Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach

Catarina Sismeiro and Randolph E. Bucklin[1]

March 2002
Revised June 2003
Revised November 2003
Revised January 2004

[1]Catarina Sismeiro is Assistant Professor, Marshall School at USC, Los Angeles, CA 90089. Randolph E. Bucklin is Professor of Marketing, Anderson School at UCLA, 110 Westwood Plaza, Los Angeles, CA 90095. The authors thank an anonymous collaborating company for providing the data used in this study. The authors also thank the editor, three anonymous JMR reviewers and Anand Bodapati, Bart Bronnenberg, Dominique Hanssens, and Keisuke Hirano for their valuable comments. Research support from the UCLA Academic Senate is gratefully acknowledged. Email correspondence to the first author at sismeiro@marshall.usc.edu.

Modeling Purchase Behavior at an E-Commerce Web Site:
A Task Completion Approach

Abstract

The authors develop and estimate a model of online buying using clickstream data from a Web site that sells cars. The model predicts online buying by linking the purchase decision to what visitors do and to what they are exposed while at the site. Predicting Internet buying poses several modeling challenges. These include: (1) online buying probabilities are usually low which can lead to a lack of predictive and explanatory power from models, (2) it is difficult to effectively account for what Web users do, and to what they are exposed, while browsing a site, and (3) because online stores reach a diverse user population across many competitive environments, models of online buying must account for the corresponding user heterogeneity.

To overcome these hurdles, the authors decompose the user's purchase process into the completion of *sequential* Nominal User Tasks (NUTs) and account for heterogeneity across visitors at the county level. Three major user tasks required to complete a purchase are: (1) completion of product configuration, (2) input of complete personal information, and (3) order confirmation with provision of credit card data. These tasks are managerially meaningful and correspond to considerable visitor loss. Using a sequence of binary probits estimated through Bayesian methods, the authors model the visitor's decision of whether or not to complete each task *for the first time*, given that the visitor has completed the previous tasks *at least once.* The propensity for completing one task is also allowed to affect the propensity for completing subsequent tasks.

Results indicate that visitors' browsing experiences and navigational behavior are predictive of task completion for all decision levels. The use of interactive decision aids, the exposure to site design characteristics, and the effort visitors exert to gather and process information are predictive of completion of the first two NUTs, but not the third. Results also indicate that the number of repeat visits *per se* is not diagnostic of buying propensity, and that offering sophisticated decision aids at the site does not guarantee increased conversion rates. The authors also compare the predictive performance of the task completion approach versus single-stage benchmark models in a holdout sample. The proposed approach provides superior prediction and better identifies likely buyers, especially early in the task sequence. Implications for Web site managers are also discussed.

Key words: Internet, E-Commerce, Choice Models, Clickstream Data, Bayesian Methods

INTRODUCTION

As the growing population of Internet users becomes familiar with the medium, activities such as searching for product information online have become widespread. In contrast, online buying is not yet as commonplace. According to a study by EMarketer (2000), in the year 2000, nearly three-fourths of Internet users browsed, researched or compared products online. However, more than 65% of those shoppers never used the Internet to make an actual purchase. This reluctance to adopt the Web as a retail channel, combined with the high number of Web site visits for information search, have produced very low visitor (or visit) to purchase conversion rates. For example, a Forrester Research (1999) study found that conversion rates are less than 2% for about 70% of the commercial Web sites surveyed, and that they typically vary from 1% to 4%.

The low rate of visit-to-purchase conversion means that predicting and understanding online buying behavior is of utmost importance for e-commerce Web site managers. Even very small changes in conversion can result in considerable increases in sales revenue. Nevertheless, little is known about the effects of what visitors do at a site, and to what they are exposed at a site, on the propensity of a user to buy online (e.g., Totty 2002). The purpose of this paper is to develop a modeling approach to predict individual online buying behavior, and to estimate and test the proposed model on clickstream data collected by the operator of a major commercial Web site. In so doing, we investigate the predictive power of visitors' actions at the site (e.g., browsing behavior, use of interactive decision aids, and information search and processing) while also controlling for differences in visitors' exposure to page design and content. We also seek to provide implications for Web managers, including potential online and offline interventions.

Our approach is based on the idea that we may improve the prediction of online purchase, at the individual visitor level, by first grouping a given Web site's activities into a small number of discrete tasks, each of which must be completed for purchase conversion to take place (e.g., selection of product, placement into shopping cart, provision of shipping information, provision of billing information). Thus, rather than predict the buy/no buy outcome directly as a function of covariates, we predict it as the product of a chain of conditional probabilities each of which corresponds to the completion of a task required for purchase. These are manifested as, for example, the probability of product selection given site visit and the probability of shopping cart placement given product selection. Each conditional probability is, in turn, made a function of covariates capturing visitor's within-site browsing behavior, site usage, and idiosyncratic page exposures. Our conditional probability approach to studying consumer e-commerce is somewhat analogous to the classic decomposition used by scanner panel researchers in packaged goods: quantity given brand choice, brand choice given category purchase, category purchase given store visit (Gupta 1988). We also note that in the packaged goods domain, predictor variables such as price and promotion have been modeled to affect each consumer decision separately and have often been shown to vary in impact and significance across those decisions (e.g., Gupta 1988, Bell, Chaing, and Padmanabhan 1999).[2]

Using data from an e-commerce Web site that sells new cars, we investigate whether our conditional probability approach to modeling purchase conversion yields superior prediction versus a null "one-step" model. The null model uses the same covariates but does not decompose the site into sequential tasks. We also investigate whether our multi-step approach can provide more diagnostic information about site performance than a single-step model. Because the same covariate (e.g., a measure of Web site usage such as average time spent per page) may be positively related to the completion of one task but negatively related (or not related at all) to the completion of another task, there is potential for the task completion approach to be more diagnostic than a one-step approach.

Since the site we analyze sells new cars, its structure differs somewhat from e-commerce retailers selling a broad assortment of low-cost items. Thus, the tasks that pertain to a new car purchase online are not exactly the same as those at widely-used e-commerce sites such as Amazon.com. Nevertheless, we view the general idea of using a chain of conditional probabilities to decompose site tasks required for purchase as transferable to other classes of products and types of e-commerce Web sites.

[2] Recently, Van Heerde, Gupta and Wittink (2003) have shown that intepretations of the elasticities calculated in studies such as these have commonly been incorrect.

Literature

Previous experimental and empirical research has demonstrated that what visitors are exposed to, and what they do in a site visit, may influence the dynamics of conversion probabilities and the likelihood that a visitor will buy online. For example, Mandel and Johnson (2002) show that page design can change preferences, and hence purchasing decisions, by influencing attribute importance. They find that users dynamically adapt their behavior to the page-by-page stimuli presented to them, even when unaware of their own adaptive behavior.

Dellaert and Kahn (1999) show that site users are able to separate their evaluations of waiting experiences while browsing a site from the evaluations of the site itself. However, when there is uncertainty about the waiting (as with the majority of page downloads), the negative feelings generated by the waiting experience carryover into the evaluation of the Web site. Therefore, the delays in page download that visitors experience may negatively affect the visitor's entire Web site experience. As a result, visitors may decide to exit the site, or to abandon the shopping cart without purchasing.

Häubl and Trifts (2000) study the use of interactive decision aids (recommendation agents and comparison matrices) and their impact on consumer decision making. They find that use of decision aids decreases the consumer's search effort for product information, reduces the size but increases the quality of consideration sets, and improves the quality of purchase decisions. This raises the possibility that the use of interactive decision aids on the Web can lead to a higher online purchase propensity.

Bucklin and Sismeiro (2003), using clickstream data recorded in server log files, examine two aspects of browsing behavior: (1) the visitor's decisions to continue browsing or to exit the site, and (2) the length of time spent viewing each page. The authors show that the browsing behavior of Web site visitors changes dynamically both within and across site visits. Their results suggest that users with time constraints are also either more efficient or more focused in their product searches, and seem to learn faster across site visits. Hence, visitors' browsing behavior (number of pages requested and time spent at the site) might indicate their interest level in online buying by revealing their time investment and search behavior at the site.

Moe (2003) takes a useful first step toward integrating content variables into the analysis of browsing and purchase. Using page-to-page clickstreams from an online store, and based on observed in-store browsing patterns that include the general content of the pages viewed (e.g., category, product, or brand pages), the author develops and tests a typology of store visits that vary in terms of purchasing likelihood. These include knowledge building, hedonic browsing, directed-buying, and search/deliberation. The work focuses on understanding the different types of online store visits and the metrics that may differentiate them. The article does not explicitly provide a model of browsing behavior or the consumer purchase decision.

Moe and Fader (2003) develop an individual probability model of visit to purchase conversion. Based on the pattern of previous visits and purchases for each site visitor, the model predicts those visits likely to convert to purchases. The model allows previous visits to affect purchases in different ways, and it also allows shopping behavior to change and evolve over time as a function of past experiences. However, within-visit activities such as browsing behavior, task completion, use of interactive decision aids, exposure to site design and site content are not studied.

Moe et al. (2003) proposed a Bayesian treed model for predicting online purchase in which site visitors are also simultaneously clustered into homogenous behavioral segments. The approach is cross-sectional in nature, linking measures of overall site usage behavior during a visit with the buy or no-buy outcome associated with a given visit.

In a recent working paper, Montgomery et al (2003) proposed a modeling approach to analyze the path users choose to follow while visiting a Web site. Visitors to an online bookseller are conceptualized as oriented toward either browsing or purchase. Their choices for the type of page to request from the site are modeled with a dynamic multinomial probit. Though the paper focuses on path analysis, the authors also show how their model can be used to predict purchase conversion. They propose simulating all of the various paths visitors might take on a given site visit and then computing the proportion of those paths which involve the order page. Predictions of purchase conversion by the same visitor on future site visits are not incorporated in the model.

Proposed Approach

Our modeling approach predicts online ordering, for a given site visitor, based on the decomposition of the Web site into sequential Nominal User Tasks (NUTs), a concept first introduced by Rebane (2001). This approach permits us to model conditional probabilities for task completion, each of which are much higher than the typical final visit-to-order conversion rate of 1% to 4%. Our objective is to improve the performance of the choice model in predicting purchase conversion versus a single-step approach. While Montgomery et al (2003) model the entire path of page choices, we identify and model the user's progress past critical points of attrition in the tasks required for purchase. Unlike previous studies, our modeling approach also allows for purchase conversions to take place in either the current or in subsequent site visits made by a given visitor.

For the Web site we studied, we identified three principle NUTs required to complete a purchase: (1) completion of product configuration, (2) input of complete personal information, and (3) order confirmation with provision of credit card data. These tasks are both managerially meaningful and corresponded to considerable, but not excessive, visitor loss at each step. Though the tasks specifically pertain to purchasing a new car, we note that the general framework also corresponds well to the tasks required for purchase of many products and services on the Web. This is especially the case where the user is required to engage in some customization of the product or service before ordering (e.g., apparel, mortgages and home equity loans, and insurance). On the other hand, our approach is not immediately applicable to sites where users' paths may branch into multiple (or parallel) task sequences, each leading to purchase (e.g., the choice to use "One-Click" versus following the conventional ordering sequence). To accommodate this, our model would need to be extended to allow for the additional complexity of user path selection.

In our approach we jointly model the first time each visitor completes each task, given that the visitor has started a site visit and completed the preceding tasks (if any). Task completion is a function of previous browsing behavior, repeat visitation, use of decision aids, and the level of information gathering and processing evidenced by the user. Our model therefore links what users do—and to what they have been exposed on the site—to the completion of each sequential nominal user task. The results from modeling each task can provide diagnostic information about site performance as well as potential insights into consumer purchase behavior on the Internet. We also allow the propensity of completing one task to affect the propensity of completing subsequent tasks. In addition, using random effects defined at the county level, we account for heterogeneity and regional differences in the Web site's competitive advantage, pricing, laws, consumers, and speed of Internet connections.

This paper is organized as follows. First we describe our modeling approach. Second, we describe the clickstream dataset we use to estimate and test the model and introduce the Nominal User Tasks (NUT). Third, we present results from estimating our model and discuss implications. Finally, we conclude, noting limitations in our research and next steps.

MODELING APPROACH

The purpose of our modeling work is to predict—and in so doing hopefully provide a better understanding of—individual-level online buying behavior. Next we present some specific challenges in modeling and predicting online purchasing behavior. We then introduce our task completion approach to modeling online purchases and describe the specific model formulation for each task.

Challenges in Modeling and Predicting Online Purchases

The prediction of Internet buying behavior poses specific modeling challenges. Perhaps because little effort is required to visit an online store, among the users that visit the site few ever place a product order. Thus, unlike visits to "bricks and mortar" retailers, an online store visit in itself might reveal little about a shopper's purchase intention. The very low level of purchase incidence per visitor, or per shopping trip, also means that one-step or single stage statistical procedures such as a logit or probit model of purchase given site visit can underestimate the individual-level purchase probabilities (King and Zeng 2001).

Another modeling challenge is that Web site visitors face few browsing restrictions. Although visitors to any (well-designed) commercial Web site face some navigational restrictions that impose a sequence of tasks before a purchase can occur (e.g., visitors must add the product to the shopping cart before buying it), visitors might request multiple pages from the site in no pre-determined sequence before buying online. Therefore, in building a model of buying behavior, it is not trivial to effectively account for what Web users do, and what they are exposed to, while browsing the site.

Finally, online stores reach a diverse user population and face a multitude of competitive environments (in practice one for each local market of their visitors). For example, an online retailer serving the entire continental U.S. serves about 40,000 different Zip code areas. Each Zip code may have distinct laws, distinct demographic and competitive makeup, and Internet access that is not equally reliable or fast. Also, online retailers may customize their offerings (e.g., price) to exploit consumer differences or as a response to regional costs. Consequently, browsing behavior and purchase decisions may differ across visitors from different regions.

A Task Completion Approach with Regional Heterogeneity

In order to address these challenges, we model the completion (or failure to complete) the sequence of key user actions which lead to a purchase. These actions correspond to Nominal User Tasks (NUTs) Web site visitors *must* complete before purchasing online. A Nominal User Task in Internet-based environments, is a set of activities Web site visitors perform to reach a certain, well-defined goal (Rebane 2001). Examples of such tasks for an e-commerce Web site include locating products, adding items to the shopping cart, proceeding to checkout, providing personal information and confirming the purchase. We assume that by completing tasks visitors reveal their level of *purchase intention.* We then predict purchase with a multilevel binary choice model of task completions. Specifically, we model the visitor's decision of whether or not to complete each task *for the first time*, given that the visitor has completed the previous tasks, if any, *at least once.* We do not model the number of times visitors complete each task, nor do we model the last time each task is completed. Our objective is to identify purchase motivated Web site visitors as early as possible in their use of the site.

The precise definition of the NUTs will depend on the structure of the Web site under analysis. To benefit from the conditioning structure created by the decomposition of the site into NUTs, and to obtain additional managerial insights that would otherwise be missed, we need to apply the choice models to managerially meaningful and carefully defined NUTs. Each NUT should then be sequential, engaging, and associated with a high, but not excessive, visitor loss (i.e., a considerable proportion of Web site users that could potentially complete the task does not do so). By decomposing user behavior into tasks that are no longer rare events (e.g., many more visitors make it from site visit to product selection than from site visit to ultimate purchase), we seek to improve the performance of the overall choice model and thereby better predict whether an online purchase is going to be made.

Task completion *per se* provides only a noisy signal of purchase intention. Purchase intentions may also differ *a priori* among visitors and may change over time, as visitors interact with the site and gather more information. For example, some Internet users may be gathering information for immediate or future purchases while others may be simply exploring the site for hedonic reasons or be engaged in non-targeted knowledge building. Consequently, to better predict purchase—and to better understand visitor behavior—our model incorporates covariates for what visitors are exposed to while browsing and what they do at the site before completing each task (e.g., browsing behavior, information search and processing).

Finally, we use the geographical location of visitors to build a random effects model in which visitors from the same locality are assumed to respond similarly. Geographic location then serves as a proxy for factors that may lead to differing competitive advantage (e.g., price advantage) or to differing experiences with the Internet environment (e.g., connection speed). This should also enable us to better predict individual level purchase probabilities for first-time visitors who have no previous browsing history with the site.

Random Utility Formulation

We model the visitor's decision of whether to complete each task, for the first time, conditional on the completion of previous tasks (if any) with a hierarchical binary probit that allows for region-specific parameters.[3] We assume that Web users will complete a certain task, for the first time, as long as its value (utility) exceeds some threshold. We also assume that the utility of completing the task is not certain but stochastically related to the value of the pages previously requested. In other words, the visitor uses current and past page characteristics to infer the utility of task completion. Therefore, we model the "attractiveness" of completing

[3]We model the visitor's decision to complete each task, for the first time, across all site visits made by that visitor. However, for frequently purchased goods, visitors may make multiple purchases across website visits. We can extend our model to accommodate this scenario. For example, using our task completion approach, we could model the first time a visitor completes each task in a given site visit.

each task, at the disaggregate level, as a function of what visitors do and what they have been exposed to online.

We define task utility as

$$y_{is}^{m} = \mathbf{x}_{is}^{m}\boldsymbol{\beta}_{s}^{m} + \varepsilon_{is}^{m} \tag{1}$$

where y_{is}^{m} is the utility associated with the first completion of task m by visitor i from region s ($m = 1, \ldots M$, where M is the number of tasks, $i = 1, \ldots N_s$ where N_s is the number of unique visitors from region s, and $s = 1, \ldots, S_m$ where S_m is the total number of regions for task m[4]). $\boldsymbol{\beta}_{s}^{m}$ is a $(p_m \times 1)$ vector of region specific parameters, $\mathbf{x}_{is}^{m}$ is a $(1 \times p_m)$ vector of covariates which includes an intercept, and ε_{is}^{m} is a normally distributed random variable with mean zero and unit variance, for identification purposes, that is $\varepsilon_{is}^{m} \sim N(0, 1)$. For each task m, the covariates are defined with respect to the base alternative—never completing the task—and we normalize the utility of this base alternative to zero. While we use region-specific covariates to handle heterogeneity in this application, these could be replaced, for example, with segment-specific covariates in other applications.

We observe whether a given visitor has completed a task, or not. Since we do not observe utilities, we assume that a positive task utility is associated with a first-time task completion and a negative utility is associated with never completing a given task. If these events are coded as

$$\begin{aligned} C_{is}^{m} &= 1 \text{ if task } m \text{ is completed for the first time by visitor } i \text{ from region } s \text{, and} \\ C_{is}^{m} &= 0 \text{ if task } m \text{ is never completed by visitor } i \text{ from region } s, \end{aligned} \tag{2}$$

for $i = 1, \ldots, N_s$, $m = 1, \ldots, M$, and $s = 1, \ldots, S_m$, then we observe C_{is}^{m} such that

$$\begin{aligned} C_{is}^{m} &= 1 \text{ if } y_{is}^{m} \geq 0 \text{, and} \\ C_{is}^{m} &= 0 \text{ if } y_{is}^{m} < 0. \end{aligned} \tag{3}$$

Because tasks are sequential—visitors cannot complete task $m+1$ if they have not completed task m (for $m = 1, \ldots M-1$ where M is the maximum number of tasks under analysis)—

[4]The number of regions for each task might differ if visitors from a region never complete certain tasks. Therefore, we have $S_1 \leq S_2 \leq \ldots \leq S_M$.

we also have the following constraints:

$$
\begin{aligned}
C_{is}^{m+1} &= 1 \implies C_{is}^{m} = 1 \\
C_{is}^{m} &= 0 \implies C_{is}^{m+1} = 0,
\end{aligned}
\tag{4}
$$

for $m = 1, \ldots, M-1$, $i = 1, \ldots, N_s$, and $s = 1, \ldots, S_m$.

As a result, we can model the likelihood contribution of each individual, l_{is}, as a string of conditional probabilities, such that

$$
\begin{aligned}
l_{is}\left(C_{is}^{1}, C_{is}^{2}, \ldots, C_{is}^{M}\right) &= P\left(C_{is}^{1}\right) P\left(C_{is}^{2} \mid C_{is}^{1}\right) P\left(C_{is}^{3} \mid C_{is}^{1}, C_{is}^{2}\right) \ldots \\
&\quad \ldots P\left(C_{is}^{M} \mid C_{is}^{1}, C_{is}^{2}, \ldots, C_{is}^{M-1}\right).
\end{aligned}
\tag{5}
$$

Our model also allows for selectivity bias across the different task levels. An individual user's higher or lower propensity to complete any given task may influence what that individual will do next. Therefore, at each decision level, we allow the latent utility from previous tasks (if existent) to be included as a covariate. We can then test if knowing whether someone is more or less likely to complete a given task helps predict future task completions.

To model heterogeneity across visitors, we adopt a random coefficient approach. A random effects model, which stochastically pools data across individuals, is frequently employed in the analysis of choice models (for a review see Allenby and Rossi 1999). Because we have only one observation per individual and task level, we will use the geographical location of each user to pool the individual-level data and estimate region-specific parameters (each region will have a specific parameter and each individual within the region will share the same parameters). Therefore, for each task m and covariate k (for $k = 1, \ldots, p_m$, $m = 1, \ldots, M$, and $s = 1, \ldots, S_m$):

$$
\beta_{ks}^{m} \sim \mathcal{N}\left(\mu_{mk}, \sigma_{mk}^{2}\right). \tag{6}
$$

In this random effects formulation we do not take into account the geographical proximity of the different regions (Bronnenberg and Sismeiro 2002). Given the large number of regions in our analysis, incorporating the potential spatial correlation in our multilevel choice model is quite cumbersome and we therefore leave this issue for future research.[5]

The approach adopted here is based on the conditional distributions of the parameters as presented in Appendices A and B. A commonly used alternative approach (e.g., Allenby and Rossi 1999) is to specify joint prior distributions for the parameters. This requires the estimation of the full variance-covariance matrix of the model parameters. Using simulated data, we compared our proposed conditional sampling algorithm against (1) the full variance-covariance approach, and (2) a simultaneous algorithm that assumes full independence across parameters. Model comparison was based on (1) in- and out-of-sample predictive performance, and (2) recovery of the true response parameters. We found that the conditional approach we present here was better in recovering the true parameter values associated with each variable for both low and high parameter correlation scenarios. Our approach also produced considerably better predictions even when hierarchical variables (i.e., more information) were included in the full covariance approach.[6]

EMPIRICAL ANALYSIS

We applied our model to the actions of a sample of visitors to a major commercial Web site in the automotive industry. The company wishes to remain anonymous. Web users visiting the site could view general company information and instructions on how to use the site, research the configuration of almost all commercially available cars and light trucks, obtain fixed "no-haggle" prices, and place an order for a vehicle with a credit card deposit (the remainder of the transaction is completed by phone, email, and fax). The site was simple and linear and contained standard search capabilities. Because the site provided users with a great deal of product and price information, it was used by visitors in various stages of

[5] We believe the assumption of spatial independence is also likely to be a reasonable one for our model. This is because omitted factors such as user demographics are unlikely to lead to significant spatial correlation in the choices users make to complete site tasks and ultimately buy online. Note that our model predicts the binary outcome of whether or not a user will order a new car online, not the specific make or model of the vehicle.

[6] We also compared the predictive performance of our conditional estimation approach versus the full covariance approach on the data used in this study. We found that the proposed conditional approach produced a higher pseudo-Bayes factor (Appendix D) than the full covariance approach. In addition, we obtained better out-of-sample predictions with the approach presented here versus the full covariance approach. We also incorporated demographic information from census data in a full covariance hierarchical model structure. Again, we found that the proposed conditional estimation produced superior pseudo-Bayes factors. Detailed results are available from the authors upon request.

the new-car purchase process (from “just dreaming”, to researching vehicle attributes and prices, to those ready to place an order). Management was interested in improving how the site ushers through those visitors who are serious about buying while capturing advertising revenue from exposures to all users. Below, we further describe the data set and explain how it was processed. We then describe each Nominal User Task and present our estimation results.

Data

The company monitored the browsing of visitors to the site from December 1, 2000 to February 15, 2001 using a proprietary system called TRACKER. The system identified each page with a unique tag. For each page requested by a user, the database recorded the page id, the cookie id, the day and time of the request, the Zip code inputted by the visitor (when required), any affiliated Web site the visitor was referred from, and information specific to each page. All database records correspond to single page requests and not hits, as would be the case in server log files, and account for refreshes and back button requests.[7] From the set of visitors monitored by TRACKER, we retained those who had completed the first task (defined below) for the first time between January 1, 2001 and January 20, 2001 as well as those who never completed any of the tasks despite visiting the site during that time period. This preserves a one-month initialization period for the model variables and allows time to observe whether visitors complete any remaining tasks after completing the first one.

From the resulting subset of visitors, we retained only those who had started a site session—defined as going beyond the home page (Bucklin and Sismeiro 2003)—and who had inputted a single U.S. Zip code.[8] We do not consider visitors living outside the continental U.S. (e.g., Puerto Rico) or visitors living in states which prohibit online car sales (e.g., the state of Texas). In practice these users cannot complete any of the critical tasks at the site. The final dataset also excludes all company personnel and Web crawlers. It contains the complete browsing behavior of 96,498 site visitors observed during the full tracking period (70

[7] Due to time sensitive data in most pages, the server forced a new page request for refreshes and back buttons.

[8] As a result, we have removed from the sample those visitors who had not begun a configuration (a Zip code is required to begin configuring any car) and those who had inputted multiple Zip codes (visitors inputing multiple Zip codes correspond to less than 5% of the visitors who started a session). This ensures that only those visitors who revealed a minimal interest in car buying were retained and we avoid arbitrarily assigning visitors to a single Zip code.

days). These visitors come from 2,045 different U.S. counties. The number of orders contained in our final sample is 1,969 which produces a conversion rate of roughly 2%. These orders constitute about 15% of the total placed on the site during the period studied.

Specific Task Definitions

Figure 1 presents a schematic drawing of the site we studied. The entire sequence of pages presented in Figure 1 corresponds to the shortest path to purchase. Visitors may request other pages not required for a vehicle order (e.g., pages containing company information, pricing policies, and consumer reports) as well as return to or repeat any of these pages multiple times. As Figure 1 indicates, we decomposed the within-site tasks into *three* NUTs. We sought to define the tasks to be sequential, engaging, managerially meaningful, and to correspond to a considerable, but not excessive, visitor loss. Table 1 presents summary statistics for each task, illustrating the considerable attrition (at least 65%) at each level.

The first task (NUT1), Completion of Product Configuration, represents the basic starting point for purchase. This NUT aggregates several minor tasks that go beyond mere price checking. For example, to complete this task the visitor must select the make, model, and trim of the car; s/he must also decide on options, color and on the purchase of extra warranty programs. Consequently, the customization activity and the effort involved in the completion of this NUT should reveal a strong interest in car buying.[9]

The second task (NUT2) corresponds to the Input of Complete Personal Information. A visitor can complete this task, which requires considerable data input effort, by filling out several detailed forms after a full car configuration. We believe such forms may make salient the concerns of Web users about on-line privacy. According to the last GVU survey (GVU 1998) more than 70% of Internet users demanded new laws to protect Internet privacy, and 77.5% percent of respondents declared that online privacy is even more important than the convenience of online buying. This raises the question of how the response of visitors to the need to input this extensive personal information may be related to browsing and site covariates, and how such response predicts final ordering behavior.

[9] The company did not retain information about the specific vehicle configured to complete NUT1. Thus, we observe the completion of this activity but not the make or model the user configured. We are therefore unable to include measures of product variety search in our study.

The third task (NUT3) is Order Confirmation with Credit Card Provision. Visitors can complete an online order, after fully configuring the car and after providing their personal information, by inputting a valid credit card number and agreeing to pay a deposit. The deposit was required from all ordering visitors, although it was generally refundable. The deposit requirement may partly explain the considerable user attrition for this task. First, it increases the payment salience of the entire transaction. Second, it separates serious buyers from those who may still be experimenting with the site. Finally, it raises potential concerns regarding online security (almost 80% of respondents of the 1998 GVU survey were either somewhat or very concerned with online security). Our approach allows us to predict response to the requirement of a credit card deposit based on visitors' previous actions on the site and exposure to Web site characteristics.

Model Variables

We model the user's decision to complete each Nominal User Task as a function of variables representing (1) browsing behavior (i.e., time and page views), (2) repeat visitation to the site (return and total number of sessions) (3) use of interactive decision aids, and (4) data input effort and information gathering and processing. These variables were extracted directly from the TRACKER database and are described in Appendix C. We also identified a series of page specific characteristics that may influence the visitor's experience with the site and be associated with task completion (clutter, dynamic content, presence of figures, number of links, and page size). Two independent judges performed the page-by-page analysis needed to construct these measures. Each user's idiosyncratic exposure to these characteristics, at each task level, is obtained by weighting these variables by the page visitation for each user.[10] In our approach, values for each covariate for NUT2 and NUT3 are computed based on user activity *following* the first-time completion of the preceding NUT. (We also estimated models where covariates were allowed to cumulate across NUTs versus being reset to zero with the completion of a NUT. In all cases, results for the reported approach produced superior fits.)

[10] This is analogous to what is done in the store choice literature as grocery shoppers buy in each trip multiple goods and are exposed to multiple marketing actions (see Bell and Lattin 1998). Note that each visitor can visit several different pages within the site before completing any of the tasks, or before leaving the site. Each user can also visit any page multiple times.

Table 2 presents the summary statistics for all covariates by NUT.

Estimation Method

We use a Bayesian approach, implemented with Markov Chain Monte Carlo (MCMC) methods, to estimate the joint model of task completion (equations 1-6). We introduce priors over the parameters that are common to all visitors (see Appendix A). We then sample from the joint posterior distributions by sampling from the full conditional distributions of the model parameters. For each independent task model we applied standard theory to obtain the full posteriors of β^m_{ks}, μ_{mk} and σ^2_{mk} ($m = 1\ldots, M$, $k = 1, \ldots, p_m$, and $s = 1\ldots, S_m$) conditional on the latent utilities (Gelfand and Smith 1990; Gelfand et al 1990). We use data augmentation methods to simulate the latent utilities conditional on the model parameters (Tanner and Wong 1987).

For the multiple task model with selectivity bias we derived the full conditional posteriors for all model parameters. Because not all of the full conditional distributions are known, we use an Adaptive Rejection Sampling (ARS) step within the Gibbs sampler to simulate from these unknown distributions (Gilks and Wild 1992). Appendix B gives the full conditional distributions and the details on the Gibbs sampler, ARS steps, and data augmentation procedures.

We monitored the chains for convergence and, after convergence, we allowed for long chains to run. On average, we burned in about 20,000 draws for each model and simulated an additional 20,000 draws for posterior analysis. We kept only the 10^{th} draw from each chain to reduce computer memory requirements. The resulting 2000 draws were used in our analysis.

Estimation Results

In what follows below, we report the results for a model with county-level heterogeneity. This choice of regional aggregation was based on managerial significance, data characteristics, and fit. First, managers believed that counties would provide a good unit of analysis with which to analyze the activity of the company (both the regional nature of competition and pricing decisions could be captured by a county-level approach). Second, observations at the level of Zip codes were often sparse, causing most region-specific parameters to be indistinguishable from the population mean parameters (we observed about 13,336 Zip codes for NUT1, 8,016 for NUT2 and 3,276 for NUT3, with a median number of observations per Zip code of 3, 2, and 1 for NUT1, NUT2 and NUT3 respectively). At the state level, the aggregation was coarse (41 states observed at each task level with a median number of observations of 1,163 for NUT1, 336 for NUT2, and 51 for NUT3).

We tested three alternative heterogeneity specifications against our proposed county-level approach by estimating each model and computing the pseudo-Bayes factors (see Appendix D). A Zip code-level model performed worse for all three tasks. A state-level model performed worse than the county-level specification for NUT1 and NUT2, and equally well for NUT3. A homogeneous model, in which all individuals and regions shared the same response parameters and intercept, performed worse than all other models.

We also tested linear and non-linear (logarithmic and quadratic) functional forms for the continuous predictor variables in the model. We selected the specifications which provided the best pseudo-Bayes factor. Table 3 gives the functional form specifications adopted for the county-level binary probit models along with the pseudo-Bayes factors.

Accounting for possible selectivity bias among the three tasks provides no pseudo-Bayes factor improvement. Table 3 shows that Selective 2 and Selective 3 have lower pseudo-Bayes factors than Full 2 and Full 3, respectively. Thus, knowing whether a visitor was more or less likely to complete a given task, for the first time, gives no additional predictive power as to whether s/he is more likely to complete subsequent tasks. Empirically, it is sufficient for these data to simply condition upon whether or not the visitor has completed the previous task (if any) and we therefore dropped selectivity bias from the models.

Table 4 presents a summary of the population-level parameters (posterior means and 95% probability intervals) for each binary probit model. Variables which did not improve the pseudo-Bayes factor were dropped from estimation and have been designated "n.s." in Tables 3 and 4. We note that the 95% probability intervals for all reported parameters cover the same sign as the population means. Because it is often difficult to gauge effect size from parameter values alone, in Table 5 we present the corresponding elasticities or, for indicator variables, attributable risk. Attributable risk is the absolute change in the probability of task completion due to the presence, versus absence, of the factor (e.g., Manski 2001).

An overview of the results shows that decomposing the online purchase process into Nominal User Tasks yields insights that would be unavailable from a single-stage approach which simply predicted overall purchase likelihood. Table 4 shows that many of the model covariates have different effect signs for different NUTs or they are predictive of the completion of one task but not of others. For example, the variable RETURN (a site exit and return prior to task completion but following completion of the preceding tasks, if any), does not improve prediction for the first task, is positively signed for the second, and negatively signed for the third. Had the model been constrained to a single-stage, it would not be possible to reveal these differences. This diagnostic capability is a key advantage of the task completion approach versus single-stage or purely cross-sectional approaches (e.g., Moe et al 2002). Later, we will also compare the predictive performance of our model versus single-stage approaches. We now discuss some of the substantive results obtained from model estimation.

Browsing Behavior. Following Bucklin and Sismeiro (2003), we capture the browsing behavior of site visitors with two variables, (1) the total time spent viewing the site (TIME) and (2) the total number of page views (TPAGE). Note that both variables are computed since the completion of the previous task, or, in the case of NUT1, from the beginning of the first recorded site visit. We find that viewing time (TIME) is not predictive of completing NUT1 but that the longer visitors use the site, given that they completed the first task, the more likely they are to complete NUT2 or, given that, NUT3 (the 95% probability interval of the posterior means of the population level parameters for TIME is $[0.753, 0.828]$ for NUT2, and $[0.823, 0.921]$ for NUT3). The posterior mean of the elasticity of TIME, on average across all counties, is 1.856 for NUT2 and 1.654 for NUT3. On the other hand, TPAGE is positively signed for NUT1 but negatively signed for NUT2 and NUT3. Collectively, these results suggest that, among users who complete the first task and have therefore configured a vehicle, purchase is more likely from those who spend more time at the site but request fewer pages. This is consistent with the "Directed Buying" behavior discussed by Moe (2002). In contrast, users who continue to request many pages after completing a vehicle configuration may still be searching or simply engaged in hedonic use of the site.

Repeat Visitation. The number of repeat visits to the site, TSESSION, did not provide improvements in the pseudo-Bayes factors for any of the task levels and was dropped from the model. Though users may browse differently the more they visit a site (Johnson, Bellman and Lohse 2003; Bucklin and Sismeiro 2003), this variable was not diagnostic of online car purchase. This finding differs from Moe and Fader (2003). They report that the more each Internet user visits a site the more likely they are to buy online. Our contrary results may be due to a difference in the type of purchase involved. Moe and Fader (2003) study books and CDs with multiple purchases per visitor while we use data on the purchase decision for a new vehicle.

Though the number of site visits is not predictive of buying decisions, we do find that a site exit and return, RETURN, predicts online car buying when combined with information on task completion. The 95% probability interval of the posterior means of the population level parameters for RETURN is $[0.120, 0.375]$ for NUT2, and $[-0.566, -0.210]$ for NUT3. The positive sign for task two indicates that visitors are more likely to complete NUT2 (and input their complete personal information) if they have previously exited the site and subsequently return, both following the completion of NUT1. On the other hand, the negative sign for task three indicates that visitors are less likely to purchase (complete task three) if they have left the site after completing task two. Table 5 shows that these are substantial effects. The average attributable risk (the average change in task completion probability) across all regions is 0.035 and -0.048 for NUT2 and NUT3, respectively.

The effects of RETURN on NUT2 and NUT3 have useful managerial implications. First, site exits, per se, do not always bring bad news (at least for the online purchase of new cars which may require extensive offline information search). Even if visitors exit the site after completing their first full car configuration (NUT1), that does not imply a lower purchase probability. Indeed, in this case it raises the probability of online purchase. In contrast, after visitors input their personal information (NUT2), a site exit becomes a negative indication even when that user returns. At this point, however, the company has the user's contact information, so individual targeting via email, direct mail, or telephone, becomes possible. Users could be ranked by predicted purchase probability (using all of the information up to their last site exit) and targeted accordingly. Of course, the success of such efforts also depends upon their ability to raise purchase probabilities sufficient to cover the costs of the additional marketing activities.

Use of Decision Aids. The implementation of electronic agents and interactive tools to aid consumer decision making has been a major consideration in the development of good online shopping environments (West et al 1999). Häubl and Trifts (2000) found positive effects from the use of decision aids on consumer's search efforts, size and quality of consideration sets, and quality of purchase decisions. In their experiments, subjects were required to make purchase decisions. Previous research has not yet addressed whether the use of decision aids actually increases the likelihood of online purchasing. The Web site we studied had implemented a comparison matrix tool that allowed visitors to select cars for side-by-side comparison (management called it "the comparator"). We find that a visitor's use of the car comparison tool, DECAID, is associated with a lower probability of completing a full car configuration (NUT1), and that it is not predictive of subsequent task completion (the 95% probability interval of the posterior means of the population level parameters for DECAID is $[-0.645, -0.544]$ for NUT1). Not only was the effect of the decision tool negative in this case, but the effect size is substantial. The average attributable risk across regions is -0.089 which corresponds to an average 58.2% decrease in task completion probabilities.

There may be two reasons for this result. First, those visitors who actually used the comparator may also be those with lower *a priori* purchase intentions. Second, the comparator may have lowered satisfaction with the site leading users to fail to complete a full product configuration. While our model cannot distinguish between the two, additional analyses may provide clues. About 16% of site visitors used the comparator before ever completing NUT1. However, these visitors used the comparator an average of only three times before ever completing NUT1 (or before leaving the site for good). If the comparator was well designed and used in aiding decision making (either to buy online or elsewhere) we would expect it to be used intensively by those who elected to use it in their decision process – especially this early in the task sequence. Instead, we observe that visitors seldom use the feature and, when they do, they use it few times. Visual inspection of the comparator also revealed several design flaws. For example, the limited space allowed for the side-by-side display of car specifications made car comparisons difficult and effortful. Therefore, it is possible that the specific implementation of this decision aid is producing the negative effect on conversion probabilities.

Our finding that offering decision aids is not a guarantee of increased conversion, and in fact negatively affects it in this case, suggests that experimental research to investigate the effect of decision aid implementation on site usage and task completion may be worthwhile. Even if use of a decision aid like the comparator is simply associated with reduced conversion probabilities for other reasons, the company could still use the comparator page to place targeted banner ads and obtain revenue from selling banner impressions. For example, the company could sell to car manufacturers placement of targeted ads given the cars visitors configure or compare.

Input Effort and Information Gathering. Our goal for this set of variables is to investigate whether the way visitors search for product information and interact with the site (e.g., by inputting data, by requesting information) reveals anything about their purchase likelihood. To capture user input effort, we use the average interactivity level of the pages requested by each visitor, INTERACT. To capture the effects of information search and processing, we defined three variables: (1) the average time visitors spent viewing each page, PAGETIME, (2) the average number of words on the pages visitors requested, WORDS, and (3) a dummy variable that indicates whether information on the company pricing policies has been accessed by the user, PRICING (e.g., what are the guarantees provided by the site, what does the price include, etc.).

The variable INTERACT correlates positively with the completion of NUT1 and NUT2 (the 95% probability interval of the posterior means of the population level parameters for INTERACT is $[1.738, 1.872]$ for NUT1 and $[0.979, 1.142]$ for NUT2). On the other hand, input effort is not predictive of the completion of NUT3. This result indicates that users who expended more effort prior to the first two tasks were more likely to purchase. (The insignificant effect for the third task may be due to the fact that effort levels are more similar across users than for the first two tasks.)

The three measures for information gathering (WORDS, PAGETIME and PRICING) are not predictive of the completion of NUT3, but are predictive for NUT1 and NUT2. Indeed, knowing the type of information visitors access while browsing the site may help identify users who are purchase motivated from those who are not. Our results indicate that visitors who access information on pricing policies before completing a full car configuration for the first time (NUT1) are less likely to ever complete the first task, whereas visitors who have completed NUT1 and then access the same type of information, are more likely to complete NUT2 (input of personal information). The 95% probability interval of the posterior means of the population level parameters for PRICING is $[-0.561, -0.478]$ for NUT1 and $[0.250, 0.457]$ for NUT2. The posterior mean of the attributable risk for PRICING is -0.075 for NUT1 and 0.050 for NUT2, on average across all counties. It is possible, then, that accessing pricing policies early in the task sequence reveals concerns with online buying but, later in the task sequence, it reveals commitment.

The longer visitors spend viewing each page (PAGETIME), and the more information is contained in the pages they requested from the site (WORDS), the more likely they are to complete a full car configuration, and the less likely they are to complete personal information forms (after fully configuring their first car). Before a first full car configuration is completed, visitors motivated by purchase may gather more information and take longer to process it compared to non-motivated users (i.e, those that never complete NUT1). In contrast, after a first car is fully configured, visitors that are still motivated (and, therefore, complete NUT2) seem to gather less information and take less processing time.

Exposure Control Variables. Exposure to page specific characteristics, download time, and site design and structure may also influence purchase decisions through their impact on Web site evaluations by site visitors. For example, Mandel and Johnson (2002) show that page design can change preferences, and hence purchasing decisions, by influencing attribute importance. Dellaert and Kahn (1999) show that when there is uncertainty about the online waiting experiences (as with the majority of page downloads), the negative feelings generated by the waiting experience carry over into the evaluation of the Web site. To control for factors such as these, we defined five variables at each NUT level (CLUTTER, DYNAMIC, FIGURES, LINKS, and PAGESIZE). These variables capture some of the differences in online experiences due to download time, aesthetic appeal, and the design of the pages visitors requested before completing each task. (Please see Appendix C for a full description.)

We find that four of the five variables are significantly related to the completion of task one, all are related to the completion of task two, while none are related to the completion of task three. CLUTTER has a sign reversal between NUT1 and NUT2, indicating that exposure to more cluttered pages is negatively associated with car configuration while positively associated with completion of personal information. DYNAMIC has a significant squared term for NUT2 and the corresponding elasticity is negatively signed for both NUT1 and NUT2. Greater user exposure to pages with more dynamic content is therefore negatively related to the completion of the first two site tasks. Similarly, exposure to greater numbers of figures is also negatively related to task completion. On the other hand, exposure to more links is positively related to task completion. In sum, these exposure control variables were successful in capturing variation in purchase probabilities due to idiosyncratic elements of site exposure. The implications for site design are circumscribed, however, because realized values for these variables result from the page view choices made by individual users.

MODEL VALIDATION

We assessed predictive performance using out-of-sample tests. We generated a holdout sample by randomly selecting one-third of the original records (32,166 visitors). The remaining two thirds (64,332 visitors) were used to re-estimate the prediction models for each NUT. Table 6 presents summary statistics for the estimation and the holdout samples. Both samples have conversion rates very close to the original dataset. Due to the random assignment of each visitor to the holdout sample, some holdout visitors were associated with counties not observed in the estimation sample (169, 228, and 135 visitors for NUT1, NUT2 and NUT3, respectively). We discarded these visitors and performed the prediction tests on the remaining records.

We evaluated three prediction models. The first model, MULTI, is our task completion approach. Using the new estimation sample, we re-estimated our best fitting models for each task level. The new sets of parameters were then used to generate predictions. The second model, SINGLE1, is a single task model of buying behavior. This is a hierarchical probit model for the prediction of online orders. It accounts for heterogeneity at the county level, incorporates visitors' actions and exposure to Web site design and structure, but it does not include information on task completion. We estimated the model on the new estimation sample and selected the final formulation for this model based on pseudo-Bayes factors. The third model, SINGLE2, is the SINGLE1 model with two additional covariates. These are dummy variables that represent the completion of the first and second NUTs, respectively. This model allows us to test whether or not it is the structure of the multilevel approach—and not simply the information contained in task completion—that enhances prediction. (Both task dummies provided pseudo-Bayes factor improvements.)

We use each model to predict vehicle orders before the completion of *each* task. This enables us to compare how the models perform as users progress through the site and more information about each visitor becomes available. For our multilevel approach, to predict orders prior to the completion of NUT1 and/or NUT2, we need to scale down the task completion probabilities. To do this, we can use either the conversion rates observed in the estimation sample for the corresponding counties or estimate county-specific conversion rates with an intercept-only probit model. In this case, we chose to use the intercept-only probit because it shrinks the conversion rates to the mean for counties which have few observations.

Figures 2a, 2b, and 2c present lift charts for each model prior to the completion of each NUT. The data in the charts were tabulated by sorting the purchase probabilities for all holdout visitors, as predicted by the models. We then took the 10% of all (holdout) Web site visitors with the highest predicted probability and determined how many of these visitors ordered a vehicle. We repeated this procedure for 20% of the visitors, 30%, and so on. We then computed and plotted the fraction of online purchases that each model would have been able to capture at the different targeting percentages. Each graph also shows the expected performance of randomly sampling the visitors (CHANCE).

Our proposed modeling approach, MULTI, outperformed both the SINGLE1 and SINGLE2 models in terms of lift (the lift lines corresponding to the MULTI model are always above all others). For example, Figure 2a shows that by targeting the best 20% of all holdout Web site visitors before the first task (NUT1) is ever completed, we are able to capture about 56% of online buyers if we use the MULTI approach. If we use the SINGLE1 and SINGLE2 approaches we are able to capture only 38% and 34% of buyers, respectively. The MULTI approach also dominates the alternative models before the completion of NUT2 and NUT3. We note that SINGLE1 and SINGLE2 are not easy-to- beat models; they perform well when compared to chance. By comparing the performance of the SINGLE2 model with the performance of MULTI and SINGLE1 we also find that it is not the knowledge of task completion *per se* that is allowing the MULTI model to do well. It is the combination of model structure *and* task completion information that enables the MULTI approach to more accurately predict who is going to buy from the site. (Interestingly, though SINGLE2 uses more information than SINGLE1, in out-of-sample prediction this information either hinders or does not change performance.)

We also assessed the performance of the MULTI model versus the same model without county-level heterogeneity in the parameters. The model without heterogeneity was also estimated using the new estimation sample of two-thirds of the original visitors. Figures 3a, 3b, and 3c present lift charts for MULTI and the model with no heterogeneity computed on the holdout sample visitors. The lift charts show that incorporating county-level heterogeneity yields improvements in the predictive power of the model, especially for the case prior to NUT1 (Figure 3a).

To further evaluate the predictive performance of the proposed approach, we also report several traditional benchmarks for predictive accuracy. These are mean squared error (MSE), the overall predictive hit-rate (based on correct predictions of orders and non-orders), and the hit-rate computed solely for vehicle orders. For both hit-rates we used a 50% cut-off value.[11] Table 7 presents the predictive performance for the various models computed prior to each task. The task completion approach (MULTI) performs better than the SINGLE1 model for all performance metrics. MULTI is also better than the SINGLE2 model in all cases except that SINGLE2 edges out MULTI on overall hit-rate and the non-orders hit-rate before NUT2. Note that this follows from the fact that SINGLE2 predicts that no visitor will order online before the completion of NUT2; this raises the overall and non-order hit-rates at the expense of the order hit-rate.

[11] A 50% cut-off value is intuitively appealing but may not be optimal. The advantage of lift charts is that they implicitly compare model performance for multiple cut-off points.

We believe that the hit-rate for vehicle orders is the best model comparison for managers to use. First, online purchases are low probability events. With a conversion rate of 2.1%, if we measure predictive accuracy based on overall hit-rates, a "very good" predictive model would simply predict that no purchase will ever occur. We would then obtain an overall hit-rate of 97.9%, even though the "model" provides the company with no useful information. In addition, a correctly predicted purchase has much greater value than (1) a correctly predicted no-purchase (true negative), and (2) an incorrectly predicted purchase (false positive). Companies are likely to gain more from the prediction of purchases versus the prediction of no-purchases. Though MULTI provides a seemingly small absolute improvement in order hit-rate for NUT1 and NUT2, the gains are meaningful because of the low probabilities of online ordering and the demonstrated improvement in targeting potential provided by the multilevel modeling approach (e.g., as evidenced by the lift charts).

The results reported in Table 7 also show that the predictive performance of all three models improves as more information becomes available about each user. Early in the task sequence, before a full car configuration has been completed, the hit-rate for orders is quite low. However, the MULTI model is able to better predict car orders (0.3% against 0.1% for the SINGLE models). As users navigate the site and complete tasks, this provides additional information that can be used to predict purchase likelihood. By the last stage, all models show that this information yields better predictions. The MULTI model performs very well (and much better than the single models) at the end of the sequence task, after visitors have inputted the personal information (i.e., following NUT2).

Discussion

The strong performance of the multi-task approach highlights its potential as a managerial tool. As we noted above, the company can predict the purchase likelihood for each visitor and then decide on targeted promotions or communications. For example, pop-up ads constitute one of the most profitable types of Web advertising, but they are also particularly disruptive. We illustrate how the company could the use the proposed approach to better allocate exposures to pop-up ads, say from car manufacturers, to minimize their impact on the company's main business. We assume that the company has estimated the multi-task model on the estimation sample and will target visitors from the holdout sample (as before, visitors from counties not observed in the estimation sample are removed). To keep the example simple we assume the media buyer wants to show one pop-up per unique user. The company has two possibilities: (1) randomly assign pop-up exposures, or (2) show pop-ups to visitors with the lowest predicted purchase probability. If visitors who have not yet completed NUT1 are sorted with respect to purchase probability, then the company would end up showing pop-up ads to one actual buyer for the first 1,000 or 2,000 exposures, and to three actual buyers for the first 3,000 exposures (versus 21, 42 and 63, respectively, with random assignment). With targeted exposures for visitors who have completed NUT1, but have not yet completed NUT2, the company would show pop-ups only to non-purchasers for the first 1,000 exposures, to two purchasers for the first 2,000 exposures, and to three purchasers for the first 3,000 exposures (instead of 69, 138 and 207 actual car buyers, respectively, with random exposures).

The benefits of the proposed approach can also be seen if we compare the expected loss in gross margin from the two alternatives. We assume that the average gross margin per car is $1,500 and that a pop-up exposure reduces purchase probability due to lower satisfaction and possible site exits. Table 8 shows the expected loss of gross margin for the holdout sample under the random and targeted exposure scenarios for 2% and 0.5% purchase probability reduction. The targeted exposure scenario is better across all conditions; it is also the only potentially profitable one if the company could not charge more than about $2 per thousand pop-up exposures (the current average cost for this advertising). Hence, pop-up ad exposures might profitably be sold if the company does not expose visitors who have already completed NUT2 (before NUT3 in Table 8), *and targets* visitors who are the least likely to buy online.

CONCLUSION

The goal of this research has been to build a predictive model of online buying by visitors to a commercial Web site. We identified three major challenges in predicting online buying: (1) a potential lack of predictive and explanatory power due to very low conversion rates for online purchasing, (2) the difficulty of effectively accounting for what Web users do, and to what they are exposed, while browsing the site, and (3) the need to account for a diverse user population and diversity of competitive environments faced by online stores.

We proposed a conditional probability approach in which we decompose the activities at a Web site into *sequential* Nominal User Tasks (NUTs) that must be completed for a purchase to take place. Instead of predicting buy or no buy given a site visit, we predict the probability of completion for each task that is required for purchase, in sequence. We identified three NUTs for the e-commerce Web site under study: (1) completion of product configuration, (2) input of complete personal information, and (3) order confirmation with provision of credit card data. These tasks were both managerially meaningful and corresponded to considerable, but not excessive, visitor loss. Because each NUT, when conditioned on the preceding one (if applicable), is associated with higher completion rates, the model is not forced to predict rare events. The resulting binary probit specification jointly models the first time each user completes each task, given that the user has started a site visit and completed the preceding tasks (if applicable).

To account for what visitors do at the site, we included covariates for browsing behavior, repeat visitation, use of decision aids, input effort and information gathering and processing. We also constructed measures of exposure to site design and structure. These incorporate the visitor's idiosyncratic experience at the site (i.e., the set of pages requested) and the page-specific characteristics.

We applied our task completion model to clickstream data provided by a major commercial Web site in the automotive industry. While we present an application to a specific site, the general idea behind our approach should be widely applicable. Conditioning the choice model on intermediate events prior to final purchase may be applied to any e-commerce site that requires users to complete certain tasks, in sequence, prior to purchase. Placing items in a shopping cart, completing shipping and billing information, and providing credit card information are sequential tasks common to a wide variety of consumer oriented e-commerce sites. In our case, the first task of vehicle configuration corresponds to a more elaborate set of activities than simple shopping cart placement. Nevertheless, product customization and specification by the user also characterizes e-commerce sites offering computers, electronics, apparel, and financial services (e.g., insurance, mortgages, home equity loans).

Because the site we analyzed sells new cars, we observed at most one purchase per visitor. To account for the heterogeneity in Web site user population and the multitude of competitive environments this online retailer faced, we allowed for county-specific parameters through a random-effects specification. The company management believed a county-level specification would account for the regional nature of competition and pricing decisions. We also tested other regional specifications—state and Zip code level—as well as a homogeneous model and found that the county-level approach performed best.

Estimation and prediction results showed that significant variation in task completion behavior could be explained by what visitors do while browsing the site and by their exposure to page specific characteristics. Our results confirmed that the proposed conditioning approach significantly reduced error when predicting whether new Web site visitors were, or were not, going to buy online (even when little information was available, i.e., early in the task sequence). Using a holdout sample of site visitors, we also tested the proposed model against two single-stage benchmark models. We found that the multi-level task completion approach provided superior targeting capability (as evidenced by lift charts) as well as better measures of predictive accuracy (mean squared error and hit rates).

In addition to improvements in prediction, our results also shed light on the relationship between task completion and predictor variables capturing visitors' browsing behavior, use of the site, and exposure to page characteristics. These relationships can be useful diagnostic tools in the evaluation of site performance. In many instances, we found that the same covariate had a positive effect on completion for one task, but a negative effect on another task (or was not predictive at all). Thus, combining the information on task completion with visitors' activities on the site (e.g., type of information accessed), can help us better distinguish purchase motivated users from others. Changes in effect signs and/or effect sizes could not be revealed in a single stage model of Web site purchase where model coefficients do not change as users proceed through tasks. For example, a site exit and return is not predictive of completing task one, it is positively associated with completing the second, and negatively associated with completing the final task. Accessing pricing policy information is negatively related to the completion of task one, but positively related to the completion of task two. Use of the decision aid on the site is negatively related to completing the first task, but has no effect on completing the other tasks. Additional reversals occur for several other model covariates.

From a substantive perspective, our findings suggest that the more time and effort visitors invest in the site, the more likely they are to eventually buy at the site (as evidenced by the positive effect of total time spent and the user's input effort on interactive pages). In contrast to Moe and Fader (2003), who study the online purchase of books and CDs, we obtain no relationship between the extent of repeat site visitation and new car purchase. We also find that the offering of a decision aid at the site does not increase conversion rates, as previous research might suggest (Haubl and Trifts 2000). Indeed, there is a negative effect of decision aid use on the completion of task one. Examples like these suggest many avenues for further research on consumer behavior in e-commerce settings, using clickstream data, experiments, or both.

In addition, this study offers several implications for Web site managers. For example, our results suggest that the company should contact (e.g., by email) visitors who have exited the site without ordering if they have inputted personal information. This is because such visitors are less likely ever to order on their own. Also managers may consider increasing advertising exposure for those visitors who remain in intensive search mode after completing the first task (e.g., continue to request many pages, spend a long time per page, and request pages that are information intensive). By "personalizing" the ad content of the site, managers can augment their revenues with the sales of ad impressions (or clicks) and, at the same time, minimize any interference with the main line of business. In Table 8, we provided a specific example of the potential use of the model for pop-up ads.

There are a number of limitations to our work. First, we did not model the number of times each task is completed by the visitor, nor did we incorporate all possible information about the content of the Web pages viewed. For example, the data do not tell us which specific vehicle a user has configured to complete NUT1, so we cannot incorporate measures of product variety search. We also did not investigate the effects of promotional campaigns nor did we directly account for the actions of competitors. To account for such factors, a more complete dataset (e.g., "single source") would be required. Instead, our approach can be applied to the data companies already collect using clickstream tracking software. Also, we have analyzed only the buying decision for new cars. As a result, some of our substantive findings with respect to browsing and exposure variables may not hold for the purchase decisions of small-ticket items that are bought more frequently. Finally, our random effects formulation does not take into account the geographical proximity of the different regions (Bronnenberg and Sismeiro 2002). If significant spatial correlation exists in our data, incorporating geographical structure into our model of task completion would allow us to better predict the behavior of visitors from counties we have previously observed. In addition, we would be able to predict the behavior of visitors from counties that were never previously observed. Developing new models to address these issues is worthwhile for future research.

Appendix A: Priors

For each probit model we specify Normal conditional priors for the between-subject parameters (μ_{mk}, $k = 1, \ldots, p_m$, $m = 1, \ldots, M$) with zero mean, and Inverse Gamma priors for their conditional variances $\left(\sigma^2_{mk},\ k = 1, \ldots, p_m,\ m = 1, \ldots, M\right)$:

$$\mu_{mk} \sim \mathcal{N}\left(a_0, a_1\right) \text{ and } \sigma^2_{mk} \sim \mathcal{IG}\left(\frac{s_0}{2}, \frac{s_1}{2}\right) \text{ for , } k = 1, \ldots, p_m,\ m = 1, \ldots, M.$$

Both the setting of prior hyperparameters and the nature of the prior distribution have the potential to influence the posterior distribution of the individual-specific parameters. In practice, we take very diffuse priors to induce a mild amount of shrinkage. Therefore we take $a_0 = 0$, $a_1 = 50$, $s_0 = 2$ and $s_1 = 2$. These represent diffuse (non-informative) but proper distributions.

Appendix B: Conditionals and Simulation Algorithm

B.1. Independent Probit

1. Set starting values for the unknown parameters.

2. Simulate the utilities, y^m_{is} for $m = 1, \ldots, M$, $i = 1, \ldots, N_s$, and $s = 1, \ldots, S_m$, given the parameter values:

$$\begin{aligned} y^m_{is} \,|\, \mathbf{X}^m, \boldsymbol{\beta}^m &\sim \text{Truncated } \mathcal{N}\ \left(\mathbf{x}^m_{is}\boldsymbol{\beta}^m_s, 1\right) \\ \text{s.t. if } C^m_{is} &= 1 \text{ then } y^m_{is} \geq 0 \\ \text{if } C^m_{is} &= 0 \text{ then } y^m_{is} < 0. \end{aligned}$$

3. Draw the individual specific utility parameters, β^m_{ks} for $m = 1, \ldots, M$, $k = 1, \ldots, p_m$, and $s = 1, \ldots, S_m$, from the following conditional posteriors:

$$\begin{aligned} \beta^m_{ks} \,|\, \boldsymbol{\beta}^m_{-ks}, \mathbf{Y}^m, \mathbf{X}^m &\sim \mathcal{N}\left(\overline{\mu}_{mks}, \overline{\sigma}^2_{mks}\right) \\ \overline{\sigma}^2_{mks} &= \left(\sum_{i=1}^{N_s} \left(x^m_{kis}\right)^2 + \frac{1}{\sigma^2_{mk}}\right)^{-1} \\ \overline{\mu}_{mks} &= \overline{\sigma}^2_{mks}\left(\sum_{i=1}^{N_s} x^m_{kis} * \overline{y}^m_{kis} + \frac{\mu_{mk}}{\sigma^2_{mk}}\right) \\ \overline{y}^m_{kis} &= y^m_{is} - \sum_{l \neq k} \beta^m_{ls} x^m_{lis}. \end{aligned}$$

4. Draw the between-subject means, μ_{mk} for $k = 1, \ldots, p_m$ and $m = 1, \ldots, M$, from the following conditional posterior distributions:

$$\begin{aligned} \mu_{mk} \,|\, \boldsymbol{\beta}_{mk}, \sigma^2_{mk} &\sim \mathcal{N}\left(\overline{a}_{0mk}, \overline{a}_{1mk}\right) \\ \overline{a}_{1mk} &= \left(\frac{S_m}{\sigma^2_{mk}} + a_1^{-1}\right)^{-1} \\ \overline{a}_{0mk} &= \overline{a}_{1mk}\left(\sum_{s=1}^{S} \frac{\beta^m_{ks}}{\sigma^2_{mk}} + \frac{a_0}{a_1}\right). \end{aligned}$$

5. Draw the between-subject variances, σ^2_{mk} for $k = 1, \ldots, p_m$ and $m = 1, \ldots, M,$, from the following conditional posterior distributions:

$$\begin{aligned}
\sigma^2_{mk} \,|\, \boldsymbol{\beta}_{mk}, \mu_{mk} &\sim \mathcal{IG}\left(\frac{\overline{s}_{0mk}}{2}, \frac{\overline{s}_{1mk}}{2}\right) \\
\overline{s}_{1mk} &= s_1 + S_m \\
\overline{s}_{0mk} &= s_0 + \sum_{s=1}^{S} \left(\beta^m_{ks} - \mu_{mk}\right)^2 .
\end{aligned}$$

6. Repeat steps 2–5.

B.2. Dependent Probits

To test for selective bias among the different task levels, we allow the individual propensity of completing each task to depend on the utility, of preceding tasks (if any), y^{-m}_{is}, through the function $g(\cdot)$ and the $(p_{2m} \times 1)$ vector of parameters $\boldsymbol{\beta}^m_{2s}$. We have chosen to use a n-order polynomial specification for $g\left(\mathbf{y}^{-m}_{is}, \boldsymbol{\beta}^m_{2s}\right)$ with no intercept (for identification purposes). Hence we define our vector of covariates as $\mathbf{x}^m_{is} = \left[\begin{array}{cc} \widetilde{\mathbf{x}}^m_{is} & \mathbf{y}^{-m}_{is} \end{array}\right]$ where $\mathbf{y}^{-m}_{is}$ is a $(1 \times p_{2m})$ vector that contains the powers of the utilities of preceding tasks; $\widetilde{\mathbf{x}}^m_{is}$ is a $(1 \times p_{1m})$ vector of covariates. $\boldsymbol{\beta}^m_{is} = \left[\begin{array}{cc} \boldsymbol{\beta}^{m\prime}_{1s} & \boldsymbol{\beta}^{m\prime}_{2s} \end{array}\right]'$ is now a $(p_m \times 1)$ vector of parameters where $p_m = p_{1m} + p_{2m}$. We believe that such specification can adequately capture the nature of the dependence among the different task levels as it can be seen as an n^{th} order approximation to a nonlinear functional relationship.

To obtain unbiased, but not efficient, posterior results we can simply use the MCMC steps as described previously. Because the full conditionals for all tasks levels except the last one are now not known, to obtain efficient results we apply a Adaptive Rejection Sampling algorithm to draw for the individual specific parameters, $\boldsymbol{\beta}^m_{is}$ (Gilks and Wild 1992).

Appendix C: Description of Model Variables

Each covariate was computed for all tasks ($m \in \{1, 2, 3\}$), all visitors ($i = 1, \ldots, N_s$), and all regions ($s = 1, \ldots, S_m$). For NUT2 and NUT3 we used the visitor's behavior from the moment s/he completes the preceding task (NUT1 and NUT2, respectively). If it is a cumulative variable, we reset the counter to zero when the visitor completes the preceding task. If the covariate is a weighted average of page characteristics, we use only the visitor's page visitation pattern after the previous task has been completed. For NUT1 all covariates are defined from the moment each visitor starts the first site session.

D.1. Browsing Covariates (for each task m, $m \in \{1, 2, 3\}$)

$TIME^m_{is}$: Total time visitor i from county s spent browsing the site after completing the previous task, if applicable. (The duration of the last page view of a session is unrecorded in the clickstream. The time at the site is therefore the sum of all page view durations up to the last page view of a site visit.)

$TPAGE_{is}^{m}$: Total number of pages requested by visitor i from county s after completing the previous task, if applicable.

D.2. Repeat Visitation (for each task m, $m \in \{1,2,3\}$)

$RETURN_{is}^{m}$: Dummy for a return session; for NUT2 and NUT3 this dummy variable takes the value 1 if visitor i from county s has exited the site (terminated a session) and returned after completing the previous task (and takes the value 0 otherwise); for NUT1 this dummy variable takes the value 1 if visitor i from county s has exited the site (terminated a session) and returned after starting his first site visit (and takes the value 0 otherwise).

$TSESSION_{is}^{m}$: Total number of sessions made by visitor i from county s. A page requested by a visitor starts a new session if it is requested after an idle period of 30 minutes or more (Catledge and Pitkow 1995).

D.3. Use of Decision Aids (for each task m, $m \in \{1,2,3\}$)

$DECAID_{is}^{m}$: Dummy for the use of the decision aid; for NUT2 and NUT3 this dummy variable takes the value 1 if visitor i from county s has used the car comparison matrix at least once after completing the previous task (and takes the value 0 otherwise); for NUT1 this dummy variable takes the value 1 if visitor i from county s has used the decision aid at least once after starting his first site visit (and takes the value 0 otherwise).

D.4. Input Effort and Information Gathering (for each task m, $m \in \{1,2,3\}$)

$INTERACT_{is}^{m}$: Average interactivity level for the pages requested by visitor i from county s. Each page was rated by two independent judges from 1 (very low interactivity) to 5 (very high interactivity). The variable is a weighted average based on the pages requested by the visitor. Pages rated low in interactivity are pages with no or few links and with no input requirements. Pages rated very high in interactivity require a high level of input (e.g., complex forms) and an intensive processing effort by the user.

$PRICING_{is}^{m}$: Dummy for the request of the pricing policies page; for NUT2 and NUT3 this dummy variable takes the value 1 if visitor i from county s has requested the pricing policies page at least once after completing the previous task (and takes the value 0 otherwise); for NUT1 this dummy variable takes the value 1 if visitor i from county s has requested the pricing policies page at least once after starting his first site visit (and takes the value 0 otherwise).

$PAGETIME_{is}^{m}$: Average view duration per page for visitor i of county s, (computed after completing the previous task, if applicable).

$WORDS_{is}^{m}$: Average number of words present on the pages requested by visitor i from county s (computed after completing the previous task, if applicable).

D.5. Exposure Control Covariates (for each task m, $m \in \{1, 2, 3\}$)

$CLUTTER_{is}^{m}$: Average level of clutter for the pages requested by visitor i from county s. Each page was rated by two independent judges from 1 (low clutter) to 3 (high clutter). A weighted average is then computed based on the pages requested by the corresponding visitor.

$DYNAMIC_{is}^{m}$: Percentage of pages requested by visitor i from county s that require online access to a database. Pages that require access to a database are created dynamically online. These pages provide the information requested by the user (e.g., the monthly payments for the loan given certain specifications selected by the user) and therefore take longer to load.

$FIGURES_{is}^{m}$: Average number of figures for the pages requested by visitor i from county s.

$LINKS_{is}^{m}$: Average number of links for the pages requested by visitor i from county s.

$PAGESIZE_{is}^{m}$: Average level of page size for the pages requested by visitor i from county s. Each page was rated by two independent judges from 1 (short page)to 3 (long page). A weighted average is then computed based on the pages requested by the corresponding visitor.

Appendix D: Variable Selection

Bayes Factors are widely used to make comparisons among statistical models. Given the complexity of our model, however, Bayes Factors are computationally nontrivial. We therefore follow Gelfand (1996) and use the cross-validation predictive densities and the *pseudo-Bayes factor* (PSBF) to perform model selection.

The pseudo-Bayes factor for Model 1 (M_1) with respect to Model 2 (M_2) is given by

$$PsBF_{12} = \frac{\prod_{r=1}^{NT} d_r\left(M_1\right)}{\prod_{r=1}^{NT} d_r\left(M_2\right)},$$

where $d_r\left(M_i\right)$ is the cross-validation predictive density of model i for the r^{th} observation and NT is the total number of observations. The Monte Carlo estimate of d_r for a given model M_i can be approximated by

$$\widehat{d}_r\left(M_i\right) = B\left(\sum_{s=1}^{B} \frac{1}{f\left(u_r \mid \Theta^{(s)}, M_i\right)}\right)^{-1},$$

where $f\left(u_r \mid \Theta^{(s)}\right)$ is the likelihood contribution of the r^{th} observation given the s^{th} parameter draw of the Gibbs sampler, $\Theta^{(s)}$, B is the number of retained MCMC simulations after a burn-in period and thinning of the chain, and u_r is the dependent variable under analysis. In the case of the choice model, $f\left(u_r \mid \Theta^{(s)}\right)$ corresponds to the probit likelihood. We have performed all the computations in their logarithmic form, to achieve better numerical accuracy. We have then used the log of the pseudo-Bayes factor for model comparison.

References

Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (March-April), 57–78.

Bell, David R. and James M. Lattin (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket" Shoppers Prefer EDLP," *Marketing Science*, 17 (1), 66–88.

Bell, David R., Jeongwen Chiang, V. Padmanaban (1999), "The Decomposition of Promtional Response: An Empirical Generalization," *Marketing Science*, 18 (4), 504-526.

Bronnenberg, Bart J. and Catarina Sismeiro (2002), "Using Multimarket Data to Predict Brand Performance in Markets for Which No or Poor Data Exist," *Journal of Marketing Research,* 39 (1), February, 1–17.

Bucklin, Randolph E. and Catarina Sismeiro (2003), "A Model of Web Site Browsing Behavior Estimated on Clickstream Data," *Journal of Marketing Research, 40 (3),* August, 249-267.

Catledge, Lara D. and James E. Pitkow (1995), "Characterizing Browsing Behaviors on the World Wide Web," *Computer Networks and ISDN Systems*, 27(6), 1065–1073.

Dellaert, Benedict and Barbara E. Kahn (1999), "How Tolerable is Delay? Consumers' Evaluations of Internet Web sites After Waiting," *Journal of Interactive Marketing*, 13 (1), Winter, 41–54.

Emarketer (2000), www.emarketer.com.

Forrester Research (1999), www.forrester.com.

Forrester Research (2001), www.forrester.com.

Gelfand, A.E. (1996), "Model Determination Using Sampling Based Methods," in Gilks, W.R., S. Richardson, D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 145–162.

Gelfand, A.E. and A.F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelfand, A.E., S.E. Hills, A. Racine-Poon and A.F.M Smith (1990), "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *Journal of the American Statistical Association*, 85 (June), 972–985.

Gilks, W. R. and P. Wild (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41 (2), 337–348.

Gupta, Sunil (1988), "Impact of Price Promotions on When, What and How Much to Buy," *Journal of Marketing Research,* 25 (November), 342-356.

GVU (1998) http://www.gvu.gatech.edu/user_surveys/

Häubl, Gerald and Valerie Trifts (2000), "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science*, 19 (1), Winter, 4–21.

Johnson, Eric J., Steven Bellman, Gerald L. Lohse (2003), "Cognitive Lock In and the Power Law of Practice," *Journal of Marketing*, 67 (April), 62-75.

Montgomery, Alan, Shibo Li, Kannan Srinivasan, and John C. Liechty (2003), "Modeling Online Browsing and Path Analysis Using Clickstream Data," working paper, Graduate School of Industrial Administration, Carnegie Mellon University.

King, Gary and Langche Zeng (2001), "Logistic Regression in Rare Events Data," *Political Analysis*, 9 (2), February, 137–163.

Mandel, Naomi and Eric J. Johnson (2002), "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," *Journal of Consumer Research*, 29 (2), 235-245.

Manski, Charles F. (2001), "Nonparametric Identification Under Response-Based Sampling," in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, eds. C. Hsiao, K. Morimune, and J. Powell, New York: Cambridge University Press.

Moe, Wendy (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstreams," *Journal of Consumer Psychology* 13 (1-2), 29-40.

Moe, Wendy, Hugh Chipman, Edward I. George, and Robert E. McCulloch (2003), "A Bayesian Treed Model of Online Purchasing Using In-Store Navigational Clickstream," working paper, McCombs School of Business, University of Texas, Austin.

Moe, Wendy W. and Peter S. Fader (2003), "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science*, forthcoming.

Rebane, George J. (2001), "COSMO: Overview and Technical Approach," Technical Report #TR0103-1, Bizrate.com.

Tanner, Martin A. and Wing H. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (June), 528–549.

Totty, Michael (2002), "So Much Information...," *Wall Street Journal*, December 9, R4.

Van Heerde, Harald J., Sachin Gupta, and Dick R. Wittink (2003), "Is 3/4 of the Sales Promotion Bump Due to Brand Switching? No, it is 1/3" *Journal of Marketing Research,* forthcoming.

West, Patricia M., Dan Ariely, Steve Bellman, Eric Bradlow, Joel Huber, Eric Johnson, Barbara Kahn, John Little, and David Schkade (1999), "Agents to the Rescue?" *Marketing Letters*, 10 (3), 285–300.

Table 1
Summary Statistics for Nominal User Task Attrition Rate

Tasks	Total Users	Users Completing Task	Percentage of Users Completing Task	Counties
Complete Product Configuration (NUT1)	96,498	29,238	30.30	2,045
Input Complete Personal Information (NUT2)	29,238	5,753	19.68	1,446
Order Confirmation with Provision of Credit Card Data (NUT3)	5,753	1,969	34.23	689

Table 2
Summary Statistics for Covariates

	NUT1		NUT2		NUT3	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Browsing Covariates						
TIME (minutes)	24.44	33.78	16.05	33.89	12.57	30.37
TPAGE	14.92	18.30	11.13	21.06	7.77	16.75
Repeat Visitation						
RETURN	0.33	0.47	0.31	0.46	0.26	0.44
TSESSION	1.78	1.82	2.34	2.86	2.71	3.24
Use of Decision Aids						
DECAID	0.16	0.36	0.03	0.18	0.02	0.14
Input Effort and Information Gathering						
INTERACT	3.30	0.27	3.58	0.52	4.43	0.71
PRICING	0.09	0.28	0.05	0.21	0.03	0.18
PAGETIME (minutes)	1.71	1.29	1.25	1.48	1.65	2.45
WORDS	75.92	47.93	48.88	49.37	161.99	56.32
Control Variables						
CLUTTER	1.55	0.15	1.40	0.21	1.81	0.24
DYNAMIC	0.60	0.19	0.84	0.19	0.88	0.19
FIGURES	5.63	0.87	4.03	1.12	3.69	1.11
LINKS	7.49	1.44	7.19	1.11	7.82	0.95
PAGESIZE	1.95	0.09	1.64	0.31	1.96	0.13

Table 3
Model Specifications and Fits*

	Null 1	Full 1	Null 2	Full 2	Selective2	Null 3	Full 3	Selective3
Browsing Covariates								
TIME	–	n.s.	–	Log	Log	–	Log	Log
TPAGE	–	Log	–	Log	Log	–	Log	Log
Repeat Visitation								
RETURN	–	n.s.	–	Dummy	Dummy	–	Dummy	Dummy
TSESSION	–	n.s.	–	n.s.	n.s.	–	n.s.	n.s.
Use of Decision Aids								
DECAID	–	Dummy	–	Dummy	Dummy	–	n.s.	n.s.
Input Effort and Information Gathering								
INTERACT	–	Linear	–	Linear	Linear	–	n.s.	n.s.
PRICING	–	Dummy	–	Dummy	Dummy	–	n.s.	n.s.
PAGETIME	–	Log	–	Linear	Linear	–	n.s.	n.s.
WORDS	–	Log	–	Log	Log	–	n.s.	n.s.
Control Variables								
CLUTTER	–	Linear	–	Log	Log	–	n.s.	n.s.
DYNAMIC	–	Linear	–	Quadratic	Quadratic	–	n.s.	n.s.
FIGURES	–	Linear	–	Linear	Linear	–	n.s.	n.s.
LINKS	–	Linear	–	Log	Log	–	n.s.	n.s.
PAGESIZE	–	n.s.	–	Log	Log	–	n.s.	n.s.
Utilities								
NUT1 Utility	–	–	–	–	Linear	–	–	Linear
NUT2 Utility	–	–	–	–	–	–	–	Linear
Number of Covariates	1	11	1	14	15	1	4	6
Contribution	-58,762.6	-32,730.3	-14,337.8	-7,463.7	-7,525.0	-3,667.4	-1,703.8	-1,751.4
Log PsBF	–	**26,023.3**	–	**6,874.1**	6,812.8	–	**1,963.6**	1,916.0

* All models include intercept; Null models are intercept-only models with county heterogeneity; Selective models include the utility of the preceding tasks as covariates (utility of NUT1 for NUT2, and the utility of NUT1 and NUT2 for NUT3).
n.s. – did not provide an improvement of log of PSBF; Removed from final estimation.

Table 4
Population Level Parameters
(posterior means and 95% probability intervals)*

	NUT1	NUT2	NUT3
Intercept	-6.97 [-7.26,-6.79]	0.35 [0.08,0.71]	0.53 [0.43,0.62]
Browsing Covariates			
TIME	n.s.	0.79† [0.75,0.83]	0.87† [0.82,0.92]
TPAGE	0.09† [0.06,0.11]	-0.69† [-0.75,-0.62]	-1.77† [-1.87,-1.66]
Repeat Visitation			
RETURN	n.s.	0.29 [0.12,0.38]	-0.38 [-0.57,-0.21]
Use of Decision Aids			
DECAID	-0.59 [-0.65,-0.54]	n.s.	n.s.
Input Effort and Information Gathering			
INTERACT	1.81 [1.74,1.87]	1.07 [0.98,1.14]	n.s.
PRICING	-0.52 [-0.56,-0.48]	0.35 [0.25,0.46]	n.s.
PAGETIME	0.93† [0.84,1.05]	-0.06 [-0.09,-0.02]	n.s.
WORDS	0.96† [0.93,1.0]	-0.51† [-0.57,-0.45]	n.s.
Control Variables			
CLUTTER	-0.60 [-0.71,-0.52]	1.35† [1.10,1.58]	n.s.
DYNAMIC	-0.41 [-0.51,-0.27]	-6.83 [-7.14,-6.47]	n.s.
DYNAMIC^2	n.s.	3.35 [3.06,3.72]	n.s.
FIGURES	-1.28 [-1.31,-1.25]	-0.98 [-1.05,-0.92]	n.s.
LINKS	0.54 [0.52,0.55]	1.55† [1.38,1.75]	n.s.
PAGESIZE	n.s.	-0.65† [-0.88,-0.35]	n.s.

* We report on the best fitting models for each task level (Full1, Full2, and Full3 from Table 4).
† Variable enters the model in the log form.
n.s. – did not provide an improvement of log of PSBF; Removed from final estimation.

Table 5
Elasticities
(posterior means and 95% probability intervals)*

	NUT1	NUT2	NUT3
Browsing Covariates			
TIME	n.s.	1.86 [0.27,3.23]	1.65 [0.11,2.72]
TPAGE	0.15 [0.02,0.30]	-1.70 [-2.86,-0.25]	-3.39 [-5.69,-0.23]
Repeat Visitation			
RETURN†	n.s.	0.04 [0.00,0.08]	-0.05 [-0.14,-0.00]
Use of Decision Aids			
DECAID†	-0.09 [-0.17,-0.02]	n.s.	n.s.
Input Effort and Information Gathering			
INTERACT	11.39 [2.14,19.12]	9.28 [1.69,14.21]	n.s.
PRICING†	-0.08 [-0.14,-0.02]	0.05 [0.00,0.10]	n.s.
PAGETIME	1.77 [0.36,2.86]	-0.18 [-0.79,-0.02]	n.s.
WORDS	1.71 [0.36,2.76]	-1.07 [-1.80,-0.16]	n.s.
Control Variables			
CLUTTER	-1.59 [-2.83,-0.31]	3.31 [0.49,5.72]	n.s.
DYNAMIC	-0.59 [-1.17,-0.14]	-2.47 [-4.99,-0.23]	n.s.
FIGURES	-15.21 [-25.63,-2.48]	-9.46 [-19.07,-1.04]	n.s.
LINKS	7.06 [1.71,10.99]	3.83 [0.57,6.57]	n.s.
PAGESIZE	n.s.	-2.26 [-3.83,-0.34]	n.s.

* We report on the best fitting models for each task level (Full1, Full2, and Full3 from Table 4).
† Attributable risk: the change in probability of task completion due to the presence, versus absence, of the factor.
n.s. – did not provide an improvement of log of PSBF; Removed from final estimation.

Table 6
Summary Statistics for the Estimation and Holdout Samples

	Task	Total Users	Users Completing Task	Percentage of Users Completing Task	Counties
Estimation (2/3)	1	64,332	19,437	30.2	1,914
	2	19,437	3,794	19.5	1,271
	3	3,794	1,301	34.3	564
Holdout (1/3)	1	32,166	9,801	30.5	1,601
	2	9,801	1,959	20.0	957
	3	1,959	668	34.1	430

Table 7
Predictive Performance*

	Before NUT1			Before NUT2			Before NUT3		
	Single 1	Single 2	Multi	Single 1	Single 2	Multi	Single 1	Single 2	Multi
MSE** Overall	0.022	0.022	0.020	0.065	0.069	0.056	0.214	0.209	0.086
Hit-Rate (%)									
Orders	0.1	0.1	0.3	0.9	0.0	2.9	28.2	45.2	83.7
Non-Orders	99.9	99.9	100.0	99.6	99.8	99.6	89.8	80.7	92.1
Overall	97.8	97.8	97.9	92.8	93.0	92.9	68.2	68.2	89.1

* Used a 0.5 probability cut-off (a car ordering is predicted when the probability is at least 0.5).
** MSE is mean squared error.

Table 8
Expected Gross Margin Loss in Car Sales for 1,000 Pop-up Exposures*

	Percentage Reduction in Purchase Probability	Exposure Scenario Random	Targeted
Before NUT1	2.0%	-$606.70	-$1.29
	0.5%	-$151.67	-$0.32
Before NUT2	2.0%	-$2,008.03	-$0.34
	0.5%	-$502.01	-$0.08
Before NUT3	2.0%	-$10,287.30	-$1,048.38
	0.5%	-$2,571.82	-$282.10

* Assumes an average gross margin of $1,500 per car. The expected gross margin loss for the targeted case was computed based on the expected purchase probability of the 1,000 visitors forecast as least likely to buy. For the random case the overall purchase rate was computed using the expected value for visitors at the different task levels (e.g., for the random exposure scenario before NUT1, -$606.70 = (1,301/64,332)×1,000×$1,500×2%).

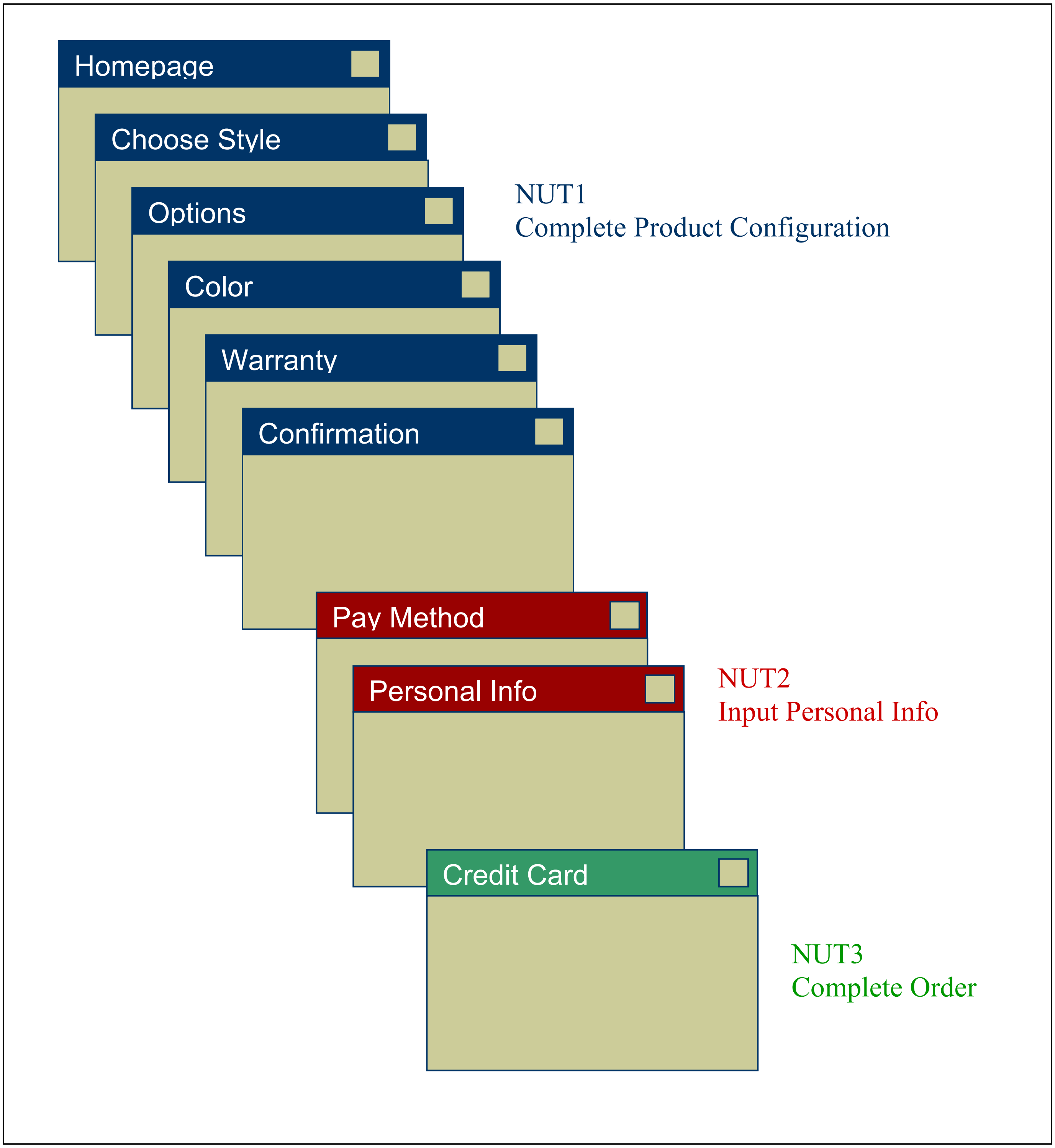

Figure 1: The Nominal User Tasks (NUTs)

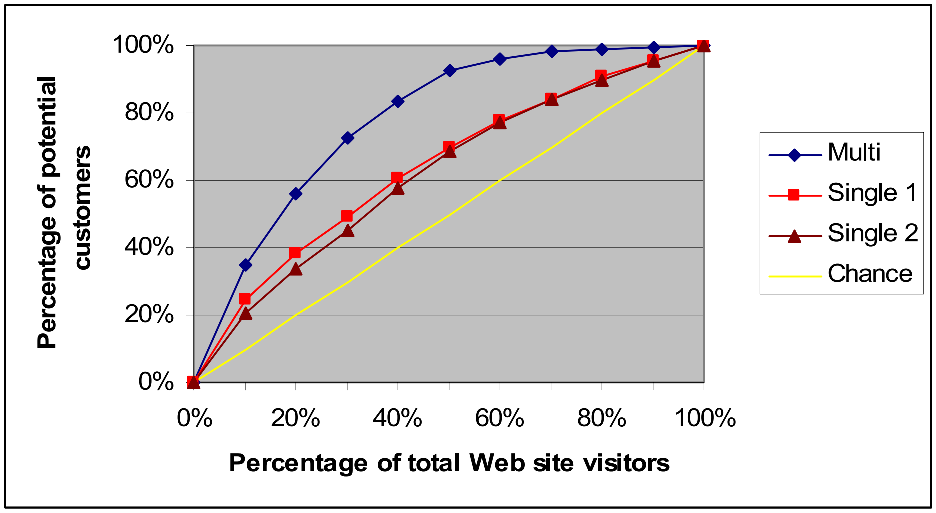

Figure 2a: Lift Chart for Order Prediction Before NUT1

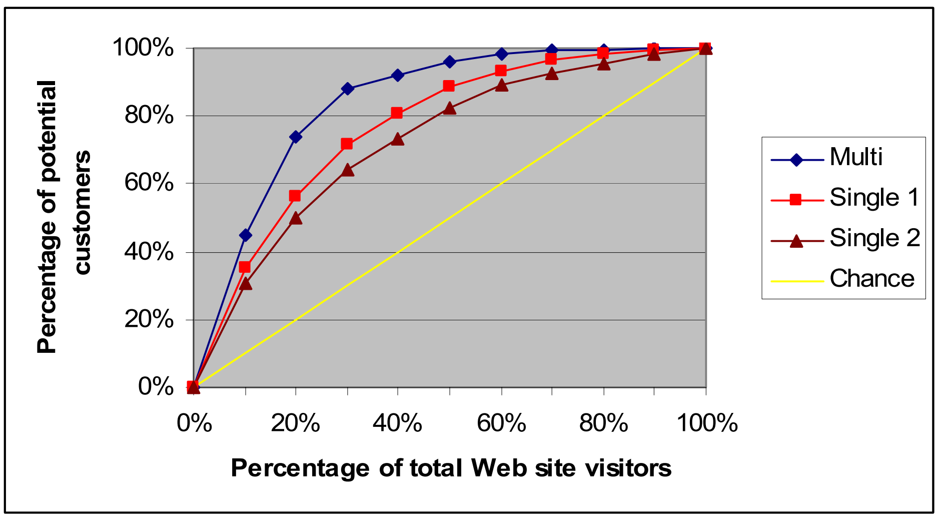

Figure 2b: Lift Chart for Order Prediction Before NUT2

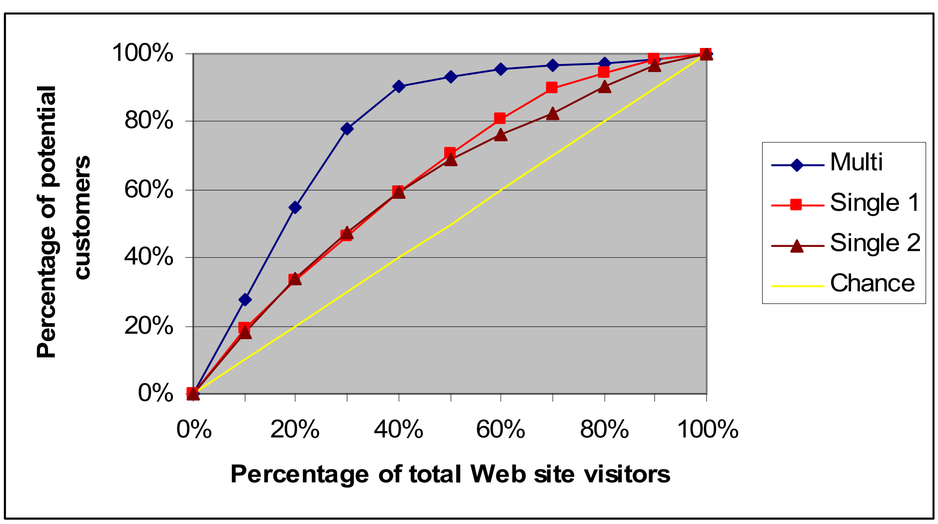

Figure 2c: Lift Chart for Order Prediction Before NUT3

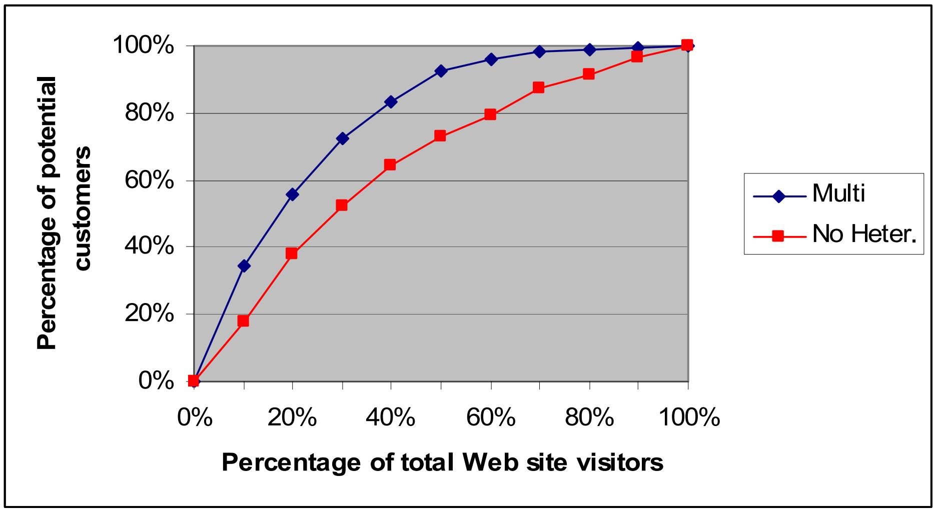

Figure 3a: Lift Chart for Order Prediction Before NUT1

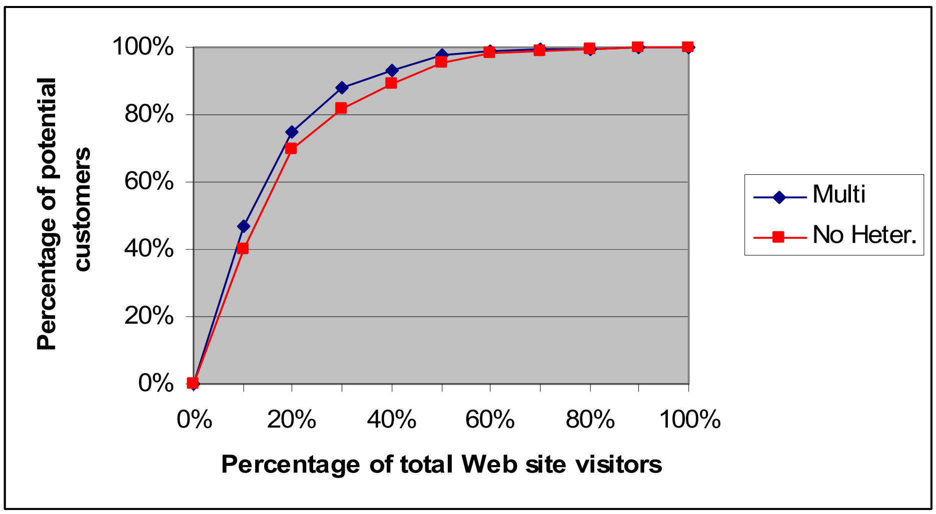

Figure 3b: Lift Chart for Order Prediction Before NUT2

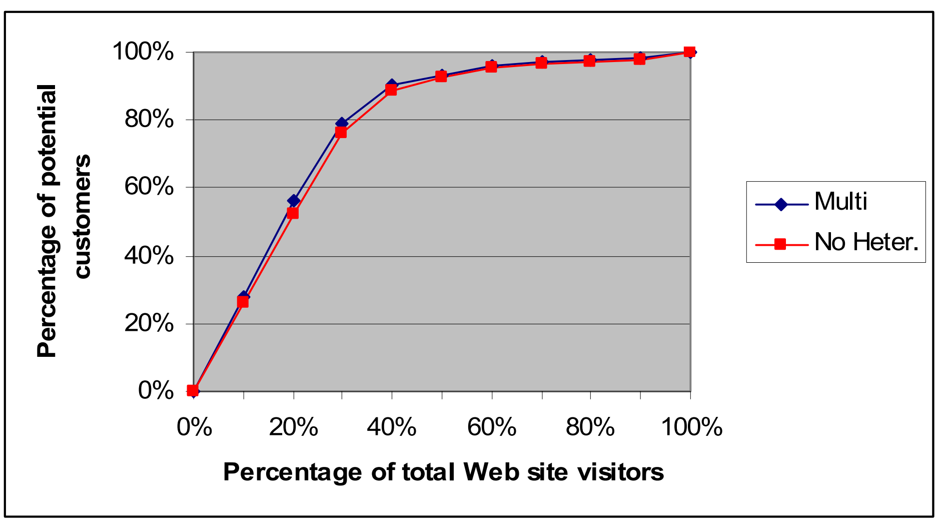

Figure 3c: Lift Chart for Order Prediction Before NUT3