A Two-Stage Model of the Promotional Performance of Pure Online Firms

Jianan Wu, Victor J. Cook, Jr., Edward C. Strong
A. B. Freeman School of Business, Tulane University, New Orleans, Louisiana 70118
{jianan.wu@tulane.edu, victor.cook@tulane.edu, bigfoot@tulane.edu}

Internet firms frequently employ a two-stage approach to promotional activities. In Stage 1, they attract customers to their websites through advertising. In Stage 2, firms generate sales transactions or sales leads through their website.

Comprehensive assessment of the promotional performance of pure online firms requires the study of Stage 1 and of Stage 2 jointly. In this paper we develop a joint two-stage conceptual and econometric model for assessing website promotion on three important dimensions: (1) how advertising response can be measured by linking media schedules to website log files; (2) how advertising and website characteristics jointly affect the desired system outcome of the promotion; and (3) whether the joint investigation of advertising response and desired system outcomes is essential to assess the results of website promotion.

Three general findings follow from application of our model to a pure online firm’s campaign to generate sales leads through print advertising. First, advertising and website characteristics affect sales leads in different ways. A characteristic may influence sales leads directly, or indirectly, or both. Second, assessing advertising effectiveness in an online environment may not require costly survey research data. Instead, secondary data available from website log files may be used for such assessment. Third, the interaction between the first and second stages of our two-stage model can lead to misspecifications that produce misleading inferences. This occurs because the unobserved characteristics in generating website visits and sales leads may be correlated.

**Key words:** e-commerce; website traffic; sales leads; two-stage model; sample selection

**History:** V. Sambamurthy, Senior Editor; Nirup Menon, Associate Editor. This paper was received on December 12, 2002, and was with the authors 23.75 months for 3 revisions.

1. Introduction

Firms may be motivated to establish an online presence to achieve many objectives, including image building, generating sales leads, and transactions. These ends cannot be fulfilled unless customers know of and visit the firm’s website. Pure online firms and lesser-known offline firms must promote their websites to their target audience in various advertising vehicles in online and/or offline media. For an “Internet company that does not have a ‘brick and mortar’ counterpart, the ability of the website to ‘promote’ the company is likely to be a crucial factor in the company’s success” (Agarwal and Venkatesh 2002, p. 182). According to *Advertising Age* (adage.com), dot-coms spent $7.4 billion on advertising in 1999.

Online firms follow essentially the same two-stage model of advertising. In the first stage, firms attract customers to their websites through advertising either in traditional media such as print, radio, and television, or via online media such as banners, search engines, e-mails, and affiliate programs (Hoffman and Novak 2000). In the second stage, when potential customers visit their website, firms provide them with carefully selected content and friendly navigation to create an environment conducive to generating the desired information system outcomes such as further inquiry, purchase, Web satisfaction, information quality, and system quality (Straub et al. 2002, p. 230).

These pure online firms share the same goal at Stage 1, but expect different system outcomes at Stage 2. In general, for firms in B-2-C or C-2-C markets with low customer involvement, the desired system outcome is a sales transaction. For other firms, the desired system outcome may be sales leads or some general expression of interest (Halliday 2001). In either situation, desirable system outcomes go beyond
merely inducing website traffic like the number of visits or number of hits.

Measuring and understanding the behavior of website visitors and its linkage to the desired system outcome has attracted the attention of many scholars (e.g., the Special Issue on Measuring e-Commerce in Net-Enabled Organizations of *Information Systems Research*). The importance of research on “effective information processing techniques for analyzing customers’ electronic behavior patterns such as clickstream patterns, system outcomes such as further inquiry, purchases” (Straub et al. 2002, p. 230) has also gained recognition from the IS researchers (e.g., Straub et al. 2002). Indeed, published studies have contributed significantly to understanding customer online behavior, both in generating traffic (e.g., Hoffman and Novak 2000) and transactions within electronic stores (Hoque and Lohse 1999, Kraut et al. 2000, Telang et al. 2001, Agarwal and Venkatesh 2002, Chen and Hitt 2002, Wu and Rangaswamy 2003, etc.). However, these studies are inadequate to assess the promotional effect in the two-stage business model in online markets because the effect generated before the customers’ arrival at a firm’s website (i.e., in Stage 1) may confound the effect generated at the website (i.e., in Stage 2). A critical research question is: How do we disentangle the effects of the two stages so that we can comprehensively assess website promotion?

In this paper, we seek to answer that research question for pure online firms1 by studying the two stages jointly. We develop a joint two-stage conceptual and econometric model to address three important dimensions of promotional effectiveness: (1) how advertising characteristics (e.g., ad exposure frequency, ad placement, etc.) affect website visits; (2) how website characteristics (e.g., website stickiness, serial position, etc.) affect the desired system outcome of a promotional campaign; and (3) how advertising characteristics and website characteristics jointly affect desired system outcomes.

This paper contributes to the literature in three ways. First, we develop a conceptual framework for the two-stage promotion model at the micro level. Second, we develop a methodology to measure advertising response in an online environment that does not require costly survey research data. Instead, it utilizes available secondary data from website log files, ad characteristics, and website design. Third, a customized econometric model based on our conceptual framework is developed to jointly assess promotional performance. This solution extends current discrete bivariate models (Greene 1998, Boyes et al. 1989) in several important aspects. The empirical analysis shows that the correlation between advertising response and sales leads is explained not only by the observed characteristics, but also by unobserved advertising and website factors that often are ignored. This suggests that the misspecification of single-stage website promotion models and separate two-stage models is likely to yield misleading results.

The paper is organized as follows. In §2, we develop the conceptual framework, generate the related hypothesis, and discuss the issues in our study design for the two-stage promotional model. In §3, we develop an econometric model to test our hypothesis, based on our conceptual framework. The empirical analysis using data from one pure online firm (Thestyle.com) appears in §4. Finally, in §5 we discuss some managerial implications and directions for future research. The empirical analysis using the conceptual framework applied to a pure online firm’s campaign to generate sales leads through print advertising suggests important insights: Advertising and website characteristics affect sales leads in different ways. Some advertising characteristics, like single versus double issues of a magazine, influence sales leads indirectly via their effect on website visits. Some website characteristics, like average time per page, influence sales leads directly. However, other characteristics, like response latency, affect sales leads both indirectly and directly. These insights matter to management by showing where to put an ad on a magazine page in order to increase website traffic. For example, if at zero marginal cost website traffic can be increased just 2% as a result of placing the ad flush with the right-hand page of a magazine rather than on the left-hand page, the profit impact of this action is significant.

---

1 The question applies to firms with offline presence as well, but only online firms have access to the log file data needed to conduct such an investigation. See §5 for more discussion of this opportunity.
2. Conceptual Framework Development

A. Conceptual Framework and Hypothesis Development

The conceptual framework for the two-stage promotional model based on the consumers’ multistage decision-making processes is presented in Figure 1. We discuss the details of the consumers’ decision process and develop the associated hypothesis along their decision processes.

**Stage 1: Advertising Exposure to Website Visits.** An individual decides to visit the website if her expected utility after seeing an ad is beyond a certain threshold. A widely adopted theory of advertising response stems from utility theory in economics (e.g., Wu and Rangaswamy 2003, Moe and Fader 2004, etc.). This theory suggests that advertising response is determined by both an individual’s expected utility derived from the ad and her threshold utility (or reservation utility). Utility above the threshold leads to a favorable response (e.g., website visit). Holbrook and Lehmann (1980) and Naccarato and Neuendorf (1998) show that advertising response is multidimensional. It may take the form of brand recall, a cognition dimension; or brand preference, an attitude dimension; or brand choice, a behavioral dimension. In our conceptual framework, we specify the advertising response as the action of website visits (e.g., whether a website visit is generated or how many website visits are generated, etc.)—a behavioral dimension.

The visitor’s expected utility is affected by ad/media characteristics and dynamics. Broussard (2000) suggests that higher frequency of ad exposure leads to stronger advertising responses. Consider the case of a double issue of a magazine. A double issue remains on sale twice as long as a single issue. Being seen on the newsstand twice as long provides an opportunity for readers to experience a memory trigger from first exposure that sends them to a website. A double issue also remains in the reader’s inventory twice as long. Remaining in the reader’s inventory for twice as long increases the chance a visitor will be sent back to the website for more information. The frequency of ads increases brand recall, and thus the expected number of website visits. Therefore, we have:

**Hypothesis 1A (H1A).** Frequency of ad exposure has a positive impact on website visits.

**Figure 1** A Dynamic Two-Stage Promotion Model of Pure Online Firms

<table>
<thead>
<tr>
<th>Stage 1 characteristics (X)</th>
<th>Stage 2 characteristics (Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Advertising vehicle</td>
<td>• Stage 1 response (N)</td>
</tr>
<tr>
<td>• Advertising exposure</td>
<td>• Product images and copy</td>
</tr>
<tr>
<td>• Ad content and frame</td>
<td>• Website layout and design</td>
</tr>
<tr>
<td>• Advertising dynamics</td>
<td>• Interactive tools</td>
</tr>
<tr>
<td>• Online capabilities</td>
<td>• Security controls</td>
</tr>
<tr>
<td>• Information desirability</td>
<td>• Browsing functions</td>
</tr>
<tr>
<td>• Audience</td>
<td>• Website dynamics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Responses of Stage 1 (N)</th>
<th>Responses of Stage 2 (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Number of visits</td>
<td>• Close browser/go to other websites</td>
</tr>
<tr>
<td></td>
<td>• Submit form/send e-mail</td>
</tr>
<tr>
<td></td>
<td>• Make a purchase</td>
</tr>
</tbody>
</table>

- Repeat visits
The information processing of an ad updates the visitor’s expected utility (Edell and Staelin 1983). The most salient attributes in ad processing are content and frame characteristics. Frame characteristics pertain to the medium itself and attract the attention of consumers. Content characteristics relate to the subject matter and presentation. Although both frame and content characteristics have an impact on the effectiveness of advertising (Naccarato and Neuendorf 1998), it has been consistently demonstrated that frame characteristics explain greater variance in print advertising effectiveness. Advertising studies (Holbrook and Lehmann 1980, Naccarato and Neuendorf 1998) show that content variables do not perform nearly as well as frame variables in predicting the effectiveness of advertising.

Hypothesis 1b (H1b). Attractive placement of the ad has a positive impact on website visits.

In particular, the marketing literature suggests that ads with more attractive placement in a medium will be more effective. For example, in print advertising, ads placed on the right-hand page create greater brand recall than those placed on the left-hand page of a magazine (e.g., Diamond 1968, Naccarato and Neuendorf 1998). Keller (1991) also shows that competitive advertising produces interference effect and significantly reduces recall of brand claims.

Hypothesis 1c (H1c). Competitive advertising has a negative impact on website visits.

If a reader does not visit in the current period, the advertising effect may be carried over in the reformulation of expected utilities in a way that triggers visits in the next period, even though the effect wears out over time. Advertising carryover effects have been extensively studied in marketing literature and are often captured in lagged effects (e.g., Clarke 1976, Blattberg and Jeuland 1981). As such, we have:

Hypothesis 1d (H1d). Advertising has a diminishing carryover effect on website visits.

An individual’s threshold is affected by information desirability and search cost. An individual’s threshold is her reservation utility for visiting a website and may be affected by multiple factors. Higher information desirability and/or lower search costs may imply a lower threshold for a website visit. For example, individuals with a low-speed Internet connection incur higher search costs and hence have a higher threshold for visiting a website.2 This situation is similar to Palmer’s (2002) “download delay” for a Web server. Ad response latency (i.e., advertising response time), as a surrogate for the effects of both advertising characteristics and ad response threshold, has been studied in marketing literature (Tybje 1979). Haaijer et al. (2000) found that incorporating response latency significantly improved the prediction of choice behavior in conjoint analysis. Response latency has also been widely studied in the psychology literature. Shorter response latency to a stimulus often suggests stronger association and higher accessibility to the stimulus (Fazio 1990). Attitudes with higher accessibility are more likely to guide behavior than attitudes with lower accessibility (Fazio 1986). Therefore, we have:

Hypothesis 1e (H1e). The visitor’s response latency has a negative impact on website visits.

Stage 2: Website Visits to Desired System Outcomes. A desired system outcome is obtained from the visitor if her modified utility is beyond a certain threshold. The Stage 2 model again follows utility theory, as described in the Stage 1 model. If the visitor is pleased with what she has experienced on a website, she may produce a desired system outcome such as adding the product in her consideration set by requesting more information or making a purchase (Straub 2002). The nature of a promotional campaign depends on the business model (B-2-B, B-2-C, and C-2-C) and the degree of customer purchase involvement (high versus low). Classic marketing theory (e.g., Kotler 1991) suggests that consumers in B-2-C markets with low customer purchase involvement often exhibit habitual buying behavior in purchase decisions. A promotional campaign in such markets often targets sales transactions. For example, online stock-trading firms like etrade.com, ameritrade.com, and datek.com advertise aggressively to attract visitors to their websites. The firms expect that customers will sign up for their services and pay a transaction fee for each trade the customers execute. On the contrary, the purchase decisions in B-2-B markets often

2 We thank one of the reviewers for bringing online capability to our attention.
involve multiple decision makers/stages and exhibit complex buying behavior. A promotional campaign in such high customer purchase involvement markets primarily aims to generate sales contacts/leads rather than sales transactions per se. For example, the job-finding firm monster.com, which advertised during both the 2001 and 2004 NFL Super Bowls, expects that their information system generates matches, not the eventual transactions, made between job seekers and employers. Our application concerns various forms of informational inquiries generated by the website (see §4).

The visitor’s expected utility is modified by website characteristics and dynamics. The factors that influence the degree to which the visitor’s utility is modified include (1) the Stage 1 response and (2) website characteristics. The linkage between advertising response and sales or sales leads has been well established in the marketing literature. Little (1979) found that sales move dynamically upward when advertising increases and downward when advertising decreases. We have:

**Hypothesis 2a (H2a).** The Stage 1 advertising response has a positive impact on desired system outcomes.

In e-commerce literature, Moe and Fader (2004) found that cumulative websites visits are good predictors of purchases. In particular, Bucklin and Sismeiro (2003) found that website stickiness predicts purchase behavior well.

**Hypothesis 2b (H2b).** A visitor’s website stickiness has a positive impact on desired system outcomes.

The website characteristics include: brand names and images (Degeratu et al. 2000); website layout and design characteristics like size, display format, serial position, ease of use/navigation (Devaraj et al. 2002, Koufaris 2002); the availability of interactive tools such as sorting, searching, and personal lists (Wu and Rangaswamy 2003); and website security variables such as privacy, trust, and reputation (Urban et al. 2000). Hoque and Lohse (1999) found that the serial position of information at a webpage attracts more attention. As such, we have:

**Hypothesis 2c (H2c).** The serial position on a webpage has a positive impact on desired system outcomes.

The **Joint of Stage 1 and Stage 2.** Indirect and direct effects on desired system outcomes. As we discussed above, the impact of website characteristics on desired system outcomes is direct, while that of advertising characteristics is indirect (via correlations between two stages). However, response latency, as a proxy for the impact from both advertising characteristics and consumers’ threshold response in Stage 1, may have not only an indirect impact on desired system outcome as stated by H1e (because of its association to advertising characteristics), but also a direct impact parallel to that stated for advertising response strength by H2a (because of its association to Stage 2 threshold and the possible correlation between the thresholds of the two stages). As such, we have:

**Hypothesis 3 (H3).** Response latency has both direct and indirect negative impact on desired system outcomes.

It is worth pointing out that although some of the hypotheses regarding either website visits only (H1a–H1e) or desired system outcomes only (H2a–H2c) may have been examined separately before in the marketing and information systems literature via single-stage models, they have never been tested in a joint two-stage promotional model. We next discuss our research design to empirically test these hypotheses.

**B. Field Study Design**

Thestyle.com is an online retailer providing an original collection of designs built on the core concepts of de Stijl (Friedman 1982). The products featured on the website during the ad campaign quite literally were “not for sale.” They are museum-quality works of art as well as functional furniture. The owner’s advertising objective was to generate sales leads on the website³ that might yield negotiated contracts for made-to-measure products sold on special commission. The products were handmade from rare hardwoods, delivered with a portfolio of construction photos for each step in the process, and assembled

³ In the context of generating sales leads, repeat visits to a website are not equivalent to repeat purchases. These are fundamentally different constructs because it is possible for a visitor not to make a purchase on the site. Repeat visits should be thought of as continuing information search rather than product trial. We thank one of the reviewers for bringing this to our attention.
on the collector’s site. Contract negotiations typically took several months, and the lead time on delivery was about 16 months. The designer-collector relationship in this application is similar to B-2-B markets with a long buying cycle, where sales leads and conversion rates, rather than transactions, are the focus of marketing efforts (e.g., Seibel Systems and GE Medical Imaging).

Thestyle.com’s advertising campaign was designed by the authors and implemented as a field experiment. Two important decisions were involved. First, selecting the media vehicle for the campaign. The New Yorker magazine was chosen based on the consideration of the targeted segment of the market. Subscribers to The New Yorker magazine fit the customer profile better than any other print vehicle: They had high disposable incomes; a large percentage were empty-nest urban professionals who owned two or more residences, and had taken two or more trips to Europe in the past year. Importantly, they also owned one or more pieces of collector-quality modern art. This product involvement increased the likelihood that readers would recognize the important link between the designs featured in ads for Thestyle and the early 20th century Dutch design group known by the name of their magazine: de Stijl. More than any other magazine, readers of The New Yorker were predisposed to find the designs featured in the ads attractive. To identify sales leads for Thestyle.com from its upscale target segment, an advertising campaign was executed in The New Yorker magazine from March 13, 2000, to December 11, 2000, in five periods. The actual ads used in this campaign and the website referenced in the ads are provided in Figure 2. During the advertising campaign, Thestyle.com recorded its system log files and sales leads. The website was not advertised on any search engines in order to avoid confounding visitor response to the advertising campaign. We also made sure from the log files that none of the visits were directed from popular search engines.

Second, selecting an advertising schedule for the campaign. This decision involved four considerations:

1. the use of level versus pulsed insertions, 
2. the size of the individual print ads, 
3. the interinsertion time, and 
4. the control of content characteristics. 

The first of these issues is well researched. The general conclusion is that pulsing schedules are more effective than level schedules in situations where advertising budgets are constrained (Strong 1974). More recently, Bronnenberg (1998) also shows that pulsing policy is optimal if the demand follows a discrete and interpretable Markov process with concave transition probabilities in advertising, and the advertising budget is constrained. Given that the advertising budget would allow far less than one insertion in every issue of the magazine, a pulsing schedule was designed.

The second decision regarding the size of the ads was a bit more complex. Learning theory (Ebbinghaus 1913) suggests that the effectiveness of an exposure increases less than proportionately with the area of the stimulus. Specifically, Ebbinghaus proposed that the effectiveness of a stimulus increases as the cube root of its area. Given this notion, two exposures of an ad of a particular size are more effective than a single exposure of an advertisement twice that area. This logic does not support using an advertisement larger than the minimum size required by The New Yorker (approximately 2 1/2 inches square), so it was decided to use 10 exposures of the minimum-size ad. The

---

4 All of Thestyle’s products are constructed from uniform square sticks of high-density woods joined orthogonally by compression with threaded steel rods and brass nuts. As a result, every product may be constructed for use and deconstructed for shipping and storage. The structural components are guaranteed for life.

5 There is a lengthy discussion in the marketing literature regarding whether an ad campaign should be an even policy (constant spending over time) or pulsing (scheduling exposures unevenly over the same time period). Besides Strong (1974) and Bronnenberg (1998), there are other arguments that pulsing policy is optimal: First, Sasieni (1971) shows that pulsing (or more precisely chattering) is better than even policy if the response curve is S-shaped (see also Mahajan and Muller 1986). Marketing literature widely accepts the threshold argument in advertising literature. This threshold argument leads to increasing marginal return (convex) in the initial segment of the response before it eventually shows the diminishing marginal return (concave). Therefore, an S-shaped response curve is theoretically sound. Although there is no direct empirical evidence of S-shaped response curve, most researchers believe that “the efficient level of advertising generally lies on the concave part of the S-shaped response function” (Bronnenberg 1998). In practice, most firms employ a pulsing policy in their advertising decisions. Second, built on the classic goodwill accumulation advertising model of Nerlove and Arrow (1962), Naik et al. (1998) show that pulsing strategy generates greater total awareness than the continuous advertising when the effectiveness of advertisement varies over time.
third consideration was selecting the interval between exposures in the pulsing schedule. In a proprietary application of the scheduling model developed in Strong (1974), the underlying principle that makes pulsing or ad flight more effective was applied to time intervals less than a week. In that application, scheduling of two ads with very little temporal space between them (minutes) appeared to be even more effective than scheduling them with a week between ads. Given that the cost of the ad campaign remains constant, this suggests it would be more effective to schedule two ads within the same issue of a vehicle than to schedule them in consecutive issues. This led to purchasing a schedule that called for the appearance of ads in precisely the same location on two consecutive right-hand or left-hand pages in each of five editions of The New Yorker. The two pages were held relatively constant in terms of position in all five editions of the magazine. The fourth consideration was to control the potential interaction between content and frame characteristics. To control for this interaction, our advertising campaign held constant content variables like layout, appeal, brand name, and copy blocks.

3. Econometric Model Development

A. Model Specification

The joint nature of our proposed two-stage promotion framework for pure online firms makes the existing single-stage methodology inappropriate for testing our hypotheses simultaneously. Therefore, we develop a customized econometric model (ORD) to test our hypothesis based on our conceptual framework.
In Stage 1, let $N_{ij}^*$ be visitor $i$'s expected utility. $N_{ij}^*$ is unobservable. We only observe $N_{ij}$, the number of website visits an ad generated for a visitor $i$ in a given advertising period $j$. We specify $N_{ij}$ as the measure of the websites visits in our conceptual framework. This specification is similar to the approach used by Hanssens and Weitz (1980). They specified advertising response as the number of telephone inquiries generated. This specification also matches the conceptualization of continuous awareness proposed in the classic advertising model by Nerlove and Arrow (1962) and its extensions (Naik et al. 1998). Following Gupta (1988), we posit that $N_{ij}$ (count data) is generated in the manner of an ordered probit model. Therefore, $N_{ij} = n$ if and only if $\theta_{n-1} \leq N_{ij}^* \leq \theta_n$ for some thresholds $\theta_n$ ($n = 0, 1, 2, \ldots$) with $\theta_0 = -\infty$. We further posit that the latent utility $N_{ij}^*$ is a linear function of some observed Stage 1 drivers $X_{ij}$ in period $j$ and some unobserved characteristics captured by a random error $\xi_{ij}$, i.e.,

$$N_{ij}^* = X_{ij}\alpha + \xi_{ij}. \quad (1)$$

We elected to use the additive linear specifications in (1) to model a visitor’s expected utility for two reasons. First, linear specifications provide a good approximation of the true formation of utilities in both stages. There is ample evidence in the econometrics and marketing literature to support the linear assumption. Second, linear specifications are widely used in the literature and are simple to deal with in model estimation. The choice modeling literature in econometrics and marketing in the past two decades followed the linear specification in consumers’ utility formation (e.g., Wu and Rangaswamy 2003).

Let $L_{ij}^*$ be visitor $i$’s modified expected utility in Stage 2. Again, $L_{ij}^*$ is unobservable. We only observe $L_{ij}$, an index indicating whether visitor $i$ submits an RSVP form or sends e-mail ($L_{ij} = 1$) or not ($L_{ij} = 0$). We specify $L_{ij}$ as the measure of the desired system outcomes in our framework. We formulate Stage 2 as a binary probit model in which a desired system outcome is generated if the modified expected utility $L_{ij}^*$ is beyond a certain threshold (which is assumed to be 0 because it confounds the anchor point of $L_{ij}^*$), i.e., $L_{ij} = 1$ if and only if $L_{ij}^* > 0$. With the same consideration as in Stage 1, we posit that $L_{ij}^*$ is a linear function of some observed Stage 2 drivers $Z_{ij}$ in the period and some unobserved random errors as

$$L_{ij}^* = Z_{ij}\gamma + \eta_{ij}. \quad (2)$$

The binary specification of Stage 2 response is appropriate because the online retailer in our application was concerned only with whether the offline advertising and online website design jointly generated sales leads. However, our model can be easily modified to apply to other Stage 2 responses. For example, the desired system outcome of the website promotion could be the number of orders from a visitor in a period. In this case, the Stage 2 equation (2) can be specified as an ordered probit similar to the specification of Stage 1 equation (1). Alternatively, if the desired system outcome of the website promotion is sales revenues per customer in a period, then the Stage 2 equation (2) can be specified as a truncated normal distribution. In either case, the estimation can be carried out in a similar way with little difficulty.

So far, the Stage 1 drivers and Stage 2 drivers are specified at the period level, which is silent on the possible carryover effects for advertising response and website response across periods. For example, at Stage 1, potential customer $i$, who visited the website several times in period $j - 1$ may visit fewer times in period $j$. Or, in Stage 2, a visitor $i$ who did not generate a desired system outcome in period $j - 1$, but visited the website more times in period $j$, may be more likely to generate a desired system outcome in period $j$. To take these issues into consideration, we posit some general dynamic effects across periods. Specifically, we posit that (1) visitor $i$’s advertising response in period $j - 1$ ($N_{ij-1}$) has an impact on his or her advertising response ($N_{ij}$) in period $j$, and (2) both visitor $i$’s advertising response in period $j$ ($N_{ij}$) and the website response in period $j - 1$ ($L_{ij-1}$)
have an impact on website response in period \( j (L_{ij}) \). We use the conventional lagged-modeling approach (e.g., Clarke 1976) to formalize this idea:

\[
\begin{align*}
N^*_{ij} &= X_{ij} \alpha + \nu_N N_{ij-1} + \xi_{ij}, \\
L^*_{ij} &= Z_{ij} \gamma + \kappa_N N_{ij} + \kappa_L L_{ij-1} + \eta_{ij}.
\end{align*}
\]  

(3) (4)

Note that in Equations (3) and (4), the random errors \( \xi_{ij} \) and \( \eta_{ij} \) capture the unobserved components (e.g., missing explanatory variables) of the advertising response and desired system outcome for visitor \( i \) in period \( j \). If these unobserved components share some common attributes (e.g., the psychographics of visitor \( i \)), then \( \xi_{ij} \) and \( \eta_{ij} \) may be correlated. For this consideration, we then assume that \( \xi_{ij} \) and \( \eta_{ij} \) follow an i.i.d. bivariate normal distribution as

\[
\begin{pmatrix}
\xi_{ij} \\
\eta_{ij}
\end{pmatrix} \sim BIV(0, \Sigma), \quad \text{and} \quad \Sigma = \begin{pmatrix}
\sigma_N^2 & \rho \sigma_N \sigma_L \\
\rho \sigma_N \sigma_L & \sigma_L^2
\end{pmatrix}
\]  

(5)

is the variance-covariance matrix. Our model is completely specified by Equations (3)–(5).

### B. Estimation Specification

Note that Stage 1 equation (3) is an ordered probit model and the Stage 2 equation (4) is a probit model. We fix \( \sigma_L = 1 \) and \( \sigma_N = 1 \) for model parameter identification. Further, we must fix \( \theta_0 = 0 \) because \( \theta_0 \) is confounding the constant term in Equation (3). Also note that our model is a recursive simultaneous equation system (Greene 1998). Using the reduced form of the simultaneous equation system, it is easy to see that the structural parameter \( \kappa_N \) is also identified if the sets of independent variables of Equations (3) and (4) are not identical. This condition is indeed satisfied because \( N_{ij-1} \), which shows up in Equation (3), does not appear in Equation (4).

Our model has a censoring issue as well. If visitor \( i \) does not visit the website in period \( j \) (i.e., \( N_{ij} = 0 \)), then visitor \( i \) can never return a desired system outcome (i.e., \( L_{ij} = 0 \)). This censoring characteristic leads to sample selection bias in parameter estimation if it is not properly dealt with (Boyes et al. 1989). Therefore, we partition the sample log-likelihood into two parts:

\[
\begin{align*}
LL = & \sum_i \sum_j \ln f(N_{ij} = n, L_{ij} = l) \Big|_{n \geq 1} \\
+ & \sum_i \sum_j \ln f(N_{ij} = n, L_{ij} = l) \Big|_{n = 0}.
\end{align*}
\]  

(6)

The first term of (6) is calculated as

\[
\begin{align*}
\ln f(N_{ij} = n, L_{ij} = l) \bigg|_{n \geq 1} &= l \ln \Pr(N_{ij} = n, L_{ij} = l) + (1 - l) \ln \Pr(N_{ij} = n, L_{ij} = 0) \\
&= l \ln(\Pr(N_{ij} \leq n, L_{ij}^* \geq 0) - \Pr(N_{ij} \leq n - 1, L_{ij}^* \geq 0)) \\
&\quad + (1 - l) \ln(\Pr(N_{ij} \leq n, L_{ij}^* < 0) - \Pr(N_{ij} \leq n - 1, L_{ij}^* < 0)) \\
&= l \ln[(\Phi(\tau_{ijn}) - \Phi_2(\tau_{ijn}, v_{ij}; \rho)) \\
&\quad - (\Phi(\tau_{ijn-1}) - \Phi_2(\tau_{ijn-1}, v_{ij}; \rho))] \\
&\quad + (1 - l) \ln[\Phi_2(\tau_{ijn}, v_{ij}; \rho) - \Phi_2(\tau_{ijn-1}, v_{ij}; \rho)],
\end{align*}
\]  

(7)

where \( \tau_{ijn} = \theta_0 - (X_{ij} \alpha + \nu_N N_{ij-1}) \) and \( v_{ij} = -(Z_{ij} \gamma + \kappa_N N_{ij} + \kappa_L L_{ij-1}) \), and \( \Phi_2 \) is the CDF for the standard bivariate normal distribution. The second term of (6) is calculated as

\[
\begin{align*}
\ln f(N_{ij} = n, L_{ij} = l) \bigg|_{n = 0} &= \ln(1 - \Phi(X_{ij} \alpha + \nu_N N_{ij-1})).
\end{align*}
\]  

(8)

Plug (7) and (8) into (6), and we obtain the sample log-likelihood. We use the maximum likelihood estimate (MLE) to estimate our model, i.e., for given data \( (N_{ij}, L_{ij}, X_{ij}, Z_{ij}) \), we estimate parameters \( (\alpha, \gamma, \theta, \nu, \kappa, \Sigma) \) by maximizing sample log-likelihood (6). This joint MLE is consistent and efficient. The covariance matrix of the estimator is obtained using the Hessian of Newton-Raphson type of gradient methods in nonlinear optimization procedures. The estimates when two stages were separately estimated using random starting values were used as the starting values for our model.

Bivariate probit models have been studied before in the literature. For instance, Greene (1998) studied a cross-sectional simultaneous bivariate probit model without censoring, and Boyes et al. (1989) studied a cross-sectional censored bivariate probit model without simultaneity. It is important to recognize that our proposed model extends these bivariate probit models in several aspects. Our model (1) is both a censored and simultaneous equation model; (2) has an ordered probit model (rather than a binary probit) in Stage 1; and (3) is both cross-sectional and longitudinal.
C. Alternative Model Specifications
While it might be standard to model a dichotomous variable $L_{ij}$ (Stage 2 response) as a probit model, there are other alternatives for modeling the count variable $N_{ij}$ (Stage 1 response). We considered three Poisson distribution-based alternatives in which each has a different rationale. The first alternative (POS) models $N_{ij}$ as a Poisson process because it has been extensively used to model random count variables (e.g., Greene 2000). The second alternative, negative binomial distribution (NBD), is an extension of POS by taking into consideration that potential customers may be heterogeneous in their visit rates. This model, commonly referred to as the negative binomial, has also been extensively adopted in marketing and economics literature to study consumer heterogeneity. The third alternative (HUR) considers that the zero count in $N_{ij}$ is qualitatively different from the nonzero count. For example, a visitor may have failed to visit the website because they did not see the ad and not because they saw the ad but chose not to visit. As such, the zero count reflects the no-exposure effect, while the nonzero count reflects the exposure effect of the ad. These two effects are qualitatively different.8

4. Empirical Analysis
A. Measurements
To test the hypotheses, we specified the measurements for the constructs developed in our conceptual framework. The Stage 1 specifications include: $SHELF_{ij}$, $RIGHT_{ij}$, $OTHCL_{ij}$, $N_{ij}^L$, $FSTAV_{ij}$, $PTIME$.9 The Stage 2 specifications include: $NCURR_{ij}$, $STIME_{ij}$.10

8 See the appendix for the development of these competing models. To have faster convergence, we estimated each stage of ORD separately and used the estimates as the starting values for all four competing models.
9 We thank one of the reviewers for this suggestion. The period time length is part of the specification of the three benchmarking models as well.
10 The Web-log files document only the timing of clicks, not what a customer actually did between clicks. We constructed a proxy for the true time spent. However, this proxy is a reasonable approach in our model specification. Under certain regularity conditions, some discrepancy between the proxy and the true time may not matter. Suppose $T$ is the true time spent and $P$ is the proxy variable. Let $D = P - T$ be the discrepancy. We have $P = T + D$. Assume that $D \sim N(a, b)$ is normally distributed and independent of the unobserved Stage 2 error. Replace $P$ by $T + D$ in Equation (4). We now have a new Equation (4) in which (1) the parameter $a$ will be absorbed into the new constant term, (2) the parameter $b$ will be absorbed into the variance of the new error term, and (3) the coefficient of $P$ in the old Equation (4) will be identical to that of $T$ in the new Equation (4). In other words, under these regularity conditions, the estimates in $P$ and $T$ would be the same.
The value of possible event els with the equal probability model, in which each model structure, we compared each of the four mod-

but ORD is by far the best among the four. show that HUR is the worst, NBD and POS are better, likelihood, AIC, and BIC. All the indices unanimously benchmark these models using various criteria: log-

HUR are in-between with 15 free parameters. We simultaneous one with 19 parameters. The NBD and parameters, while the ORD model is the least par-

all four models are given in Table 3. Among these variables to less than 0.2. Taking these facts into consideration, we dropped L\_LAG\_ij in our empirical analysis.

is 6. Therefore, we estimated five thresholds in our model (ORD).

The goodness-of-fit for the calibration results for all four models are given in Table 3. Among these models, POS is the most parsimonious with 14 free parameters, while the ORD model is the least par-

The goodness-of-fit for the calibration results for all four models with its restricted submodel, in which all the independent variables except the constants are dropped. The value of \( \rho^2_C \) indicates that the independent-variable specifications in both stages sign-

All the indices unanimously show that HUR is the worst, NBD and POS are better, but ORD is by far the best among the four.

To assess the improvement of model fit due to the specification of the independent variables, we also compared each of the four models with its restricted submodel, in which all the independent variables except the constants are dropped. The value of \( \rho^2_C \) indicates that the independent-variable specifications in both stages sign-

Four criteria were used to make our judgment on validation: (1) predicted log-likelihood, (2) \( \rho^2_L \) and \( \rho^2_C \), (3) root mean square error (RMSE), (4) U statistics. For the managerial consideration, we employed two addi-

Note.

1. The decomposition of the total time visitor \( i \) spent on the website in period \( j \) into three components (NCURR\_ij, STIME\_ij, NPAGE\_ij) reduces the correlation among these variables to less than 0.2.

2. In our application, the observation of the sales leads was a rare event (less than 1%). This fact leads to the serious lack of variation of the variable, which further leads to very unstable parameter estimation. Taking these facts into consideration, we dropped L\_LAG\_ij in our empirical analysis.
highly unbalanced in our data set (i.e., the proportion of nonleads (99.12%) is much bigger than the proportion of leads (0.88%)), the conventional hit ratio based on the naïve cut-off probability does not represent the true validity of the models and gives little predictive power. To account for this characteristic of the data, we developed a new validity measure, the adjusted hit ratio. In this measure, the cut-off probability is calculated as the average predicted probability of the calibration sample, i.e.,

\[
\text{Cut-off probability} = \frac{1}{N} \sum_{i,j} \hat{L}_{ij},
\]

\[N = \text{calibration sample size}.
\]

The adjusted hit ratio not only exhibits greater validity, but also gives much higher predictive power. This cut-off probability is managerially useful for online firms to select the right customers to target, especially when the targeting costs are substantially greater than the expected benefits, as was true in our application. However, in other situations, one may just rank each visitor using the predicted lead generation probability \(\hat{L}_{ij}\) and pick up the top “n” prospects.\(^1\)

Cross-sample validation results of the four competing models also are reported in Table 3. When the two stages of the models are examined jointly, we find that (1) the log-likelihood, \(\rho_0^2\), and \(\rho_c^2\) are consistent with

\(^1\) We thank one of the reviewers for this suggestion.
Table 3 Model Goodness-of-Fit and Cross-Sample Validation

<table>
<thead>
<tr>
<th>Goodness-of-fit</th>
<th>Model calibration</th>
<th>Model validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POS</td>
<td>NBD</td>
</tr>
<tr>
<td>No. of customers</td>
<td>2,069</td>
<td>2,069</td>
</tr>
<tr>
<td>No. of observations</td>
<td>10,345</td>
<td>10,345</td>
</tr>
<tr>
<td>No. of free parameters</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>AIC</td>
<td>0.946</td>
<td>0.947</td>
</tr>
<tr>
<td>BIC</td>
<td>0.956</td>
<td>0.958</td>
</tr>
<tr>
<td>$\rho^2_0$</td>
<td>0.810</td>
<td>0.810</td>
</tr>
<tr>
<td>$\rho^2_C$</td>
<td>0.251</td>
<td>0.249</td>
</tr>
</tbody>
</table>

Cross-sample validation

<table>
<thead>
<tr>
<th>Advertising response</th>
<th>Desired system outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POS</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.957</td>
</tr>
<tr>
<td>Total RMSE</td>
<td>—</td>
</tr>
<tr>
<td>U statistics</td>
<td>0.571</td>
</tr>
<tr>
<td>Total U statistics</td>
<td>—</td>
</tr>
<tr>
<td>Cut-off probability</td>
<td>—</td>
</tr>
<tr>
<td>APP for leads</td>
<td>—</td>
</tr>
<tr>
<td>APP for nonleads</td>
<td>—</td>
</tr>
<tr>
<td>AHR for leads</td>
<td>—</td>
</tr>
<tr>
<td>AHR for nonleads</td>
<td>—</td>
</tr>
</tbody>
</table>

Notes. AIC = $[-2 \cdot LL + 2 \cdot K]/N$ and BIC = $[-2 \cdot LL + \ln(N) \cdot k]/N$, where $K$ = number of free parameters and $N$ = number of observations. LL is the log-likelihood value. $\rho^2_0 = 1 - LL/LL(0)$ and $\rho^2_C = 1 - LL/LL(C)$, where LL(0) is the log-likelihood of the equal probability model and LL(C) is the log-likelihood value for the model in which all the covariates except the constant terms are dropped. RMSE is the square root of the average of the squared difference between the actual values and the predicted values. U statistics is a modified Theil’s U (e.g., Gupta 1988). APP for nonleads and leads are the average predicted probabilities for $L = 0$ and $L = 1$. AHR for nonleads and leads are the adjusted hit ratio for $L = 0$ and $L = 1$. Total RMSE and total U statistics are the RMSE and U statistics with the two equations calculated together.

those reported in the model calibration; and (2) the total RMSE and total U statistics strongly suggest that ORD has the best predictive performance among the four models. When the two stages are examined separately, we further find that the superior performance of ORD is largely due to its performance at Stage 1. This is not a surprise because all of the four models have an identical specification at Stage 2. A further look at the predictive power at Stage 2 outcome revealed two interesting observations. First, the average predicted probability for leads from the calibration sample is between 0.078 and 0.086 for the four models. This fact indicates that a conventional cut-off probability at 0.5 would indeed predict each observation as a nonlead with certainty (100%), and hence would give no predictive power regardless of which model was used. On the other hand, using the cut-off probability suggested by Equation (9), each of the four models can discriminate the leads and nonleads satisfactorily with an adjusted hit ratio above 98% for leads and 81% for nonleads. Second, the ORD model seems to have the better discriminating power than the other three competing models. The ORD model produces the largest gap between the average predicted probabilities for leads and nonleads. In terms of adjusted hit ratio, the same conclusion holds. Based on these results, we selected the ORD model for the rest of our empirical analysis.

C. Hypothesis Testing and Managerial Implications

Parameter estimates of the ORD model are reported in Table 4. Stage 1 hypotheses are supported. First, a double issue generates more website traffic than a single issue (the coefficient of SHELF is significantly
Table 4 Parameter Estimates and Marginal Effects on Sales Leads Generation of Model ORD and Hypothesis Testing

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Indirect marginal effect</th>
<th>Direct marginal effect</th>
<th>Total marginal effect</th>
<th>Hypothesis</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONST</td>
<td>-2.131***</td>
<td>(0.050)</td>
<td></td>
<td></td>
<td></td>
<td>H1a</td>
<td>Yes</td>
</tr>
<tr>
<td>SHELF</td>
<td>0.196***</td>
<td>(0.024)</td>
<td>0.002</td>
<td></td>
<td>0.002</td>
<td>H1a</td>
<td>Yes</td>
</tr>
<tr>
<td>RIGHT</td>
<td>0.930***</td>
<td>(0.062)</td>
<td>0.010</td>
<td>0.010</td>
<td>-0.006</td>
<td>H1b</td>
<td>Yes</td>
</tr>
<tr>
<td>OTHCL</td>
<td>-0.547***</td>
<td>(0.042)</td>
<td>-0.006</td>
<td></td>
<td>-0.006</td>
<td>H1c</td>
<td>Yes</td>
</tr>
<tr>
<td>N_LAG</td>
<td>-0.311***</td>
<td>(0.062)</td>
<td>-0.075</td>
<td></td>
<td>-0.075</td>
<td>H1d</td>
<td>Yes</td>
</tr>
<tr>
<td>FSTAV</td>
<td>1.689***</td>
<td>(0.043)</td>
<td>-0.406</td>
<td></td>
<td>-0.406</td>
<td>H1e</td>
<td>Yes</td>
</tr>
<tr>
<td>PTIME</td>
<td>2.257***</td>
<td>(0.075)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THETA_1</td>
<td>2.934***</td>
<td>(0.094)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THETA_2</td>
<td>3.486***</td>
<td>(0.107)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THETA_3</td>
<td>3.862***</td>
<td>(0.130)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THETA_4</td>
<td>4.180***</td>
<td>(0.179)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THETA_5</td>
<td>4.441***</td>
<td>(0.253)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONST</td>
<td>-3.390***</td>
<td>(0.418)</td>
<td></td>
<td></td>
<td></td>
<td>H2a</td>
<td>Yes</td>
</tr>
<tr>
<td>NCURR</td>
<td>0.222**</td>
<td>(0.082)</td>
<td>0.038</td>
<td>0.038</td>
<td></td>
<td>H2a</td>
<td>Yes</td>
</tr>
<tr>
<td>STIME</td>
<td>0.358**</td>
<td>(0.152)</td>
<td>0.062</td>
<td>0.062</td>
<td></td>
<td>H2b</td>
<td>Yes</td>
</tr>
<tr>
<td>NPAGE</td>
<td>0.418***</td>
<td>(0.094)</td>
<td>0.072</td>
<td>0.072</td>
<td></td>
<td>H2b</td>
<td>Yes</td>
</tr>
<tr>
<td>PRSVP</td>
<td>0.689***</td>
<td>(0.135)</td>
<td>0.033</td>
<td>0.033</td>
<td></td>
<td>H2c</td>
<td>Yes</td>
</tr>
<tr>
<td>FSTAV</td>
<td>-0.196***</td>
<td>(0.078)</td>
<td>-0.406</td>
<td>-0.406</td>
<td>-0.440</td>
<td>H3</td>
<td>Yes</td>
</tr>
<tr>
<td>Error correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHO (ρ)</td>
<td>0.108**</td>
<td>(0.058)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-3,780.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Significant at 0.05. ***Significant at 0.01.

Stage 2 hypotheses are also supported. First, those visitors who are most likely to become sales leads have the following characteristics: (1) they made more visits to the website (the coefficient of NCURR is significant and positive), providing support for Hypothesis H2a; and (2) they spent more time on each page (the coefficient of STIME is significant and positive) and they downloaded more pages each time they made a visit (the coefficient of NPAGE is significant and positive), providing support for Hypothesis H2b. Second, the serial positioning of the link to the response form within the homepage navigation menu is important in generating sales leads. A visitor is more likely to submit an RSVP form if the link is placed at the top of the page than at the bottom of the page (the coefficient of PRSVP is significant and positive). This finding provides support for Hypothesis H2c. The decomposition of the stickiness into knowledge of viewing time and number of pages downloaded made it possible for the owner of Thestyle.com to sort leads into high-yield visitors and low-yield visitors. In this way, he tailored personal
Table 4 indicates that the Stage 1 advertising response (N) and the Stage 2 sales leads (L) are positively correlated in two different ways: (1) a positive correlation via the observed structural parameter \( \kappa_N \) (the coefficient of NCURR is significant and positive), and (2) a positive correlation via the unobserved errors (the correlation parameter RHO of the error terms is significant and positive). These two findings suggest that although observable consumer-specific advertising and website characteristics (e.g., response latency, session time, etc.) can partially characterize the correlation between Stage 1 and Stage 2 expected utilities, the often unobserved characteristics of visitors in our two-stage promotional model (e.g., threshold effects, consumer demographics and psychographics) contribute to such correlation. This attribution may bias the assessment of the drivers of desired system outcomes when the correlation between the unobserved errors is not introduced.

We further find that the quicker visits have both a positive direct effect (the coefficient of FSTAV in Stage 2 is significant and negative) and a positive indirect effect (the coefficient of FSTAV in Stage 1 is significant and negative), providing empirical support to Hypothesis H3. The hypothesis testing results are summarized in Table 4.

To further compare the relative impact of different independent variables on sales leads, we calculated the marginal effects (Greene 1998). In our model, the conditional mean of sales leads is

\[
E(L_{ij} | X_{ij}, Z_{ij})
\]

\[= \sum_{n>0} \Pr(L_{ij} = 1, N_{ij} = n | X_{ij}, Z_{ij})
\]

\[= \sum_{n>0} \left\{ \left( \Phi(\tau_{ijn}) - \Phi_2(\tau_{ijn}, v_{ij}; \rho) \right) - \Phi_2(\tau_{ijn-1}, v_{ij}; \rho) \right\}, \quad (10)
\]

where \( \tau_{ijn} = \theta_n - (X_{ij} \alpha + v_N N_{ij-1}) \) and \( v_{ij} = -(Z_{ij} \gamma + \kappa_N N_{ij} + \kappa_1 L_{ij-1}) \), and \( \Phi_2 \) is the CDF for the standard bivariate normal distribution. The marginal effects of independent variable X or Z on the sales leads L can be obtained in either of two ways, depending on the scale properties. If the independent variable is continuous, marginal effects can be calculated by evaluating the partial derivative of the conditional mean of L in (10) at the sample mean. If the independent variable is dichotomous, marginal effects may be calculated by evaluating the difference between the two conditional means and the sample means of other independent variables with the estimated model parameters. Those independent variables specified in Stage 1 will have an indirect marginal effect on sales leads, while those independent variables specified at Stage 2 will have a direct marginal effect. For those independent variables that enter both Stage 1 and Stage 2 (i.e., FSTAV), their total marginal effects on sales leads L will be the sum of the direct marginal effect from Stage 2 and an indirect effect from Stage 1, which is carried through the endogenous variable N and the random error correlation \( \rho \). The marginal effects are reported in Table 4. First, FSTAV has much higher marginal effect (−0.440) on sales leads, which suggests that advertising characteristics surrogated by response latency have a stronger impact on sales leads than website characteristics. Second, these variables can affect sales leads in different ways. For instance, while the magazine’s shelf life SHELF has a significant indirect impact (0.002) only, response latency FSTAV has significant impact both indirectly and directly. Surprisingly, FSTAV’s indirect impact (−0.406) is much greater than its direct impact (−0.034) on generating desired system outcomes.

5. Conclusion

We proposed a two-stage conceptual and econometric model to jointly assess website promotion performance for pure online firms. To the best of the authors’ knowledge, this is the first study on website promotion assessment that investigates the effects of traditional media advertising and website presence in an integrated framework. As such, our research has several limitations. The testing of our hypothesis is based on a single study. More empirical research is needed to establish the generalization of the empirical findings. Second, we remained silent on impact of advertising content characteristics on generating website visits because of the research design and ad budget constraints. Third, the measurements we use in this study were limited by the technology available in 2000 as well as by the owner’s resistance to the use of cookies. We anticipate that better measurements will emerge as information technology continues to advance.
Our research also contributes to the e-commerce literature in important aspects. Two important managerial findings emerge from this empirical study of a pure online firm’s campaign to generate sales leads through print advertising. First, advertising and website characteristics affect sales leads in different ways. Some advertising characteristics, like single versus double issues of a magazine, influence sales leads indirectly via their effect on website visits and two-stage correlations. Some website characteristics, like average time per page, influence sales leads directly. Still other characteristics, like advertising response latency, affect sales leads both indirectly and directly. These findings generate insights for positioning ads in print vehicles as well as for website design for online firms. Second, some of the measurements we developed are directly actionable. For example, consumers spending more time per page should be treated differently from the rest of the visitors because they are more likely to generate desired system outcomes. Also, putting an RSVP at the top of a Web page generates more sales leads directly, while putting an ad on the right page of a magazine generates more website traffic, which consequently generates more sales leads.

Our joint two-stage econometric model offers several innovations that are important in the analysis of online data. First, the model is designed to deal with the simultaneity and censoring issues. These issues are intrinsic in the data structure of all pure online firms’ promotional campaigns. Our model demonstrates several ways to extend current discrete bivariate models (Greene 1998, Boyes et al. 1989): (1) it is a simultaneous bivariate model with censoring; (2) it is an ordered probit model in Stage 1 and a binary probit model Stage 2; (3) it is longitudinal. Second, our model specification is flexible. It allows the impact of advertising on desired system outcomes to be assessed separately as an indirect effect in Stage 1 via the website traffic and as a direct effect in Stage 2. This feature of our model provides better and more complete assessment of the promotional performance for online firms. Third, our approach permits the modeling of the relationships between website traffic and desired system outcomes in two different ways: as a coefficient of the observed website traffic and as a correlation of the unobserved errors. Fourth, our model can be used in situations in which survey research is not possible or is too costly. These conditions often hold in the case of new products with limited marketing budgets. The data required in our study illustrates this point. The advertising schedule and customer online shopping data are readily available at virtually no additional cost for online firms. In our application, we rely on customers’ IP addresses to identify unique visitors. This approach is appropriate for our particular application because products are high-priced collectors’ items and repeated visits are infrequent. In our data set, repeated visits are between 1% and 4% within a given period and only 2% for the five periods aggregated together. In theory, IP addresses alone are not enough to guarantee the uniqueness. The impact of this imperfect identification on the estimate in our application can be an issue for websites that carry frequently purchased products. In such cases, cookie technology should be implemented to help identify unique visitors.

There are several promising ways to extend the current research. First, our model can be extended to encompass those net-enabled organizations that employ websites in their business strategies (Straub et al. 2002). For instance, Zufryden (2000) observes that a new film’s promotion strategy often consists of two components: (1) producers broadcast a trailer on television (or in newspapers) for an upcoming film and invite moviegoers to visit the film’s website, and (2) when moviegoers visit the film’s website they obtain detailed information about it (e.g., plot, stars, and trailers). Using aggregate data, Zufryden (2000) found that website visits have a significant positive influence on box-office performance. Second, information technology has made tracking of advertising effectiveness an important service for many organizations (e.g., www.hitslink.com). Our model can be extended to those firms with an online presence that employ multiple advertising vehicles to drive website traffic by simultaneously combining online with offline media. The opportunities to extend the research in this direction increase significantly.

---

12 One way to reduce the impact of the imperfect identification of unique visitors in our data set is to treat the Stage 1 response as a binary variable. Although the ORD model provides better goodness-of-fit than the binary treatment, the estimates of the coefficients are qualitatively indifferent, which indicates that such impact is minimal in our application.
with the rapid growth of consumer access to broadband and the shift of mass TV advertising to targeted streaming-video ads.

Acknowledgments
The authors thank the 2001 ICIS Conference reviewers and audience and the 2002 Marketing Science Conference audience for providing valuable comments for an earlier version of this paper.

Appendix

A. The POS Model
In the model POS, we posit that a potential customer \( i \) arrives at the website following a Poisson distribution with rate \( \lambda_{ij} \) (per unit of time). Let the period length be \( t_j \) (number of days), then \( N_{ij} \) follows a Poisson distribution with rate \( \lambda_{ij} t_j \). Because \( \lambda_{ij} > 0 \), we further posit that \( \ln \lambda_{ij} \) is a linear function of some observable Stage 1 drivers \( X_{ij} \) in the period (the conventional log-linear specification). Therefore, Equation (3) in the ORD model is replaced by

\[
\Pr(N_{ij} = n) = \frac{\left(\lambda_{ij} t_j\right)^n e^{-\lambda_{ij} t_j}}{n!} \quad \text{and} \quad \ln \lambda_{ij} = X_{ij} \alpha + \nu_{ij} N_{ij-1}.
\]

(A1)

Notice that \( N_{ij} \) is a count variable; it will be hard to directly specify the correlation of \( N_{ij} \) and \( L_{ij} \). Therefore, we first transfer the count variable \( N_{ij} \) into a continuous standard normal variable \( \xi_{ij} \) (e.g., van Ophe 2000). Let

\[
\tau_{ij} = \Phi^{-1}\left(\sum_{k=0}^{n} \Pr(N_{ij} = k)\right) = \Phi^{-1}(\Pr(N_{ij} \leq n)) \quad \text{and} \quad \Pr(N_{ij} = n) = \int_{\tau_{ij}}^{\infty} \phi(\xi) d\xi,
\]

(A2)

where \( \phi \) and \( \Phi \) are the PDF and CDF of the standard normal distribution. In this way, the correlations of \( \xi_{ij} \) and \( \eta_{ij} \) will be specified just as in Equation (5) of the ORD model.

The likelihood function is formed similar to Equation (8) of the ORD model as

\[
LL = \sum_i \sum_j \ln f(N_{ij} = n, L_{ij} = l)|_{n \geq 1} + \sum_i \sum_j \ln f(N_{ij} = n, L_{ij} = l)|_{n = 0},
\]

(A3)

in which the first term is

\[
\ln f(N_{ij} = n, L_{ij} = l)|_{n \geq 1} = l \ln \Pr(N_{ij} = n, L_{ij} = 1) + (1 - l) \ln \Pr(N_{ij} = n, L_{ij} = 0) = l \ln \left(\Phi(\tau_{ij}) - \Phi(\tau_{ij} - \nu_{ij} \rho_1)\right) - \Phi(\tau_{ij-1} - \nu_{ij} \rho_0) - (1 - l) \ln \left(\Phi_2(\tau_{ij} - \nu_{ij} \rho_0) - \Phi_2(\tau_{ij-1} - \nu_{ij} \rho_0)\right)
\]

(A4)

where \( \tau_{ij} = \theta - (X_{ij} \alpha + \nu_{ij} N_{ij-1}) \) and \( \gamma = -(Z_{ij} \gamma + \kappa_N N_{ij} + \kappa_{L_{ij-1}}) \) and \( \Phi_2(\cdot) \) is the CDF for standard bivariate normal distribution. The second term is

\[
\ln f(N_{ij} = n, L_{ij} = l)|_{n = 0} = \ln(\Phi(\tau_{ij})) = \ln(\Phi(\Phi^{-1}(e^{-\lambda_{ij} t_j}))) = \ln(e^{-\lambda_{ij} t_j}) = -\lambda_{ij} t_j.
\]

(A5)

B. The NBD Model
In this model, we generalize the POS model by introducing an unobserved heterogeneity effect \( \nu_{ij} \) (following a Gamma distribution) into the Poisson rate \( \lambda_{ij} \) (e.g., Greene 2000),

\[
\ln \mu_{ij} = \ln \lambda_{ij} + \ln w_{ij} = X_{ij} \alpha + \nu_{ij} N_{ij-1} + \ln w_{ij} \quad \text{and} \quad g(w_{ij}) = \frac{\theta^\theta e^{-\theta w_{ij}} w_{ij}^{\theta-1}}{\Gamma(\theta)}.
\]

(B1)

This is the negative binomial distribution (NBD) for modeling the Stage 1 advertising response. Therefore, the NBD model is specified by replacing (A1) in the POS model by

\[
\Pr(N_{ij} = n) = \frac{\Gamma(\theta + n)}{\Gamma(n + 1) \Gamma(\theta)} \theta^\theta r^\theta (1 - r)^n \quad \text{and} \quad r = \frac{\lambda_{ij}}{\lambda_{ij} + \theta}.
\]

(B2)

The remaining specification of the model and the log-likelihood are similarly formed as in the POS model.

C. The HUR Model
In this model, we assume that the zero outcome of the data generating process of visits for potential customer \( i \) in period \( j \), \( N_{ij} \), is qualitatively different from the positive ones (called hurdle Poisson model). A binary probability model determines whether a zero (with probability \( p \)) or a nonzero outcome (with probability \( 1 - p \)) occurs. If the nonzero outcome occurs, it follows a truncated Poisson distribution with rate \( \lambda_{ij} \). The HUR model is specified by replacing (A1) in the POS model by

\[
\Pr(N_{ij} = 0) = p, \quad 0 < p < 1 \quad \text{and} \quad \Pr(N_{ij} = n) = \frac{(1 - p)(\lambda_{ij})^n e^{-\lambda_{ij} t_j}}{(1 - e^{-\lambda_{ij} t_j})^n}.
\]

(C1)

where \( \ln \lambda_{ij} = X_{ij} \alpha + \nu_{ij} N_{ij-1} \) if \( n > 0 \). The remaining specifications of the model and the log-likelihood are formed as in the POS model.

References


