Banner Advertising as a Customer Retention Tool in Customer Relationship Management

Puneet Manchanda
Jean-Pierre Dubé
Khim Yong Goh
Pradeep K. Chintagunta

January 2002
This version: September 2003

Puneet Manchanda is Associate Professor of Marketing, Jean-Pierre Dubé is Assistant Professor of Marketing, Khim Yong Goh is a doctoral student and Pradeep K. Chintagunta is Robert Law Professor of Marketing at the Graduate School of Business, University of Chicago. The authors would like to thank France Leclerc, Stijn M. J. van Osselaer, Wendy Moe and Young-Bean Song for feedback; John Mracek, Steve Findley, Nancy Niu and Susan Gertzis for help on data issues; two anonymous firms for providing the data and the Kilts Center for Marketing at the Graduate School of Business, University of Chicago, for research support. Manchanda and Dubé would also like to acknowledge research support from the Beatrice Foods Faculty Fund at the University of Chicago. All correspondence may be addressed to the first author at the University of Chicago, Graduate School of Business, 1101, East 58th Street, Chicago, IL 60637 or via e-mail at puneet.manchanda@gsb.uchicago.edu.

Abstract

One of the major advances of the digital economy is the facilitation of building and managing individual customer relationships – a process usually referred to as "customer relationship management" or "CRM". For a typical web site selling frequently-purchased consumer items, the most important stage of CRM is customer retention. This is because the long-term viability of a website is based on its ability to retain a significant customer base. In this study, we focus on a hitherto unexplored question – does banner advertising have a role to play in the customer retention phase of CRM. Using a rich behavioral database consisting of individual customer purchases at a web site along with individual advertising exposure, we measure the impact of banner advertising on customer retention (via purchase acceleration).

We formulate a model of individual purchase timing behavior as a function of advertising exposure. We model the probability of a current customer making a purchase in any given week (since last purchase) via a survival model. The duration dependence in the customers' purchase behavior is captured through a flexible, piecewise exponential hazard function. The advertising covariates enter via a proportional hazards specification. These covariates, richer than have typically been used in past research, consist of strictly advertising variables such as weight and quality as well as advertising/individual browsing variables represented by where and how many pages on which customers are exposed to advertising. Our model also controls for unobserved individual differences by specifying a distribution over the individual customer advertising response parameters. We do this by formulating our model in a hierarchical Bayesian framework. This also allows us to provide some insights into where the returns from targeted banner advertising are the highest and the extent to which the returns are higher compared to no targeting.

Our results show that the number of exposures, number of websites and number of pages on which a customer is exposed to advertising all have a positive effect on customer retention. Interestingly, increasing the number of unique creatives to which a customer is exposed lowers the customer retention probability. We also find evidence of considerable heterogeneity across consumers in response to various aspects of banner advertising. The extent of heterogeneity shows that the returns from targeting individual customers are likely to be the highest for the weight of advertising (the number of advertisements that they were exposed to in a given week) followed by the number of sites that they are exposed to advertising on. To demonstrate the value of the obtained individual response parameters, we carry out a simple experiment in which we compare sales response with and without targeting. We show that, relative to no targeting, targeting results in significant increases in the effectiveness of banner advertising on customer retention and hence, on profitability. Finally, in terms of the broader area of research on the effects of (any type of) advertising, we provide somewhat unique evidence that advertising does affect the purchase behavior of current, in contrast to new, customers.

Keywords: *Advertising Response, Banner Advertising, E-commerce, Internet Retailing, Targeting, Micromarketing, Survival Models, Hierarchical Bayesian Models, Markov Chain Monte Carlo methods*

Introduction

One of the major advances of the digital economy is the facilitation of building and managing individual customer relationships – a process usually referred to as "customer relationship management" or CRM. Researchers have postulated that there are four stages in such a relationship – Awareness, Exploration, Commitment and Dissolution (Mohammed et al. 2002, p. 262-265). The electronic mediation of interactions between consumers and firms vastly expands the sources and quality of data available to firms, and thus the ability to understand the role of various marketing instruments in these four stages. In this research, we focus on a hitherto unexplored question – does banner advertising have a role to play in the commitment phase of this relationship. The commitment phase is characterized by firms and customers having a sense of obligation to engage in exchange transactions. The metric that we use to measure the extent of this obligation is customer retention (i.e., repeat purchasing).

There is some evidence that banner advertising helps in the first two stages of CRM. In the awareness stage, customers recognize the firm as a potential exchange partner but have not yet initiated contact. For example, the IAB On-line Advertising Effectiveness Study (1997) showed that exposure to banner advertising increased advertisement awareness, brand awareness and purchase intention. The increase in brand awareness has also been documented in academic studies such as Dahlen (2001). In the exploration stage, customers consider the possibility of exchange via trial. There exists some preliminary evidence that banner ads facilitate customer acquisition resulting from trial (Sherman and Deighton 2001). However, from the firm's point of view, the third stage – commitment – is the most important. This is because of two reasons. First, behavioral changes favorable to the store in the commitment stage on part of the customer have the highest return relative to the other stages in the short run. Specifically, this occurs when retention rates go up across the current customer base (Winer 2001). Second, as acquiring new customers is less profitable than retaining current customers, the long-run viability of a website is based on its ability to retain a significant customer base.[1] Our focus is to examine if banner advertising helps increase the retention rate by fitting a model of the (repeat) purchase probability. In particular, using a rich *behavioral* database consisting of customer purchases at a website along with individual advertising exposure,

[1] In our application, customer retention occurs through repeat purchasing operationalized in our case as purchase acceleration. As our data are obtained from a single store, we are unable to account explicitly for retention arising through lower store-switching.

we measure the impact of banner advertising on current customers' probabilities of repeat buying and whether there exists any duration dependence in the repeat purchase probabilities.[2]

Internet advertising (and banner advertising in particular) is beginning to emerge as a viable advertising medium (Silk et al. 2001). On-line advertising has become an important component of the Internet economy and the advertising industry in general (Mangalindan 2003). The total industry expenditure on digital media in 2001 was $5.6 billion (*Advertising Age*). These numbers compare favorably with expenditure on more established media such as Outdoor[3] ($2.4 billion), Radio (2.9 billion) and Cable TV ($10.3 billion). The current projections are that on-line advertising expenditures are expected to rise 12% in 2003 to about $6.6 billion. This is mainly a result of lower on-line advertising costs and improved measurement tools (*BusinessWeek* 2003). While several forms of advertising in digital environments have emerged, industry reports indicate that the majority of digital advertisements are banner advertisements (Cho et al. 2001, IAB 1999). A banner advertisement is a section of on-line advertising space that is generally 480 x 60 pixels in size. It typically consists of a combination of graphic and textual content and contains a link to the advertiser's website via a click-through URL (Uniform Resource Locator), which acts as a web address. Given the magnitude of the Internet advertising sector, and banner advertisements in particular, measuring the effectiveness of banner advertising is likely to be of considerable interest to both academics and practitioners.

Interestingly, the effectiveness of banner advertisements has been debated since the early stages of Internet commerce.[4] Websites hosting on-line ads have been pushing for traditional "exposure" based metrics, such as "impressions" served, to allow them to charge for each banner exposure. However, difficulties in measuring on-line impressions precisely have caused much dissatisfaction amongst managers resulting in reluctance to commit funds to banner advertising (Hoffman and Novak 2000). Moreover, advertisers, who prefer to pay based on the performance of their ads, feel that impressions generally overstate advertising effectiveness. Instead, advertisers have been pushing

---

[2] Note that repeat purchase behavior is crucial to merchants selling frequently purchased consumer non-durable products and services. It is likely to be less important in product categories such as "big-ticket" items where the interpurchase intervals are fairly long.

[3] This term refers to advertising media such as billboards and hoardings.

[4] Note that electronic commerce via private value-added networks (VANs) has existed since the mid 1980s. Our focus is specifically on Internet commerce, which only emerged after the commercialization of the World Wide Web during the mid 1990s.

for heuristic metrics of performance such as "click-through", which indicates when a web surfer clicks through to the advertiser's URL via the banner. However, the effectiveness of click-through as a valid measure is also being called into question (Briggs 2001, *BusinessWeek* On-line 2001a, Song 2001). The fact that typical click-through rates are quite small in magnitude, 0.5% on average (Sherman and Deighton 2001, Dahlen 2001, Warren 2001), has led practitioners to believe that banners are ineffective. Moreover, click-through is a measure of a *visit* to the website. Since there is considerably evidence that only a small proportion of visits translate into final purchase (Moe and Fader 2003), click-through may be too imprecise for measuring the effectiveness of banners served to the mass market. These studies therefore, underscore the importance of investigating the impact of banner advertising on actual purchase behavior. The focus on click-through has also strengthened the belief that banner ads influence the awareness and exploration stages of CRM, i.e., they are regarded as customer acquisition tools rather than customer retention tools. For example, 9 out 10 of the members of the Direct Marketing Association state that banner advertisements are more effective for customer acquisition rather than customer retention (*DMA Statistical Fact Book* 2002).

Given our unique behavioral data, we investigate a hitherto unexamined role of banner advertising – specifically, its effect on customer retention. In particular, we examine whether, given a temporal interval since the last purchase, a customer makes a purchase at the website of interest and how this decision is influenced by banner advertising. We formulate a model of individual purchase timing behavior as a function of advertising exposure. We model the probability of a current customer making a purchase in any given week (since last purchase) via a survival model. Effectively, a purchase represents "failure" while no purchase represents "survival". The duration dependence in the customers' purchase behavior is captured through a flexible, piecewise exponential hazard function (Wedel et. al. 1995). The advertising covariates enter via a proportional hazards specification. We use a much richer set of covariates than has been typically used in past research (where advertising is only measured as the amount of exposure). Specifically, the covariates we use consist of strictly advertising variables such as weight and quality as well as advertising/individual browsing variables represented by where and how many pages on which customers are exposed to advertising. Our proposed model also controls for unobserved individual differences by specifying a distribution over the individual customer advertising response parameters. We do this by formulating our model in a hierarchical Bayesian framework. This also allows us to provide some insights into where the returns from targeted banner advertising are the highest and the extent to which the returns are higher compared to no targeting.

In terms of the broader area of research on the effects of (any type of) advertising on individual consumers, our work adds to the studies that have investigated the effects of advertising on purchase timing and incidence behavior in at least two ways. First, it documents the effect of more facets of advertising than has been in studies with individual data (as described above). Second, a banner advertisement is a different form of advertising relative to a standard ad in terms of visual quality, attention-getting ability and creative execution. Thus, our findings complement the findings of the effect of advertising at the individual level described in previous research. Our main finding is that, contrary to popular belief, banner advertising can act as a customer retention tool. This is because our modeling approach allows for acceleration in purchase response as a function of banner advertising exposure. While this finding is new in this context, it is not entirely unexpected. For example, some prior research has documented that the effects of advertising are not necessarily immediate (IAB 1997, Smith and Swinyard 1982, Vakratsas and Ambler 1999). From a managerial perspective, banner advertising acts as a relationship builder in the sense that it has a positive effect on purchase probabilities in any given week (since the last purchase) over and above the duration dependence effects. These results also suggest indirectly that click-through is a relatively poor measure of advertising effectiveness as it accounts for a very small proportion of overall purchases.

We find that the number of exposures, number of websites and number of pages on which a customer is exposed to advertising all have a positive effect on customer retention. Interestingly, increasing the number of unique creatives to which a customer is exposed lowers the customer retention probability. In general, the effect sizes of banner advertising on purchase are in the same order of magnitude as the effects sizes of traditional advertising. We also find evidence of considerable heterogeneity across consumers in response to various aspects of banner advertising. The extent of heterogeneity shows that the returns from targeting individual customers are likely to be the highest for the weight of advertising (the number of advertisements that they were exposed to in a given week) followed by the number of sites that they are exposed to advertising. Using the individual response parameters, we carry out an experiment that demonstrates, even under very simple targeting approaches, there are significant increases in the effectiveness of banner advertising on customer retention and hence, profitability. Finally, in terms of the broader area of research on the effects of (any type of) advertising, we provide somewhat unique evidence that advertising does affect the purchase behavior of current, in contrast to new, customers.

The structure of the paper is as follows. We first briefly discuss prior work in this and related areas. We then give an overview of the data. We present the details of the models next. We then discuss the results and the managerial implications of our findings. We conclude the paper with a discussion of the limitations of the present study and provide directions for future research.

Literature Review

The role of how marketing tools fit into CRM is an emerging area of research (Winer 2001). Our specific focus in this paper is the role of banner advertising in a digital environment such as the Internet. However, our study also builds on a long tradition in marketing of estimating (conventional) advertising response models using individual level data. We therefore discuss the relationship between our study and previous studies in both domains.

First, we provide an overview of academic research in Table 1. Most of the academic (see studies by Dahlen, Cho et al. and Gallagher et al. in Table 1) and industry research on advertising in digital environments has focused on measuring changes in brand awareness, brand attitudes, and purchase intentions as a function of exposure (as against the effects of banner advertising on actual purchase behavior). This is usually done via field surveys or laboratory experiments using individual (or cookie) level data. Thus, the focus has really been on understanding the role of banner advertising on the awareness stage.

Insert Table 1 about here

In contrast to studies using experimental data, Sherman and Deighton (2001) describe the process of serving banner advertisements and collecting response data in detail. They also report the results of an experiment carried out by a web advertising agency and an on-line merchant that showed that targeting advertising to specific customers and websites increases response rates and drives down the average cost-per-action (due to confidentiality restrictions, they report only broad, aggregate level findings).

As mentioned above, there is a long tradition of research in marketing that models response to advertising using conventional scanner panel data.[5] Our research builds upon this tradition by estimating a purchase incidence advertising response model with *individual* level response parameters after controlling for unobserved heterogeneity. Thus, our research complements other research that has used individual level data but has only estimated brand choice models (Tellis 1988 and Deighton et al. 1994). The managerial usefulness of brand choice models that ignore purchase incidence has been questioned by other researchers (Pedrick and Zufryden 1991, p. 112).[6] In terms of previous research that does model purchase incidence, our work extends it via a more detailed treatment of unobserved heterogeneity (e.g., Zufryden 1987 uses a summary measure) as well as the explicit incorporation of advertising covariates (e.g., in contrast to Pedrick and Zufryden 1991). Finally, in contrast to other studies which measure (individual) exposure to advertising via aggregate advertising dollars (e.g., Mela et al. 1998), we use individual banner advertising exposure.

To summarize, our research focuses on a new domain i.e., the role of banner advertising in customer relationship management (via consumer retention) on the Internet. The key differentiating managerial issue on the Internet is that firms and customers can build and manage relationships with *individual* customers in a much more cost-effective manner relative to other domains. Our research examines the influence of one marketing instrument - banner advertising - on a specific aspect of this relationship i.e., customer retention. To this end, our research uses banner advertising exposure and purchase data at the *individual* consumer (cookie) level and calibrates advertising response parameters at the individual level. This also distinguishes from previous research on advertising response using conventional panel data. Our research is also distinct from extant banner advertising research as it has largely been limited to the influence of banner ads on attitudes rather than on behavior.

---

[5] For research on the effects of advertising in conventional media, we refer the reader to two excellent review papers – Lodish et al. (1995) and Vakratsas and Ambler (1999).

[6] Note that, given our data, we are unable to model brand choice.

Findings from industry research (*Businessweek Online* 2001a, 2001b, Tran 2001, Song 2001, DoubleClick Press Release 2001, Warren 2001, Briggs 2001) show that banner advertising has attitudinal effects and that click-through is a poor measure of advertising response. These findings are generally consistent with the findings of the academic research discussed earlier. Interestingly, in addition to the attitudinal effects of banner advertising, we find a few studies that provide some informal evidence of its behavioral effects as well. In this paper, we use a formal model to investigate these behavioral effects for current customers.

Other recent modeling research in marketing has focused on describing browsing behavior (or clickstream) data. These studies, also summarized in Table 1, typically use activity data from web log files and/or surveys and therefore do not capture the effects of marketing instruments on sales (e.g., Chatterjee et al. 2002, Bucklin and Sismeiro 2003, Bhatnagar and Ghose 2003, Sismeiro and Bucklin 2002, Moe and Fader 2003). Our study complements these studies by applying a model that captures the effect of duration dependence and advertising covariates on customer retention using behavioral data.

Data

The data come from an Internet-only firm engaged in selling healthcare and beauty products as well as non-prescription drugs to consumers. The data were processed and made available to us by the advertising agency that was responsible for serving the advertisements for the firm in question. Due to the nature of the data sharing agreement between us and the two firms, we are unable to reveal the name of either firm. The data span a period of three months during the third quarter of 2000, specifically from June 11th to September 16th. The data are available at the individual cookie level. As mentioned earlier, most datasets used to investigate on-line environments usually comprise of browsing behavior only. Our data are unique in that we have individual level stimulus (advertising) and response (purchase incidence). The data are contained in two databases – the *CAMPAIGN* database and the *TRACER* database.

The *CAMPAIGN* database comprises the on-line advertisement banner exposure and click-through response originating from promotional campaigns that were run on websites. The data fields in the *CAMPAIGN* database consist of consumer data - a unique cookie identifier identifying the individual computer,[7] an indicator variable denoting consumer response to the banner advertisement (view or click),[8] and the date and time of banner view or click; and advertising data – the portal or alliance site's web page where the banner advertisement view or click occurred, and a unique key identifying the specific banner advertisement.

In terms of the websites on which the advertising was delivered, the database contains records of the company's advertising on portal and alliance websites such as, among others, Yahoo!, AOL, Women.com, iVillage.com, Healthcentral.com, and E*Trade. These sites accounted for over 80% of all advertising activity by the firm during this period. Note that though we have a unique identifier for each site on which the banner advertisement was served, we do not know the *specific identity* of each site.

Advertising activity typically consisted of a specific creative that operated over several weeks. In terms of the advertising message contained in the various creatives, we know that the majority of the messages were of the brand-building type for the website (i.e., the message consisted of the name of the website and a line describing the benefits of purchasing from the website). A limitation of the data is that we do not have information on the specific message in each banner (even though we have an indicator that tells us that one creative was different from another). This creative was delivered to websites in the form of a digital graphic, generally referred to as a GIF. These GIFs were of the usual size for banner advertisements (480 x 60 pixels).[9] New GIFs were typically released at the beginning of a calendar week i.e., on Sunday and/or Monday, reflecting media buying patterns. During the period covered by our data, there were one hundred total GIFs spread over fifteen major sites. However, the majority of exposures came from a small number of GIFs – seven GIFs accounted for about 55% of all exposures.

---

[7] We use the term consumer and cookie identifier interchangeably for the sake of exposition. However, as mentioned earlier, our data only allow us to identify a unique *computer* and not a unique *consumer*.

[8] Note that since we are working with behavioral data, we are unable to control for the exact nature of exposure. In other words, we are making the assumption that if the consumer was on a specific page and the banner appeared on that same page, s/he actually *viewed* the advertisement. However, given that banner ads are probably even lower involvement that TV ads (small size, exposure in the presence of competing information), this argues *against* our finding any effects. Our model estimates may thus be seen as the lower bound of the true estimates.This assumption is also consistent with prior research that has tried to document the effect of advertising exposure on sales for individual consumers (see detailed discussion in Deighton et al. 1994, p. 34).

[9] The majority of the firm's banner advertising used standard banner sizes. A very small proportion (< 5%) of advertising consisted of short banners (392 x 72 pixels) and vertical banners (120 x 240 pixels). Our data does not provide information on the size of a specific banner advertisement in the data.

Finally, the *TRACER* database contains the date and time of the purchase transaction for each unique cookie identifier. Note that we do not have information on visits to the site that did not result in a purchase.

We merged the *CAMPAIGN* database with the *TRACER* database using unique cookie identifiers. This resulted in 14370 unique cookies. We then examined the purchasing patterns of these cookies in the context of our discrete time formulation. Given that banner advertising activity was planned by the firm for each week, we chose the time interval to be a single week. Hence, our unit of observation is a "cookie-week." However, if there are a significant number of cookies that purchase multiple times in a single week, our model would be inappropriate. An examination of the data revealed that 99% of the 14370 cookies did not purchase multiple times in any given week. We then deleted all the cookies for which we could construct only one observation (i.e., if their purchase occurred in the last six calendar days of our data) as we would be unable to obtain individual level parameters for these cookies (we describe how we construct the weekly data for each cookie in the model specification section below). Finally, we deleted purchase transactions with blank cookies, repeat transactions (identical transactions at identical times) and observations with obvious data entry errors. This resulted in a panel of 12748 cookies with a total of 97805 observations. The number of observations in the data is the sum (over the 12748 cookies) of the total number of weeks for each cookie after the cookie's first purchase.

Out of these 97805 observations, a purchase[10] is made on 14.3% (13955) observations while there is no purchase on the remaining 85.7%.[11] This proportion compares favorably with purchase incidence studies that use scanner data, e.g., Bucklin and Lattin (1991) report that heavy buyers in the saltine cracker category purchased on 6% of all trips over a two-year period; Bucklin and Gupta

---

[10] As mentioned earlier, we only have data from one online store. Thus, when we refer to "no-purchase" cookie-weeks, we only refer to no purchase at that store. It is possible that the customer made purchases in that week at other (online and/or offline) stores.

[11] The firm provided us click-through information *only* if it resulted in a purchase. This was done to minimize the data processing effort and the size of the resulting database. The firm's advertising agency confirmed that the mean click-through rate during this period was between 0.25%-0.5% (in our sample, the click-through rate is close to 0.25%). This is consistent with the rates documented in other studies discussed earlier (though this rate is somewhat lower). In addition, it seems clear from the data that click-through only purchases are an order of magnitude smaller than purchases driven by banner advertising (1134 purchases versus 13955 purchases across all purchasers). These data, combined with feedback from the firm's executives, lead us to conclude that click-through is not an important path to purchase for customers of this website.

(1992) report a 7.3% purchase proportion across all trips in the liquid detergent category over a two-year period. In comparison with other research using data from on-line environments, these data also appear reasonable. For example, Moe and Fader (2003) report that only 851 out of 10000 panelists made at least one purchase on Amazon.com over an eight month period. The average number of purchases by these consumers in this time period was 1.85. Similarly, Chatterjee et al. (2002), are able to retain only 3611 out of 21783 registered panelists given at least three exposures over a seven month period.

Modeling the Customer-Website Relationship

We investigate the purchase behavior of customers who are exposed to banner advertising by the website. We model the potentially duration dependent purchase incidence decision – whether or not and when to buy from the website - via a semiparametric survival model. Specifically, we estimate a constant piecewise exponential hazard model in discrete time (Wedel et al. 1995). This allows the intrinsic purchase incidence probabilities, in the absence of covariates, to vary over time. The decision of when to purchase is also modeled as a function of the advertising exposure and browsing behavior variables at the individual customer level. To capture variability in individual choices, we allow for the individual response parameters to be distributed across customers. Thus, customer retention as a function of exposure to banner advertising is captured in the effect of the banner advertising covariates on the purchase probabilities in a given week since the last purchase.

Model Formulation

As noted above, our model formulation focuses on the weekly purchase decision i.e., consumers decide every week whether they plan to purchase or not as a function of the timing of their last purchase, marketing and behavioral variables as well as unobserved heterogeneity. Our model falls into the class of semiparametric survival models (Meyer 1990). In this model, the no purchase weeks for each customer are treated as the "survival" weeks while the purchase weeks are treated as the "failure" weeks. Earlier modeling research using customer browsing data has found evidence of heterogeneity (Moe and Fader 2003, Bucklin and Sismeiro 2003). We therefore account for heterogeneity using a continuous distribution over the individual customer response parameters. We cast our model in a hierarchical Bayesian framework and estimate it using Markov chain Monte Carlo methods (see Rossi and Allenby 2003 for a detailed review of such models). In general, with a few notable exceptions (Allenby et al. 1999, Lee et al. 2003), the use of proportional hazard models under the Hierarchical Bayesian framework has been somewhat limited in the marketing literature.

We now describe the specific model, the prior distribution of the unknowns, the likelihood function and the resulting posterior distributions. The main advantage of the semiparametric specification is that it does not impose a specific distributional assumption or a shape on duration dependence, i.e., the baseline hazard. In scanner data contexts, one can argue that given the preponderance of evidence showing non-monotonic hazards, a specific functional form showing such a pattern can be imposed on the data. However, no such evidence is currently available for (Internet) data such as ours. This drives our choice of the flexible functional form of the semiparametric specification.

The Semiparametric Survival Model

Let t_{ij} denote the interpurchase time for consumer i's spell *j*. Then the survivor function corresponding to this time is given by:

$$S(t_{ij}) = \exp(-\int_0^{t_{ij}} h(u)du) \tag{1}$$

Note that since our data are discrete survival data, we use a discrete time model to predict the probability of purchase. We first split the time axis into a finite number of intervals, $0 < s_1 < s_2 < \ldots < s_J$, with $s_J > y_{it}$ for all $i = 1,2,...,I$ and $t = 1,2,...,T_i$, where y_{it} represents the survival time for customer i's t^{th} observation. Thus, we have J intervals, $(0, s_1], (s_1, s_2], \ldots, (s_{J-1}, s_J]$ .Following the convention in the discrete-time semiparametric hazard function literature (e.g., Meyer 1990), we replace the integral in equation (1) for each of the J intervals by the following expression:

$$\int_{(t-1)_{ij}}^{t_{ij}} h(u)du = \exp(\lambda_j) \tag{2}$$

This represents a *piecewise exponential hazard* model where we assume a constant baseline hazard, $h_{0j}(y) = \log(\lambda_j)$, for $t_{ij} \in I_j = (s_{j-1}, s_j]$ where I_j is the indicator function. The $\log(\lambda_j)$ parameters do not correspond to calendar time but to the time interval following the last purchase.

They enable us to assess whether the data indicate duration dependence when the parameters are different for different time intervals or durations. Note that, for most cookies, we only have one observation for each of the J intervals. Thus the data support inference about the baseline hazard only at the pooled level i.e., we cannot specify a heterogeneity distribution across customers for any of the $\log(\lambda_j)$ parameters.

We then let the effect of the covariates enter multiplicatively i.e., we use a proportional hazard formulation. Let x_{pij} represent the p^{th} covariate for customer i in the time interval j. As we have repeated measures across customers (once we control for the pooled baseline hazard), the response parameters can be customer specific. Thus, equation (2) becomes:

$$\int_{(t-1)_{ij}}^{t_{ij}} h(u)du = \exp(\lambda_j + \sum_{p=1}^{P}(x_{pij} * \beta_{pi})) \tag{3}$$

The piecewise exponential model is general in the sense that it is sufficiently flexible to accommodate a wide variety of shapes of the baseline hazard. Note that if $J = 1$, the model reduces to a parametric exponential model with a failure rate $\lambda = \lambda_1$. It is also parsimonious in the sense that there is only one unknown parameter per time period.

Given equation (3), the probability of purchase ("failure") in any of the j time intervals for a customer i is given as:

$$\Pr_{ij}(purchase) = 1 - \exp(-\exp(u_{ij})) \tag{4}$$

where $u_{ij} = \sum_{j=1}^{J}(\lambda_j * I_j) + \sum_{p=1}^{P}(x_{pij} * \beta_{pi})$ where $I_j = 1$ in time interval j, 0 otherwise.

Thus, the overall log-likelihood for all the customers in the sample is

$$LL_i(\beta_i, \lambda \mid \nu, x) = \sum_{i=1}^{I}\sum_{j=1}^{J}[\Pr_{ij} * (1 - \nu_{ij}) + (1 - \Pr_{ij}) * \nu_{ij}] \tag{5}$$

where ν_{ij} is an indicator function which is equal to 1 if customer i purchases in time interval j, 0 otherwise and β_i and λ are vectors of β_{pi} and λ_j.

The Bayesian hierarchy and Inference

We cast our model in a hierarchical Bayesian framework. Given that we would like to obtain simultaneously the cross-sectional parameters for the discrete time hazards and the individual level parameters for the response coefficients, this framework is particularly appealing. Under this framework, to complete the model, we need to specify the prior distribution of the unknowns and derive the full conditional distributions.

Let $\psi_j = \log(\lambda_j)$ for $j = 1, 2, ..., n$. We assume that ψ_j are distributed multivariate normal with mean ψ_0 and variance V_ψ. We capture unobserved heterogeneity via the distribution of β_i (where β_i is the vector of the response parameters) by allowing for them to be distributed multivariate normal with mean β_0 and variance V_β i.e.,

$$\beta_i = \beta_0 + \nu_i \tag{6}$$

where $\nu_i \sim \mathrm{N}\left(0, V_\beta\right)$. The hyperparameters β_0 and V_β are distributed normal and Inverse Wishart respectively.

We now derive the full conditional distributions of the unknowns, $(\psi, \beta_i, \beta_0, V_\beta)$, using the joint density (equation 5) and the specified prior distributions. To obtain the posterior distribution of the unknowns, we then draw sequentially from this series of full conditional distributions until convergence is achieved.

The ψ are distributed $\mathrm{N}(\psi_0, V_\psi)$. Thus the full conditional distribution for ψ is given as:

$$p(\psi \mid \psi_0, V_\psi, \beta_i, y_{ij}, x_{ij}) \propto \ell\left(\psi\right) * \exp((\psi - \psi_0) * V_\psi^{-1} * (\psi - \psi_0)^{'})$$

where $\ell\left(\psi\right)$ is as given in equation (5). The full conditional distribution for ψ is known only up to a proportionality constant. We use the Random Walk Metropolis-Hastings algorithm to generate a candidate on iteration n as $\psi^{\mathrm{c}} = \psi^{(\mathrm{n}-1)} + \breve{\psi}_i$, where $\breve{\psi}$ is a draw from a multivariate normal proposal density, $\mathrm{N}(0, k_\psi \Psi)$. We set Ψ to the asymptotic variance-covariance matrix of the ψ parameters estimated using maximum likelihood on pooled data (assuming no customer level differences). k_ψ is a scalar that is chosen to achieve a reasonable acceptance rate. The acceptance probability is given by $\min\{\frac{p(\psi^{c} \mid \psi_0, V_\psi, \beta_i, y_{ij}, x_{ij})}{p(\psi^{(n-1)} \mid \psi_0, V_\psi, \beta_i, y_{ij}, x_{ij})}, 1\}$ where $p(.|.)$ is as given above. We set $\psi_0 = 0$ and $V_\psi = \text{diag}(100)$.

The $\{\beta_i\}$ are distributed $\text{N}(\beta_0, V_\beta)$ (equation 6). Thus the full conditional distribution for $\{\beta_i\}$ is given as:

$$p(\beta_i \mid \psi, \beta_0, V_\beta, y_{ij}, x_{ij}) \propto \ell(\beta_i) * \exp((\beta_i - \beta_0) * V_\beta^{-1} * (\beta_i - \beta_0)^{'})$$

where $\ell(\beta_i)$ is as given in equation (5). The full conditional distribution for β_i is known only up to a proportionality constant. We use the Random Walk Metropolis-Hastings algorithm to generate a candidate on iteration n as $\beta_i^{c} = \beta_i^{(n-1)} + \breve{\beta}_i$, where $\breve{\beta}_i$ is a draw from a multivariate normal proposal density, $\text{N}(0, k_\beta \text{B})$. We set B to the asymptotic variance-covariance matrix of the β parameters estimated on pooled data (i.e., assuming no customer level differences) using maximum likelihood estimation. k_β is a scalar that is chosen to achieve a reasonable acceptance rate. The acceptance probability is given by $\min\{\frac{p(\beta_i^{c} \mid \psi, \beta_0, V_\beta, y_{ij}, x_{ij})}{p(\beta_i^{(n-1)} \mid \psi, \beta_0, V_\beta, y_{ij}, x_{ij})}, 1\}$ where $p(.|.)$ is as given above.

The β_0 are distributed $\text{N}(\beta_{00}, V_{\beta_0})$. The full conditional distribution for β_0 is given as:

$$p(\beta_0 \mid \{\beta_i\}, V_\beta, \beta_{00}, V_{\beta_0}) = \text{N}(\hat{\beta}, \hat{V}_\beta)$$

where $\hat{\beta} = \hat{V}_\beta (V_{\beta_0}^{-1} * \beta_{00} + \sum_{i=1}^{I} V_\beta^{-1} * \beta_i)$ and $\hat{V}_\beta = (V_{\beta_0}^{-1} + \sum_{i=1}^{I} V_\beta^{-1})$. We set $\beta_{00} = 0$ and $V_{\beta_0} = \text{diag}(20)$.

Finally, we derive the conditional distribution for V_β^{-1} which is given as:

$$p(V_\beta^{-1} \mid \{\beta_i\}, \beta_0, \rho, \text{R}) = \text{Wishart}([\rho * \text{R} + \sum_{i=1}^{I} (\beta_i - \beta_0)(\beta_i - \beta_0)^{'}]^{-1}, \rho + I)$$

where I is the number of customers. We set the prior mean of $V_\beta = (\rho \mathrm{R})^{-1} = \mathrm{diag}(10)$, and the prior degrees of freedom, $\rho = \mathrm{NPAR} + 3$, where NPAR is the dimension of the β vector.

Inference about the unknowns was made by sequentially drawing from the four full conditional distributions outlined above. The C programming language was used to code the sequence of draws. The sampler was run for 50000 iterations and convergence was assessed by examining the time-series of draws. Inference was then based on every fifth draw after a burn-in period of 37500 draws. To ensure proper mixing, the tuning parameters, k_ψ and k_β, were chosen such that the acceptance rate of ψ and $\{\beta_i\}$ was 18% and 32%. These acceptance rates are based on the recommended rates in the literature (cf. Roberts et al. 1997).

Model Specification

In this section, we first discuss how we specify the baseline hazard. We then discuss how we choose and construct our advertising variables on the basis of past research.

We have thirteen calendar weeks in our data. We created the spell variables for each cookie in the following manner. We initialized the first spell for each cookie to the calendar week corresponding to the first purchase occasion. We then created purchase indicators, v_{ij} (equation 5), for each week following this initial week for each cookie. If there was no purchase in a subsequent week, the spell counter was incremented by one and the purchase indicator was set to zero. If there was a purchase, then the indicator variable was set to one. The spell counter was restarted at one for the week following the purchase week. Then thirteen indicator variables, $I_1 \ldots I_{13}$ (equation 4), were created and set to one corresponding to the spell counter for that week for that cookie. As mentioned, the indicator variables do not represent the calendar week but the number of weeks elapsed since last purchase. The coefficients of each of these variables, $\psi_j = \log(\lambda_j)$, represent the constant hazard for that week.[12] The four advertising covariates that we use (described below) are then constructed for each cookie-week.

[12] Note that due to right-censoring, ψ_{13}, represents that hazard of thirteen and higher weeks.

We postulate that the decision of whether to purchase in each week will be affected by advertising exposure (weight and quality) as well as individual differences (both observed and unobserved). We first discuss the advertising variables.

We expect that banner ads act as reminder tools and/or brand builders for current customers. In other words, they have the potential to enhance customer retention via purchase acceleration. Thus, exposure to banner advertising is likely to increase the probability of purchase (Cho et al. 2001).[13] We therefore construct the following variables:

- VIEWNUM represents the total number of advertising exposures in each week for each customer,
- ADNUM represents the number of creatives (GIFS) that the consumer was exposed to each week.

Prior research has shown that repeated exposures to an advertisement prevent the early decay of advertising effects (Pechmann and Stewart 1988, Cacioppo and Petty 1985). We therefore expect that increased exposure to advertising (VIEWNUM) should increase the probability of purchase in a given week. However, at some point the response to advertising should provide diminishing returns. As the empirical evidence in general supports a concave response to advertising weight (Lilien et al. 1992, p. 267), we use *log*(1+VIEWNUM) or LVIEWNUM in our specification.[14] In terms of the variety of creative execution, prior research has indicated that response to different creatives can be quite different (Lodish et al. 1995). It has also been shown that recall is enhanced if consumers are exposed to different creatives in the same campaign (Rao and Burnkrant 1991). However, in our case, while all the creatives essentially advertise the website, they are not part of a single campaign. We therefore have no prediction regarding the effect of ADNUM on inter-purchase times.

We also need to control for differences across consumers in terms of prior purchase behavior and browsing behavior. These differences could arise from both observed and unobserved differences. Observed differences may arise due to two kinds of variation – purely cross-sectional

---

13 As mentioned earlier, almost all of the banner advertisements advertised the site and the benefits of shopping at the site.

14 We use log(1+VIEWNUM) instead of log(VIEWNUM) to accommodate weeks when VIEWNUM=0.

variation (e.g., demographics) and cross-sectional combined with longitudinal variation (e.g., usage and browsing behavior). Usage variables capture systematic differences in customers' use of digital environments. For example, some customers may spend more time on the Internet and may therefore be more prone to buying from web merchants. Thus the individual probability of buying for such a consumer could be affected by individual browsing behavior and individual advertising exposure.[15] Our data do not contain any direct measures of Internet usage and browsing behavior. However, we use the data available to us and develop create two proxy variables that potentially reflect individual differences in Internet usage:[16]

- SITENUM represents the total number of unique websites on which the consumer was exposed to advertising each week.
- PAGENUM represents the number of unique web pages on which the consumer was exposed to advertising each week.

Note that the use of these variables controls for both across-customer (in that the means of these variables are likely to differ across-customers) and within-customer (as there may be differences in these variables for the same customer across weeks) differences. From the usage based arguments laid out above, we would expect these variables to have a positive effect on the decision to purchase each week. From an advertising perspective, prior research has shown that viewing a series of advertisements leads to higher recall and more positive attitudes (Pechmann and Stewart 1988, Zielske and Henry 1980). We therefore expect that the probability of purchase is higher for consumers exposed to advertising on many different websites (SITENUM) and pages (PAGENUM).

Taken together, these four covariates provide a richer description of individual exposure to advertising that has typically been studied in the literature. Specifically, we have data on quantity, quality and the location of exposure in contrast to just quantity. In addition, a banner advertisement

---

[15] Note that individual exposure to advertising may be systematically different across consumers if the firm and its advertising agency were strategically were targeting advertising to *individual* cookies based on prior browsing and/or purchase behavior. However, our discussions with the firm revealed that, during the time-period of our data, this was not the case as the technology to do this was still underdeveloped, resulting in a net loss as a result of targeting. However, this technology has matured and is significantly more cost-effective now.

[16] We would like to thank an anonymous reviewer for suggesting the use of the proxy variables.

is a different form of advertising relative to a standard ad in terms of visual quality, attention-getting ability and creative execution.[17] We summarize the expected signs for each of the variables in our specification in the table below (note that a *positive* coefficient *increases* the purchase probability and a *negative* coefficient *decreases* the purchase probability):

Variable	***Purchase Probability***
LVIEWNUM	+
ADNUM	?
SITENUM	+
PAGENUM	+

In conclusion, the temporal sequence of events for a typical customer in our data is as follows. Every week the consumer is exposed to some advertising spanning (possibly) different creatives. These exposures occur at (possibly) different web pages on (possibly) different websites. As a result, each week the consumer decides whether to purchase or not (given a purchase in the past). Table 2 provides the descriptive statistics for the covariates described above.

Insert Table 2 about here.

Results

Model Estimates: Duration Dependence

The parameters representing the baseline hazard in each period, ψ_j, are detailed in Table 3. As can be seen from the table, all the parameters have posterior distributions that are massed at a considerable distance from zero. This is not surprising given that these are pooled parameters.

Insert Table 3 about here.

[17] We would like to thank an anonymous reviewer for suggesting that we highlight this.

These parameters map directly to the purchase probability in a given week *j*. The higher the magnitude of ψ_j, the higher the probability of purchase. From the table, it can be seen that there is some non-monotonicity in the probability of purchase as the number of weeks since the last purchase goes up. As can be seen from Figure 1, the probability of purchase in a given week increases somewhat for the first three weeks and then remains flat until week six. After week six, we see two peaks in week seven and week ten, followed by a dip in week eleven and then an increase for the remaining weeks. This suggests that the mean interpurchase time is about seven weeks which is consistent with the category of product marketed – Health and Beauty products - by the website. The estimated survival pattern does not conform to any well known parametric survival function formulation. This provides some support for our choice of a piecewise exponential hazard formulation.

Insert Figure 1 about here

It may be useful to compare the hazard obtained under the proposed formulation with that obtained by alternative formulations. In the first alternative formulation considered, we compute the base purchase probability in the following manner. We use the constant hazard parameters, the individual response parameters and the observed data for each draw after the burn-in iterations for each observation using the draws from the posterior distribution of the unknowns. We then compute the purchase probabilities in a similar manner from a formulation that ignores the effect of the covariates (baseline hazard) i.e., it only captures duration dependence. We take the draws from the posterior distribution of the unknowns and set the covariates to zero for each observation and re-compute the purchase probabilities.

In the second alternative formulation, we assume that there is no duration dependence but there is an effect of covariates. We therefore use only one non-time varying intercept and the covariates as before (standard purchase incidence model) and compute the purchase probabilities for each observation as above. We then compute the mean weekly purchase probability. The model in first alternative formulation (no effect of covariates) under-predicts the probability of purchase in any given week relative to the model with covariates. This is not surprising as the effect of the advertising covariates is generally to increase the purchase probability (we discuss this in detail below). The model in the second alternative formulation (no duration dependence) fits the *average* purchase incidence probability across all time periods quite well but does poorly at capturing the differences in the purchase probability for a given week since the last purchase i.e., it predicts an almost flat hazard for the weeks since purchase.

To provide an overall measure of how well our model and the alternative formulations fit the data, we compute Mean Absolute Percent Deviation (MAPD) using the weekly empirical purchase probabilities in the data for each of the models. Across the thirteen weeks, we find that the proposed model has an MAPD of 1% compared to 20% for the first alternative model and 50% for the second alternative model. Thus, both duration dependence and the effect of the covariates need to be modeled to capture the underlying purchase probabilities.

Model Estimates: Advertising Covariates

We next examine the effect of covariates at the mean level i.e., the β_0 vector (see Table 4). The overall pattern of the results indicates that the advertising weight, quality and the individual browsing variables have an effect on the decision to purchase in any given week (all the posterior means are massed away from zero). These effects are as predicted. First, as advertising weight (the log of the number of advertising exposures every week – LVIEWNUM) goes up, the survival probability is lowered. In other words, greater exposure to advertising (numbers) has a positive effect on the purchase probability albeit in a manner consistent with diminishing returns. Interesting, the two main studies that have investigated the effect of advertising on repeat purchasers using individual exposure data are Deighton et al. (1994) and Tellis (1988). Neither study finds any effects of advertising on repeat customers (operationalized as the interaction between exposure and last brand chosen). Thus, these studies find no effects on repeat brand choice behavior. To the best of our knowledge, there are no other studies that have found a positive effect of advertising on current (repeat) customers. Thus, our finding seems somewhat unique in this regard.

However, the effect of advertising quality (the number of creatives that a customer is exposed in every week – ADNUM) is positive on the survival probability. So, exposure to more creatives in a week decreases the probability of purchase. This result is not surprising given that previous research has not hypothesized or documented a specific direction of the relationship. However, there is anecdotal evidence that redundancy in page layout can help consumers learn how to navigate a website more easily.[18] In the context of online advertising, redundancy (i.e. the same message is repeated consistently) may help consumers learn and retain the message of an advertised website. This fact is especially relevant given the plethora of competing banner messages to which a web surfer may be simultaneously exposed. This result is also more consistent with the Lodish et al. (1995) finding cited earlier. In addition, given that the creative content of a banner ad is relatively constrained, exposure to more creatives can lead to more fragmentation rather than reinforcement.

Insert Table 4 about here

The effect of being exposed to the site on many different websites (SITENUM) and many different pages (PAGESNUM) is to lower the survival probability. Thus, broadly speaking, the more the number of locations (sites and pages) on which the consumer is exposed to advertising, the higher the probability of purchase. Cross-sectional differences in browsing behavior across the total number of consumers could account for this effect. However, as discussed below, the fact that the individual response coefficients are all positive implies that even within consumer, exposure on a greater number of locations increases the purchase probabilities.

In terms of the differences across customers, the diagonal elements of the V_β matrix are all positive and massed away from zero in a significant manner (Table 5a). This implies that there is considerable heterogeneity across customers (a more detailed discussion on this follows). There are also interesting correlation patterns across the customer response parameters (Table 5b). First, only two of the six correlations are massed away from zero (all correlations that are massed away from zero are in **bold**). This suggests that the four advertising variables are not highly correlated at the individual level i.e., the four variables measure different facets of responsiveness to advertising. Second, the correlation in response parameters across LVIEWNUM and SITENUM is negative and massed away from zero. This implies that responsiveness to advertising is lower for consumers who are more responsive to being exposed to banner advertising on many websites. As the mean effect of both LVIEWNUM and SITENUM is positive, there are tradeoffs in developing individual level targeting based on these two variables. Third, the correlation in response parameters across ADNUM and SITENUM is negative and massed away from zero i.e., customers who are more

[18] See, for example, the Sothebys.com case (HBS case number 9-800-387 by Roger Hallowell and Abby Hansen, May 25, 2000).

responsive to different creatives are less responsive to being exposed to advertising on many websites. However, as the main effect of ADNUM seems to be negative, this correlation implies that it may be better to expose consumers to the same creative at a small number of sites for maximal response.

Insert Tables 5a-5b about here.

In summary, we find that, in our data, advertising weight and copy affect consumers' decision to visit the website and make purchases. In addition, we also find that cross-sectional differences in browsing behavior have an effect on purchase probabilities. Finally, exposure on distinct locations (sites and pages) for the same consumer also tends to increase the purchase probability. We also find that these response parameters vary across consumers and that there are some interesting correlations across these parameters.

Comparison with Null Models

We investigate the performance of our model relative to two null models. The first null model (NM1) captures the effects of covariates with individual level response parameters but does not capture duration dependence. Specifically, NM1 is a standard purchase incidence model i.e., the model has a binary logit form with an intercept term and the advertising covariates. The second null model (NM2) captures the duration dependence but does not capture the effect of the advertising covariates. We use a standard goodness-of-fit measure, $\rho^2 = 1 - [\mathrm{L(PM)} / \mathrm{L(NM)}]$, to check the performance of our model (PM) relative to the null model(s) (Ben-Akiva and Lerman 1985, p. 91). The within-sample ρ^2 for NM1 is 0.11 and that for NM2 is 0.15. We also carry out a similar comparison on a random holdout sample of 1000 customers. Using the estimates obtained from the remaining customers, we compute the log-likelihood for all three models. We find that the out-of-sample ρ^2 for NM1 is 0.05 and that for NM2 is 0.08.

These comparisons provide some measure of validation of our model. Specifically, they show that capturing both duration dependence and the effect of covariates improves both within and out of sample fit relative to a model that only captures one of the two.

Managerial Implications

In this section, we use our results to explore their implications on managerial practice. First, we compute the average effect sizes of the various advertising variables. Second, we investigate the variation in responsiveness for these advertising variables so that we can obtain an understanding of the returns to targeting.

Elasticities

To understand the extent of the effect sizes, we compute the change in probability of purchase for a ten percent change in the advertising variable for each observation and then compute the mean elasticity across observations. The mean elasticity magnitudes are detailed in Table 6. Note that even though the mean effects are small, the standard deviations indicate that they are massed away from zero. As can be seen from the table, the elasticities are in the same order of magnitude as reported for conventional advertising in the literature (e.g., Sethuraman and Tellis 1991 report an average advertising elasticity of demand of 0.10).

Insert Table 6 about here

There are some interesting implications of these findings. First, it seems that firms need to cut back on exposing customers to different creatives and stick with a smaller set of creatives. This result may also be a reflection of the fact that banner ads are limited in terms of the creative that they can deliver. Second, it seems that the weight and quality of advertising have smaller effects than where consumers are exposed to banner advertising. This reinforces the belief in the industry that delivering a consistent message across many different sites and pages is the most effective method of marketing communication on the Internet where there are many distractions on the same web page. Third, the number of websites on which a customer is exposed to advertising is somewhat more important than the number of pages on which the customer is exposed. So, firms should locate themselves on the high-traffic pages across more websites rather than spread exposures across many possibly unrelated pages.

Returns to Targeting

Given the average effect sizes and the premise that banner advertising plays a significant role in customer retention, managers may then be interested in exploiting the one-to-one targeted marketing potential of the Internet. To do this, we need to compute the returns to targeting across the four advertising/individual difference variables used in our analysis. The answer to this can be obtained by examining the heterogeneity in response parameters across the individual customers. We compute the coefficient of variation (standard deviation divided by the mean) for the distribution of the individual response parameters to describe the size of the variation in response across the four parameters. They are 0.92 (LVIEWNUM), 0.22 (ADNUM), 0.34 (SITENUM) and 0.07 (PAGENUM). These numbers imply that individuals differ the most in terms of response to the number of ads they are exposed to, followed by the number of sites on which they are exposed, the number of creatives they are exposed, and then the number of pages. The returns to targeting therefore follow the same order.

We formalize our findings on the returns to targeting through a stylized revenue (profitability) experiment. First, we classify customers as "high sensitives" (H) and "low sensitives" (L) via a median split on each of the four individual level parameters. We then bin them into four groups – HH, HL, LH and LL – using their sensitivity on the two stimuli which we identified (above) that the returns from targeting are likely to be the highest - LVIEWNUM and SITENUM. In other words, a HL customer is one who is a "high sensitive" on the amount of exposed advertising but a "low sensitive" on the number of sites on which the exposure occurs. We then choose a week sufficiently far out from the last purchase occasion, Week 9, to contrast the effect of targeted and un-targeted advertising on customer retention and hence profitability. We first assume that the firm can expose customers (who do not purchase in Week 9) to one, two or three banner advertisements. These ads can be distributed on one, two or three sites. In the un-targeted strategy, all these customers get exposed to the identical number of banner ads on the same number of sites. In the targeted strategy, we choose different levels of exposure and sites depending upon the classification of the customer. We then compute the new probability of purchase (in each condition) and take the product of that probability with the average historical dollar expenditure by that customer – the expected revenue - on the website. We then sum across all the chosen customers to obtain the total revenue in each condition. On the basis of industry feedback, we assume that every additional exposure costs the firm $0.05 and that there is a $0.02 charge for every additional website that the advertising is placed on. We computed the return for each strategy as [(Total Revenue – Total Cost)/(Total Cost)] of each strategy. The results are given in Table 7. Note that the absolute values reported in the table may or may not be meaningful. For the purpose of our stylized experiment, it is only important to focus on the relative values.

Insert Table 7 about here

There are three interesting findings from the table. First, the return is always greater for the targeted advertising strategy. This is in spite of the fact that we use a simple targeting rule such as a median split and then assign the number of exposures and number of websites in a fairly simple manner. Second, the general magnitude of the return gets smaller as more exposures are provided – the maximum return is 108.30, 141.21 and 158.74 for three, two and one exposure. This is likely due to the diminishing return nature of response to exposure. Finally, the return gets greater as the options for targeting get larger. To clarify, the incremental return for an additional three, two and one exposures are 19% (108.30/91.21), 14% (141.21/123.42) and 5% (158.74/151.78) respectively. With only one additional exposure, firms are limited in how they can target, e.g., they can decide which customers to expose to and on which site. In contrast, with three exposures, finer targeting is possible, leading to higher (relative) returns. In conclusion, even with a simple targeting strategy, the firm can reap significant benefits. Thus, this experiment demonstrates the value of obtaining individual level parameters to create profitable targeting strategies.

Discussion

Our findings above have several implications for managers. First, we do find unique evidence that banner advertising plays a role in the commitment stage of CRM. This is a unique and different finding that is different from the common belief in the industry that banner advertising only works in the awareness and exploration stages of CRM (i.e., acts as a customer acquisition tool). Specifically, we find that banner advertising increases the purchase probabilities for current customers. Second, the elasticity estimates are close to those documented for conventional advertising, suggesting that managers should expect to see effect sizes that are consistent with other forms of advertising. To the best of our knowledge, this is the first documentation of these effect sizes. Third, as the temporal gap between exposure and purchase can be as high as seven days, the increase in purchase probability is likely to be past the point of exposure. This implies that managers may be focusing on the wrong metric when they use instantaneous metrics such as click-through to measure advertising effectiveness. Fourth, our results have implications for the design and execution of banner ad campaigns. Broadly speaking, campaigns should be designed such that customers are exposed to fewer (and more consistent) creatives across many pages and websites. Fifth, given a fixed number of exposures and creatives, returns to exposure are somewhat higher for sites first and then pages. Finally, while the mean response to the number of exposures is somewhat lower, the returns to targeting on this measure are likely to be the highest. A stylized experiment shows that the returns to targeting are higher than not targeting and that these returns get relatively higher as the targeting options (in the number of exposures and the number of websites on which consumers may be exposed) get larger.

Conclusion

Our research fits into a small, but fast growing subfield of empirical research dedicated to measuring how the Internet provides new marketing opportunities in areas such as pricing, product assortment decisions and advertising. This paper is the first attempt, to the best of our knowledge, to model the effects of banner advertising on the customer retention stage of CRM. The retention stage is a crucial stage – it has a direct impact on the profitability and consequent survival of web merchants (that market frequently-purchased products). We use a unique dataset to investigate the effects of banner advertising on the weekly purchase probability of existing customers. Our main finding is that, contrary to popular belief, banner advertising can act as a customer retention tool. This is because our modeling approach allows for temporal separation between advertising and purchase behavior. We speculate that the temporal separation exists because advertising acts as a brand building tool and/or a reminder. The corollary to this finding is that measures of instantaneous behavior such as click-through may be a poor measure of advertising effectiveness. We find that both the weight and quality of advertising have an effect on customers' purchase probabilities. An interesting finding is that the more the creatives a customer is exposed to in a given week, the lower is the purchase probability. Our explanation for this is that, given a relatively simple medium like banner advertising and the amount of competing information on web page, different messages dilute the impact of advertising. Our findings also show that exposure to banner advertising on more (different) websites and web pages has a slightly larger effect on the individual purchase probabilities than the weight and quality of advertising. This may be because these two covariates contain information about both advertising exposure and individual differences in browsing behavior.

We also find evidence of considerable heterogeneity across consumers in response to advertising. In terms, the heterogeneity is highest for the ad weight response followed by the number of sites response. Thus, developing targeted communication is likely to pay the highest dividend on these two dimensions. We illustrate the managerial benefits of our approach by carrying out a stylized experiment that shows that the revenue impact of customer retention via targeted banner advertising is higher than an untargeted approach. Finally, in terms of the broader area of research on the effects of (any type of) advertising, we provide somewhat unique evidence that advertising does affect the purchase behavior of current customers.

We would also like to note some limitations of our research. These limitations arise primarily from the lack of information in our data. First, we note that our results may not apply to customers who have not purchased items at least once at this website. Second, we do not have any demographic information and other relevant behavioral metrics (such as Internet usage) on the cookies. This information may have been useful in explaining a larger part of the unobserved heterogeneity. Third, our results would have been richer if we had information on the actual message contained in each advertisement and the identity of the referral sites. Fourth, we do not have any knowledge of the other marketing variables such as price and promotion during consumers' purchase visits. Finally, our targeting exercise would be more relevant if we had data on the profit per customer and not just revenue per customer. These limitations may be addressed in future research by running formal field experiments (as in Lodish et al. 1995) or by obtaining richer datasets that provide natural variation on these dimensions.

References

Allenby, G. M., R. P. Leone and L. Jen (1999), "A dynamic model of purchase timing with application to direct marketing," *Journal of the American Statistical Association*, **94**, 365-374.

Ben-Akiva, M. and S. Lerman (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, MA.

Bhatnagar, A. and S. Ghose (2003), "An analysis of frequency and duration of search on the Internet," *Journal of Business,* forthcoming **77**(2).

Briggs, R. (2001), "Abolish Clickthrough now!," from < http://www.intelliquest.com/resources/whitepapers/abolishclick.pdf >.

Bucklin, R. E. and S. Gupta (1992), "Brand choice, purchase incidence and segmentation: An integrated modeling approach," *Journal of Marketing Research*, **34**(5), 201-215.

Bucklin, R. E. and J. M. Lattin (1991), "A two-stage model of purchase incidence and brand choice," *Marketing Science*, **10**(4), 24-39.

Bucklin, R. E. and C. Sismeiro (2003), "A model of web site browsing behavior estimated on clickstream data," *Journal of Marketing Research*, **40**(3), 249-267.

BusinessWeek On-line (2001a), "A chat with a true believer," *Special Report: Internet Advertising*, July 12, 2001.

BusinessWeek On-line (2001b), "On-line advertising: It's just the beginning," *Special Report: Internet Advertising*, July 12, 2001.

BusinessWeek (2003), "Online ads take off – again: Mainstream marketers are hiking their Internet ad spending," May 5, 2003, 75.

Cacioppo, J. T. and R. E. Petty (1985), "Central and peripheral routes to persuasion: The role of message repetition," in *Psychological Processes and Advertising Effects*, A. Mitchell and L. A. Alwitt, eds., Erlbaum, Hilsdale, NJ.

Chatterjee, P., D. L. Hoffman, and T. P. Novak (2002), "Modeling the clickstream: Implications for web-based advertising efforts," *Marketing Science*, forthcoming.

Cho, C., J. Lee, and M. Tharp (2001), "Different forced-exposure levels to banner advertisements," *Journal of Advertising Research*, **41**(4), 45-56.

Dahlen, M. (2001), "Banner advertisements through a new lens," *Journal of Advertising Research*, **41**(4), 23-30.

Deighton, J., C. M. Henderson and S. A. Neslin (1994), "The effects of advertising on brand switching and repeat purchasing," *Journal of Marketing Research*, **31**(2), 28-43.

DoubleClick Press Release (2001), "Diameter's campaign evaluation illustrates direct correlation between exposure to on-line ads and brand impact," July 30, 2001.

Gallagher, K., K. D. Foster, and J. Parsons (2001), "The medium is not the message: Advertising effectiveness and content evaluation in print and on the Web," *Journal of Advertising Research*, **41**(4), 57-70.

Hoffman, D. and T. Novak (2000), "When exposure-based advertising stops making sense (and what CDNOW did about it)," *Working Paper*, Owen Graduate School of Management, Vanderbilt University.

IAB (1997), IAB On-line Advertising Effectiveness Study, from < http://www.intelliquest.com /resources/reports/mbi_report02.asp >.

IAB (1999), IAB Internet Advertising Revenue Report, 1999 Second Quarter Results, from < http://www.iab.netreports/PwC_IAB1999Q2.PDF >.

Lee, J., P. Boatwright and W. A. Kamakura (2003), "A Bayesian model for prelaunch sales forecasting on recorded music," *Management Science*, **49**(2), 179-196.

Lilien, G. L., P. Kotler and K. S. Moorthy (1992), *Marketing Models*, Prentice Hall, NJ.

Lodish, L. M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, M. E. Stevens (1995), "How TV-advertising works - A meta-analysis of 389 real-world split cable TV-advertising experiments," *Journal of Marketing Research*, **32**(2), 125-139.

Mela, C. F., S. Gupta, and K. Jedidi (1998), "Assessing long term promotional influences on market structure," *International Journal of Research in Marketing*, **15**(2), 89-107.

Meyer, B. (1990), "Unemployment Insurance and Unemployment Spells," *Econometrica*, **58**(4), 757-782.

Moe, W. W. and P. S. Fader (2003), "Dynamic conversion behavior at e-commerce sites," *Management Science*, forthcoming.

Moe, W. W., H. Chipman, E. I. George, and R. E. McCulloch (2002), "A Bayesian Treed model of online purchasing behavior using in-store navigational clickstream,", *Working Paper*, McCombs School of Business, University of Texas at Austin.

Mohammed, R. A., R. J. Fisher, B. J. Jaworski and A. M. Cahill (2002*), Internet Marketing: Building Advantage in a Networked Economy*, McGraw-Hill/Irwin, New York, NY.

Pechmann, C. and D. W. Stewart (1988), "Advertising repetition: A critical review of wearin and wearout", in *Current Issues and Research in Advertising*, J. H. Leigh and C. R. Martin, Jr., eds., **11**(1,2), Michigan Business School, University of Michigan.

Pedrick, J. H. and F. Zufryden (1991), "Evaluating the impact of advertising media plans: A model of consumer purchase dynamics using single-source data," *Marketing Science*, **10**(2), 111-130.

Rao, U. H. and R. E. Burnkrant (1991), "Effects of repeating varied ad executions on brand name memory," *Journal Of Marketing Research,* **28**(4), 406-417.

Roberts, G.O., A. Gelman and W.R. Gilks (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms," *Annals of Applied Probability*, **7**, 110-120.

Rossi, P. E. and G. M. Allenby (2003), "Bayesian statistics and marketing," *Marketing Science*, forthcoming.

Sethuraman, R. and G. Tellis (1991), "An analysis of the tradeoff between advertising and price discounting," *Journal of Marketing Research*, **28**(5), 160-174.

Sherman, L. and J. Deighton (2001), "Banner advertising: Measuring effectiveness and optimizing placement," *Journal of Interactive Marketing*, **15**(2), 60-64.

Silk, A. J. , L. R. Klein, and E. R. Bernt (2001), "The emerging position of the Internet as an advertising medium", *Netnomics*, **3**, 129-148.

Sismeiro, C. and R. E. Bucklin (2002), "Modeling purchase behavior at an e-commerce website: A conditional probability approach," *Working Paper*, Marshall School of Business, University of Southern California.

Smith, R. E. and W. R. Swinyard (1982), "Information response models: An integrated approach," *Journal of Marketing*, **46**(1), 81-93.

Song, Y. (2001), "Proof that on-line advertising works," Atlas Institute, from < http://www. atlasdmt.com/insights/dm.asp >.

Tellis, G. J. (1988), "Advertising exposure, loyalty and brand purchase: A two-stage model of choice," *Journal of Marketing Research*, **25**(2), 134-144.

Tran, K. T. L. (2001), "A dictionary of on-line-marketing jargon," *Wall Street Journal*, April 23, 2001.

Vakratsas, D. and T. Ambler (1999), "How advertising works: What do we really know," *Journal of Marketing,* **63**(1), 26-43.

Mangalindan, Mylene (2003), "Web ads on the rebound after multi-year slump," *Wall Street Journal*, Aug 25, 2003.

Warren, S. (2001), "Winners and Losers," *Wall Street Journal*, April 23, 2001.

Wedel, M., W. A. Kamakura, W. S. DeSarbo and F. T. Hofstede (1995), "Implications for asymmetry, nonproportionality and heterogeneity in brand switching from piece-wise exponential mixture hazard models," *Journal of Marketing Research*, **32**(4), 457-462.

Winer, R. S. (2001), "A framework for customer relationship management," *California Management Review*, **43**(4), 89-105.

Zielske, H. and W. Henry (1980), "Remembering and forgetting television advertisements," *Journal of Advertising Research*, **20**(2), 7-13.

Zufryden, F. (1987), "A model for relating advertising media exposures to purchase incidence behavior patterns," *Management Science*, **33**(10), 1253-1266.

Table 1: Overview of Research on On-line Advertising and Consumer Behavior

Study	***Research Issue***	***Product Category***	***Type of Data***	***Dependent Variables***
Our study	Evaluate impact of banner advertising on consumer on-line purchase behavior	Healthcare and beauty products Non-prescription drugs	*Market data:* - Advertising exposure - Actual purchase transaction	Purchase probability
Dahlen (2001)	Study impact of brand familiarity and Internet user experience on banner ad effectiveness	Insurance product Travel services Detergent Ice cream Coffee Automobile parts	*Experimental data:* - Laboratory experimentation	Brand awareness Brand attitude Click-through rate
Cho, Lee and Tharp (2001)	Evaluate effects of different levels of forced exposure to banner ads on consumer response	Consumer brands Retail, financial, travel services	*Experimental data:* - Laboratory experimentation	Brand awareness Attitude towards ad and brand Click-through rate Purchase intention
Gallagher, Foster and Parsons (2001)	Evaluate effectiveness of web-based ads, compared to print-based ads	Tourism service Coffee Arts and crafts	*Experimental data:* - Laboratory experimentation	Brand recall and recognition Attitude towards ad and brand
Sherman and Deighton (2001)	Evaluate efficacy of target on-line advertising	Healthcare products	*Market data:* - Cookie data - Advertising site response	Web browsing frequency Purchase incidence Conversion rate
Bhatnagar and Ghose (2003)	Study relationship between on-line search patterns and type of information sought	Information	*Survey data:* - On-line survey questionnaire	Search duration Search frequency Type of information
Chatterjee, Hoffman and Novak (2002)	Model consumer click-through on banner ads	High-technology, durable goods	*Browsing data:* - Click-stream logs	Click-through response
Bucklin and Sismeiro (2003)	Model web site "stickiness" and consumer learning behavior	Automobiles	*Browsing data:* - Click-stream logs	Web page choice Visit duration
Moe and Fader (2003)	Model consumer rates of visit-to-purchase conversion	Books	*Browsing data:* - Click-stream logs	Purchase incidence (proxied) Conversion propensity
Moe, Chipman, George, and McCulloch (2002)	Model on-line purchasing behavior from clickstream data	Nutrition products	*Browsing data:* - Click-stream logs	Purchase incidence
Sismeiro and Bucklin (2002)	Model on-line purchasing behavior from clickstream data	Automobiles	*Browsing data:* - Click-stream logs	Purchase incidence

Table 2: Descriptive Statistics

Variable	*Mean*	*Std. Dev.*
LVIEWNUM	0.25	0.66
ADNUM	0.23	0.64
SITENUM	0.07	0.30
PAGENUM	0.09	0.33

Table 3: Discrete Time Hazard Effects

Variable	*Mean*	*Posterior Std. Dev.*
ψ_1	-3.57	0.057
ψ_2	-2.80	0.036
ψ_3	-2.54	0.029
ψ_4	-2.74	0.038
ψ_5	-2.48	0.035
ψ_6	-2.49	0.039
ψ_7	-1.78	0.029
ψ_8	-1.97	0.038
ψ_9	-1.98	0.038
ψ_{10}	-0.96	0.024
ψ_{11}	-1.30	0.034
ψ_{12}	-0.89	0.031
ψ_{13}	-0.31	0.032

Table 4: Expected and Estimated Effects (Covariates)

Variable	*Expected Sign*	*Mean*	*Posterior Std. Dev.*
LVIEWNUM	+	0.10	0.009
ADNUM	?	-0.25	0.017
SITENUM	+	1.55	0.065
PAGENUM	+	0.84	0.049

Table 5a: Covariance Matrix

	LVIEWNUM	ADNUM	SITENUM	PAGENUM
LVIEWNUM	0.113	-0.004	-0.238	-0.014
	(0.033)	(0.010)	(0.043)	(0.021)
ADNUM		0.103	-0.177	-0.016
		(0.011)	(0.038)	(0.048)
SITENUM			1.861	0.001
			(0.531)	(0.180)
PAGENUM				0.153
				(0.032)

Table 5b: Correlation Matrix

	LVIEWNUM	ADNUM	SITENUM	PAGENUM
LVIEWNUM	1.00	-0.03	**-0.54**	-0.09
ADNUM		1.00	**-0.41**	-0.15
SITENUM			1.00	-0.06
PAGENUM				1.00

Table 6: Elasticities

Variable	***Elasticity (Mean)***	***Elasticity (Std. Dev.)***
LVIEWNUM	0.02	0.005
ADNUM	-0.03	0.009
SITENUM	0.05	0.020
PAGENUM	0.04	0.015

Table 7: Returns to Targeting

Group/Stimulus	*Optimal Exposure/ Site*	*Revenue* ($)	*Cost* ($)	*Return* [a]
Additional Exposures = 3				
Un-targeted [b]	3/2	71989.12	780.71	91.21
Targeted [c]		67375.88	616.40	108.30
HH group	3/3			
HL group	3/1			
LH group	2/2			
LL group	0/0			
Additional Exposures = 2				
Un-targeted [b]	2/2	71573.79	575.26	123.42
Targeted [c]		54073.95	380.24	141.21
HH group	2/2			
HL group	2/2			
LH group	1/1			
LL group	0/0			
Additional Exposures = 1				
Un-targeted [b]	1/1	43945.24	287.63	151.78
Targeted [c]		42086.99	263.48	158.74
HH group	1/1			
HL group	1/1			
LH group	1/1			
LL group	0/0			

Notes:
(a) Return is computed as (Total Revenue - Total Cost) / (Total Cost).
(b) For the un-targeted scenario, there could be different strategies, e.g., 3 exposures on 3 sites versus 2 sites. We simulated all possible scenarios and picked the one with the highest return (for 3 exposures, it was 2 sites and for 2 exposures, it was 2 sites as well).
(c) This represents the banner ad placement for each group, e.g., 3/2 represents 3 exposures on 2 sites.

Figure 1: Duration Dependence