Poisson assumption; however, we obtain a linear/nornmal approximation $\tilde{q}(g_{t-1}|g_{t-1}, y_t)$ with a *Newton-Raphson* step (See Web Appendix B).

2.1 Particle Filtering/Smoothing

2.1 Simulation for $\mathbf{g}_{0:T}$:

<u>Forward Filtering Algorithm:</u> (Sampling Importance Resampling)

1. For $n = 1,..,N_s$ Sample: $\mathbf{g}_0^n \sim p(\mathbf{g}_0)$, $w_0^n = \frac{1}{N_s}$. Set $t \to 1$.

2. For $n = 1,..,N_s$ Sample: $\mathbf{g}_t^n \sim \tilde{q}(\mathbf{g}_t \mid \mathbf{g}_{t-1}, \mathbf{y}_t)$, from the approximate optimal importance function, obtained via *Newton-Raphson* as just described.

3. For $n = 1,..,N_s$, Update weights:

$$w_t^n = w_{t-1}^n \frac{\pi(\mathbf{y}_t|\mathbf{g}_t^n)p(\mathbf{g}_t^n|\mathbf{g}_{t-1}^n)}{\tilde{q}(\mathbf{g}_t^n|\mathbf{g}_{t-1}^n, \mathbf{y}_t)}, \text{ and normalize them: } w_t^n \to \frac{w_t^n}{\sum_{l=1}^{N_s} w_t^l}.$$

$\pi(.)$ and $p(.)$ are likelihood and goodwill state equations, respectively.

4. If $N_{eff} < 0.8N_s$ resample with replacement from the set $\{\mathbf{g}_t^n\}_{n=1}^{N_s}$, where w_t^n is the probability of resampling the state $\mathbf{g}_t^n$. Reset the weights $w_t^n = \frac{1}{N_s}$.

5. Set $t \to t+1$, Repeat Step 2 until end of time period (*T*). Filtered estimates of the full posterior are obtained from: $\{\mathbf{g}_{0:T}^n, \mathbf{w}_{0:T}^n\}_{n=1}^{N_s}$.

<u>Backward Sampling Algorithm:</u> (See Godsill, Doucet, and West 2004)

1. Choose $\hat{\mathbf{g}}_T = \mathbf{g}_T^n$ with probabilty w_T^n.

2. For $t = T-1$ to 1.

   - Calculate $w_{t|t+1}^n \propto w_t^n f\left(\hat{\mathbf{g}}_{t+1} \middle| \mathbf{g}_t^n\right)$ for each $n = 1,.., N_s$.
   - $\hat{\mathbf{g}}_t = \mathbf{g}_t^n$ with probabilty $w_{t|t+1}^n$.

3. The results are draws $\hat{\mathbf{g}}_{1:T} = \{\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2,,,, \hat{\mathbf{g}}_T\}$ from the full conditional posterior, $p(\mathbf{g}_{0:T} | \mathbf{y}_{1:T}, \xi)$ (See Bruce 2008 for more details).

2.0 Sampling $p(\varsigma | \mathbf{g}_t, \boldsymbol{\theta}_t, \mathbf{m}_t)$

Given draws for the state vectors $\mathbf{g}_t = \{g_{11t}, g_{12t}, ..., g_{IJt}\}$ and $\boldsymbol{\theta}_t = \{\theta_{11t}, \theta_{12t}, ..., \theta_{IJt}\}$, one can sample the fixed parameters $\zeta = \{\zeta_1, \zeta_2\}$ using basic MCMC ideas. For example, let $\varsigma_1 = \{\sigma_j, \alpha_i, \delta_{ij}, \boldsymbol{\beta}_{ij}\}$ be the vector of fixed parameters, defined in the state equation (5); and $\varsigma_2 = \{\eta_{ij}, B_{ij}\}$ be those defined in the measurement equations (6-7), chosen to have independent normal priors. Assume also that the prior on the system and observation variances $A_2 = \{\boldsymbol{\Sigma}_{ij}, \mathbf{H}_{ij}\}$ in the measurement model are each independent inverse Wisharts, with all priors chosen to be vague. The joint distribution $p(\varsigma_2, A_2 | \mathbf{m}_{1:T}, \mathbf{g}_{1:T}, \boldsymbol{\theta}_{0:T})$ is thus readily simulated via a set of normal and inverse wishart conditionals (Gamerman, 1997; Rossi et al. 2005). Finally, with conditional

distribution $p\left(\mathbf{v}_{ijt}^{g} \middle| \mathbf{v}_{ijt}^{m}\right)$ derived from measurement model, one can similarly sample conditional posteriors and $p\left(\varsigma_1 \middle| \mathbf{g}_{0:T}, \mathbf{H}\right)$.

TABLE 1: DATA SUMMARY BY AD TARGETS (MEANS)

Measures	*Format Types*	*Retargeting*	*Male*	*Female*	*Age*	*Total*
Clicks	Flash (160x600)	155.82	139.43	285.82	1382.68	2117.46
	Flash (300x250)	292.94	70.55	217.36	1453.74	2034.59
	Flash (728x90)	420.68	128.80	274.52	1704.24	2528.24
	GIF (160x600)	1.45	0.74	1.79	14.83	18.81
	GIF (300x250)	1.53	0.25	2.21	19.69	23.68
	GIF (728x90)	1.60	0.31	1.69	16.41	20.01
Impressions	Flash (160x600):					
	Product	181845.58	114546.73	149402.92	1098548.9	1544344.1
	Price	402601.12	239398.52	520822.05	1024514.6	2187336.3
	Flash (300x250)					
	Product	188393.25	85656.40	140812.78	1027470.3	1442332.74
	Price	448608.59	178961.81	412859.85	1207589.7	2248019.91
	Flash (728x90)					
	Product	265517.16	142603.98	204317.82	1475542.2	2087981.16
	Price	570619.16	276865.37	601745.19	1508415.9	2957645.10
	GIF (160x600):					
	Price	331.16	64.19	449.08	529.85	1116.22
	International*	155.09	443.08	814.88	6584.31	7997.36
	GIF (300x250):					
	Price	311.68	47.95	927.75	535.08	1580.63
	International	235.36	297.36	1203.09	14454.53	16190.34
	GIF (728x90):					
	Price	259.74	69.18	794.23	584.44	1527.88
	International	363.46	460.91	1215.66	13104.37	15144.39

T=154 Days, *International refers to blank impressions ("white spaces") served to exclude consumers in overseas markets.

TABLE 2: CLICK-THROUGH (%) BY MEDIA FROMAT AND AD TARGETS

Format Types	*Remarketing*	*Male*	*Female*	*Age*	*Total*
Flash (160x600):	0.0530	0.0394	0.0426	0.0652	0.0567
Flash (300x250)	0.0460	0.0267	0.0393	0.0651	0.0551
Flash (728x90)	0.0503	0.0307	0.0340	0.0571	0.0501
GIF (160x600)*:	0.1003	0.0906	0.0043	0.0684	0.0134
GIF (300x250)*:	0.1359	0.0068	0.0716	0.0598	0.0686
GIF (728x90)*:	0.0357	0.0727	0.0525	0.0413	0.0435

T=154 Days, *CTR – price ads

Table 3: ALTERNATE MODELS

Models	*Description*	*DIC*	*Rank*
Model 1	Dynamic Zero-Inflated Poisson (DZIP)	14256.9	1
Model 2	Dynamic Poisson (DP)	22989.4	2
Model 3	Dynamic Hurdle Poisson (DHP)	23733.6	3
Model 4	Zero-Inflated Poisson, No Dynamics: δ=0	25402.3	4
Model 5	Dynamic Zero-Inflated Negative-Bin (DZINB)	31705.8	5
Model 6	Dynamic Negative-Bin (DNB)	31978.5	6
Model 7	Zero-Inflated Negative-Bin, (ZINB), δ=0	32556.4	7
Model 8	Normal Dynamic Linear Model (NDLM)	33455.8	8

Table 4: EFFECT OF PRODUCT CONTEXTUAL VARIABLES ON IMPRESSIONS

Targets	*Retargeting*	*Male*	*Female*	*Age*
Product Ads				
Category 1	**0.0919**	0.0407	-0.0122	-0.0094
Category 2	-0.0442	-0.0284	-0.2460	-0.1262
Category 3	**0.5249**	**0.6214**	**0.8791**	**0.6189**
Category 4	**0.7994**	**0.9754**	**1.1196**	**0.9306**
Category 5	**0.6459**	**0.5762**	**0.8399**	**0.9794**
Category 6	0.3274	0.2446	0.2356	0.3586
Category 7	-0.1851	-0.3391	**-0.4695**	**-0.5563**
Category 8	**1.2745**	**1.3370**	**1.8473**	**1.3779**
Category 9	-0.0643	-0.1789	-0.3717	-0.1785
Category 10	0.1265	-0.1205	-0.0904	-0.4301
Category 11	-0.0387	-0.1015	-0.0675	0.0682
Category 12	**-0.7680**	**-0.9852**	**-1.4167**	**-0.8812**
Category 13	**-0.7346**	**-0.8950**	**-1.3714**	**-0.8087**
Category 14	-0.1503	0.4245	-0.1945	-0.6860
Price Ads				
Category 1	0.0640	0.0510	0.0346	-0.1529
Category 2	0.0960	-0.0135	0.3031	0.2283
Category 3	0.0257	-0.0559	0.2017	-0.0575
Category 4	0.1597	0.0530	0.1768	0.1298
Category 5	**0.3704**	**0.4333**	**0.8596**	**0.8727**
Category 6	-0.0762	-0.0776	0.0800	-0.0998
Category 7	-0.0525	0.0486	0.0717	-0.0942
Category 8	-0.1455	-0.1300	0.1539	-0.1491
Category 9	**0.2914**	**0.3909**	0.2642	**0.4060**
Category 10	0.0729	-0.0262	-0.1305	0.0099
Category 11	**-0.1270**	**-0.2344**	-0.2613	-0.1664
Category 12	0.1459	-0.0054	-0.2542	0.0390
Category 13	0.1426	0.2781	**-0.4545**	0.1064
Category 14	0.2922	0.2111	0.4132	0.4688

*bold 95% HPDI

Table 5: MEASUREMENT MODEL: CORRELATIONS WITH GOODWILL ERROR

Digital Format	*Message*	*Retargeting*	*Male*	*Female*	*Age*
Flash (160x600)	Product	**0.1801**	**0.2788**	**0.1596**	**0.2224**
	Price Offer	0.1342	**0.2669**	0.1810	**0.3137**
Flash (300x250)	Product	**0.2350**	**0.1737**	**0.2074**	**0.2556**
	Price Offer	0.1294	**0.2492**	**0.2788**	**0.4743**
Flash (728x90)	Product	**0.2188**	**0.2660**	**0.2077**	**0.2191**
	Price Offer	**0.1810**	**0.2387**	**0.2111**	**0.3029**
GIF (160x600)	Price Offer	0.1168	0.0035	0.1290	**0.2438**
	International	0.1155	0.0877	0.0432	-0.0592
GIF (300x250)	Price Offer	**0.2051**	0.0660	0.1416	**0.2316**
	International	0.1402	0.0255	-0.0996	**0.4237**
GIF (728x90)	Price Offer	0.1016	0.0063	**0.2059**	**0.3756**
	International	0.0930	**0.3027**	0.0583	0.0226

*95% HPDI

TABLE 6: ESTIMATES FROM PROPOSED MODEL: FORMAT, MESSAGE AND AD TARGETS

Parameters:	*Retargeting*	*Male*	*Female*	*Age*
Flash Effect	**2.4688**	**2.3394**	**2.8292**	**2.5346**
Banner Orientation- Size:				
160x600, σ_1	**-1.5515**	**-1.8996**	**-2.5996**	**-2.1887**
300x250, σ_2	**-2.0080**	**-2.1199**	**-2.7617**	**-1.3978**
728x90, σ_3	**-1.1505**	**-1.8390**	**-2.6533**	**-2.3606**
Flash Formats:				
Flash (160x600):				
Carryover Rate, δ_1	**0.5243**	**0.6180**	**0.5582**	**0.6767**
Product Offer, β_{11}	0.0025	**0.0254**	**0.0372**	**0.0303**
Price Offer, β_{12}	**0.1264**	**0.0965**	**0.1472**	**0.1164**
Flash (300x250):				
Carryover Rate, δ_2	**0.6849**	**0.7004**	**0.5996**	**0.6397**
Product Offer, β_{21}	0.0110	**0.0308**	**0.0335**	**0.0245**
Price Offer, β_{22}	**0.0960**	**0.1051**	**0.1341**	**0.0865**
Flash (728x90):				
Carryover Rate, δ_3	**0.6734**	**0.6764**	**0.5835**	**0.7547**
Product Offer, β_{31}	-0.0046	0.0068	**0.0325**	**0.0223**
Price Offer, β_{32}	**0.0784**	**0.0406**	**0.1330**	**0.0935**
GIF Formats:				
GIF (160x600):				
Carryover Rate, δ_4	**0.1190**	**0.1223**	**0.1482**	**0.1116**
Price Offer, β_{41}	-0.0179	0.0231	**0.1062**	0.0243
International, β_{42}	**0.7783**	**0.3196**	**0.8475**	**0.8196**
GIF (300x250):				
Carryover Rate, δ_5	**0.1258**	**0.2725**	**0.1916**	**0.1714**
Price Offer, β_{51}	**0.0961**	-0.0483	**0.1230**	-0.0050
International, β_{52}	**0.7774**	**0.3661**	**0.8076**	**0.7053**
GIF (728x90):				
Carryover Rate, δ_6	**0.0869**	**0.1410**	**0.1615**	**0.1288**
Price Offer, β_{61}	**0.0722**	0.0117	**0.1344**	-0.0188
International, β_{62}	**0.6507**	**0.1920**	**0.7671**	**0.7942**

*bold 95% HPDI

Table 7: 90% Depreciation and Ad Elasticity

Digital Format	*Mean (δ)*	D_{90} (Days)	*Elasticity of Product*	*Elasticity of Price Offer*
Retargeting				
Flash (160x600)	**0.5243**	4.8404	0.0043	**0.2663**
Flash (300x250)	**0.6849**	7.3075	0.0344	**0.3128**
Flash (728x90)	**0.6734**	7.0502	-0.0157	**0.2437**
GIF (160x600)	**0.1190**	2.6136	---	-0.0182
GIF (300x250)	**0.1258**	2.6639	---	**0.0983**
GIF (728x90)	**0.0869**	2.5217	---	**0.0786**
Male				
Flash (160x600)	**0.6180**	6.0277	**0.0695**	**0.2540**
Flash (300x250)	**0.7004**	7.6855	**0.1073**	**0.3708**
Flash (728x90)	**0.6764**	7.1155	0.0209	**0.2440**
GIF (160x600)	**0.1223**	2.6234	---	0.0354
GIF (300x250)	**0.2725**	3.1651	---	-0.0621
GIF (728x90)	**0.1410**	2.6805	---	0.0140
Female				
Flash (160x600)	**0.5582**	5.2118	**0.0885**	**0.3471**
Flash (300x250)	**0.5996**	5.7505	**0.0968**	**0.3051**
Flash (728x90)	**0.5835**	5.5284	**0.0784**	**0.3210**
GIF (160x600)	**0.1482**	2.7032	---	**0.1233**
GIF (300x250)	**0.1916**	2.8483	---	**0.1362**
GIF (728x90)	**0.1615**	2.2761	---	**0.1595**
Age				
Flash (160x600)	**0.6767**	7.1221	**0.0956**	**0.3660**
Flash (300x250)	**0.6397**	6.3907	**0.0682**	**0.2412**
Flash (728x90)	**0.7547**	9.3869	**0.0939**	**0.3933**
GIF (160x600)	**0.1116**	2.5918	---	----
GIF (300x250)	**0.1714**	2.7789	---	----
GIF (728x90)	**0.1288**	2.6430	---	----

*bold 95% HPDI, Elasticity Evaluated at Posterior Draws, $Elasticity = \frac{\partial \lambda}{\partial a}\frac{a}{\lambda} = \frac{a\beta f'(a)}{(1-\delta)}$

Table 8: ALLOCATIONS -- Actual and Model Based Impressions (1.0e+8)

Digital Format	*Exposure*	*Actual Impressions*	*Model-Based*	*% Change (t-1|t)(1)*
Flash (160x600)	Product	**2.3783**	**2.3053**	-3.0676
	Price Offer	**3.3685**	**4.5319**	34.5372
Flash (300x250)	Product	**2.2212**	**1.6987**	-23.5230
	Price Offer	**3.4620**	**3.4985**	1.0540
Flash (728x90)	Product	**3.2155**	**2.3561**	-26.7275
	Price Offer	**4.5548**	**4.7973**	5.3249
GIF (160x600)	Price Offer	**0.0017**	**0.0037**	115.2477
GIF (300x250)	Price Offer	**0.0024**	**0.0091**	279.9628
GIF (728x90)	Price Offer	**0.0024**	**0.0072**	201.8751
Targeted Segments				
Retargeting	------	**3.17005**	**3.8480**	21.3872
Male	------	**1.59892**	**1.07883**	-32.5273
Female	------	**3.12948**	**3.2943**	5.2674
Age	------	**11.31338**	**10.9905**	-2.8541
Message Content				
Product		**7.8150**	**6.3601**	-18.6164
Price		**11.3920**	**12.8477**	12.7806

Figure 1: Digital Ad Formats

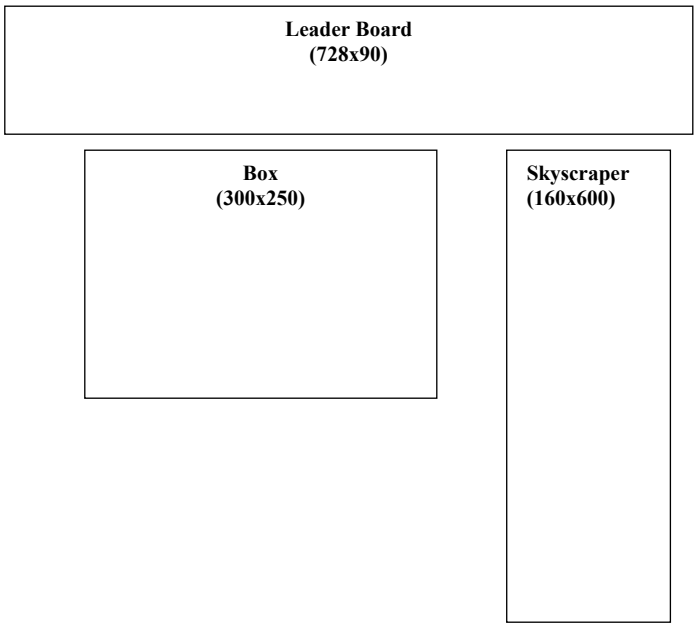

Figure 2: Plot of Clicks by Ad Formats

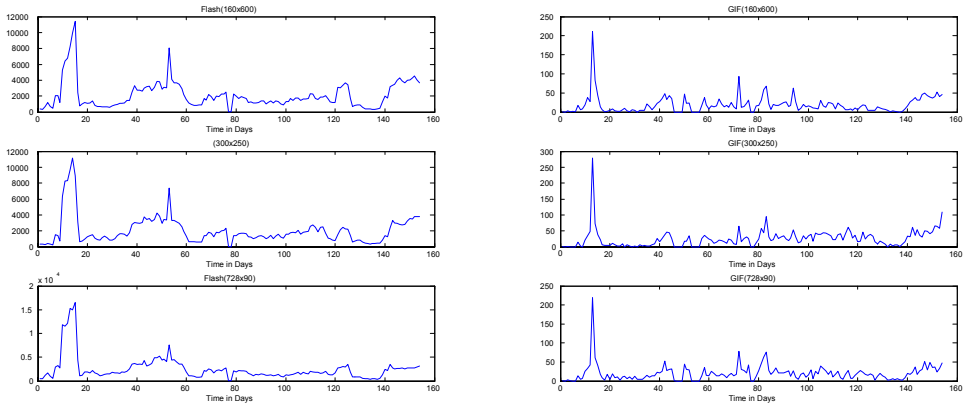

Figure 3: Clicks by Ad Formats and Targets

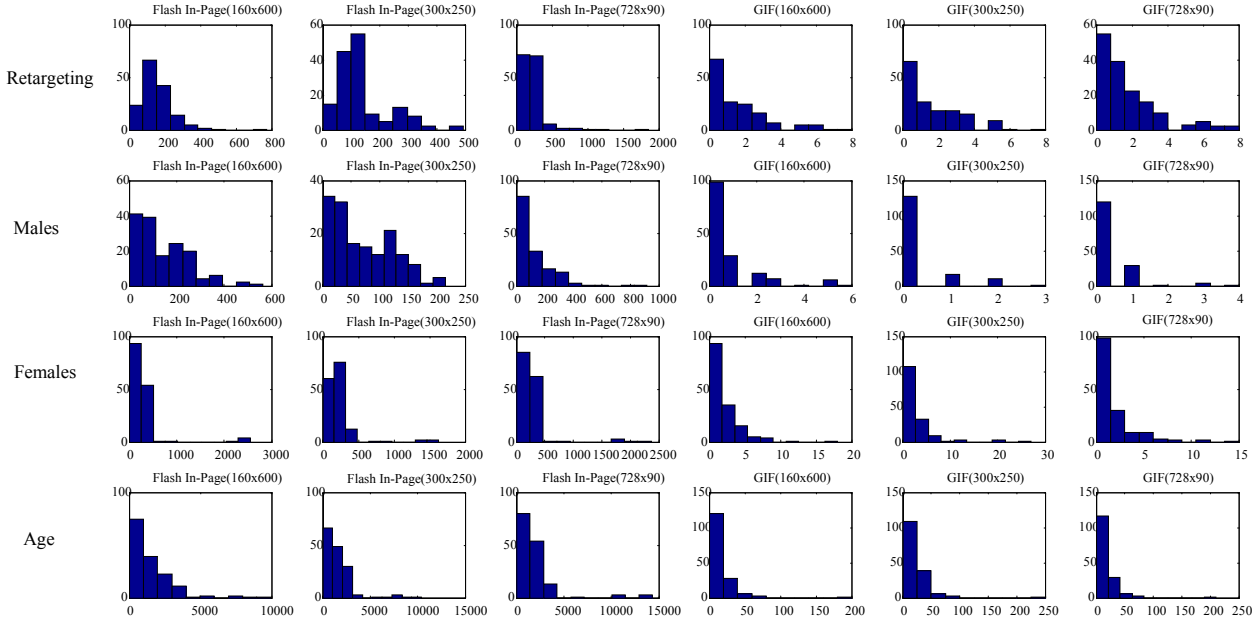

Figure 4: Plot of Clicks vs. Posterior Means (Ad Formats and Targets)