A Study of Consumer Switching Behavior
Across Internet Portal Websites

August 2002

Sukekyu Lee, Fred Zufryden, and Xavier Drèze*

* Sukekyu Lee, Assistant Professor of Marketing, Hankuk University of Foreign Studies, College of Business and Economics, Seoul, Korea 130-081. Fred Zufryden is the Ernest W. Hahn Professor of Marketing at the University of Southern California, Marshall School of Business, Los Angeles, California 90089-1421. Xavier Drèze is Visiting Professor of Marketing at the University of California, Los Angeles. The authors acknowledge and thank Milward Brown Interactive Inc., for providing the data used in this study. They would also like to thank Sivaramakrishnan Siddarth, at the University of Southern California, for his comments on a previous version of this paper.
ABSTRACT

This paper describes an empirical study of the dynamics of consumer switching behavior across major Internet portal Websites. The study methodology is based on a multinomial Logit-Markov model framework that characterizes consumer switching-behavior. We propose a tractable Maximum Likelihood procedure to estimate the switching model’s parameters on the basis of online panel data.

This empirical study highlights the potential managerial implications of the model in the context of E-commerce. In particular, the results of the study are shown to provide potential managerial insights regarding the strength of a portal Website, relative to its competitors, in terms of its ability to attract and retain visitors, as well as the relative vulnerabilities of competing portals from which visitors are drawn. In addition, the study examines the effect and impact of causal factors (e.g., visitor gender, past Internet usage, log on time, and time spent on previous sites), on portal Website switching behavior and loyalty.

[Key Words: Internet Browsing Behavior, Website Choice Behavior, Web Panel Data Analysis, Market Structure Analysis, Multinomial Logit Models, Markov Process, Maximum Likelihood Estimation]
INTRODUCTION

The recent growth of the Internet medium has led to numerous new business opportunities. Among these, search engine sites (i.e., Yahoo!, Excite, Infoseek, Lycos), virtual communities (e.g., AOL, Geocities), ISP’s (Internet Service Providers), Internet browser companies (Microsoft and Microsoft Network, Netscape), and online shopping companies (e.g., Amazon.com) have become relatively important players on the Internet.

It has been suggested that Internet giants are seeking to become portal Websites. Portals are sites that provide general Internet capabilities and serve as a gateway to additional information [26]. The strategic importance of portals to marketers is evidenced by the research of Li, Liechty and Montgomery [37]. They report that in their data, the average reach of the top three portals is 53% while the average reach of the top three web sites in four other categories (Auction, Entertainment, Retailing, and Services) is only 18%. Thus, increased competition is expected among top portals on the Internet. It has been suggested that a key to winning this battle is for a site to develop “stickiness.” Stickiness is Web jargon for the ability of a Website to induce “surfers” to stay on a Web site for a long period of time and come back to use the site’s services at later dates.

As suggested by Shapiro and Varian [45], Internet portals are currently spending large amounts of money to keep old customers satisfied and generate new online customers. New visitors can be drawn from two sources: “current Internet users” mainly visiting other competing sites, and “non-Internet users” who have never used the Internet before.

As of yet, no study has investigated the dynamics of competition among main portals. Consequently, a primary objective of this study is to explore the dynamics of consumer switching behavior, and stickiness, among major portal sites on the Internet. Through the analysis of switching behavior, and stickiness,
behavior, it is shown how managers of portal sites may potentially determine the relative market strength of their companies, in terms of consumer visits to their sites, as well as potentially measure the impact of their marketing strategies on the Websites’ switching probabilities and market shares. In particular, it will be shown how the results of this study may suggest marketing strategies that are designed to enhance a portal Website’s ability to draw consumers from competing sites as well as for retaining consumers.

BACKGROUND FOR EMPIRICAL STUDY

In this study we focus on major portal Websites because they account for a large portion of the overall traffic that is generated on the Internet. They are also of interest because of the fierce nature of the competition between them, in the quest of becoming a leader of the E-commerce world. In addition, the investigation of the market structure and switching patterns among the main portals was expected to be of interest not only to managers, from a marketing strategic standpoint, but also to researchers, from a theoretical perspective.

Currently, there are various sources of commercially available online Web panel data (e.g., Mediametrix, NetRatings, Milward Brown). These data provide a potentially rich source of consumer information, especially to study consumer switching behavior on the Internet. In this study we used a set of online panel data provided by Milward Brown, a leading Web measurement research firm. The data is based on observations of individual panel members recruited by using a random digit dialing method and weighted to represent the Internet population. The data set consists of more than 800,000 Web browsing activities of about 50,000 online individual panel members and covers more than 1,000 different Websites. Since this data set is based on an online “panel,” it does not suffer from the common data

---

2 According to reports from Forrester Research [17,18], Yahoo! AOL, and MSN alone currently enjoy 15% of all Internet traffic and 45% of the advertising spend. Thus, although the portal sites investigated here comprise less than 20% of total traffic on the Internet, they earn more than 60% of total online advertising expenditures. In addition, these sites represent highly competitive E-companies, vying for a successful position among all E-commerce sites.
limitation of log file analysis in identifying unique visitors.\(^3\) Thus, we were able to unambiguously distinguish multiple visits by individual Web users from single visits by multiple Web users.

A shortcoming of this panel approach is that it is only representative of home Internet usage at the exclusion of ‘at work’ usage. This is a common problem of online panels due in part to the fact that the participation in the panels requires tracking software to be installed on a user’s computer. The tracking software follows the user’s “click stream” by recording all user activities and transmitting back to the panel company. Given the security risks, companies are often reluctant to allow their employees to install such software.

For each individual panel member, we determined the corresponding chronological sequence of his/her Website visits, including consecutive visits made to portal Websites. For the purpose of defining site visits within the data set, we used the traditional “half-hour rule.” This rule is used by most click-stream analysis software (e.g., WebTrends or Microsoft IIS). That is, we considered any two page requests to the same Website that are spaced by less than half an hour as belonging to the same site visit, while two consecutive page requests that are spaced by more than half an hour were treated as belonging to two distinct site visits. For each visit, we tabulated data on the particular Website visited and on potential explanatory variables (i.e., the Website that was visited previously, as well demographic and situational variables. These variables are discussed in greater detail in a subsequent section dealing with the “Estimation of the Switching Behavior Model.”

In the following sections, we first provide an overview of the related literature. Next, we propose a methodological approach based on a first-order Markov brand switching model and a related Maximum Likelihood estimation (MLE) technique for its implementation. We then describe empirical results derived from the application of our methodology to the study of Website portal switching behavior, as well as a validation of the proposed model that is conducted on the basis of analysis and holdout data.

\(^3\) See Drèze and Zufryden [13], for a review of problems associated with tracking individuals using log file data.
This is followed by a discussion of potential managerial implications of our results. Finally, we conclude with a discussion of some of the limitations of our study as well as their implications for future research.

OVERVIEW OF RELATED LITERATURE

In this study, we examine the switching behavior patterns of consumers across Internet portal Websites as well as the potential determinants of switching behavior. Although, there is no precedence for such a study in the Internet literature, there have been numerous marketing studies that have dealt with stochastic brand choice models and, in particular, the modeling of brand switching behavior. The keen interest in this area has been due, in part, to the potential managerial insights that may be derived from the study of brand switching patterns. For example, among its wide applications, switching pattern research has been used to evaluate and predict the market performance of packaged goods as a function of marketing mix and consumer demographic variables [54, 55], to study market structures [7], in Conjoint analysis [39], and for new product evaluation [48].

The study of stochastic models of brand choice behavior has received considerable attention in the marketing literature. A major focus of early models, proposed in the 60’s and 70’s, dealt with alternative ways of characterizing the “order” of the brand purchase behavior process. This led to the examination of the potential effect of present purchase behavior on future purchase probabilities (i.e., “purchase event feedback”). Among the first types of models considered were 1) zero-order models [3, 19, 32], 2) first-order Markov models [15, 25, 27], and 3) infinite-order models [6, 36, 51].

Aside from a few exceptions that considered the effect of a single marketing variable [28, 38, 46] early stochastic models generally did not consider the effect of explanatory variables, such as marketing mix and consumer segmentation variables, on brand choice probabilities. Consequently, despite their usefulness in describing the brand choice behavior process, their practical application and actionability was limited from a managerial perspective. With this major limitation in mind, subsequent models sought
to enhance the managerial usefulness of prior models by relating brand choice probability to marketing mix variables.

In an early attempt to consider explanatory variables in stochastic brand choice models, Jones and Zufryden [30] and Zufryden [52] suggested using a binomial logit-based brand choice model formulation to examine the market performance of packaged goods. These studies included price and consumer demographic data as explanatory variables. However, a limitation of these early models is that they defined a market in terms of only two brand states (i.e., a brand of interest versus all other competing brands) and thus did not provide a complete competitive view of the market. Wagner and Taudes [49] later proposed an extension of previous zero-order models by developing a multi-brand stochastic model framework where the mean purchase rate from a Poisson process is expressed as a function of marketing mix variables and time. Numerous recent studies have followed to further investigate the relationship between brand choice probability and explanatory variables using consumer panel data [5, 10, 11, 29, 32, 33, 34, 43].

A stream of research that relates most closely to the methodology that is proposed in the present study has sought to incorporate explanatory variables within first-order Markov brand choice processes. Zufryden [53] developed a logit-based Markov model approach that relates explanatory variables to transition probabilities in a two-brand market. The study was later extended to the multi-brand case by using a multinomial-logit-based model in Zufryden [55]. Other marketing studies have focused on the relationship between brand switching patterns and explanatory variables in various marketing contexts. For example, several studies have relied on brand switching to examine market competition and define market structures [7, 31]. Givon and Horsky [20] developed a two-state first-order Markov model that incorporates advertising and pricing effects. Mahajan, Green, and Goldberg [39] have shown that multi-brand switching models can be used to measure cross-price demand relationships in Conjoint Analysis studies. Urban, Hauser, and Roberts [48] utilize a Markov process to model custom flows across states as a function of marketing variables in the analysis of the launch of new products.
In the latter studies of brand switching behavior, a first-order Markov chain framework has been shown to provide a viable modeling approach. Among the appealing properties of the first-order Markov model are its ability to characterize consumer behavior in multi-brand (or more generally multi-state) market situations, its consideration of the effect of previous purchase events, and its ability to describe competitive behavior. In addition, first-order Markov models can provide potentially useful diagnostic outputs describing consumer switching as well as brand loyalty measures. Furthermore, the latter models can provide useful predictive outputs that include short and long-term market share predictions as a function of explanatory variables. In this study, we show how the model can be based on a tractable Multinomial Logit model framework whose parameters can be estimated by means of available statistical procedures. All of these features make it an attractive theoretical modeling approach that also has potential relevance from a managerial perspective.

In implementing the latter Markov models, a central issue is the estimation of model parameters. However, a review of the estimation procedures used in the research noted in the preceding paragraphs reveals some limitations and logical inconsistencies. Carpenter and Lehmann [7], do not specifically define the “last brand choice” as an independent variable in their formulation but rather use a coding scheme based on pair-wise brand similarities. Subsequently, a restricted generalized least squares (GLS) technique is proposed that draws on the results of Nakanishi and Cooper [42] to estimate model parameters. However, as noted in Zufryden [55], the estimation procedure proposed by Carpenter and Lehmann [7] does not enforce the logical consistency of the underlying model (e.g., the sum of the conditional probabilities across each row of a first-order Markov switching matrix must be equal to 1). In contrast, Zufryden [55] proposes an estimation procedure that is based on a restricted weighted least squares procedure and enforces the logical consistency of the Markov model. Consequently, in addition to yielding probabilities that are properly constrained, between 0 and 1, the resulting conditional transition probabilities of the estimated first-order Markov model appropriately sum to unity for each row of the transition matrix. Nevertheless, this procedure also suffers from drawbacks. It is best suited for the
incorporation of categorical explanatory variables (e.g., demographics) and is difficult to apply to continuous variables (e.g., price). Furthermore, it requires a large amount of data to permit model estimation and avoid the tendency for sparse observations within observational cells. Given the limitations of previous estimation techniques, this study proposes an MLE technique. This approach can be shown to provide an attractive estimation procedure as it easily permits the consideration of both continuous and categorical explanatory variables in first-order Markov switching models. In addition, MLE avoids the potential problem of sparseness of data within cells, and leads to parameter estimates with appealing statistical properties. The issues of methodology development and implementation are discussed in the following sections.

**STUDY METHODOLOGY**

We now turn to the formulation of a methodological framework to study consumer switching behavior across portal Websites on the Internet. The proposed methodology is based on a first-order Markov switching behavior model whose transition probabilities are functions of explanatory variables. In contrast to zero-order models, it should be noted that this model approach provides a general framework that can be used to describe the transition probabilities in various contexts, from a given state (i.e., Website) to another.

In the interest of generalizing our approach, we assume that there are $J$ Websites (states) that a consumer can switch to at his/her $i^{th}$ visit. These $J$ alternatives are mutually exclusive and collectively exhaustive. Using the notation of Zufryden [55] and following the logistic formulation described in Diebold, Lee and Weinback [14], Filardo [16], and Gray [21], we define the $J \times J$ matrix of transition probabilities for a first-order Markov process as:

\[
p_{ij} = \frac{\exp(b_{ij})}{\sum_{k=1}^{J} \exp(b_{ik})}, \quad \text{for } j = 1, 2, \ldots, J
\]

by setting vector $b_j$ to 0 (zero vector) as a normalization.

---

4 (1) can be obtained directly from the standard multinomial logit form: $p_j = \frac{\exp(b_{ij})}{\sum_{k=1}^{J} \exp(b_{ik})}$, for $j = 1, 2, \ldots, J$ by setting vector $b_j$ to 0 (zero vector) as a normalization.
\[ P_s^i(j \mid k) = \frac{e^{\beta_j X(k) + \lambda_j v_s + \epsilon_j}}{1 + \sum_{l=1}^{J-1} e^{\beta_l X(k) + \lambda_l v_s}} \quad \text{for } j=1,2,3, \ldots, J-1, \text{ and } k=1,2,\ldots, J, \]  

where, we define the transition probability \( P_s^i(j \mid k) \) as the probability that a consumer facing situation \( s \), will switch to the site \( j \) at the \( i^{th} \) visit given that his/her previous visit was to site \( k \). In addition, we define the following variables and parameters:

- \( v_s = (v_{s1}, v_{s2}, \ldots, v_{sM}) \) is a vector of \( M \) explanatory variables, corresponding to a particular situation \( s \) (e.g., a particular setting of user-specific variables such as demographics, user behavior, marketing mix at a given choice occasion, etc.), with \( s=1,2,\ldots,S \).
- \( x = (1, x_1, x_2, \ldots, x_{J-1}) \) is a previous state (i.e., last site visited) specification vector consisting of binary variables with \( x_k = 1 \) if a consumer was formerly in state \( k \) (i.e., if Website \( k \) was chosen at the previous visit occasion), and 0 otherwise, for sites \( k=1,2,\ldots,J \).
- \( X(k) \) = Particular setting of vector \( x \).
- \( \beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{J-1}) \) is a vector of parameters corresponding to \( x \).
- \( \lambda_j = (\lambda_{j1}, \lambda_{j2}, \ldots, \lambda_{JM}) \) is a vector of parameters corresponding to the situational variables \( v_s \),

and \( \epsilon_j \) = a random error term.

Note that each individual is assumed to have the same transition probability given that s/he faces the same explanatory variables. However, each individual will face different settings of the explanatory variables because s/he faces different situations, including different Internet use patterns, marketing mix, and possesses different demographic characteristics. Thus, modeling individual differences in the explanatory variables s/he faces captures some level of heterogeneity among individuals.

Since \( \sum_{j=1}^{J} P_s^i(j \mid k) = 1 \), in order to enforce logical consistency, we now define conditional probabilities \( P_s^i(j = J \mid k) \) corresponding to the last state \( j=J \) as:
\[ P_s^j (j = J \mid k) = 1 - \sum_{j=1}^{J-1} P_s^j (j \mid k) = \frac{1}{1 + \sum_{l=1}^{J-1} e^{\beta_j X(k) + \lambda_j \psi_j}.} \] 

(2)

Thus, the above model specification is logically consistent as it satisfies the condition that the resulting switching matrix is stochastic (i.e., that the sum of conditional transition probabilities equals unity for each row, or \[ \sum_{j=1}^{J} P_s^j (j \mid k) = 1 \] over each row), and since (2) is a multinomial logit form, all the conditional probabilities are properly range-constrained between 0 and 1.

Now, assuming that the sum of the error term \( \sum_{j=1}^{J-1} \epsilon_j = 0 \) and \( \epsilon_j \) follows the standard extreme distribution [22, 23], we can take the log of the ratio of (1) and (2) and thereby obtain the following generalized multinomial logit model of the switching behavior, in a linear form, as a function of explanatory variables:

\[ \ln \left( \frac{P_s^i (j \mid k)}{P_s^i (J \mid k)} \right) = \beta X(k) + \lambda \psi + \epsilon. \] 

(3)

ESTIMATION OF THE SWITCHING BEHAVIOR MODEL

We now develop an MLE procedure to estimate the parameters of the first-order Markov model developed above, based on the matrix of conditional transition probabilities. MLE procedures have commonly been applied to multinomial logit models in the brand choice modeling literature [4, 9, 23, 24, 40].

In our empirical illustration, we considered \( J = 5 \) mutually exclusive and collectively exhaustive site states (\( j = 1, 2, ..., J \)). These states include Yahoo!, AOL, Netscape, and other main portal sites (i.e., Excite, Microsoft, MSN, Geocities, and Infoseek which were classified within an “Other Portals” category). These portal sites were selected in view of their relative importance on the Internet.
addition, other Websites not included in the portal categories were grouped into a “Non-portal site”
category.

We assume that observations are independent of each other and that a given observation depends
on the particular situation \( s \) we are considering. Thus, the likelihood of observing the sample data is
found by taking the product of the probabilities at every site visit. Since we consider the odds ratio of
\( P_s^i(j/k) \) to \( P_s^i(J/k) \) in (3), the likelihood function is given by the following product:

\[
L = \prod_{i=1}^{N} \prod_{s=1}^{S} \prod_{j=1}^{K} \prod_{k=1}^{J-1} \frac{P_s^i(j|k)}{P_s^i(J|k)}.
\]

(4)

Now, based on the relationship in (3), (4) becomes:

\[
L = \prod_{i=1}^{N} \prod_{s=1}^{S} \prod_{j=1}^{K} \prod_{k=1}^{J-1} \left( e^{\beta_s X(k) + \lambda_s v_i + \varepsilon_j} \right).
\]

(5)

Then, taking the logarithm of both sides of (5), leads to the following log likelihood function:

\[
\ln(L) = \sum_{i=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{K} \sum_{k=1}^{J-1} \beta_s X(k) + \lambda_s v_i + \varepsilon_j.
\]

(6)

The resulting log likelihood function (6) has a linear form, in terms of parameters, and can be
used to readily estimate the first-order Markov model by maximizing \( \ln(L) \) by appropriate choice of the
parameter vectors, \( \beta_s \) and \( \lambda_s \).

To operationalize our model, we first defined explanatory variables to reflect the last site visited
(prior state) as follows:

\[
X_1 = \begin{cases} 
1 & \text{if a consumer’s previous state was Yahoo!}, \\
0 & \text{otherwise}.
\end{cases}
\]

\[
X_2 = \begin{cases} 
1 & \text{if a consumer’s previous state was AOL}, \\
0 & \text{otherwise}.
\end{cases}
\]

\[
X_3 = \begin{cases} 
1 & \text{if a consumer’s previous state was Netscape}, \\
0 & \text{otherwise}.
\end{cases}
\]

\[
X_4 = \begin{cases} 
1 & \text{if a consumer’s previous state was one of the “Other Portals”}, \\
0 & \text{otherwise}.
\end{cases}
\]
These dummy variables were defined as above to permit the consideration of a first-order Markov process, reflecting switching among the various portal Websites, within a logit model framework [55].

In addition, we considered the following situational variables:

\[ X_5 = \begin{cases} 1 & \text{if a consumer visits a site during the evening (6pm-midnight),} \\ 0 & \text{otherwise.} \end{cases} \]

\[ X_6 = \begin{cases} 1 & \text{if a consumer visits a site on a day during the first half of the week (Mon through Wed),} \\ 0 & \text{otherwise.} \end{cases} \]

\[ X_7 = \text{a continuous variable representing how long a consumer stayed at the previous Website.} \]

\[ X_8 = \begin{cases} 1 & \text{if a consumer is a “heavy” Internet user,} \\ 0 & \text{otherwise.} \end{cases} \]

\[ X_9 = \begin{cases} 1 & \text{if a consumer is male,} \\ 0 & \text{otherwise.} \end{cases} \]

Our choices of situational variables were constrained by data availability. Although the Milward Brown panel provided a rich source of Internet behavior data, it was limited in that it did not incorporate observations on marketing variables, such as promotions and banner ads, that may significantly affect Internet visiting behavior [50]. Nevertheless, subject to data availability, our choices of situational variables were guided by results of previous studies. For example, it has been shown that males and females exhibit differences in their Internet browsing behavior and that gender is one of the best predictors of Internet usage [41]. Men are shown to access the Internet more frequently and to be more persistent than women. Furthermore, Li et al [37] report that men use portals with greater frequency than women. Thus, in an attempt to allow for such differences, we defined a gender variable \( (X_9) \) as a major demographic variable in our study.

In consumer purchase behavior studies, there has been considerable attention given to “heavy users” as a prime target consumer segment [35, 47]. Furthermore, the “heavy half theory” has stressed the

---

5 The criterion to determine “heavy” vs. “light” user groups was based on the total number of prior visits. Heavy users were defined as those who made at least 11 visits per month. This usage classification criterion was based on an evaluation of the frequency distribution of visits observed in the calibration data sample.
importance of “heavy users” by suggesting that 20% of consumers in a product category account for 80% of all product purchases [44, 47]. With regard to Internet website visiting patterns, Drèze and Huss herr [12] find that expert users process web pages faster and more sparingly (e.g., they only process the portion of web pages that is most relevant to their task) than novice users. Thus it was likewise expected that heavy and light Internet users could exhibit differences in switching behavior. To reflect these potential effects, we defined the “heavy” Internet user variable (X_8) to classify Internet users on the basis of the magnitude of their Internet visiting frequency.

A major objective of Websites is to provide information value to their visitors. In this regard, it is expected that the amount of time spent by visitors on a Website should affect the amount of information that will be derived from the Website. For this reason, a common Website objective is to maximize the amount of time visitors spend browsing the pages within a site [13]. In addition, as visitors gain familiarity with a site, by extensive exploration and time spent on the site, it is expected that subsequent browsing behavior may be affected. For example a visitor may be more likely to return to a site, in order to derive further informational value or gain additional benefits from the site. Conversely, a visitor who has spend a significant amount of time on a site may decide to switch to another Website because s/he may feel that s/he has obtained sufficient information value from the former site and that no additional information value will be gained by returning and spending additional time on the site. In an attempt to capture the latter effects, we defined X_7 to represent the amount of time spent by a visitor on the Website that was visited at the previous occasion.

Audience viewing data for standard media (e.g., A.C Nielsen TV viewing reports) show that audience viewing patterns over time vary substantially, in terms of size and demographics. In fact, advertising managers are cognizant of these variations in developing their TV media plans and schedules. On the Internet, past studies have similarly shown that Internet usage behavior varies substantially over time [41]. Thus, Internet usage may vary according to time of day and day of the week. In particular, it is expected that Internet users may spend more time on the Internet during the evening hours of the day, or
during weekends, when more leisure time is available to them for browsing. Consequently, we defined two dummy variables to reflect these potential time effects. First, $X_5$ was defined to represent the time of day during which a visitor browsed the Internet. In addition, we defined $X_6$ to reflect the day of the week in which the Internet browsing occurred.

**STUDY RESULTS**

The switching model described in (6) was estimated using the proposed MLE method and implemented by using the CATMOD procedure of SAS. The results of the estimation are summarized in Tables 1 and 2. In Table 1, we note that all the variables investigated here are statistically significant ($p < 0.05$). Moreover, the log likelihood ratio indicates that the model is statistically significant ($p < 0.0001$) and provides a good fit to our data.

Table 2 provides a summary of the estimation results for each explanatory variable including the relative magnitudes of each parameter and the corresponding levels of statistical significance. Here, estimation results show that 32 out of a total of 40 parameters are statistically significant at the $\alpha=0.05$ level while 27 out of a total of 40 parameters are significant at the $\alpha=0.01$ level. In particular, the coefficients along the diagonal of the transition matrix, corresponding to variables $X_1$ through $X_4$, tend to be the most statistically significant. This implies that the switching probability to a particular site is significantly affected by a consumer’s previous state. This empirical finding supports the notion that consumer browsing behavior follow a *first-order* Markov process as opposed to a *zero-order* (memory-less) process. This finding is consistent with previous studies [54, 55]. Thus, we conclude that the former site variables, which represent the site visited at the previous visit occasion, are useful in predicting what site a consumer will visit next.

**Interpretation of Model Parameters**

The magnitude of model parameters reflects the relative importance in the contribution of particular explanatory variables to the switching probability. It is recalled that this study uses the
logarithm value of the odds ratio of the conditional switching probabilities (a portal Website relative to a non-portal Website) to formulate the relationship with the explanatory variables. Thus, for an indicator variable such as \( X_1 \) (previous state was Yahoo!), which can take on values 0 or 1, by taking \( e^\beta \), we obtain the change in the ratio of \( P_s^f(j/k) \) to \( P_s^f(J/k) \) controlling for the other covariates. For example in Table 2, we have \( b_{11} = 4.4168 \). This yields \( e^{b_{11}} = 82.8 \). Therefore, controlling for the other covariates, the ratio of the conditional probability of visiting Yahoo! to the probability of visiting a non-portal Website is 82.8 times greater for those whose previous visit was to Yahoo! than for those who previously visited a non-portal Website. To translate this increase in probability ratio to an actual transition probability, we can go back to equation (1) and plug-in the estimated values of all the \( b_{ij} \) parameters. Table 3 illustrates such an exercise for a scenario where we assumed that all other covariates were equal to 0 (i.e.,

\[
P[j \mid k] = \frac{e^{\beta_{0} + \beta_{1} X_1}}{1 + \sum_{l=1}^{J-1} e^{\beta_{0} + \beta_{l} X_1}}.
\]

As one can see, the 82.8 fold increase in the ratio of the probability of going to Yahoo! over the probability of going to a non-portal Website translates in a change in conditional probability from 0.028 to 0.650. In other words, the probability (stated in percent terms) of visiting Yahoo! after visiting another Website is 2.8% while the probability of revisiting Yahoo! if the last Website visited was Yahoo! is 65%. Whether the last site visited is Yahoo! rather than a non-portal Website also affects the probability of visiting the other portals. Thus, Table 3 shows that the probability of visiting AOL decreases from 2% to 0.8%; the probability of visiting Netscape increases from 1.6% to 1.9%; and the probability of visiting another portal decreases from 9.3% to 8.3%.

This interpretation can be applied to the other dummy variables in this study such as the “previous state” (e.g., AOL, Netscape or Other Portals), “Logon time,” “Logon day,” and “Heavy Internet users.” For instance, the coefficients for “Logon time” show that logging on in the evening, rather than during other times of the day, increases the probability of visiting Yahoo!, and decreases the probability of visiting Other Portals. It is also noted that heavy users are less likely to use portals than
light users (parameters $b_{81}$, $b_{82}$, $b_{83}$, and $b_{84}$ all have negative signs). Finally, in contrast to males, the female segment has a greater propensity to use AOL and Other Portals, and a lower propensity to visit Netscape.

For a continuous variable such as “Total time spent at the previous site,” we can use the transformation $100(e^{\beta}-1)$, which gives the percentage increase in the expected ratio of conditional switching probability for each one-unit increase in the variable. For example, consider the effect of how long a consumer stayed at the previous Website. For Yahoo!, $b_{71} = -0.0318$ and consequently, $100(e^{-0.0318}-1) = -3.13$. According to this result, each 1-minute increase in time spent at the previous site is associated with a 3.13% decrease in the expected switching probability to the Yahoo! site during the next visit, holding other covariates constant.

Note that Yahoo! has greater consumer retention power ($b_{11} = 4.4168$) than AOL ($b_{22} = 4.3711$). Consumers who visit Yahoo! come from AOL and Netscape sites as well as other main portal sites. Particularly, other main portal sites ($b_{41} = 1.4172$), such as “Excite,” “Infoseek,” “Microsoft,” and so on, appear to provide the most significant source of switching to Yahoo!.

Similarly to the results of Zufryden [55], we find the constant terms all negative in sign. These constant terms reflect the contributions to each site’s switching probability from consumers whose last visit was made to a regular Website as opposed to a portal. In sum, the model parameters provide potentially useful diagnostic information about the market structure for main portal sites as well as for competitive strategy in a dynamic market. Some of the managerial implications of our results are discussed further in a subsequent section.

**Model Validation**

To provide additional evidence of the descriptive and predictive goodness of fit of the switching model, we examined the switching model results specific to two explanatory variables: gender and visit frequency. Thus, actual switching matrices were developed based on a segmentation of consumers in one of four groups: Heavy user-Male, Heavy user-Female, Light user-Male, and Light user-Female segments.
Based on these segments, we contrasted the estimated switching probabilities we derived from our model with the actual values observed from our data. The results corresponding to our five states are summarized in Table 4. These results show that the coincidence between actual and estimated switching probabilities is very good. The Chi-square test of estimated and actual switching probabilities for each segment also suggests that the estimated probabilities and the actual probabilities are not statistically different, implying a good overall descriptive model fit.

We developed and estimated the switching model on the basis of calibration data (covering a time span of four weeks, from 2/11/98 – 3/10/98). The model calibration period consisted of a total of 488,000 browsing events made by 4,500 different individuals panel members from a nation wide U.S. sample. We also conducted a predictive validation of the model on the basis of holdout data (covering a span of three weeks, from 3/11/98 to 4/1/98). The holdout data consisted of 350,000 browsing events made by 2,800 different Web users. For purposes of predictive model validation, we compared the estimated switching probabilities obtained from the calibration data (2/11/98 – 3/10/98, 4 weeks) with the actual switching probabilities observed in the holdout data.

By updating the initial portal site market shares observed during the calibration period with the estimated Markov switching matrix [54, 55], we predicted actual portal site shares for the holdout period. Table 5 provides market share predictions, as well as estimated steady state shares, from our predictive test. Here, it is noted that the model produces reasonably accurate future-market share forecasts. Overall, the results show average absolute errors of 12.5% to 39.1% from the actual values.

Managerial Implications

In this study of consumer switching behavior across major portal Websites, our empirical results suggest potentially useful insights for understanding Website switching and loyalty behavior. Thus, the study provides potentially useful information about the determinants of portal Website choice behavior and loyalty formation. This information has managerial relevance for the analysis of market structure and market dynamics and the development of effective promotion strategies on the Internet. In particular, the
following provides illustrations of some of the managerial questions that may be investigated on the basis of our empirical results:

1. **What is the relative strength of my Website in attracting visitors?** – Marketing managers can measure the relative market strength of their sites relative to their competitors by evaluating the magnitudes of the switching probabilities. For example, since the ratio of the probability of switching to Yahoo! from AOL to the probability of switching to AOL from Yahoo! is greater than 1, we can say that more customers are switching from AOL to Yahoo! than are switching from Yahoo! to AOL. Thus, as Yahoo! is drawing more customers from AOL than AOL is from Yahoo!, it is expected that AOL will ultimately lose market share to Yahoo! in the long run.

2. **How loyal are visitors to my site in relation to competing sites?** - The diagonal elements of the switching matrix may be defined as measures of “site loyalty” or “stickiness.” For example, the empirical results of Table 2 show that all sites have similar loyalty patterns, in terms of their ability to retain visitors (i.e., coefficients lie within the range .43711 to .48122). However, based on the relative sizes of their respective coefficients, Netscape ($b_{33}=4.4601$) and Other Portals ($b_{44}=4.8122$) appear to achieve the highest loyalty among the site alternatives considered. Further, AOL (with coefficient .43711) exhibits the lowest loyalty level.

3. **From which competing site(s) can I draw the most visitors?** – Each Website will tend to attract visitors from others to a varying degree. Table 2 provides insights about the competing sites that may provide the best sources for attracting new visitors. For example, the coefficients of Table 2 suggest that Yahoo! draws more visitors from Netscape ($b_{31}=1.8828$) and Other Portals ($b_{41}=1.4712$) than from the other Websites. In contrast AOL draws most from Other Portals ($b_{42}=1.6133$), while Netscape draws most from Yahoo! ($b_{13}=1.4677$) and Other Portals ($b_{43}=1.5658$). In turn, Other Portals draw more from Netscape ($b_{34}=1.4770$) and AOL ($b_{24}=1.3149$). Thus, the parameters of the switching model can also provide insights about the structure of competition within an online market - particularly for main Internet players like
portals. In particular, managers can evaluate the expected changes in the individual-level switching probability with respect to changes in the explanatory variables over time for specified market segments (e.g., see Table 5).

4. **How do logon time and day affect visits to my site?** - From Table 2, and the coefficients for the variable $X_5$, it may be noted that Yahoo! draws more visitors from those logging on in the evening (6PM-midnight). In contrast, AOL, Netscape and Other Portals draw more visitors from those logging on earlier in the day. However, the latter results were not statistically significant in our study in the case of AOL and Netscape. In the case of logon day ($X_6$), it is noted that logging on early in the week has a positive impact on visiting behavior for each Website. However only Netscape and Other Portals were shown to have significant coefficients.

5. **How does total time spent on previous sites affect visits to my site?** – The coefficients of variable $X_7$ in Table 2 indicate that the more time is spent on the previously visited site, the less likely is a subsequent visit to Yahoo!. In contrast, the more time spend on the previously visited site, the more likely a subsequent visit to AOL, Netscape and Other Portals. These results may suggest that Yahoo! is not viewed as adding much value, beyond that derived from previous visits to competing Websites. This is in contrast, to subsequent visits to AOL, Netscape or Other portals, which do appear to be viewed to provide additional value.

6. **Are heavy Internet users more likely to visit my Website?** – Table 2 indicates that the coefficients for variable $X_8$ were negative for each portal state. This clearly suggests that heavy Internet users are less likely to visit portal sites. This is likely because they are more sophisticated users and therefore rely less on information and features that are provided by portal sites. Perhaps they might have developed a library of bookmarks, or know the URL of the sites they want to visit without relying as much on portals to consolidate the information for them.

7. **Is my site more attractive to customers of particular demographics (e.g., gender)?** – Here, our empirical results indicate that AOL, Netscape and Other Portals all seem to be more attractive to
female visitors (i.e., coefficients corresponding to $X_9$ are all negative). In contrast, Yahoo! tends to be more attractive to males (i.e., $b_{y1}$ is positive). However the latter coefficient was not shown to be statistically significant.

8. **What overall strategies can I use to increase visitors to my site?** – Our empirical results suggest potential differential strategies for alternative Websites. For example, in the case of Yahoo!, our empirical results suggest that it may wish to target current Netscape and Other Portal users. Perhaps one way to do this is to feature banner ads or links on sites that are frequently accessed by such users. In addition, the importance of visitors logging on during the evening suggests a corresponding timing of promotions, such as banner ads, to reach visitor prospects. For example, companies, such as DoubleClick, make it possible to specifically time the delivery of banner ads. Moreover, from a segmentation standpoint, our results indicate that Yahoo! should focus on a - Light user-Male target. In contrast, the strategic implications for AOL differ to a large extent. Here, our results suggest that it should target Other Portal users and focus on a Light user-Female segment. In the case of Netscape, our results suggest that it should target Yahoo! and Other Portal users, time promotions early in the week, and target a Light user-Male segment.

9. **How will my strategies affect the future performance of my Website?** – The steady-state probabilities that may be derived from our model (e.g., see Table 5) provide us with interesting insights about expected future Website performance. For example, with respect to the impact of market segmentation strategies, note how the share of non-portal sites changes depending on whether the visitors are light or heavy Internet user. Whereas light users rely on portals for 60% of their visits, heavy users rely on portals for only 20% of their visits. Again this indicates that heavy users either use the Internet for different purposes than light users, or perhaps that they are more knowledgeable about where things are online and, consequently, do not need to rely on portals as much as light users. One can also use the steady-state probabilities to better understand the positioning of the three portals studied. For example, our results suggest that Yahoo!
dominates AOL in every segment except the Light user-Female segment. Similarly, Netscape dominates or equals AOL on all but the Light user-Female segment. Yahoo! seems to be appealing to all user groups, while AOL appeals to the Light user-Female, and Netscape to a Light user-Male segment.

CONCLUSION

Existing Internet companies have sought to maintain their current positions or expand through mergers and/or acquisitions of new Internet companies. The success of Internet companies will depend on which company has the most accurate information about Internet users. In this regard, a major objective of this study was to provide insights from a switching model to help managers enhance their understanding of customer loyalty and switching behavior across portal Websites. As such, the results from our study can be used to potentially enhance the current and potential market positioning of portal sites on the Internet.

From a methodological perspective, this study utilizes a multivariate choice model, based on a multinomial logit-Markov framework, for analyzing portal Website switching behavior. The switching behavior model was expressed as a function of explanatory variables that includes the site that was previously visited as well as Internet use behavior and demographic variables. The model also captures individual heterogeneity in switching probabilities across consumers by expressing switching probabilities as a function of possible situations the consumers may face. The model was implemented on the basis of a MLE estimation procedure that was applied to commercial online panel data. As such, this study extends approaches used in choice models. In particular, this study extends research by Carpenter and Lehmann [7] and Zufryden [55] that utilize GLS-based methods to estimate multi-brand Markovian stochastic choice models.

Despite the potential merits of our study, certain study limitations should be noted. First, our study investigates conditional switching behavior. That is, we do not attempt to predict when an Internet
user will be active. Rather, we predict what site (or site category) will be visited given that the user is active. Thus, based on the occurrence of a given visit, we provide a conditional probability that a particular Website will be selected given the setting of the situational variables. Hence, a natural extension of our study would be to jointly study the visit frequency and conditional switching behavior, within a comprehensive model structure [54].

The only aspects of the previous visits that we take into account are which site was visited and the duration of the visit. However, there is clearly more information obtainable from the data gathered during each visit. Portals such as Yahoo! or AOL are very heterogeneous in their content and site visitors can use them for many different purposes. What visitors do on the site should tell us something about their likelihood of revisit or switching. For example, a user who visits Yahoo! to check stock quotes or play games is probably more likely to return to Yahoo! than a user who visits Yahoo! merely to search for information. Indeed, this second user is likely to find a link on Yahoo! that will send him/her to another Website where the information is located - thereby creating an immediate switch to another site. Hence, an understanding of within-visit behavior might further contribute to the explanation of switching patterns.

In our study, we capture consumer heterogeneity in a classical way by considering user characteristics and visit history to capture differences across consumers. Alternatively, one could potentially use hierarchical Baysian methods to model consumer heterogeneity [1, 2, 8]. However, because hierarchical Bayesian methods are more complex and considerably less tractable with respect to implementation, they were not considered in this study.

Finally, although this study suggested a flexible general model framework, that can incorporate relevant marketing mix variables, given limitations in our data, an empirical evaluation of the latter’s effects was not undertaken here. Hopefully, as more comprehensive online data including information about marketing mix variables becomes more readily available, the proposed model can be refined so as to include marketing effects.
Nevertheless, despite its limitations, this study illustrates how the proposed switching model framework may be used to provide potentially valuable insights and a better understanding of how consumers choose portal Websites. As shown in this study, this can be instrumental in the formulation of marketing strategies to enhance a portal Website’s ability to draw and retain consumers. Consequently, it is expected that future model refinements should provide further behavioral insights as well as aid managers in more effectively evaluating and planning their marketing efforts on the Web.
REFERENCES


## Table 1: Maximum Likelihood Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1086.25</td>
</tr>
<tr>
<td>X1 Former Yahoo!</td>
<td>4</td>
<td>6055.61</td>
</tr>
<tr>
<td>X2 Former AOL</td>
<td>4</td>
<td>3219.61</td>
</tr>
<tr>
<td>X3 Former Netscape</td>
<td>4</td>
<td>3975.69</td>
</tr>
<tr>
<td>X4 Former Other Portals</td>
<td>4</td>
<td>0196.61</td>
</tr>
<tr>
<td>X5 Logon Time (6pm-midnight)</td>
<td>4</td>
<td>21.71</td>
</tr>
<tr>
<td>X6 Logon Day (Mon-Wed)</td>
<td>4</td>
<td>13.18*</td>
</tr>
<tr>
<td>X7 Total Time Spent</td>
<td>4</td>
<td>369.69</td>
</tr>
<tr>
<td>X8 Heavy User</td>
<td>4</td>
<td>210.76</td>
</tr>
<tr>
<td>X9 Male</td>
<td>4</td>
<td>25.33</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>2336</td>
<td>4121.03</td>
</tr>
</tbody>
</table>

* Italicized value is significant at the 0.05 level. All other values are significant at the 0.01 level.
### Table 2: Summary of Estimated Parameters from MLE Method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intercept</th>
<th>Yahoo Site</th>
<th>AOL Site</th>
<th>Netscape Site</th>
<th>Other Portal Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln[P(Yahoo!</td>
<td>Xk) / P(non-portals</td>
<td>Xk)]</td>
<td>-3.4149</td>
<td>-3.7585</td>
<td>-3.9804</td>
</tr>
<tr>
<td>ln[P(AOL</td>
<td>Xk) / P(non-portals</td>
<td>Xk)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln[P(Netscape</td>
<td>Xk) / P(non-portals</td>
<td>Xk)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln[P(Other portals</td>
<td>Xk) / P(non-portals</td>
<td>Xk)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.4149</td>
<td>-3.7585</td>
<td>-3.9804</td>
<td>-2.2092</td>
<td></td>
</tr>
<tr>
<td>(0.1910)</td>
<td>(0.1953)</td>
<td>(0.1960)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former Yahoo</td>
<td>4.4168</td>
<td>0.3489</td>
<td>1.4677</td>
<td>1.1563</td>
<td></td>
</tr>
<tr>
<td>(0.0570)</td>
<td>(0.2570)</td>
<td>(0.1371)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former AOL</td>
<td>0.4294</td>
<td>4.3711</td>
<td>0.0575</td>
<td>1.3149</td>
<td></td>
</tr>
<tr>
<td>(0.2495)</td>
<td>(0.0775)</td>
<td>(0.3086)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former Netscape</td>
<td>1.8828</td>
<td>0.3219</td>
<td>4.4601</td>
<td>1.4770</td>
<td></td>
</tr>
<tr>
<td>(0.1196)</td>
<td>(0.3224)</td>
<td>(0.0710)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former Other Portals</td>
<td>1.4172</td>
<td>1.6113</td>
<td>1.5658</td>
<td>4.8122</td>
<td></td>
</tr>
<tr>
<td>(0.1092)</td>
<td>(0.1285)</td>
<td>(0.1173)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logon Time (6pm-midnight)</td>
<td>0.0867</td>
<td>-0.0258</td>
<td>-0.0631</td>
<td>-0.1909</td>
<td></td>
</tr>
<tr>
<td>(0.0526)</td>
<td>(0.0730)</td>
<td>(0.0631)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logon Day (Mon- Wed)</td>
<td>0.0597</td>
<td>0.1052</td>
<td>0.1732</td>
<td>0.0862</td>
<td></td>
</tr>
<tr>
<td>(0.0521)</td>
<td>(0.0698)</td>
<td>(0.0610)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Time Spent</td>
<td>-0.0318</td>
<td>0.0906</td>
<td>0.1157</td>
<td>0.0170</td>
<td></td>
</tr>
<tr>
<td>(0.0079)</td>
<td>(0.0083)</td>
<td>(0.0077)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy User</td>
<td>-0.4134</td>
<td>-0.6824</td>
<td>-0.5414</td>
<td>-1.3667</td>
<td></td>
</tr>
<tr>
<td>(0.1825)</td>
<td>(0.1796)</td>
<td>(0.1830)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.0439</td>
<td>-0.2927</td>
<td>0.1071</td>
<td>-0.0865</td>
<td></td>
</tr>
<tr>
<td>(0.0528)</td>
<td>(0.0693)</td>
<td>(0.0625)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Bold type values denote significance at the 0.01 level. Italicized values indicate significance at the 0.05 level. The numbers in parentheses are standard errors of the estimates.
Table 3: Sample Switching Probabilities

| Site          | P[site | Non-Portal] | P[site | Yahoo!] |
|---------------|----------------|-------------|
| Yahoo!        | 0.028          | 0.650       |
| AOL           | 0.020          | 0.008       |
| Netscape      | 0.016          | 0.019       |
| Other Portals | 0.093          | 0.083       |
| Non-Portals   | 0.833          | 0.239       |
Table 4: Actual vs. Estimated Switching Probability Matrices

### Actual Switching Probability Matrix*

<table>
<thead>
<tr>
<th>State**</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy user-Male</td>
<td>0.6253</td>
<td>0.0034</td>
<td>0.0229</td>
<td>0.0256</td>
<td>0.3228</td>
</tr>
<tr>
<td>Light user-Male</td>
<td>0.8108</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0811</td>
<td>0.1081</td>
</tr>
</tbody>
</table>

### Expected Switching Probability Matrix*

<table>
<thead>
<tr>
<th>State**</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy user-Male</td>
<td>0.6014</td>
<td>0.0078</td>
<td>0.0246</td>
<td>0.0299</td>
<td>0.3365</td>
</tr>
<tr>
<td>Light user-Male</td>
<td>0.6305</td>
<td>0.0127</td>
<td>0.0322</td>
<td>0.0846</td>
<td>0.2400</td>
</tr>
</tbody>
</table>

* Summary of Chi-Square Tests for Each Cell

<table>
<thead>
<tr>
<th>Cell</th>
<th>Chi-Sq V</th>
<th>df</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell1</td>
<td>1.000000</td>
<td>16</td>
<td>1.0000</td>
</tr>
<tr>
<td>Cell2</td>
<td>1.000000</td>
<td>16</td>
<td>1.0000</td>
</tr>
<tr>
<td>Cell3</td>
<td>1.000000</td>
<td>16</td>
<td>1.0000</td>
</tr>
<tr>
<td>Cell4</td>
<td>0.999995</td>
<td>16</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

** State: 1=Yahoo!, 2=AOL, 3=Netscape, 4=Other Portals, 5=Non-portal sites.
Table 5: Evaluation of Future Share Predictions for Holdout Sample

<table>
<thead>
<tr>
<th>Site</th>
<th>Heavy user-Male Segment</th>
<th>Light user-Male Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
<td>Estimated</td>
</tr>
<tr>
<td></td>
<td>Share</td>
<td>Share</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>0.0629</td>
<td>0.0539</td>
</tr>
<tr>
<td>AOL</td>
<td>0.0169</td>
<td>0.0195</td>
</tr>
<tr>
<td>Netscape</td>
<td>0.0325</td>
<td>0.0343</td>
</tr>
<tr>
<td>Other Portals</td>
<td>0.0735</td>
<td>0.0906</td>
</tr>
<tr>
<td>Non-Portals</td>
<td>0.8143</td>
<td>0.8016</td>
</tr>
<tr>
<td>Mean Absolute% Error</td>
<td>12.05%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Heavy user-Female Segment</th>
<th>Light user-Female Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
<td>Estimated</td>
</tr>
<tr>
<td></td>
<td>Share</td>
<td>Share</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>0.0661</td>
<td>0.0496</td>
</tr>
<tr>
<td>AOL</td>
<td>0.0203</td>
<td>0.0304</td>
</tr>
<tr>
<td>Netscape</td>
<td>0.0253</td>
<td>0.0291</td>
</tr>
<tr>
<td>Other Portals</td>
<td>0.1052</td>
<td>0.0987</td>
</tr>
<tr>
<td>Non-Portals</td>
<td>0.7832</td>
<td>0.7922</td>
</tr>
<tr>
<td>Mean Absolute% Error</td>
<td>19.43%</td>
<td></td>
</tr>
</tbody>
</table>