State Dependence at Internet Portals

Avi Goldfarb
Joseph L. Rotman School of Management
University of Toronto
Toronto, ON M5S 3E6
Canada
agoldfarb@rotman.utoronto.ca.

This study offers evidence of the existence of switching costs on the Internet. It uses more flexible methods than previously possible to separate switching costs from serially correlated unobservables at Internet portals. The data contain nearly 1,000 observations per household, allowing for household-specific regressions that control for all household-specific heterogeneity. The results show that households exhibit switching costs. The loyalty generated by these costs drives a large fraction of portal visits and generates considerable revenues; however, these revenues are not large enough to justify the losses incurred by Internet portals in the 1990s while building market share. The results also suggest that random coefficients models overestimate true state dependence.

1. Introduction

Consumers who have purchased a particular brand in the past often face (real or perceived) costs of purchasing a different brand. Klemperer (1995) and others have shown that these "switching costs" give firms market power over their repeat customers. If switching costs are large enough, they may imply that current market shares drive future profitability. Furthermore, switching costs may discourage entry, dampen incentives to differentiate, and generally reduce competitiveness.

The perception that significant switching costs would provide above-normal returns in the long run led many Internet businesses in the 1990s to incur short-run losses in pursuit of market share. Fast Company

This research was supported by the Social Science Research Council's Program in Applied Economics Predissertation Fellowship, by a Plurimus Corporation Fellowship, and by a SSHRC grant number 538-02-1013. Thanks to Plurimus Corporation for providing me with the clickstream data, and to David Barth, Shane Greenstein, Ig Horstmann, Charles Manski, Sridhar Moorthy, Robert Porter, two anonymous referees, the coeditor, and many seminar participants for helpful comments. All remaining errors are my own.

Journal of Economics & Management Strategy, Volume 15, Number 2, Summer 2006, 317–352

magazine (Darsey, 1999, p. 198) noted, "The irony is that you don't have to be the best. You just need to be there fast." Many economists, however, argued that switching costs on the Internet should be negligible. Shapiro and Varian (1999, p. 110) emphasize that the competition is just one mouseclick away, and that switching from reading information on one web site to another involves even less effort than switching magazines. Similarly, Gandal (2001, p. 1105) claims that "there are little (if any) consumer switching costs" at Internet portal web sites. This leads him to conclude that competition is fierce and first-mover advantages are short-lived. If these authors are correct and switching costs are low, there is little reason to suffer short-run losses to gain customers. Without the glue of switching costs, web sites need to find other ways to protect themselves from fierce competition, for example, by differentiation. If industry insiders are correct and switching costs on the Internet are large, then introductory offers and entry-deterring behavior become useful tactics.

This paper contributes to this discussion by seeking to identify the extent of switching costs in the Internet portal market.[1] The paper examines data on every web site visited by 2,651 households from December 27, 1999 to March 31, 2000, a total of 2,645,778 observations. Switching costs are identified by testing the null hypothesis that the current web site visit is independent of the previous web site visit, even controlling for all household-level heterogeneity. Because of the richness of the data set, household-level heterogeneity can be controlled with a separate regression for every household. The results suggest that users of Internet portals do face switching costs. Simulations using the regression coefficients indicate that switching costs generate between 11% and 15% of market share for the most popular portals. This result stands in contrast to Gandal (2001), who finds no evidence of switching costs for search engines. At the same time, these switching costs estimates are not nearly as large as those found in other markets. Shum's (2004) estimates of the effect of switching costs on purchase probability in ready-to-eat breakfast cereals are at least six times larger than the estimates for Internet portals in this paper. Keane (1997) suggests that switching costs in the ketchup market generate quantity increases of 75% or more. Although nontrivial, switching costs for Internet portals are small relative to consumer products and not large enough to justify the large losses incurred by many Internet companies.

1. The paper focuses on the Internet portal market because it is large and competitive. Over 25% of all web site visits in the data are to Internet portals, making it by far the largest category. According to Nielsen/Netratings, Yahoo's web site had 81 million different home users in December 2004; its market capitalization in February 2005 was $49 billion.

There are a number of possible explanations for switching costs online. Johnson, Bellman, and Lohse (2003) emphasize the cost of thinking as the key determinant of switching costs. They label this as "cognitive switching costs." Klemperer (1995) also describes a number of possible reasons for switching costs that could be relevant in the portal market: physical investment (personalization), transactions costs (search time), learning, brand loyalty, and cognitive dissonance problems. This paper does not distinguish between these alternative explanations, an identification task that is beyond even the detailed data used here.

The identification of switching costs is notoriously difficult in any analysis of consumer behavior. The core problem is to separate the direct effect of past choices ("true state dependence") from serially correlated unobservables ("spurious state dependence").[2] Heckman (1981, p. 115) makes the point best when he argues that if individuals have different preferences, "and if these differences [in individual preferences] are not properly controlled, previous experience may appear to be a determinant... of future experience solely because it is a proxy for temporally persistent unobservables that determine choices." True state dependence, Heckman argues, is best identified with a truly exogenous shock that forces individuals to choose something that they otherwise would not have chosen.[3] Such shocks are typically not available. Therefore, switching costs are usually identified by testing the null hypothesis that, controlling for household-level heterogeneity, the current choice is independent of the previous choice. The previous literature has resorted to distributional assumptions on household-level heterogeneity to try to separate product preferences from state dependence. For example, Jain, Vilcassim, and Chintagunta (1994) assume that there are a small number of household types and that households of the same type have the same preferences. Keane (1997) assumes household preferences are normally distributed.[4]

2. Throughout the paper, the terms "true state dependence," "switching costs," and "loyalty" are used interchangeably to make the arguments clearer. The econometric method identifies true state dependence. The implications of true state dependence relate to the economic concept of switching costs and the marketing concept of loyalty.

3. Greenstein (1993) and Israel (2005) measure switching costs using an exogenous shock. Greenstein exploits changes in hardware compatibility to identify switching costs in computer purchases. Israel exploits age-dependent changes in insurance prices to identify switching costs (which he labels "true tenure dependence") in automobile insurance.

4. Despite their importance to online strategies, there has been no systematic econometric study to separately identify online switching costs from serially correlated unobservables. Gandal (2001) and Chen and Hitt (2002) do assess the significance of online switching costs. Gandal finds little evidence of switching costs; however, because of data constraints he does not control for household preferences. Chen and Hitt, relying on the strong assumption that new and existing customers have the same average preferences,

In this paper, however, the rich data set means that distributional assumptions on household-level differences are not necessary. Although the identification still relies on household controls rather than Heckman's ideal of an exogenous shock, this paper identifies switching costs more precisely than much of the previous literature. The results of a comparison with the various random coefficients models previously used in the literature show that random coefficients models overestimate the switching costs relative to household-specific regressions.[5] The random coefficients models that assume normal distributions over as many coefficients as possible perform best. This implies that structural discrete choice models that rely on random coefficients for identification of substitution patterns (e.g., Berry, Levinsohn, and Pakes, 1995) should assume normal distributions over consumer heterogeneity in as many coefficients as is feasible.

The next section presents the general model and econometric specification. Section 3 describes the data set and variable construction. Section 4 presents the results of the main model, robustness checks, comparisons with random coefficients models, measures of the magnitude of the true state dependence, and correlations between true state dependence and household characteristics. Section 5 concludes that users likely face economically significant true state dependence at Internet portals.

2. General Model and Estimation

2.1 Model

The decision to visit a web site is modeled as a standard discrete choice problem. Users choose the portal that they expect will give them the highest utility. The choice set consists of three types of web sites. First, there are the top 18 Internet portals in the data. This group consists of all portals with over 0.5% market share in the portal category.[6] Second, there are a number of fringe portals in the data with very small market shares. In the analysis, these are combined into a composite 19th portal. Third, there are destination web sites that are visited without a prior

find evidence for switching costs for online brokers. Neither of these papers uses rich household-level controls. Furthermore, neither estimates the size of the switching costs.

5. Household-specific regressions are also less computationally intensive. Estimation of random coefficients models was limited to six firms and 100 households. For the household-specific regressions, 18 firms and 2,651 households were used.

6. There was a natural cutoff between the 18th and 19th most common portals in the sample. The 19th is a local Pennsylvania portal with only 0.5% market share that likely ranked highly because of the peculiarities of the data. Furthermore, any portal with less than 0.5% market share is certainly fringe.

portal visit. The expected utility for household i from visiting any portal j (top 18 or fringe) at time t is[7]

$$Eu_{ijt} = X_{ijt}\beta_{ij} + \varepsilon_{ijt}, \tag{1}$$

where X_{ijt} includes a loyalty variable and time-varying characteristics of web site j and household i, β_{ij} is a vector of household-specific coefficients that may vary by web site, and ε_{ijt} is an idiosyncratic error term. Users also have an outside option of not visiting a portal at all and instead going directly to a destination web site with utility:[8]

$$Eu_{i0t} = Z_{i0t}\gamma_i + \varepsilon_{i0t}, \tag{2}$$

where Z_{i0t} includes characteristics of the destination goal of the online session and time-varying characteristics of household i, γ_i is a vector of household-specific coefficients, and ε_{i0t} is an idiosyncratic error term.[9]

To estimate the probabilities of visiting the portals properly, it is necessary to make assumptions about the joint distribution of the error terms. A nested logit error structure is used for three reasons. First, and most important, it allows simulations to include an outside good: going directly to a web site. This ensures that an improvement in portal quality can increase the portal market size. Second, the nested logit avoids the independence of irrelevant alternatives (IIA) problem between the nests. Although controlling for household-specific heterogeneity alleviates some of the IIA problem, IIA may exist in the form of unobserved heterogeneity across online sessions by a given household.[10] Third, the nested logit allows inclusion of the fringe portals in the estimation in a sensible way. The problem is that fringe portals are consistently much worse in measured performance than other portals. Therefore, using them as an aggregated alternative in the bottom nest along with the 18 other portals would make it seem that households often choose a

7. It is expected utility based on the expectations of the user, not those of the observer, that is of interest. Because portals are experience goods, the user does not know in advance the utility gained from a visit. The user forms expected utility based on past experience at the portal. For example, the user does not know how long she will spend at the web site. The user does, however, have an expectation of how long it will take based on her past experience at that web site.

8. Portals are defined as distinct from destination web sites. For example, YahooNews is considered to be a distinct site from the Yahoo portal.

9. The appendix details the calculation of destination goals.

10. IIA suggests that any two brands with identical market shares will be equally affected by price changes of any other brand. Berry, Levinsohn, and Pakes (1995, p. 847) use the example of Yugo and Mercedes having the same market share, and therefore, being equally affected by a change in the price of any third car. Although IIA still exists for the top 18 portals at the household level, the IIA controls are better than most previous studies due to the household-level brand preferences.

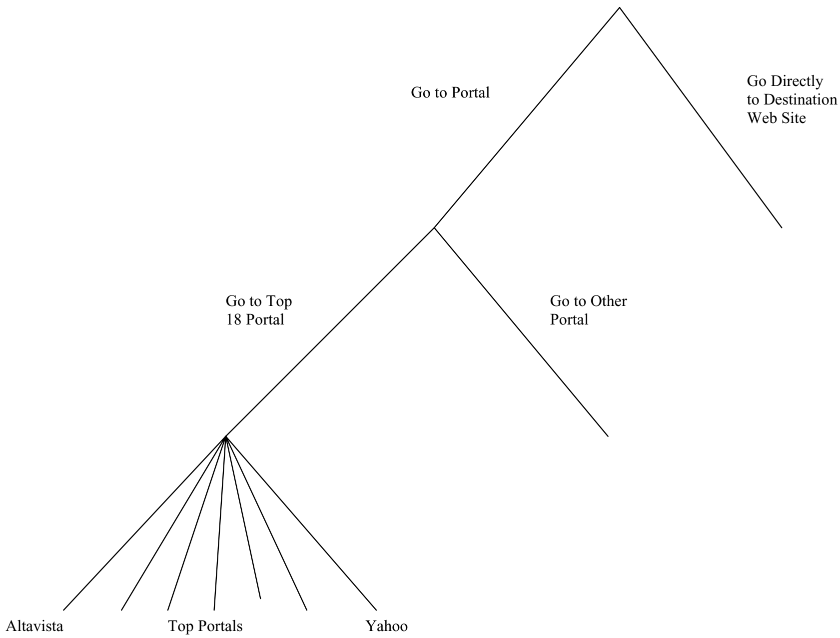

FIGURE 1. NESTING STRUCTURE

portal with poor measured performance. This will skew the estimated impact of measured performance downward. In fact, fringe portals are by definition used infrequently. Having a separate nest from the top 18 portals avoids skewing results in this way.

The assumed nesting structure is displayed in Figure 1. This structure implies that the expected utility from visiting one from the top portals ($j = 1, \ldots, 18$) is defined by

$$Eu_{ijt} = \delta_i \sigma_i (X_{ijt}\beta_{ij} + \varepsilon_{ijt}), \tag{3}$$

where δ_i is the inclusive value coefficient for the middle nest, σ_i is the inclusive value coefficient for the bottom nest. To be consistent with utility theory, σ_i and δ_i are constrained to be between zero and one through the functional forms: $\sigma_i = \exp(\underline{\sigma}_i)/(1 + \exp(\underline{\sigma}_i))$ and $\delta_i = \exp(\underline{\delta}_i)/(1 + \exp(\underline{\delta}_i))$. The expected utility from visiting one of the fringe portals ($j = 19$) is

$$Eu_{i19t} = \delta_i (X_{i19t}\beta_{i19} + \varepsilon_{i19t}). \tag{4}$$

The utility from visiting a destination web site ($j = 0$) is still represented by (2).

There are two important assumptions embedded in this setup. First, consumers are assumed to view each web site as a separate company. Therefore, web sites that are owned by the same parent (such as America Online (AOL) and Netscape, or Hotbot and Lycos) are not grouped together in any way. Brand-fixed effects should control for many of the issues, especially because the market structure changed little over the period in question (no mergers or bankruptcies occurred among the top 25 portals).

Second, there are no forward-looking dynamics in this model. Consumers are assumed to (i) ignore the lock-in effect when they decide which web site to visit and (ii) not consider the option value of visiting a web site to learn its characteristics and make better future choices. While allowing for dynamics would be ideal, the dynamic model becomes intractable without unrealistic assumptions. If consumers do anticipate lock-in, then the estimates of true state dependence will be biased toward zero. In other words, nonmyopic consumers will switch portals to overcome potential harm due to true state dependence. In the data, however, these appear as switches resulting from features and the idiosyncratic error. Therefore, this bias does not change the result of rejecting the hypothesis of zero true state dependence. It will, however, bias the estimated impact on visits and revenues toward zero.

2.2 Estimation

To derive the likelihood function for a particular household, first define the inclusive value for the bottom branch,

$$IV_{B_{it}} = \ln\left(\sum_{j=1}^{J_i} \exp(X_{ijt}\beta_{ij})\right) \tag{5}$$

and the value for the middle branch,

$$IV_{M_{it}} = \ln(\exp(X_{i19t}\beta_{i19}) + \exp(\sigma_i IV_{B_{it}})). \tag{6}$$

The probability of household i visiting portal j given that it visits one of the main portals is

$$P_{ijt}(X_{ijt}, \beta_{ij}) = \frac{\exp(X_{ijt}\beta_{ij})}{\sum_{k=1}^{J_i} \exp(X_{ikt}\beta_k)}. \tag{7}$$

If instead household i visits a fringe portal the probability, given that the household visits a portal, is

$$P_{i19t}(IV_{iB_{it}}, X_{i19t}, \beta_{i19}, \sigma_i) = \frac{\exp(X_{i19t}\beta_{i19})}{\exp(X_{i19t}\beta_{i19}) + \exp(\sigma_i IV_{iB_{it}})}. \tag{8}$$

Here, the denominator includes the inclusive value for the bottom nest. It is this implication of the nested logit error structure that ensures the composite fringe portal does not bias the results. The probability of visiting any one of the main portals given that a household visits a portal is

$$P_{iBt}(IV_{iB_{it}}, X_{i19t}, \beta_{i19}, \sigma_i) = \frac{\exp(\sigma_i IV_{iB_{it}})}{\exp(X_{i19t}\beta_{i19}) + \exp(\sigma_i IV_{iB_{it}})}. \tag{9}$$

Equations (8) and (9) represent the probability of visiting top and fringe portals, conditional on visiting any portal. Therefore, they sum to one. If household i visits any portal, the probability is

$$P_{iMt}(IV_{iM_{it}}, Z_{i0t}, \gamma_i, \delta_i) = \frac{\exp(\delta_i IV_{iM_{it}})}{\exp(Z_{i0t}\gamma_i) + \exp(\delta_i IV_{iM_{it}})}. \tag{10}$$

Alternatively, the probability of going directly to a destination web site is

$$P_{i0t}(IV_{iM_{it}}, Z_{i0t}, \gamma_i, \delta_i) = \frac{\exp(Z_{i0t}\gamma_i)}{\exp(Z_{i0t}\gamma_i) + \exp(\delta_i IV_{iM_{it}})}. \tag{11}$$

The probabilities in the top nest, represented by (10) and (11), rely on the inclusive values for visiting a portal (equation (6)) and the characteristics of the goal of the search (Z_{i0t}). The likelihood function for household i is as follows:

$$\prod_{t=1}^{T_i} \left(\left(P_{iMt} \left(P_{iBt} \prod_{j=1}^{J_i} P_{ijt}^{d_{ijt}} \right)^{1-d_{i19t}} P_{i19t}^{d_{i19t}} \right)^{1-d_{i0t}} P_{i0t}^{d_{i0t}} \right), \tag{12}$$

where d_{i0t} is equal to 1 if the household goes directly to a destination web site and 0 otherwise, d_{i19t} is equal to 1 if the household goes a fringe portal and 0 otherwise, and d_{ijt} is equal to 1 if the household goes directly to portal j and 0 otherwise. The estimation for each household uses full information maximum likelihood.

Note that the general framework here is similar to that used in random coefficients models. There are, however, several important differences. First, the typical random coefficients model of true state dependence does not allow for heterogeneity in variables aside from the brand preferences (e.g., Jain, Vilcassim, and Chintagunta, 1994). There are, however, a few studies in the literature have allowed for heterogeneity in some of the other coefficients (e.g., Seetharaman, Ainslee, and Chintagunta, 1999). Second, existing models impose distributional assumptions on the nature of the allowed heterogeneity. As a result, the possible shape of the heterogeneity distribution and, consequently, each household's degree of true state dependence may not be separately and accurately measured. Here, household-specific regressions are estimated and therefore do not impose a distribution on household-level heterogeneity.[11]

Because the method allows for household-specific estimation, there were some households for which a lack of data prevented an estimate of the full structure. For instance, 28% of sample households never visit a fringe portal. If this is the case, the middle nest is dropped and the model is calculated with only one nest. Similarly, there are two households who only visit top 18 portals, and no other web sites at all. For them only the bottom nest is calculated. There are also a large number of households who only visit one portal or who do not visit enough portals for the regression to be meaningful (less than 10 degrees of freedom at the bottom nest). A standard logit model is calculated for these households where they choose between visiting any portal and going directly to a web site.

3. Data

3.1 Raw Data Sources and Description

The raw data set, courtesy of Plurimus Corporation, consists of 3,228,595 web site visits over 399,613 sessions by 2,651 households from December 27, 1999 to March 31, 2000. A total of 2,645,778 of these visits occur when a household either goes to a portal or goes directly to a destination web site without visiting a portal.[12] Web sites that are visited after a portal visit are part of a separate decision process than the one explored here. Also included in the initial data set are the arrival and departure times

11. Several other studies have recommended regressions on the time-varying dimension of a panel data set. For example, Fischer and Nagin (1981) explored whether taste parameters vary across individuals in an experimental setting. They then compared random coefficients models to models without heterogeneity, and argued that the individual-specific regressions identify tastes more accurately. Pesaran and Smith (1995) performed separate regressions on employment functions in 38 separate industries. They concluded that the "lesson for applied work is that when large T panels are available, the individual micro-relations should be estimated separately." Elrod and Haubl (1998), however, described two shortcomings of individual-specific regressions. First, they argue that they are inefficient. This is not a large issue in a data set with nearly 1,000 observations per household. Second, they cite Davey (1991) in arguing that the true population variance is overestimated because the variance in the estimates is the sum of the true variance plus the estimation error variance. The means of the coefficients, however, are consistent, and the variance of the heterogeneity is accurately measured at the sample level. Therefore, the main results of this study relating to the estimated mean values on true state dependence are not affected by this criticism of individual-specific regressions.

12. The remaining visits are to a destination web site immediately subsequent to a portal visit. As discussed in the previous section, because this study examines portal choice, the outside good is only consumed when households go directly to a web site.

TABLE I.
CLICKSTREAM DATA SAMPLE

User	Host	Start Time	End Time	Bytes from	Bytes to	No. of Pages Viewed at Host
1	com.yahoo	14MAR00:08:42:55	14MAR00:08:45:28	196593	34484	3
1	com.allrecipes	14MAR00:08:45:28	14MAR00:08:50:59	65825	656	12
1	com.ivillage	14MAR00:08:55:00	14MAR00:09:09:48	541337	72005	53
1	com.allrecipes	18MAR00:12:27:10	18MAR00:12:34:46	75403	4454	5
1	com.allrecipes	21MAR00:18:31:01	21MAR00:18:36:51	75873	658	2
1	com.excite	28MAR00:13:13:59	28MAR00:13:15:04	105884	4006	4
1	com.adobe	28MAR00:13:15:06	28MAR00:13:19:38	70732	11988	9
1	gov.nara	28MAR00:13:19:38	28MAR00:13:21:57	1259	2340	1
1	gov.nara	28MAR00:13:34:09	28MAR00:13:38:00	60155	9074	13
1	com.allrecipes	30MAR00:16:44:18	30MAR00:16:52:05	86186	1857	4

for a web site visit and the number of pages viewed. Table I displays a sample of the raw clickstream data.

Plurimus, which no longer operates independently, had an anonymizing technology that allowed them to collect information about users without the users' permission. The users are anonymous and the data cannot be traced to any specific person. Plurimus was regularly audited by PricewaterhouseCoopers to ensure that it exceeded the privacy requirements of the FCC guidelines. Unlike volunteer panel data, behavioral records from anonymous users are not biased by the wish to be seen in a socially desirable light. Moreover, there is no selection bias into the sample itself, yielding a sample from a broader spectrum of socioeconomic status than is typically available from panel studies.

The data set, however, has some limitations. First, the geographic distribution of the sample is not representative. New York, Chicago, and Los Angeles are underrepresented. Roughly half the sample comes from the Pittsburgh area. Another quarter is from North Carolina and another eighth from Tampa. This problem is not severe to the extent that portals are a national product.

The second limitation is that AOL subscribers are not included. Because AOL subscribers made up roughly 50% of all American home Internet users in 2000, this could skew the results. Preliminary surveys commissioned by Plurimus show that AOL users have similar habits to other Web users when not on AOL web sites. The data do, however, undercount visits to the AOL portal.

Third, the data set contains information on few users at work. Online habits at work likely differ from those at home; however, according to a study by Nie and Erbring (2000), 64.3% of Internet users use the Internet primarily at home; just 16.8% use it primarily at work. Few data sets contain reliable at-work panel data.

The fourth limitation is that the data are collected at the household level rather than at the individual level, and there is no information on the household composition. If two people in a given household have different habits, this will show up as one person with varying habits. There is also no household size data that would allow me to look only at households with one member. While this makes it difficult to assess the extent of learning over time, it is a standard problem in consumer panels.

Fifth, the data do not contain information on households from the first time they go online. Therefore initial conditions are potentially a problem. However, this problem is partially alleviated by the law of large numbers because of the number of observations per household in the data set. More than 95% of the households in the final data set make more than 30 choices. The mean household makes 998 choices and the median household makes 515 choices. All regressions have at least 10 degrees of freedom.

Together, these five data limitations mean that results should be extended to different geographic distributions, AOL users, and at-work users with caution. Furthermore, the fourth and fifth limitations make any study of learning behavior infeasible with this data. It is an open topic for further research. Still, despite these limitations, market shares measured using the Plurimus data are similar to those measured by MediaMetrix, Nielsen/Netratings, and PC Data Online.

To measure possible media effects on behavior, the clickstream data set is linked with a data set of "media mentions" for each of the relevant companies. If a portal is mentioned on network television news (ABC, CBS, or NBC), in the *Wall Street Journal*, in the *New York Times*, or in *USA Today* on a given day or the day before, then the media mentions variable is equal to one. Otherwise, it is equal to zero. Unfortunately, the data do not reveal whether an individual was actually watching or reading a given medium. It is likely, however, that mentions in these media are highly correlated with mentions in other media, such as local newspapers.

3.2 Data Description

Following Hargittai (2000, p. 233), an Internet portal is defined as "any site that classifies content and primarily presents itself as a one-stop point-of-entry to content on the Web." Portals, such as Yahoo, Altavista, and MSN have search engine capabilities, but they also have other features. There are few, if any, pure search engines remaining. Because this paper focuses on the portal as a starting point and not as a destination, portals are defined by main pages, directory pages, and search pages, not by email and shopping pages.

Table I shows 10 lines of raw data. Using only this information, the following variables were constructed: goal of search, view length at the portal, repeated search, total visits over the sample period to the destination web site, destination type, total time spent online over the sample period, percentage of all web site visits in each of 11 categories, and various measures of true state dependence.[13] There is insufficient variation in portal features, such as auctions and number of pages indexed, for these variables to be included in the regressions. Brand dummy variables and experience variables therefore control for features. Table II provides variable definitions, and the Appendix provides a detailed description of variable derivations. Table II also shows which variables were assigned to which nest. Variables were assigned to nests using the Baye's Information Criterion estimated on an aggregated (panel) model using a subset of the data.[14]

True state dependence is identified by whether the bottom nest variable *Loyalty* (*Last Session*) is significantly different from zero. This is a dummy variable for whether the portal was visited during the previous online session. This definition is used because it is more easily interpreted as a switching cost than other measures, such as weighted measures, a within-session measure, or a day-to-day measure. Weighted measures consist of weighted averages of past choices. These measures mix many visits together, confounding the idea of whether an individual has switched. Within-session measures use a dummy for whether the portal was previous portal visited, even if the previous visit occurred in the same online session. This measure is particularly susceptible to be misspecified because the data are at the household level. If different household members use different portals, then repeated use of a portal within a session may reflect these preferences rather than a switching cost. This issue is discussed in detail in Section 4.3. Day-to-day measures define the loyalty by whether the portal was visited on the previous day. Defining loyalty day-to-day may ignore web site visits that occurred on the same day that drive loyalty. Section 4.3 shows that the main results

13. Goldfarb (2002) presents the results of a questionnaire that informed the construction of the variables.

14. This is admittedly suboptimal. Ideally, variables would be selected using the main model. However, the computational burden of that model is high, and testing the dozens of variables that could potentially be included would take years of computer time. Consequently, the variables were included on the basis of a weaker method.

TABLE II.
VARIABLE DESCRIPTIONS AND SUMMARY STATISTICS

Variable	Description	Mean	SD	Minimum	Maximum
Bottom Nest					
Loyalty (last session)	Dummy variable equal to one if went to the same portal during the previous online session	0.0190	0.137	0	1
Last search repeated	Dummy variable for previous search at a portal followed by another search	0.106	0.307	0	1
Last view length	Length of time of the last search at a portal	57.1	171.4	0	91,226
Media mentions	Mentioned in major US media that day or the day before	0.168	0.374	0	1
Missing data	Dummy variable equal to one if have not yet visited that portal	0.491	0.500	0	1
Portal dummy variables (17 in total)	Dummy variables for the portals in the data set	N/A	N/A	0	1
Middle Nest					
Loyalty (last session, fringe)	Dummy variable equal to one if went to the same portal during the previous online session	0.0243	0.154	0	1
Other portal	Dummy variable equal to one if visit a fringe portal	N/A	N/A	0	1
Top Nest					
Times visited destination	Total times visited the destination website over the course of the sample	390.0	1485.0	1	16,699
Information destination	Dummy variable equal to one if destination is news, sports, weather, women's, technology, or general info.	0.288	0.453	0	1
Ecommerce destination	Dummy variable equal to one if destination is a shopping or an auction web site	0.177	0.382	0	1
Communication destination	Dummy variable equal to one if destination is email, chat, or other communication	0.185	0.388	0	1
Adult destination	Dummy variable equal to one if destination is pornography	0.173	0.378	0	1
Other web site	Dummy variable equal to one if go directly to destination	N/A	N/A	0	1
	Number of bottom nest (per household)	299.5	463.3	0	6810
	Number of middle nest (per household)	19.5	85.6	0	2850
	Number of top nest (per household)	679.0	1049.2	0	17,000
	Number of brands (per household)	7.9	3.9	0	18

are robust to two distinct weighted averages of past visits, to whether the portal was visited immediately beforehand, and whether the portal was visited the previous day.[15]

Tables II and III contain descriptive statistics for the data set used in estimation. Of particular interest in Table II is the average number of observations per household and the number of different portals visited per household in each nest. The average household visits one of the top 18 portals 299.5 times, one of the fringe portals 19.5 times, and another web site 679.0 times. On average, a household visited 7.9 different portals over the course of the sample, with 54% of these visits at their most frequently visited portal. This provides more than sufficient data to conduct household-specific regressions.

The top part of Table III shows that Yahoo is the most visited portal with roughly one third of all portal visits, and it has the lowest rate of repeat visits (which are interpreted as failures). Looksmart searches are fastest, and Go2net searches are slowest. Go2net searches are likely slowest because it is a "meta-search" engine that presents results from other search engines.

The bottom part of Table III summarizes the characteristics of the households. Age, education, household size, income, percentage of married, and percentage of renting are collected at the census block level. The other characteristics are derived from the observed behavior of the households. The total time online variable shows that the average household spends 15 hours per month online.

4. Results

4.1 General Results of the Household-Specific Regressions

Table IV presents the results of the household-specific regressions. The first column reports the (unweighted) mean of the coefficients, and the second column reports the standard error of that mean calculated as in Pesaran and Smith (1995).[16] Given that the means are calculated from 2.6 million observations, it is not surprising that all are significantly different from zero. The third column displays the standard deviation of the coefficients and the fourth column displays the number of households for which the coefficient can be calculated. The fifth

15. The calculation of the weighted averages is discussed in the Appendix.

16. In particular, the regression coefficients are assumed to be independent across households. Therefore, the variance of the mean is equal to the mean of the variance. Formally,

$$s\hat{e} = \frac{1}{\sqrt{I}} sqrt\left(\sum_{i=1}^{I} \frac{1}{N_i} \mathrm{var}(\hat{\beta}_i)\right).$$

Table III.
Portal and Household Characteristics

Portal	Percentage of Share of All Portal Visits	Average Time Spent at Site (in Seconds)	Percentage of Searches Repeated	Percentage of Days with Media Mentions
Altavista	4.0	109.7	11.9	5.2
AOL	4.3	93.9	10.6	82.3
Ask Jeeves	1.1	146.1	16.8	3.1
Excite	5.1	93.4	10.1	15.6
Go	2.0	138.9	10.1	15.6
Google	0.6	104.9	18.6	0
Go2net	1.5	280.3	11.1	3.1
Goto	1.5	94.2	23.9	1.0
Hotbot	1.8	90.3	18.8	1.0
Infospace	0.6	161.9	17.6	2.1
Iwon	2.6	152.0	14.8	1.0
Looksmart	0.7	70.1	31.5	0
Lycos	2.5	96.2	29.7	16.7
MSN	17.4	116.7	9.7	6.4
Myway	2.2	153.0	11.5	0
Netscape	10.6	114.0	9.6	13.5
Snap	1.7	91.0	13.0	7.3
Yahoo	33.1	96.7	5.1	58.3
Fringe portals	6.7	193.1	15.3	0

	Mean	SD	Minimum	Maximum
Data from Observed Behavior				
Total time online (seconds)	164,344	231,017	10	4,571,000
Nonsearch visits	679.0	1049.2	0	17,000
% Adult web sites	9.86	16.46	0	100
% Brochureware	3.09	5.43	0	100
% Classified web sites	1.31	4.44	0	100
% Communication web sites	2.37	16.20	0	100
% Entertainment web sites	4.79	7.22	0	75.51
% Finance web sites	4.93	10.31	0	100
% Information web sites	8.83	10.30	0	100
% Search web sites	29.80	16.0	0	100
% Shopping web sites	6.18	7.46	0	100
% Technology web sites	3.65	5.85	0	100
% Women's web sites	0.744	1.76	0	27.14
Data from Census Block Level Information				
Average age	38.74	8.55	13	71.30
Average education	13.85	1.42	8.9	16.10
Average household size	2.53	0.369	1	3.88
Average income	46,582	22,174	4,999	190,132
% Married	49.73	0.113	0	100
% Renting	10.43	0.122	0	100

Table IV.
General Model Coefficients

Variable	Unweighted Mean of Coefficients (a)	Standard Error of Mean (b)	Standard Deviation of Coefficients (c)	Number of Households (d)	Percent Significantly Positive at 95% Level (Two-Sided Test) (e)	Percent Significantly Negative at 95% Level (Two-Sided Test) (f)
Bottom Nest						
Loyalty (last session)	1.19	2.22E−05	2.66	2361	69.97%	4.28%
Last search repeated	−0.191	3.77E−05	5.90	2214	8.49%	16.49%
Log(last view length)	0.0477	4.44E−05	0.560	2361	0.76%	0.72%
Media mentions	0.0390	2.81E−05	2.13	2361	13.68%	9.19%
Missing data	1.05	5.49E−05	3.62	2361	24.14%	12.79%
Altavista	−1.61	3.24E−05	2.69	1367	12.95%	54.43%
AOL	−1.58	2.94E−05	2.70	1507	11.21%	57.66%
Ask Jeeves	−2.29	2.59E−05	2.40	829	6.63%	65.86%
Excite	−1.65	3.69E−05	2.85	1304	13.42%	55.75%
Go	−2.09	3.85E−05	2.73	1153	9.37%	61.58%
Go2net	−1.87	2.21E−05	2.53	507	13.21%	61.54%
Google	−2.16	3.25E−05	2.64	519	6.74%	62.04%
Goto	−2.05	2.51E−05	2.72	1340	5.75%	60.90%
Hotbot	−1.70	3.27E−05	2.45	1064	9.12%	57.42%
Infospace	−2.40	2.74E−05	2.34	863	3.24%	64.54%
Iwon	−1.75	3.33E−05	2.88	529	16.07%	58.41%
Looksmart	−2.38	2.47E−05	2.15	851	2.94%	64.98%
Lycos	−1.81	2.69E−05	2.48	1346	8.62%	59.29%
MSN	−0.165	4.06E−05	3.42	2011	31.48%	37.54%
Myway	−1.35	5.90E−05	3.67	432	18.98%	52.08%
Netscape	−0.799	3.20E−05	3.45	1529	26.55%	45.85%
Snap	−2.29	3.09E−05	2.53	639	7.67%	65.41%

Middle Nest						
Loyalty (last session)	0.232	4.77E−05	4.89	1917	41.05%	11.95%
Other portal	−1.46	3.00E−05	2.35	1917	10.07%	53.63%
σ	5.16	2.68E−04	10.35	1918	N/A	N/A
Top Nest						
Log(times at destination)	−0.392	3.15E−05	2.97	2635	3.68%	23.19%
Information destination	2.71	6.81E−05	5.83	2621	65.32%	3.40%
Ecommerce destination	3.02	6.58E−05	6.54	2591	61.56%	3.51%
Communication destination	2.52	6.11E−05	6.80	2532	56.67%	4.11%
Adult destination	3.06	5.81E−05	6.99	1888	58.37%	1.27%
Other web site	0.861	7.02E−05	4.71	2635	43.64%	10.06%
$\underline{\delta}$	5.52	2.22E−04	9.04	1918	N/A	N/A
Log likelihood	−773.34	N/A	1329.45	2651	N/A	N/A
Number of observations	998.03	N/A	1385.10	2651	N/A	N/A

All mean coefficients are significantly different from zero at the 99% confidence level and sixth columns report the percentage of households for whom the coefficient is significantly positive and the percentage for whom the coefficient is significantly negative (at the 95% confidence level in a two-sided test). Although the magnitudes of the coefficients themselves are uninformative in the nested logit model, their distributions and their significance are both informative.[17]

The coefficient on the *Loyalty* (*Last Session*) variable in the bottom nest is used to identify true state dependence. As stated earlier, and following the literature, true state dependence is defined as a rejection of the null that this coefficient is equal to zero. Figure 2 shows the distribution of this coefficient across households. It shows that the *loyalty* coefficients are centrally distributed with a narrower distribution than the normal. Furthermore, Figure 2 and Table IV each show that the coefficient on *loyalty* is generally positive. In particular, in the 2,361 regressions that calculate this coefficient, it is significantly positive 69.97% of the time, and significantly negative only 4.28% of the time (at the 95% confidence level in a two-sided test). While a small fraction of users display negative state dependence, suggesting a preference for variety, most households are habitual.

Slightly over 25% of the households have coefficients on *loyalty* that are not significant. There are two possibilities for these households. First, there may simply not be enough power in the test at the household level; or second, they may in fact display no true state dependence or preference for variety. These explanations are not separately identified.

Much of the literature on product choice assumes coefficients are constant across households; however, this assumption is not supported here. Figure 3 shows the distribution of brand preferences. Only those households who visit Yahoo at least once are included, so that all households in the figure have the same base. These coefficients represent the preferences relative to Yahoo. The densities show that the distributions are all close to normal, and that preferences vary considerably from household to household. This is not only true of the brand dummy

17. Furthermore, the magnitude can be compared across models that rely on data drawn from the same distribution. The distributions of the coefficients are meaningful under a basic normalization that is also made in panel models. In particular, the probability of choosing portal j at time t is $F(X_{jt}\beta + \alpha_j + \varepsilon_{ij} > X_{kt}\beta + \alpha_k + \varepsilon_{kt})$ for all $k \neq j$. Let D^h be a dummy variable for portal h: $F((X_{jt} - X_{kt})\beta + (D^j_{jt} - D^j_{kt})\alpha_j + (D^k_{jt} - D^k_{kt})\alpha_k > \varepsilon_{ik} - \varepsilon_{ij})$. Now call D^h a part of X and α a part of β, and normalize by σ, the standard deviation of $(\varepsilon_{kt} - \varepsilon_{jt})$:

$$F\left((X_{jt} - X_{kt})\frac{\beta}{\sigma} > \frac{\varepsilon_{ik} - \varepsilon_{ij}}{\sigma}\right).$$

Under this normalization, the coefficients are all on the same scale and are therefore comparable.

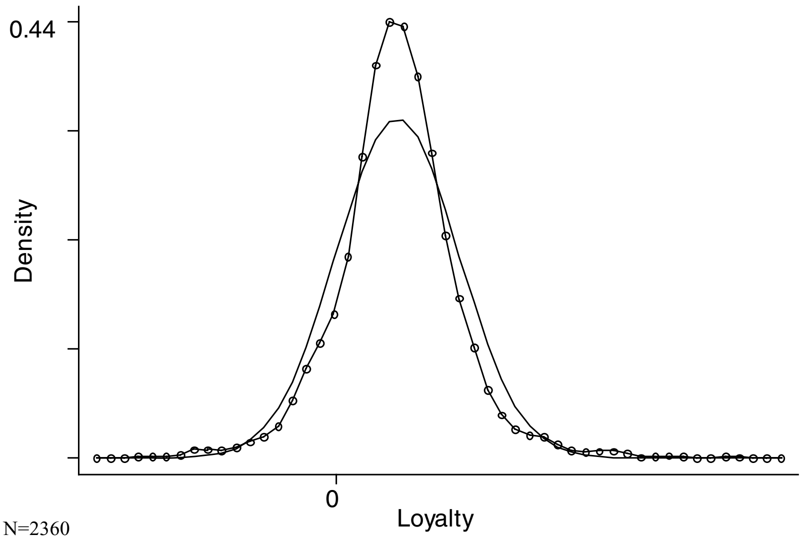

FIGURE 2. [18]*COMPARING THE KERNEL DENSITY ESTIMATE OF THE LOYALTY COEFFICIENT TO A NORMAL DISTRIBUTION*

coefficients; all of the coefficients in Table IV display considerable heterogeneity across households. The assumption of constant coefficients is not supported for any of the coefficients estimated.

In summary, the *loyalty* coefficient is usually positive, suggesting true state dependence.

4.2 Marginal Effects

The previous section shows that the mean *loyalty* coefficient is significantly different from zero. It does not give a sense of the economic importance of this measured true state dependence. This section presents different measures of the marginal impact of true state dependence on web site market shares and revenues. Table V provides simulations on the impact of true state dependence on Internet portals by simulating the impact of changes in past behavior on current choices. These simulations do not take into account the reactions by portal operators that such changes may induce in advertising and content, and therefore do not represent an equilibrium. They do, however, give an idea of the

18. The kernel density estimates (dotted lines) in Figures 2 and 3 use the Epanachnikov kernel and optimal bandwidth, but results are robustto specification. A solid line shows the normal distribution.

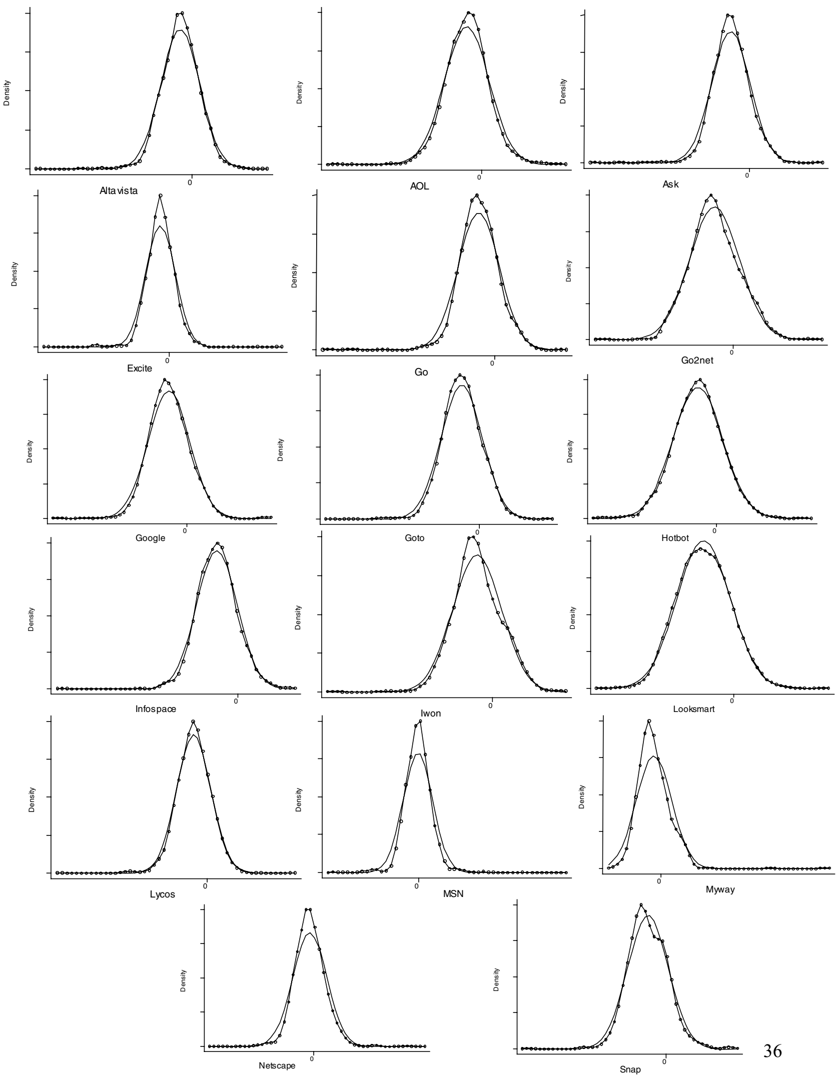

FIGURE 3. [19]KERNEL DENSITY ESTIMATES OF BRAND-FIXED EFFECTS

magnitude of the effects, but should be interpreted as elasticities rather than as prescriptions for strategic action. There are three main results in this section. First, the marginal effects are not trivial. Second, switching

19. Only households that visit Yahoo are included to keep the base of comparison constant.

Table V.
Marginal Effects: The Impact of Variable Changes on Market Shares (Standard Errors in Parentheses)

Portal	(a) Predicted Shares (%)	(b) Simulated Share of Web Site in the Week Following an Extended Shutdown Such That All Users Have No History at the Web Site (%)	(c) Effect of Loss of History in Percentage Terms $\frac{(a)-(b)}{(a)}$	(d) Estimated Weekly Value of True State Dependence to Portals[a]	(e) Simulated Probability Visit the Web Site Given Visited That Web Site Last Period (%)	(f) Market Share After 10% Increase in Own Repeated Searches (%)
Altavista	1.29% (0.00841)	1.11% (0.0738)	13.95%	$230,574	2.04% (0.0209)	1.29% (0.00910)
AOL	1.41% (0.00860)	1.24% (0.0354)	12.06%	$217,764	2.40% (0.0185)	1.41% (0.00754)
Ask Jeeves	0.329% (0.00324)	0.298% (0.00367)	9.42%	$39,710	0.569% (0.00520)	0.329% (0.00381)
Excite	1.69% (0.00527)	1.49% (0.00361)	11.83%	$256,193	4.97% (0.00831)	1.68% (0.00437)
Go	0.616% (0.00723)	0.537% (0.00586)	12.82%	$101,196	1.50% (0.0205)	0.616% (0.00743)
Go2net	0.498% (0.00486)	0.459% (0.00349)	7.83%	$49,958	0.760% (0.0126)	0.496% (0.0121)
Google	0.205% (0.00363)	0.181% (0.00372)	11.71%	$30,743	0.262% (0.00511)	0.205% (0.00749)
Goto	0.456% (0.00726)	0.425% (0.0123)	6.80%	$39,710	0.882% (0.0284)	0.455% (0.00521)
Hotbot	0.601% (0.00449)	0.552% (0.0145)	8.15%	$62,767	1.10% (0.0102)	0.600% (0.00536)
Infospace	0.183% (0.00398)	0.177% (0.0103)	3.28%	$7,686	0.316% (0.00817)	0.183% (0.00650)
Iwon	0.838% (0.00442)	0.711% (0.00641)	15.16%	$162,683	2.50% (0.00981)	0.838% (0.00876)
Looksmart	0.202% (0.00333)	0.191% (0.00349)	5.45%	$14,091	0.365% (0.00412)	0.202% (0.00336)
Lycos	0.819% (0.00572)	0.754% (0.00372)	7.94%	$83,263	1.32% (0.0291)	0.818% (0.00417)
MSN	5.62% (0.0140)	4.96% (0.0141)	11.74%	$845,437	11.30% (0.0134)	5.59% (0.0121)
Myway	0.692% (0.00437)	0.610% (0.0144)	11.85%	$105,039	0.892% (0.00719)	0.692% (0.0167)
Netscape	3.45% (0.00928)	2.98% (0.0113)	13.62%	$602,054	6.81% (0.0123)	3.44% (0.0122)
Snap	0.564% (0.00410)	0.497% (0.00682)	11.88%	$85,825	0.710% (0.00462)	0.564% (0.00681)
Yahoo	10.60% (0.0132)	9.03% (0.0911)	14.81%	$2,011,116	19.06% (0.0185)	10.57% (0.0112)
Other portal	1.97% (0.00715)					
Other web site	68.05% (0.0122)					

[a] Assumes 43.3 million online households during the relevant time period and a value to the web site of 4.01 cents per visit. This is a measure of the marginal effect of previous customers on current revenues and is based on the difference between the visit probabilities in columns (a) and (b).

costs are much larger for web sites with established loyalty programs, such as Iwon. Third, although the marginal effects are not trivial, they are probably not large enough to justify the enormous losses sustained by many of these web sites to build market share.

Column (a) of Table V provides the predicted market shares of the various brands based on imposing the estimated coefficients on the actual data. As expected, these shares correspond closely to the observed shares in the data. The other columns of the table show the results of a number of thought experiments. Columns (b), (c), and (d) simulate the effect of a loss of history at a particular web site. In particular, they assume that all consumers do not visit that web site in the previous period. Column (b) then simulates market shares for the following week, column (c) shows the relative size of the effect on market share, and column (d) provides a rough monetary estimate of the effect.[20] Though clearly hypothetical, these simulations suggest the marginal impact of the true state dependence on the portals. One way to think of this simulation is to imagine an extended web site shutdown during which users visit other web sites. Then, when the web site of relevance comes back online, assuming users' underlying preferences for the web site are unchanged, the marginal effect of the *loyalty* coefficient can be simulated because the previous web site visit was no longer at that web site.

There are two main results of columns (b), (c), and (d). First, the effect of state dependence is nontrivial. The simulations suggest that a temporary loss of customers can reduce market share by 3%–15%. Second, web sites with features aimed at inducing loyalty have higher switching costs. The second result is best seen with a comparison of Iwon and Lycos. Both portals have similar predicted shares; however, Iwon has a rich loyalty program and Lycos does not. The temporary shock described above has a much larger impact on Iwon than Lycos. Iwon loses 15.16% of its share, more than any other portal. Lycos only loses 7.94% of its share. After Iwon, Yahoo and Altavista have the highest switching costs. These web sites successfully implemented many features aimed at increasing loyalty. For example, Yahoo and Altavista have highest percentage of their users on their own email.[21]

20. This is based on a number of industry characteristics and assumptions. Advertising is typically paid on a per-view basis. Combining revenue data from J. Walter Thompson Company for nine portals from January to September 2000 with the visits data in this study, the average portal gets roughly 4.01 cents per visit. Furthermore, at the time of this study, Plurimus estimates showed that there were 43.3 million online households. Revenue changes were estimated based on combining these estimates. The dollar values are included to give an idea of the absolute size of the demand elasticity. They rely on several assumptions and do not represent equilibrium behavior. However, the monetary values are informative about the importance of true state dependence.

21. Email account providers could be determined in the clickstream data.

On the other hand, the web sites that have the lowest switching costs (Infospace, Looksmart, and Goto) had few features aimed at loyalty. Loyalty programs and free email seem to have been at least partly successful in generating switching costs.

Column (e) shows a simulation aimed at determining the marginal effect in a different way. It shows the simulated probability of a visit to the web site given that the web site was visited in the previous period. The simulated probability is based on all data in the sample. Again the estimated marginal effects are not trivial. Furthermore, Iwon seems to benefit disproportionately from loyalty.

Although the gains due to state dependence are significant, they are probably not large enough to justify the losses incurred by many of these web sites to build market share. For example, Excite lost $43 million in 1996 and $30 million in 1997. Yet, the simulations in column (d) show that switching costs give Excite only $256,193 per week. Even if the loyalty did not dissipate over time, it would take 5.5 years for switching costs to pay off the losses. Similarly, Lycos lost $52 million in fiscal 1999, it only gains $83,263 per week due to switching costs.

The marginal effect of the loyalty variable is also lower than that found by other researchers in other markets. Shum (2004) finds that recent purchase of a brand raises current purchase probability by 20 times. The comparable estimates for Internet portals in column (e) range from 26% for Snap to nearly 300% for Iwon and Excite. Using flexible random coefficients models in the ketchup market, Keane (1997) estimates an effect of recent purchase on sales of 75% or more.[22] This compares to the 3%–15% result for Internet portals described in columns (b), (c), and (d) of Table V.

In summary, the marginal effects are large though probably not large enough to justify the losses that the web sites sustained in the late 1990s. Furthermore, the benefits of loyalty accrued disproportionately to web sites with loyalty programs or other features that may induce state dependence.

4.3 Alternative Specifications

This section presents a number of alternative specifications of the main model to test the robustness of the base model presented in Table IV and Section 4.1. The first part of the section discusses the identification of true

22. Keane (1997) does not present this exact number. He estimates the switching cost in the ketchup market to be equivalent to a drop in price of 5% to 27%. He also estimates the effect of an approximately 45% drop in price to increase sales by 313%. Combined, and given a linear structure on price effects, these results suggest an effect on sales of 75% or more.

state dependence rather than spurious correlation because of clustering of individual usage within the household or because of clustering of goals. The second part explores the choice of loyalty measure.

Models 1 through 6 of Table VI show that the *loyalty* coefficient is still significant, and it changes little, even controlling for a number of sources of spurious correlation. To reduce the computational burden of conducting many supplemental regressions, these were run on 100 households rather than the full sample of 2,651 households.[23] Model 1 shows the results on this subsample of the standard model used in Tables IV and V.

One source of spurious correlation may arise because the data are at the household level.[24] If individuals within the household use different web sites and individual usage is correlated over time, then within-household heterogeneity may appear as state dependence. For example, children who use Yahoo may use the Internet between 6 PM and 9 PM, a stay-at-home parent who uses Lycos may be online 9 AM–3 PM, and a working parent who uses Hotbot may be online 9 PM–11 PM. Model 2 limits the data to all visits between 9 AM and 3 PM on weekdays (not including December 27–January 2) to try to get closer to individual usage. Other time periods are unlikely to meet the requirement of having only one household member online. The general results change little in Model 2. The average *loyalty* coefficient falls slightly. The marginal effect of the "loss of history" simulation described in the above section also has a small decrease. Overall, however, the measured true state dependence still matters.

Another possible source of the measured true state dependence is clustering of goals in time. For example, suppose individuals typically visit ecommerce web sites in December, information web sites in January, and adult web sites in February. Also suppose that MSN is best for e-commerce searches, Ask Jeeves is best for information searches, and Altavista is best for adult web site searches. In this case, the finding of true state dependence in Table IV may be a function of this clustering of goals in time. Models 3 through 6 of Table VI estimate the main model by goal. Again, the loyalty coefficient is significant. The magnitude of the marginal effect in column (f) also changes a little.

As discussed earlier, the loyalty variable is defined by a dummy variable on whether the portal was visited in the previous session,

23. I believe this is enough data to ensure robustness. To ensure that the regressions could be estimated for all of the households, each of the households included in these regressions had at least two visits in each subsample. Therefore these are, on average, households with more visits.

24. As discussed earlier, this is one of the reasons for defining loyalty by the previous session.

TABLE VI.
ROBUSTNESS CHECKS USING SUBSAMPLES OF THE DATA AND DIFFERENT LOYALTY DEFINITIONS—USING RANDOMLY SELECTED GROUP OF 100 HOUSEHOLDS

Model	Model Description	(a) Unweighted Mean Loyalty Coefficient[a]	(b) Percentage of Positive at 95% Confidence Level (Two-Sided Test)	(c) Percentage of Negative at 95% Confidence Level (Two-Sided Test)	(d) Predicted Yahoo Share (%)	(e) Simulated Share of Yahoo in the Week Following an Extended Shutdown Such That All Users Have No History at Yahoo (%)	(f) Effect of Loss of History in Percentage Terms[a] $\frac{(d)-(e)}{(d)}$	(g) Log Likelihood	(h) Number of Households
Subsamples[b]									
1	Standard model on these households	1.15	80%	4%	9.00%	7.65%	15.00%	−55,360	176,966
2	9 AM–3 PM Weekdays	1.01	63%	3%	9.58%	8.22%	14.20%	−28,914	91,748
3	Communication goals only	0.91	48%	3%	13.68%	11.41%	16.59%	−11,304	37,224
4	Ecommerce goals only	1.03	45%	1%	7.08%	6.15%	13.14%	−9,536	30,004
5	Information goals only	1.05	51%	2%	8.55%	7.26%	15.09%	−13,818	44,364
6	Adult goals only	0.96	48%	4%	3.07%	2.61%	14.98%	−8,790	32,096
Loyalty Definitions[b]									
7	Used previous session loyalty, standard model	1.08	75%	3%	10.45%	8.88%	15.02%	−35,050	70,251
8	One period lag loyalty, standard model	0.774	74%	2%	10.45%	8.37%	19.90%	−33,399	70,251
9	GL loyalty, standard model[c]	1.63	75%	3%	10.46%	7.37%	29.54%	−32,292	70,251
10	GL loyalty by session, standard model[c]	0.635	48%	3%	10.45%	8.82%	15.60%	−35,493	70,251
11	Used previous day loyalty, standard model	0.141	41%	2%	10.44%	9.76%	6.51%	−35,825	70,251

[a] Results in these columns are significantly different from zero at the 99% confidence level.
[b] In models 1–6, the 100 households all had at least two visits in each subsample. In models 7–11, the households were selected randomly.
[c] For models 9 and 10, GL loyalty is set to its maximum (0.99) for Yahoo in column (e) and to zero for Yahoo in column (f).

because it is more easily interpreted as a switching cost than weighted measures and because there are fewer confounding factors than within-session definitions and day-to-day definitions. In the offline context, however, other loyalty definitions are used because they are often more appropriate. Models 7 through 11 of Table VI show that the core results are robust to a number of loyalty measures more commonly used in the literature. The results are based on a subset of 100 households taken from the full data set. Model 7 provides a basis of comparison by replicating the earlier results on the subset of households. Model 8 defines loyalty as a dummy variable for whether the previous visit to a portal was to that portal. It includes within-session visits. Model 9 uses Guadagni and Little's (1983) loyalty measure, a weighted average of the variable used in Model 8 (see the Appendix for details). This is the model that has the best explanatory power as measured by the log likelihood. In Model 10, Guadagni and Little's variable is redefined to ignore within-session activities.[25] Model 11 defines loyalty by a dummy variable for whether that portal was visited during the previous day that the household was online. While the measured true state dependence falls as the visits relevant to the loyalty variable are further back in time, it is always significantly different from zero at the 99% confidence level. True state dependence is a significant factor in this market.

In summary, Table VI shows that the general results are robust to specifications that help control for individual usage clustering within the household, for clustering of goals in time, and for different definitions of the *loyalty* variable.

4.4 Switching Costs and Household Characteristics

Switching costs may be a function of household characteristics. This section explores whether some types of households display higher switching costs than other types. Any differences by household type could be used to inform web sites about the long-term consequences of attracting different types of consumers. In particular, Table VII shows

25. Guadagni and Little's (1983) loyalty measure, used in Model 9, has the most explanatory power; however, I focus on the *used last session* measure because of the concerns over within-session serial correlation and interpreting a weighted measure as a switching cost. While Guadagni and Little's measure is generally larger than the *used last session* measure, the distributions are similar and are significantly positive the same number of times. More details on Guadagni and Little's (1983) loyalty measure are contained in the Appendix.

TABLE VII.
OLS REGRESSIONS OF LOYALTY COEFFICIENTS ON HOUSEHOLD CHARACTERISTICS (STANDARD ERRORS IN PARENTHESES)

Variable	(a)	(b)	(c)
Total time online	8.46e−07***	8.29e−07***	
	(2.15e−07)	(2.17e−07)	
Nonsearch visits			2.34e−04***
			(4.98e−05)
% Adult web sites	1.69***	1.84***	1.27***
	(0.354)	(0.378)	(0.373)
% Brochureware	2.17**	2.30**	2.02**
	(0.929)	(0.938)	(0.927)
% Classified web sites		0.595	
		(1.13)	
% Communication web sites	1.44***	1.55***	1.23***
	(0.359)	(0.375)	(0.364)
% Entertainment web sites		0.478	
		(0.728)	
% Finance web sites	0.668	0.832	0.519
	(0.520)	(0.544)	(0.522)
% Information web sites	1.31**	1.43***	1.06**
	(0.521)	(0.536)	(0.523)
% Shopping web sites	1.37**	1.45**	1.17*
	(0.689)	(0.701)	(0.688)
% Technology web sites	1.22	1.33	1.02
	(0.857)	(0.863)	(0.857)
% Women's web sites		2.90	
		(2.82)	
Average education	0.0652*	0.0656*	0.0598*
	(0.0344)	(0.0347)	(0.0344)
% Married		0.0230	
		(0.439)	
Constant	−0.833	−0.971*	−0.636
	(0.508)	(0.584)	(0.506)
N	2,361	2,361	2,361
R^2	0.0202	0.0208	0.0226

Column (a) contains all variables that satisfied F-tests for inclusion.
***Significant at a 99% confidence level.
**Significant at a 95% confidence level.
*Significant at a 90% confidence level.

the results of regressing true state dependence coefficients on household characteristics. The summary statistics of the measured household characteristics are in Table III. *Average education* and *percentage married* come from census block demographics. Other census block demographics such as *average income*, *average age*, and *percentage of renting* were found to be uncorrelated with the loyalty coefficients. The other variables in the regressions depend on the observed web site habits of the households. For example, if 10% of household *i*'s visits are to entertainment web sites, then *percentage to entertainment web sites* is equal to 0.10.

Column (a) presents the main results of this section. It contains all variables that satisfied F-tests for inclusion, column (b) includes *percentage to classified*, *percentage to entertainment*, and *percentage to married*, and column (c) presents an alternative proxy for experience. The R^2 at the bottom of each column shows that none of these models have much explanatory power. The result is consistent with Rossi and Allenby (1993), who found that most household characteristics do not explain coefficients well.

Columns (a) and (b) show that *total time online* over the entire sample is the most important included factor in true state dependence in terms of significance. It is also important in economic terms: an increase in total time online by 1 hour per week is equivalent to a 3% increase in the value of the coefficient. This result does not change if the log of *total time online* is taken. Households that spend more time online have *higher* true state dependence coefficients. Goldfarb (2002) found that time online is a good proxy for experience. Combining these two results suggests that more experienced users have higher true state dependence. This is contrary to a common perception that new users are timid and afraid to switch, while experienced users are comfortable at any web site. Instead, it suggests that experienced users face true state dependence independent of abilities and preferences.

The following three explanations are consistent with this result. First, and most likely given the above results, is Johnson, Bellman, and Lohse's (2003) concept of cognitive switching costs. As a user's comfort with a web site increases, true state dependence rises. This is a rational explanation for habit persistence. The result that more experience is correlated with more lock-in is also somewhat consistent with Beggs and Klemperer (1992): switching is too costly for older consumers, so companies only compete for the new generation.

A second possibility is that more experienced users are more likely to personalize their web pages; however, only a small portion of the online population uses personalized web pages. A third possibility is that users who spend more time online do so because they are not good at navigating the Internet. Consequently, what appears to be experience is actually incompetence. This implies more searches at the same web site because these individuals less able to find what they seek.

Column (c) of Table VII explores this hypothesis by replicating Model 1 with a different measure of experience. In this column, experience is measured by the number of nonsearch visits rather than total time online. The qualitative effect does not change, and it is therefore unlikely that time online is measuring incompetence.

Table VII also reveals that households that visit more communication web sites, more adult web sites, more information web sites, or more shopping web sites have lower measured true state dependence. In column (a), a 1% increase in visits to communication web sites from the mean is correlated with the largest increase in true state dependence, 0.237%. A 1% increase in visits to adult, information, or shopping web sites is correlated with an increase in the loyalty coefficient of 0.116%, 0.0804%, or 0.0589%, respectively. To increase switching costs, portals could try to attract users that prefer these types of web sites.

In summary, Table VII shows that household characteristics do little to explain loyalty. The characteristic with the most explanatory power is online experience, proxied by total time online.

4.5 Measuring True State Dependence Under Different Models

Because of data constraints, most estimates of state dependence are derived from panel methods. This section compares the panel regressions previously used in the literature to household-specific regressions to better inform future panel studies. It shows that the panel regressions typically used in the literature overestimate true state dependence relative to household-specific regressions. Table VIII compares the household-specific regressions in Model 1 with the panel methods in Models 2 through 8. While the coefficients are not directly comparable to the household-specific regressions in Model 1 (or to each other) due to normalizations, they show the same pattern as the marginal effects in column (j).

For computational reasons, the panel models could only be estimated using 6 firms rather than 18. Each model was estimated using full information maximum likelihood. The loyalty coefficient in the household-specific regressions of Model 1 is slightly lower here than when all 18 firms are included.

Model 2 assumes no unobserved household heterogeneity in brand preferences or the loyalty coefficient. It estimates a loyalty effect that is much higher than the one found with household-specific regressions. If unobserved household heterogeneity is not considered,

TABLE VIII.
COMPARISON OF DIFFERENT METHODS—USING 100 HOUSEHOLDS AND THE TOP SIX PORTALS

Model	Model Description: Each Uses GL Loyalty and Six Portals	(a) Weighted Mean Loyalty Coefficient*	(b) Median Loyalty Coefficient	(c) 25th Percentile Coefficient	(d) 75th Percentile Coefficient	(e) Log Likelihood	(f) Users for Loyalty Estimate	(g) N	(h) Simulated Probability Visit Yahoo Given Visited Yahoo Last Period (%)	(i) Simulated Share of Yahoo in the Week Following an Extended Shutdown Such That All Users Have No History at Yahoo	(j) Effect of Loss of History in Percentage Terms $\frac{(h)-(i)}{(h)}$
1	Household-specific regressions	1.19	1.09	0.515	1.78	−36,786	85	70,251	10.76%	9.50%	11.70%
2	Panel, no random coefficients	2.35	2.35	2.35	2.35	−54,208	100	76,908	10.65%	7.66%	28.08%
3	Panel, binomial brand random coefficients	2.44	2.44	2.44	2.44	−50,920	100	76,908	10.98%	8.25%	24.84%
4	Panel, binomial brand and loyalty random coefficients	2.25	1.73	1.73	2.65	−50,986	100	76,908	10.93%	8.90%	18.58%
5	Panel, trinomial brand and loyalty random coefficients	1.21	0.956	0.500	2.43	−46,118	100	76,908	10.84%	9.33%	13.92%
6	Panel, normal brand random coefficients	1.55	1.55	1.55	1.55	−43,224	100	76,908	10.82%	9.18%	15.12%
7	Panel, normal brand and loyalty random coefficients	1.45	1.45	1.37	1.53	−43,219	100	76,908	10.82%	9.37%	13.43%
8	Panel, normal all bottom nest random coefficients	1.32	1.32	1.25	1.39	−43,156	100	76,908	10.78%	9.35%	13.26%

*All results in this column are significantly different from zero at the 99% confidence level.

true state dependence appears to be overestimated. Models 3, 4, and 5 present loyalty results from discrete distribution random coefficients models.[26] Models 6, 7, and 8 present loyalty results from normal distribution random coefficient models. All these models overestimate the effect of true state dependence. Even Model 8, the most flexible and best-fit random coefficients model estimated, with independent normal distributions assumed on all bottom nest coefficients, estimates true state dependence to be at a higher level than the household-specific regressions. The effect of a loss of history is 13.26% in the random coefficients regression in Model 8; it is only 11.70% in the household-specific regressions on this data.

As described previously, Figures 2 and 3 show that the distributions of the household-specific brand and loyalty coefficients are close to normal. This suggests that assuming a normal distribution is better than a discrete distribution. Table VIII is consistent with this hypothesis. The results presented in Models 7 and 8 are closer to those in the household-specific model than are the others, although both still overestimate true state dependence. Model 5 also does a good job approximating the effect. Therefore, if there is insufficient data to conduct household-specific regressions, the results suggest that assuming a normal distribution on as many coefficients as is feasible will provide the least biased result. Of particular importance is allowing for heterogeneity in the loyalty coefficient as well as the brand-fixed effects. Models that do not allow for heterogeneity in the loyalty coefficient perform much worse. These results differ from those found by Abramson, Andrews, Currim, and Jones (2000) in a Monte Carlo study. They found that discrete heterogeneity does not affect the loyalty coefficient and that continuous heterogeneity has few problems. In this real-world application, both overestimate measured true state dependence.

In summary, in this application, random coefficients models overestimate the degree of true state dependence relative to household-specific regressions. However, when there is insufficient data to perform household-specific regressions, assuming normal distributions on (at least) the brand coefficients and the loyalty coefficient is the next best.

26. The algorithm used in estimation of the bottom nest is largely based on Jain, Vilcassim, and Chintagunta (1994). Starting with a binomial distribution of preferences, they keep adding types until Baye's information criterion fails. Columns 3, 4, and 5 follow this method, but use two independent discrete distributions: one on brand preferences and one on loyalty. Two distributions are used because loyalty needs to be uncorrelated with brand preferences to be separately identified. Baye's Information Criterion fails at more than three types of each, totaling nine types.

5. Conclusion

This paper has rejected the hypothesis that the previous web site visited does not affect the current web site choice. Using the same identification argument as previous studies on state dependence, but with a more flexible econometric framework, the results suggest that switching costs matter in online markets.

The existence of switching costs in online markets would partially justify the land-grab mentality that characterized the early period of Internet growth. Theoretical models of switching costs imply an early period of intense competition. Still, the results suggest that switching costs drive no more than 15% of market share for Internet portals. This is probably not large enough to justify the losses that many Internet companies incurred in the late 1990s.

From a strategic perspective, portals should look for ways to increase switching costs, without turning off customers, to increase revenues and move to profitability. The portal with the richest loyalty program, Iwon, received an especially large percentage of its visits because of switching costs. Portals without loyalty programs, personalization, or email, such as Infospace and Looksmart were barely affected by switching costs. Portals can also try to focus their attention on users who are more likely to display higher switching costs, which, the results suggest, are likely to be the more experienced users, and users that visit more communication, information, and shopping web sites.

A comparison of econometric methods of estimating switching costs showed that random coefficient methods do not allow for sufficient heterogeneity both in brand preference and in the switching coefficient. Consequently, these methods overestimate switching costs. If data are limited so that household-specific regressions are not feasible, this study suggests that models that allow for more heterogeneity are better. Models that do not allow for heterogeneity in the loyalty coefficient appear to be particularly flawed. This has implications for the structural discrete choice models commonly used in industrial organization (Berry, Levinsohn, and Pakes, 1995; Nevo, 2001; etc). These models often rely on randomly distributed coefficients for identification. For example, Nevo (2001) and Dube (2005) use random coefficients to identify substitution patterns, and Ackerberg and Gowrisankaran (2004) use random coefficients to identify network effects. The results of the household-specific regressions in this paper suggest that the normal distribution typically assumed in these models is reasonable. Furthermore, there is considerable heterogeneity in the loyalty coefficient. Therefore, there is little basis for the common econometric assumption that households differ in brand preferences only, and not in true state dependence (e.g., Shum, 2004; Israel, 2005).

In summary, this paper has shown that switching costs are significant in the Internet portal market, and users differ in these costs. Still, the profits generated by these switching costs are not large enough to justify the substantial losses incurred by many Internet companies in the late 1990s.

Appendix

Knowing the goal of search is important for knowing whether a search fails. The goal of search was determined by the category of the site following a visit to a portal, if that next site was visited within 5 minutes of the end of the portal visit. If the goal of the search is another portal, then the goal of the first search is considered to be the same as the goal of the second. If no site is visited within 5 minutes of the end of a portal visit, then the search is considered to have no known goal; 23.4% of all searches have no known goal. Most of these occur because many people return to a portal page before logging off the Internet. These are not considered in the proxy variable for failed searches, *last search repeated*. The goals were divided into roughly 100 overlapping categories, including news, music, email, shopping for computers, automotive information, and travel.

An online session end is defined as a 15 minute or more break between web site downloads. It is important to identify sessions to derive *used last session* loyalty.

The view length at a portal is the time of departure minus the time of arrival (in seconds). Because it is time spent during previous visits that is important for whether a household returns to that portal, this study only reports results from a one period lag on *last view length*. More complicated functions of past time spent do not yield qualitatively different results in a subsample of the data.

Search failure is proxied by *repeated search*. If a household visited two portal sites in a row, and there was less than 5 minutes between visits, then the first search is considered a failure. Furthermore, if the household conducts a search and then searches again for the same goal (at the same site or at a different one) within 5 minutes of the first search, then the *repeated search* variable is equal to one. While 5 minutes is arbitrary, extending the time to 10 minutes or shortening it to 3 minutes did not change the number of repeats much. As with time spent, it is whether previous searches at a site were repeated that matters. Also as with time spent, more complicated functions of past repeated searches do not yield qualitatively different results. A search is only identified as repeated if it occurred in a previous session. This avoids confusion over the use of a browser's back button. This variable is labeled *last search repeated*.

How much time a household's previous visit to a portal took and whether that search was repeated are only observed when the household has visited that portal previously in the data set. Because many households visit a portal for the first time relatively late in the data set, these variables are missing for a large number of observations. Therefore there is a dummy variable for missing data. One minus the missing data variable is interacted with the view length of previous search and the previous search repeated variables. This overcomes the significant potential bias of assuming a value for the missing data or of ignoring it entirely. This missing data dummy has no economic interpretation.

The variable *total visits to destination* proxies familiarity with the destination. It is calculated as the total number of times a household visits the destination web site over the course of the sample period.

Destinations were divided into five types: *information*, *ecommerce*, *communication*, *adult*, and *other*. These destination type dummy variables are in the household-specific regressions (*other* is the base).

On a subsample of the data set with an aggregated model, the Baye's Information Criterion suggested that *total visits to destination* and *destination type* only mattered for the choice of whether to use a portal, and not the portal choice itself. These variables are therefore only included in the top nest. They are interacted with a dummy variable for the outside good being chosen.

Household characteristics are based on the online habits of the households: *total time online* and percentage of all web site visits by that household in 12 categories. The data also contain household-level demographic characteristics based on the census block.

One of the robustness checks for the loyalty variable uses Guadagni and Little's (1983) loyalty measure. This analysis mimics Guadagni and Little's methodology for constructing their "loyalty" variable almost exactly. In their paper, loyalty is considered to be a weighted average of past purchases of the brand, treated as dummy variables. Let $portsame_{ijt} = 1$ if household i visited portal j in its previous search and zero otherwise.

$$GL\ loyalty_{ijt} = \omega GL\ loyalty_{ijt-1} + (1 - \omega) portsame_{ijt}. \tag{A1}$$

Rather than estimate ω by maximum likelihood, which would significantly complicate the computational problem, they calibrate ω based on dummies for lags of length 1–10 (Fader, Lattin, and Little, 1992 do estimate this by maximum likelihood). Guadagni and Little's (1983) procedure was followed using a quarter of the data set as a multinomial logit for the bottom nest. The model was estimated with 10 lagged dummy variables, and the value of ω that minimizes $\sum_{t=1}^{10} |\omega^{t-1} - \beta_{-t}|$, where β_{-t} is the coefficient of the tth lagged dummy of *portsame*, is $\omega = 0.782$. This number is held constant for each household. Like Keane (1997), $GL\ loyalty_{ij1}$, the first observation for each household, is set to zero. The initial condition bias is mitigated by the large number of observations per household. The results also show *portsame* alone as a loyalty variable. The *GL loyalty* variable has more explanatory power in the household-specific regressions than *portsame*. In a recent study, Abramson, Andrews, Currim, and Jones (2000) also find *GL loyalty* to be the best fit.

The *GL Loyalty for sessions* variable is constructed similarly except that the most recent visit is considered to be the last visit of the previous session.

The $portsame_{ijt}$ variable is defined to depend on the previous portal visited of any kind, not just the previous of the main 18 portals used in this study. Therefore, if a household visits Yahoo then Top9.com and then Yahoo again, $portsame_{ijt}$ on the second visit to Yahoo is equal to zero.

References

Abramson, C., R.L. Andrews, I.S. Currim, and M. Jones, 2000, "Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice," *Journal of Marketing Research*, 37, 410–426.

Ackerberg, D. and G. Gowrisankaran, 2004, "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," *Working Paper*, Washington University.

Beggs, A. and P. Klemperer, 1992, "Multi-Period Competition with Switching Costs," *Econometrica*, 60, 651–666.

Berry, S., J. Levinsohn, and A. Pakes, 1995, "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Chen, P.-Y. and L.M. Hitt, 2002, "Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry," *Information Systems Research*, 13, 255–274.

Darsey, D., 1999, "The People behind the People behind e Commerce," *Fast Company*, June, 185–198.

Davey, K.S., 1991, "Predicting Shares from Preferences for Multiattribute Alternatives," Unpublished Doctoral Dissertation, Vanderbilt University.

Dube, J.P., 2005, "Product Differentiation and Mergers in the Carbonated Soft Drink Industry," *Journal of Economics and Management Strategy*, 14, 879–904.

Elrod, T. and G. Haubl, 1998, "An Extended Random Coefficients Model, with Application to Metric Conjoint Analysis," *Working Paper*, University of Alberta.

Fader, P.S., J.M. Lattin, and J.D.C. Little, 1992, "Estimating Nonlinear Parameters in the Multinomial Logit Model," *Marketing Science*, 11, 372–385.

Fischer, G.W. and D. Nagin, 1981, "Random Versus Fixed Coefficient Quantal Choice Models," in C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: The MIT Press, pp. 273–304.

Gandal, N., 2001, "The Dynamics of Competition in the Internet Search Engine Market," *International Journal of Industrial Organization*, 19, 1103–1117.

Goldfarb, A., 2002, "Understanding Internet Clickstream Data," *Working Paper*, University of Toronto.

Greenstein, S., 1993, "Did Installed Base Give an Incumbent Any (Measurable) Advantage in Federal Computer Procurement," *RAND Journal of Economics*, 24, 19–39.

Guadagni, P.M. and J.D.C. Little, 1983, "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

Hargittai, E., 2000, "Open Portals or Closed Gates? Channeling Content on the World Wide Web," *Poetics*, 27, 233–254.

Heckman, J., 1981, "Statistical Models for Discrete Panel Data," in C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: The MIT Press, pp. 179–195.

Hsiao, C., 1986, *Analysis of Panel Data*, Cambridge, UK: Cambridge University Press.

Israel, M., 2005, Tenure Dependence in Consumer-Firm Relationships: An Empirical Analysis of "Consumer Departures from Automobile Insurance Firms," *RAND Journal of Economics*, 36, 165–192.

Jain, D.C., N.J. Vilcassim, and P.K. Chintagunta, 1994, "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data," *Journal of Business and Economic Statistics*, 12, 317–328.

Johnson, E.J., S. Bellman, and G.L. Lohse, 2003, "Cognitive Lock-In and the Power Law of Practice," *Journal of Marketing*, 57, 62–75.

Jones, J.M., and J.T. Landwehr, 1988, "Removing Heterogeneity Bias from Logit Models Estimation," *Marketing Science*, 7, 41–59.

Keane, M.P., 1997, "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, 15, 310–327.

Klemperer, P., 1995, "Competition When Consumers Have Switching Costs," *Review of Economic Studies*, 62, 515–539.

Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, UK: Cambridge University Press.

Nevo, A., 2001, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Nie, N.H. and L. Erbring, 2000, "Internet and Society," *Working Paper*, Stanford Institute for the Quantitative Study of Society.

Pesaran, M.H. and R. Smith, 1995, "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113.

Rossi, P.E. and G.M. Allenby, 1993, "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30, 171–182.

Seetharaman, P.B., A. Ainslie, and P. Chintagunta, 1999, "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research*, 36, 488–500.

Shapiro, C. and H.R. Varian, 1999, *Information Rules: A Strategic Guide to the Network Economy*, Boston: Harvard Business School Press.

Shum, M., 2004, "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market," *Journal of Economics and Management Strategy*, 13, 241–272.