Information Acquisition and Consumer Choice[*]

Preliminary Draft – INCOMPLETE – Comments sought! January, 2011

T. Bresnahan, T. Landvoigt, and +P.-L. Yin[†]

We examine individual choice and information gathering by demanders in a technologically dynamic market, and make our empirical application to mass-market use of the Internet. Technologically dynamic markets, in which newly-invented product varieties expand the choice set, present demanders with an information acquisition problem as well as a choice problem. We model this as a simple repeated two-stage problem for a rational consumer. In stage I of each period, the consumer decides whether or not to learn the characteristics of any new products in the choice set. In stage II, the consumer decides on a product. Over time, the attractiveness of information rises in our model because consumers have rational expectations about the rate of improvement of new varieties, though they do not know the realization of the improvement to their own utility from characteristics of a new variety until they engage in costly information acquisition. The consumers' initial choice before information acquisition may be determined either by a past choice or through a default or "opt out" choice being set for the consumer.

Empirically, we model consumers as heterogeneous both in the costs of information acquisition and in the net-of-cost benefits of new products. We work with a dataset on internet browser adoption which is unique in two ways. First, it has information relating not only to consumer choice but also to consumer information. We are able to verify the correct answer to questions consumers have answered about their choice, and use consumer error as an observable indicator of incomplete consumer information. Second, we also have a great deal of information about which consumers were faced with a default or "opt-out" choice and what that choice was. These two kinds of information go directly to the central concerns of our model. We are particularly interested in the impact of heterogeneity in information acquisition costs. High information cost consumers will tend to be less responsive to improvements in new varieties, for example, and more likely to stay with a default or "opt out" choice. We show that, for plausible observable predictors of information costs, these implications hold and are quantitatively strong.

---

[*] This version of the paper is preliminary in a number of senses. There are few citations, and we do not yet thank the many colleagues who have given us valuable comments.

[†] Stanford University, Stanford University, and MIT, respectively.

1. –Introduction

Technologically dynamic markets, in which newly-invented product varieties expand the choice set, present demanders with an information acquisition problem as well as a choice problem. A consumer who has not recently investigated available product varieties will not know what choices are available; there may be (time) costs of investigating new product varieties to learn how much utility each will yield. The implication for demand is at least three fold. First, there is a distinction, perhaps a quantitatively important distinction, between the demand for a new product if all consumers were fully informed and the demand for the same product taking into account consumers' need to gather information.

Second, the timing of consumer learning adds dynamic elements to demand. New varieties may have lower demand for a while because consumers do not yet know about them. This raises the demand for product varieties which have already been chosen, creating an inertial effect. A subset of consumers with high information-gathering and product-testing costs may optimally decide to become informed only slowly. That behavior will further slow the movement of demand to new options and away from existing ones.

Third, costly consumer information gathering also gives a demand advantage for products which are a default choice. If consumers have to "opt-out" of the default choice by gathering information about other options, this lowers demand for other options and raises demand for the default – to a quantitatively important degree if information gathering is costly enough for sufficiently many consumers. This feature of demand gives suppliers a motive to seek to have their product "placed" as the default choice. Many high-tech industries have active corollary markets for "product placement" as a default choice. For example, consumers who buy a new computer typically are offered a number of products and services bundled with it, including try-to-buy software from antiviral to word processing, "free" software such as a browser or advertising-supported games, and internet service provider signups. The product placement is valuable; no economist will be surprised that software and services firms pay operating system (OS) suppliers or computer manufacturers for placement. Other industries have similar corollary markets;[1] the important economic point is that when the behavior of demanders makes demand

[1] More generally, many consumer decisions create similar opportunities for related goods and services. A consumer who buys an iPhone finds some apps on it and more made default through the app store (which you can be

for a default product very different from demand for other products, suppliers may pay for placement as the default product.

In this paper, we focus on the demand side. We model a simple repeated two-stage problem for a rational consumer. In stage I of each period, the consumer decides whether or not to learn the characteristics of any new products the choice set. In stage II, the consumer decides on a product. We use a rational-ignorance framework in which consumers decide whether to be informed and, if informed, whether to make a new product choice in a forward-looking way. We solve the consumer's dynamic programming problem and apply it directly to estimation, letting consumers vary both in their information processing costs and in their net valuation of new products in the choice set.

This simple structure has three main benefits from our perspective.

First, it can be implemented empirically, allowing us to draw the quantitative distinction between demand in the ordinary (fully-informed) sense and demand taking into account information costs. We make our empirical application to the demand for browsers in the late 1990s, a time and industry in which the supply-side made significant efforts to gain "default" product placement because it was highly influential on demanders' choices. Bresnahan & Yin (2005) showed that browser product placement, in particular distribution of browsers with new computers, was very influential on both consumers' brand choice (Internet Explorer (IE) vs. Netscape (NS)) and consumers version choice (IE2 vs. IE3). In this paper, we focus on version choice. The "non-default" product to which consumers must opt-out is the newest version of their brand of browser.

Second, the model's structure permits us to investigate deeply the demand benefits accruing to a default product. We can infer from a seller's willingness to pay for default placement that it influences behavior; we can tell from the quantification of demanders' response to default placement whether there is a significant impact of the distribution convenience arising from default placement. Those reduced-form inferences, however, do not tell us why demanders

---

sure charges app developers for placement). Even consumers who have a new baby in hospital may be offered photography services, baby formula, and other products (and even not-for-profit hospitals accept "donations" for product placement). These arrangements do not bind consumers to a particular choice in each of the related categories, but they do create a default choice. Many consumers take the default choice, and even continue to choose the same brand long after. A default choice lies somewhere in between having already chosen a particular product (so that search would be needed to find an alternative) and receiving an advertising message (so that search is cheaper for a particular product).

respond to default placement. Three obvious stories leap immediately to mind. (1) Demanders may bear some transaction costs of acquiring a non-default product. In our application, for example, demanders bear the costs of downloading a new and better browser. (2) Even perfectly informed demanders may view the default product and alternatives as providing similar value. In that case, modest transaction costs associated with switching from the default will have large impacts on demand. (3) Demanders may have information-processing costs associated with learning of the newest versions. In our empirical work, we will conclude that this third, information-based, explanation is far more quantitatively important than the other two.

A final advantage of our model is that it addresses policy problems of growing importance and scientific questions of growing importance. In a number of policy arenas, such as the privacy of consumer information, the central distinction between marketing schemes is whether consumers are compelled to opt out of a plan versus opt in. Understanding whether the distinction between opt-in and opt-out is driven by informational advantages versus other frictions goes to the core policy question of whether consumers have consented to a choice about which they were informed. From a scientific perspective, those of us who study information technology industries today are working in an era when cutting-edge demand, long associated with professionalized buyers automating white-collar work, now serves the consumer. The industrial marketing of IT has been critical in the past; the consumer marketing of IT may be critical in future.

2. Model

Since the work of McFadden (1974, 1978) a standard approach has applied to consumer discrete choice. Consumer i chooses choice j from choice set J and picks the maximum over J of V_{ij}. V_{ij} is the (indirect) utility to i of choice j, i.e., it is the benefit to i of choosing j net of any acquisition cost. The consumer knows the choice set and the utilities V_{ij}, but the econometrician does not know V_{ij}. The heterogeneity across consumers in V_{ij} determines the aggregate demand curve for choice j and is thus of great interest. Typically, research proceeds by making assumptions about the distribution of V_{ij}, including its dependence on characteristics of chooser i (z_i), characteristics of choice j (x_j), and deep demand parameters θ.

One could undertake any number of interesting counterfactuals with estimates of such a model. In the context of technical change, a particular focus is the addition of new elements to the choice set. Once the model is estimated, one could calculate the uptake of a new product, say j', with specific features, say $x_{j'}$, as $Pr_i[V_{ij'} \geq V_{ij}]$. The essential element of this calculation is the use of the model of heterogeneity across i to calculate the demand behavior for the new variant.[2]

A related standard approach applies when the consumer's problem is dynamic. In this case, the consumer's problem is to make product choices over time given that preferences may change, product attributes may change, or the choice set may include choice j' at time t_1 but not at time t_0 ($t_1 > t_0$). The flow (indirect) utility in a particular time period t if consumer i has made choice j is u_{ijt} which now has a t subscript referring to the time of consumption. A dynamic element is introduced because there is a one-time utility loss from adopting product j' at time t, which is called $\delta_{ij't}$. Thus a consumer who uses product j for her lifetime of t=0,…,T has discounted payoff $\sum_{t=0,T} \beta^t u_{ijt}$ while one who switches to j' at time τ has discounted payoff $\sum_{t=0,\tau-1} \beta^t u_{ijt} + \beta^\tau \delta_{ij'\tau} + \sum_{t=\tau,T} \beta^t u_{ij't}$. Now the consumer's payoff function defined dynamically (abusing notation slightly by reusing V this time as the value function rather than the utility function but adding the time subscript) as the value function for a consumer who begins time t using product j is

$$V(j)=\max\{u_{ijt} + \beta\, V(j),\ u_{ij't} - \delta_{ij't} + \beta\, V(j')\}$$

This model, too, has an approach to the expected rate of uptake of a new product. For an individual consumer, a new product must be better than an old product (as in the static model) and also must pay for the costs of switching to it over the period of time the consumer uses it (which may be truncated by the arrival of an even better product in the future). Thus, the demand for a new product in cross section is determined by heterogeneity across consumers in costs of adopting it, flow benefit of using it, and expected use period (expected arrival of an even better alternative in the future).

Our approach is closest to this standard dynamic model. However, we add an element we think is important for modeling new variants in industries with technological progress. In our framework, the consumer need not have perfect knowledge about new product variants. Even after a product has been introduced into the market place, a consumer may not automatically know its price or product characteristics, and thus may not automatically have an assessment of the utility of adopting it. Of course, this also means that the consumer is not fully informed

[2] At this juncture, once again, we note this version is very preliminary and does not have a careful review of the accomplishments of the tremendous earlier models on which we are building or a careful statement of what is new in our treatment. For now, we strive for a careful statement of what our model is.

about new products that will be available in the future, or the precise utility that she will receive from those (as yet unknown) new products or product characteristics. This adds an information-gathering element to the consumer's problem. We argue that this information-gathering element is not the same, for a number of analytical purposes, as consumer valuation, and that heterogeneity across consumers in information-gathering is also not the same as heterogeneity in valuation.

Our model has a strong-rational ignorance flavor. We do not assume that consumers who are uninformed about the latest product variants are irrational or imperfectly foresighted, but rather that their costs of becoming informed are greater than the expected benefit of learning of a valuable new product and adopting it, where the expected benefit is calculated according to the true model of new product introductions.

Our model requires three concepts of the choice set facing consumer i. The first is the true list of products available in the market at time t, $\mathcal{J}_t$. The second is the subset (not necessarily a strict subset) of $\mathcal{J}_t$ that has been observed by consumer i as of time t, $J_{ti} \subseteq \mathcal{J}_t$. The last is consumer i's knowledge that permits her to form a probability assessment of what she would learn if she were to observe the true choice set. Our approach is to posit a true stochastic process for the expansion of $\mathcal{J}_t$. over time and for the characteristics (X) of new elements of $\mathcal{J}_t$. We endow consumers with knowledge of the stochastic process but not of the realization. Thus, they can form a rational expectation of the value of making an observation and then choosing a new product.

Let the consumer's knowledge of the choice set at time t be J_{ti}, a subset of $\mathcal{J}_t$. We assume that consumer i will learn about a new product after introduction with constant hazard rate ψ. At each time t, the consumer's decision process consists of two phases: (i) first, the consumer either observes new elements of the choice set and their characteristics (or does not observe), learning what flow utility she will receive from each choice, and then (ii) the consumer chooses whether to choose a different product from the set she has observed (or not, staying with her existing choice).

The relationship between our model and a standard choice model can be seen in this simple diagram.

Option 1	Option 2	Option 3
x_1	x_2	x_3
$u_i(x_1)$	$u_i(x_2)$	$u_i(x_3)$

The diagram shows three potential choices for user i. Suppose that all three options have in fact been introduced to the market, so that $\mathcal{J}_t=\{1,2,3\}$ but that user i has, by time t, observed only the first two options, so that $J_{it}=\{1,2\}$. This means that if the user were to make a product choice at this time, she would be choosing only between options 1 and 2. We also give the consumer information, I_t, which lets her form a probability assessment of whether product 3 has in fact been introduced and, if so, of $E[u_i(x_3)|\, I_t]$. We show option 3 in gray to visually depict the idea that, conditional on product 3 in fact lying in $\mathcal{J}_t$, a consumer will have a probability assessment less than 1 that 3 is available.

2.1. Specific Application: “newest version of product”

We begin with the simplest version of this model in which new versions of a single product are introduced over time, such as iPhone 1, 2, 3, 4 or Netscape Navigator 1, 2, 3, 4. Time is discrete. Think of one time period as one month. The choice space is also simple, where there is only one brand of a product, and a new and improved version appears in the marketplace at a specific time. Consumers do not necessarily know of a new and improved version. They do, however, have rational expectations about the arrival process of new versions.

2.1.1. Utility and initial conditions primitives.

The flow utility function is fundamental. For now we suppress variation across consumers (i.e., no subscript i) and focus instead only on the decision problem of a representative consumer. Each period in which a consumer uses the version of the product that was released into the market at time t, with characteristics x_t, yields utility $u(x_t)$.[3] We start the consumer off at time t_0 using product 0 with characteristics x_0. For now, we do not distinguish between the case in which the consumer starts with the outside good and the case in which the consumer has been given an initial default choice.

[3] In our empirical model consumers are heterogeneous and subscript i names the consumer. For the remainder of this theoretical section we will drop the subscript i and simply write about “the” consumer.

2.1.2. Evolution of the best available technology

We begin with a model of how the true choice set $\mathcal{J}_t$ is determined. In this application, the choice set expands over time as new versions are introduced.

The best available product in the market (not necessarily observed by the consumer) at time t has characteristics X_t. We assume that X_t follows a scalar time series process which updates only if a new version is introduced at time t (s_t=1) and not if there is no new introduction (s_t=0). We assume that s_t follows a Markov chain with transition matrix P. If there is a new product introduction into the market, its improvement in X is given by ε_t, where ε_t is i.i.d. with positive support.

$$X_t = X_{t-1} + s_t\,\varepsilon_t \qquad (1)$$

While the consumer does not automatically observe s or ε, we assume that she knows the distribution function for ε, F(), and the Markov matrix P. The specific functional form assumptions we make for these distributions are described below.

2.1.3. Consumer's opportunities, information, and state variables

The consumer begins with a technology x_0 at time 0. In every subsequent period, the consumer either observes the most current technology or not. A consumer who observes at time t1 learns the current technical level and has the opportunity to choose it. After that period, if the consumer does not observe again, she continues to know the market technology level as of time t1 and continues to have the opportunity to choose it. It will not, however, be rational for consumers to choose a market technology level other than at the time it is first released. Thus the consumer's state variables at time t are

x_t	Technical level of product the consumer is using at the beginning of time t
τ_t	Number of periods since the last time the consumer observed the true state
$\tilde{x}_t$	$\tilde{x}_t = X_{t-\tau}$ -- Technical level at last observation
$\tilde{s}_t$	$\tilde{s}_t = s_{t-\tau}$ -- release status at last observation

This notation covers all the relevant cases. Consider a consumer who observes a product at time t1, which had been introduced earlier than that date, does not adopt it, and then does not observe again for τ periods. At time t=t1+τ, she has $x_t = X_{t0}$; τ_t=t-t1; $\tilde{x}_t = X_{t1}$; $\tilde{s}_t$=0 (=s_{t1}). If the same consumer were instead to adopt it, x_t would be updated to X_{t1} as well.

2.1.4. Consumer's Decision Problem.

The consumer's problem can be characterized by dynamic programming. At each date t, there are two phases to the consumer's decision problem. First, the consumer receives a draw of a Bernoulli random variable that is one with probability ψ. In case the draw is equal to one, the consumer gets to observe the current technology level. A consumer who observes can additionally decide to choose the current version by paying a cost of φ. We interpret the cost φ as including the price of the new technology level and any adjustment costs such as the costs of downloading new software or installing it on a computer. This cost φ can be spread out over several periods of usage, so the condition for choosing the new technology is not the same as $u(X_t)- \varphi > u(x_t)$ but instead depends on the dynamic program.

Let the value function at the beginning of the period (before potentially observing) be given by $W(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)$. It is easiest to work backwards by defining two value functions, called $V^O(x_t, s_t, X_t)$ for the case of observation, and $V^N(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)$ for the case of non-observation. The arguments of $V^O(x_t, s_t, X_t)$ are technological level of the consumer's existing choice and the release status and technological level as of time t – which are observed by the consumer. The arguments of $V^N(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)$ are the same as of W() because the consumer's information set does not change.

The case of non-observation is simple. The consumer's information state does not change and the consumer keeps using the same product as before. Thus

$$V^N(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)=u(x_t)+\beta\, W(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t+1) \tag{2}$$

where β is the consumer's discount factor.

The case of observation involves choice by the consumer and, whether the choice is of the newest technology or not, an updated information state. That is, $\tilde{x} = X_t$ because the consumer has observed it (and similarly for $\tilde{s}_t$). In particular, we have

$$V^O(x_t, s_t, X_t)=\max\{u(x_t)+\beta\, W(x_t, X_t, s_t, 0),\ u(X_t)- \varphi +\beta\, W(X_t, X_t, s_t, 0)\} \tag{3}$$

The consumer's decision to chose X_t is called d_t and is given by

$$d_t=1 \text{ iff } u(x_t)+\beta\, W(x_t, X_t, s_t, 0) < u(X_t)- \varphi+\varepsilon_t +\beta\, W(X_t, X_t, s_t, 0) \tag{4}$$

We will use this in estimating the consumer's decision. For now we merely note that the decision by an informed consumer to choose the newest technology, in our formulation, compares the one-time adjustment cost, $\varphi+\varepsilon_t$, to the value of having the newest technology now and in the future evaluated for a fully-informed $(X_t, s_t, 0)$ consumer. It is best to understand the decision not to choose the latest technology as the decision to continue to use x_t and to wait for a future product introduction. The consumer also must take into account the i.i.d. shock to her costs of acquiring the technology, ε_t. Since this shock is i.i.d, it does not enter W(), though of course the distribution of ε_t in the future affects the decision to choose a new product today.

We can now write out W() based on the likelihood that the consumer becomes informed and the distribution of the information should she get it, which is

$$W(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)=(1-\psi)\, V^N(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)+ \psi E[V^O(x_t, s_t, X_t)|\, \tilde{x}_t, \tilde{s}_t, \tau_t] \qquad (5)$$

Finally, we allow for the consumers' decision whether or not to observe the current state. Let the effort expended by a consumer on data gathering be e and the resulting probability of observing the state be $\varphi(e)$ with $\varphi(0)=0$. A consumer has an effort cost of γ and picks e to maximize W(), so that, when et is not zero it solves

$$-\psi\,(e)\, V^N(x_t, \tilde{x}_t, \tilde{s}_t, \tau_t)+ \psi'(e_t)E[V^O(x_t, s_t, X_t)|\, \tilde{x}_t, \tilde{s}_t, \tau_t]= \gamma \qquad (6)$$

This means that a consumer's optimal program is given by e_t and d_t.

For now, we take the probability of observing the technology state in any given period, φ_i, to be a given feature of a consumer, not a choice variable.

3. Dataset

We employ individual level data on browser use from Georgia Institute of Technology's Graphics Visualization and Usability (GVU) Center's online surveys of web usage. These surveys were conducted biannually in April and October of each year. We employ data from 7 waves of the survey (surveys 4-10) from Oct 1995 through Oct 1998. The survey asked questions about the web browser and operating system respondents were using, how long the respondent had been using the internet, how frequently she used the internet, and how much time she spent browsing. It also asked a number of demographic questions regarding age, gender, education, income, occupation, and location.

3.1. Sample

The original sample size of survey respondents across all seven surveys was 96,974. Each wave, the number of respondents ranged between 23,348 to 10,108, except for the last wave, for which there were only 5,022 respondents. During the final year of the survey, there was a drop in participation, likely due to the anticipation by the survey managers that the survey process was about to terminate, and their resulting lack of incentive to recruit respondents. Any analysis sample size will be smaller for each year because we exclude some observations for which there is missing data. Further, we were not able to match all visitors to the survey site from its weblog to survey responses, so the number of usable observations dropped to 93,670. Missing demographic information also was cause for elimination from our sample, resulting in the loss of more observations; this varies by analysis and is discussed more below. In this paper, we have used missing-data dummies for missing demographics and kept the observations. Definitions and sources for all variables are listed in Appendix Table 8.

In a number of cases, several fields which were asked on the same page of the survey are missing. These indicate rather than refusing to answer, the survey respondent may have simply skipped a page of the survey. If we were to also throw out these observations, it would reduce our sample to 60,390. We focus on the set of users that used Internet Explorer (versions 1-5) or Netscape (versions 1-4) browsers. The elimination of duplicate entries within and across surveys, the elimination of observations on platforms and browsers which are not of interest in this study, and the elimination of observations where the version of Windows was not specified brings the total observations without missing data to 47, 899. Summary statistics for these observations are reported in Appendix Table 9..

A number of these responses were by the same individual in different waves (383). Since the incidence of repetition is small, we do not attempt to exploit the limited panel data structure. We simply treat these as if they were unique individuals. Our final sample of respondents was distributed across the survey waves as shown in Table 1.

We matched the survey responses with weblogs from the GVU servers that recorded the user-agent field of each respondent as they accessed the survey online. The user-agent field is a code sent by a respondent's computer to the server hosting the survey so that the survey can be rendered in the appropriate way given the respondent's particular brand of browser and operating system. This user-agent field thus reveals the browser brand, browser version and operating system that the respondent is using.

The sample is heavily weighted towards Netscape users: 87% of the respondents over all waves of the survey are using some version of Netscape. The sample was also heavily weighted to users of Windows 95: nearly half the sample (22288) were using Windows 95, nearly a quarter each were using Macintosh (12293) or Windows 3.1 (12909), and only 409 respondents were using Windows 98. (We discarded observations using other kinds of computers and other kinds of browsers.) Dummy variables for the platforms are constructed for the Macintosh, Windows 3.1, 95 and 98 platforms as indicated in the user-agent field (*AMAC, AWIN31, AWIN95* and *AWIN98*).

In addition to the automatic recording of the browser and operating system used to fill in the survey, the survey asks the user several questions about these same facts, i.e., browser and operating system. Comparing these answers to the user-agent field thus allows us to confirm whether the respondent's answers on current browser and platform usage matched the actual browser and platform being used. In particular, surveys 8 & 9 ask, "Do you think you are using the most up-to-date version of your browser?" Survey 10 asks a slightly different question, "For your **primary** browser, do you think you are using the most up-to-date version?" We first check whether the browser indicated in the user-agent field is indeed the newest version of its brand at the time of the survey, recorded as the dummy *NEWEST*, and then check whether the respondent's answer matches, recording the match under the dummy *RIGHT*. There is possibly some discrepancy between what browser/operating system the respondent was using at the time they filled out the survey and what they thought of as their "primary" browser. Respondents could choose from responses listed in Table 2 if they were not certain. In fact, users who said they were certain were significantly more likely to be right than those who answered they were "not so certain" or "I think I am using…." We anticipate exploiting this variation, together with the informative answer "don't know" in a future treatment of consumer information. In this paper, we use only *RIGHT*.

The survey also asks respondents to select their "primary computing platform" from the following list: DOS (984 cases), NT (824), Windows (9143), Windows 95 (22628), Macintosh (11704), Macintosh 8 (557), Windows 98 (471), Don't know (270), and other (1318). We are able to construct a dummy variable *OSWRONG* for when the reported computing platform conflicts with that listed in the user-agent field. Again, there may be some discrepancy between what operating system the respondent was using at the time they filled out the survey and what they thought of as their "primary" OS.

3.1.1. Missing Data

There are two sources of missing data across all the surveys. The first comes from the survey respondent. In a number of the questions (e.g. in response to "What is your age?"), the respondent has the option to respond with "Rather Not Say". The respondent may also simply leave the answer blank and decline to answer. However, even for the same question, the ability to respond with “Rather Not Say” can change from survey to survey. A second source of missing data also comes from the differences in surveys. Some questions are simply not asked in some waves of the survey, so data may be missing for an entire year on a particular regressor.

In our model, we will account for refusals to respond via separate dummy variables. For refusals to respond, the dummy = 1 if the respondent consciously chose not to respond, and zero otherwise. Thus, the coefficient for that dummy will capture the estimated mean value of responses from those who choose not to respond. The specific names of the dummies and the regressors for which these dummies are relevant are described in the next section.

3.1.2. Regressors

A summary of all survey variables and their definitions can be found in Appendix A. This section discusses the relevant regressors employed in our model.

Several regressors control for user demographics in the choice of browser and version. The dummy *MALE* records if the respondent was male.. The age of the respondent in years is a continuous regressor in years denoted *AGE*. *DAGE* is a dummy for the 711 individuals who declined to reveal their age. The income of the respondent *INC* is another continuous variable; *DINC* is the dummy for the 7042 respondents who did not want to reveal their income. *PAYWORK* indicated that internet access is paid for by work. We combine *PAYWORK* with the dummy for those who do not know who pays for access, *PAYDK*, to form *PAYWORK_DK*. The base case category is that the respondent pays for internet access, i.e., the respondent has control over and knowledge of internet access.

We have two variables measuring the respondent’s internet usage: *HOURS* reflects the number of hours per week the respondent uses the browser. In the same manner, *USE* measures the frequency with which the respondent uses the browser (once a month, once a week, etc.). We convert *USE* into times per month to include it as a regressor. A dummy *DHOURS* picks up 1905 missing answers to both frequency of use and hours. We create a dummy for whether the occupation involves computers (*OCCCOMP)*, so the base case category is any other occupation. *YNET1* is a dummy that indicates whether the user had been online for more than a year, so the base category is less than a year of online experience.

The last few respondent characteristics directly affect their fixed costs of getting a browser that did not come with their computer. The log of speed of their internet access, *LSPEED*, was recorded in the log of modem and internet access speeds in kbaud. When speed was unknown, *LSPEED* was recorded as 0, and a dummy *DSPEED* indicated that the speed was unknown to 5,113 respondents.

Dummies are also included for whether the user is using Internet Explorer (*IE*), and *SURVEY* denotes the wave of the survey.

3.1.3. Cases – When is t, how big is τ

To define the initial conditions for our model for each respondent in our sample, we designate the last time we know a respondent has observed the state of the world as an interval between t_{0_begin} and t_{0_end}. The reason this is an interval is because we do not know exactly the specific date on which the respondent first got online or got her computer, but we do have bounds. Respondents were asked, "How long have you been on the Internet?" and given several time intervals from which to choose (0-6months, 6 months-1year, 1-3 years, 4-6 years, over 7 years). This variable, *YNET,* is the same one used to define the regressor *YNET1*.

The early bound, t_{0_begin}, is defined as (1) the earliest date in an interval when the user said she got on the Internet (based on the survey data and *YNET*) OR (2) the date of the release of the first version of the browser available on the OS that the respondent is using at the time she answered the survey, OR (3) the earliest date at which the respondent's OS was available in the marketplace, whichever is later. The later bound, t_{0_end}, is defined as (1) the latest date in the interval when the user said she got on the Internet (again based on survey date and *YNET*) OR (2) the last date when the OS that the respondent is using was the newest of its brand, whichever is later. We then designate the difference between t_{0_begin} and t_{0_end} (divided by 30 and rounded to translate the units into months) as t_{0_LENGTH}.

The combination of survey waves and possible years on the Internet choices and browser-platform combinations results in 124 different observed initial conditions for the respondents in our sample. (See Table 3).

We treat the level of technology X_t as an index, and rather than estimate it, assign a separate value for each version of each brand of browser using data and a variation on regressions in Bresnahan & Yin (2005). Essentially, using a different dataset, we regress the logit of the market share of the newest version of a browser within its brand on a version dummy, the time since introduction of the newest version, and various controls for the distribution of browsers. The coefficient on the version dummy is our estimate of X_t. We estimate this coefficient for versions 2-5 of each brand of browser, and assume version 1 has a value = 1, which seems consistent given the estimates for versions 2-5.

3.2. Econometric Specification

We estimate our model for web browser demand in this dataset of Internet-using consumers. We use the model to predict whether the consumer is using the newest version of their brand of browser. We also use it to predict whether they are well-informed. We base our application of the model on also observing some initial condition information for each consumer.

3.2.1. Model for *NEWEST*

The connection between our model and the dependent variable *NEWEST* is simple and direct. Suppose we observe consumer i at time t using browser technology x_t. Unlike the consumer, who might be incompletely informed, we can check to see what was the newest commercially released version of their browser, X_t and the date it was released. More generally, we observe the history of all the realizations of new product introductions (in an obvious notation) X_{hi}, S_{hi} up to the time we observe consumer i. We also observe a good deal of information about initial conditions for each consumer. For the moment, assume that we observe the consumer's initial conditions perfectly, i.e., we observe a time t_{i0} at which the consumer had state variables x_{ti0}, $\tilde{x}_{ti0}$, $\tilde{s}_{ti0}$, τ_{ti0} and the actual state of the technology was X_{ti0}. Call this time and this list of consumer and market state variables I_{i0}. Finally, we assume that each consumer, i, is characterized by her own download cost and information gathering parameters φ_i and ψ_i.

There are a number of different ways the consumer can come to have the newest version of the browser. All involve the consumer having observed the choice set at some time since the browser was introduced and having chosen to download and use it. If the browser has been released for more than one period, this event can have occurred at different times. Further, the history of earlier releases, observations by this consumer, and choices by this consumer effects their more recent decisions through the dynamic program. Finally, the model treats X_{hi}, S_{hi} as random variables whose distribution is determined by the distribution of ε and by P.

Taking all those effects in, it is straightforward to use the model (by brute force, or directly) to calculate the unconditional probability that a consumer who began with initial conditions I_{i0} and has type φ_i and ψ_i will have the newest version of the technology. Call this variable $\Pr(x_t = X_t | I_{i0}, \varphi_i, \psi_i)$. That is simply the implication of the model for events at time t given model assumptions. We also need to condition on the realization of product introductions of i's brand of browser, since those are data to us (if not to the consumer.) We thus calculate, again directly using the model $\Pr(x_t = X_t | I_{i0}, \varphi_i, \psi_i X_{hi}, S_{hi})$.

There are two final steps. First, we do not actually observe φ_i, ψ_i but instead have an econometric model of them. Let the observable data about person i be z_i and assume an econometric model such that the distribution of φ_i, ψ_i is $G(\varphi_i, \psi_i | z_i, \theta)$ where θ are the parameters we seek to estimate. Second, our initial conditions information sometimes determines a set of possible initial conditions $\mathcal{I}_{i0}$ instead of a singleton I_{i0}. This occurs because, for example, our initial conditions information determines the range of dates at which a consumer might have gotten their computer and thus a range of dates at which they might have gotten an earlier browser. Thus we need to sum over all of the elements of the set $\mathcal{I}_{i0}$.

Taking account of these two steps gives us the likelihood that observable consumer i has the newest version of their brand of browser. First, let

$$\Pr(x_t = X_t | \mathcal{I}_{i0}, \varphi_i, \psi_i X_{hi}, S_{hi}) = \sum_{j \text{ in } \mathcal{I}_{i0}} \Pr(x_t = X_t | j, \varphi_i, \psi_i X_{hi}, S_{hi}) / \text{card}(\mathcal{I}_{i0}). \qquad (7)$$

Then

$$\Pr(x_t = X_t | \mathcal{I}_{i0}, z_i, \theta, X_{hi}, S_{hi}) = \int \Pr(x_t = X_t | \mathcal{I}_{i0}, \varphi_i, \psi_i X_{hi}, S_{hi})\, dG(\varphi_i, \psi_i | z_i, \theta) \qquad (8)$$

3.2.2. The special case of Pr(Newest=1)

There are some users who have probability 1 of having the newest browser of their brand. Consider, for example, a consumer who has just bought a new computer on which the newest version of their brand of browser was distributed. For such a user, we assign a probability 1 to their having the newest browser (they all do in fact have it, so we are never taking ln(0)). We include them in the likelihood calculation because of the "right" equation. A more complicated user has a $\mathcal{I}_{i0}$ such that, for some of the possible initial conditions, they have probability 1 of having the newest browser. For such a user, $\Pr(x_t = X_t | j\ldots.)=1$ only for some j in $\mathcal{I}_{i0}$. Our approach is to assign probability 1 for some of the elements of the sum in equation (7) and calculate the probability for the other elements.

3.2.3. Econometric model of $dG(\varphi_i, \psi_i | z_i, \theta)$

We parameterize the download cost φ_i and observation hazard ψ_i of individual i as functions of individual characteristics, z_i. The per- period observation hazard ψ_i is naturally bounded on the interval [0,1]. For the download cost φ_i, assume that for all i, there exists an upper bound φ_{max} such that at $\varphi_i = \varphi_{max}$ it would never be optimal to download. Further note that given the structure of the problem, at the lower bound φ_i=0 it would always be (weakly) optimal to download. Hence we have $\varphi_i \in [0, \varphi_{max}]$.

To express the observation hazard and download cost in terms of individual covariates z_i, we adopt the following specification:

$$\psi_i = [1+\exp(-z_i\,\theta_p)]^{-1} \qquad (9)$$

$$\varphi_i = \varphi_{max}\,[1+\exp(-z_i\,\theta_d)]^{-1}, \qquad (10)$$

Thus this gives us $G(\varphi_i, \psi_i \mid z_i, \theta)$, and we are now in a position to maximize the likelihood with respect to θ. We have the marginal likelihood for "newest" above, at eqn (8) and we now need to multiply it for the likelihood function for "right" conditional on "newest."

3.2.4. Model of *RIGHT*

A consumer who is imperfectly informed about her choice set will make "mistakes" compared to a perfectly informed consumer. We will use the quoted form "mistakes" as a shorthand. It does not mean an irrational consumer, for in our model a consumer can be rationally ignorant and make "mistakes" as a result of her lack of perfect information.

More specifically, an imperfectly informed consumer in our model will make "mistakes" because she does not know the extent of her choice set. Our data also includes a question that contains information about the boundary of the choice set; we know the correct answer to that question, though the consumer might not. We therefore treat a factually incorrect answer as an information "mistake" in a sense parallel to our model, i.e., as an indication that the consumer is imperfectly informed.

It is helpful for our interpretation here that the specific question relates knowledge of the boundary of the choice set to choice: "Are you using the newest version of your brand of browser?" To interpret factually incorrect answers to this question as an indicator of consumers who are more likely to make "mistakes" requires two steps. First ,we need the (certainly correct) assumption that consumers who do not know the correct answer to the question can only answer it correctly with some given probability. Second – and here is where we leave the theory of consumer choice and enter the "theory" of consumer statements – we need to assume that people who are better informed about the factual truth are more likely to answer the question correctly.[4]

[4] It is important to point out at this juncture that no theory of consumer choice offers a solution to the problem "what do people say." In particular, behavioral theories of what people do or of what they know are not theories of what they say. So our need to add a model for what people say does not follow from our rational ignorance approach.

Thus, we add a second dependent variable which equals one if the consumer correctly answers the question "are you using the most recent version of your browser." This outcome is clearly not independent, either economically or statistically, from the consumers' decision to download the latest version of the browser. Accordingly, we specify a descriptive model for whether consumers are right in what they say (with some restrictions between its parameters and those of the model for becoming informed):

$\Pr(\mathit{RIGHT}_i)=\mathrm{logit}(\alpha_0 + \alpha_1 z_i \theta_p + \alpha_2 \mathrm{newest}_i)$

3.3. Some Assumed Constants

We make some assumptions about a few of the values in the model. These include

Constant	Value		Source
β	.95		standard
φ_{max}	10		Larger than largest PDV(product upload)
x_t, X_t	2*v		Regression models based off Bresnahan & Yin (2005)
Transition probabilities	.9	.1	
	1	0	

3.4. Identification of Net Value vs. Information Gathering Costs

The model puts a great deal of structure on the demand for *NEWEST,* and in principle we could attempt to discern information-gathering from demand in the ordinary sense by that alone. As a formal matter, the two sets of coefficients on z are separately identified. The information-gathering type of a consumer and the demand-type of a consumer interact with the consumer's initial conditions in a radically different way. We shall explore that element of identification by first examining a descriptive statistics model in which all of the z variables are interacted with a measure of initial conditions. Above and beyond that, however, we have separate information about consumer informedness through the observable dependent variable *RIGHT*. We get further identification of the distinction between a consumer's demand type and her information gathering type by letting the information gathering type predict both *NEWEST* (through the dynamic optimization model) and *RIGHT*.

4. Empirical Results

In this section, we report estimates of descriptive statistics models and of our structural model for a sample of users that is restricted in two ways. Specifically, we look at Internet Explorer browser users running (some version of) the Windows operating system.[5]

Second, our structural model applies to all of these users in the same way, because the economics of opt-in vs. opt-out is broadly the same for all of these users. Beginning with Internet Explorer 1, the latest version of IE was always distributed with new Windows computers. Thus, the users in our sample either had to make the choice of downloading the newest version of their brand of browser or had it distributed with their computer when they bought it. That is, all of them either opted-in to a newer version of the browser than came with their computer or kept the default.

Extension either to Netscape users running any kind of computer or to IE users on Macintosh computers would add both more interesting variation in opt-in / opt-out and, because of that, more variation in the way the model applies to the data. Whether the latest version of NS was distributed with new Windows computers changed over time, as did the distribution of IE with Macintosh computers.[6]

There is considerable observable variation in initial conditions even within this restricted sample, depending on the exact date of the survey, the date of release of the newest version of the browser, and the probability that someone observed on the survey date bought their computer since the date of release of the newest version of their browser. That probability ranges, in sample, from zero to one, with a number of different values in between. Further variation in initial conditions arises because our users first began using the Internet at different times. Our strategy in defining the sample and estimates in this paper is to keep that useful variation, so that our model is well identified both in the descriptive estimates and in the structural model, while staying in a framework within which the mapping from model to data is simple and broadly the same for all observations in the sample.

4.1. Means, etc.

In Table 4, we report descriptive statistics of the data used in estimation. Even restricting our sample, we have 5,556 observations. Just over 40% of them have the newest version of their

[5] We also drop 47 very early adopters of Windows 98. These are very odd observations, and not numerous.
[6] These distribution arrangements are documented in Bresnahan and Yin (2005).

browser (0.402) and 56% of them state correctly whether they have the newest version (0.562). Just as a threshold point, 44% (1-.562) of these consumer cannot correctly state the choice they have made; this is encouraging for a model of consumer information like ours.

The table also reports the regressors we use and the variables we use to set initial conditions for the structural model. We observe these consumers over a range of dates in the late 1990s, called *SURDATE* in the table. Given those dates, these consumers use all three of the Windows operating systems available in this time period, Windows 3.1 (and a few 3.0), Windows 95, and Windows 98, with the largest number using the Windows 95 operating system. Also given the dates, the newest version of their brand of browser (IE) at the time we observe them ranges from IE1 to IE4.

A variable of particular interest is *ADV_NEWEST*. This is the probability, conditional on the date we observe the user, the operating system that they are using, and the introduction date of the newest IE, that this user bought a computer with the newest IE version on it. We calculate it based on external data on the purchases of computers. This variable is 0 for consumers running Windows 3.1, and 1 for consumers running Windows 98. The low mean for *ADV_NEWEST* (.11) arises primarily because most of our consumers are Windows 95 users. Browser versions appeared very quickly within the time that Win95 was the operating system on new computers. A consumer in one of the later surveys running Win95 is likely, conditional on everything we know about her (i.e., the date and that she is running Win95) to have bought her computer since the newest version of IE was released.

4.2. Descriptive Bivariate Probits

In Table 5, we report a bivariate probit in which the dependent variables are *NEWEST* and *RIGHT*. This model does not permit an estimate of the separate roles of consumer information costs and of consumer valuation of new browsers net of download costs, but it is a familiar statistical procedure and offers some quite interesting clues.

We observe the dependent variable *NEWEST* for everyone in our sample. However, we observe the dependent variable *RIGHT* only for the three surveys in which users were asked what their choice was. The likelihood function is the joint normal probability of (*NEWEST*, *RIGHT*) for users for whom we observe both variables and the marginal probability of NEWEST for the users for whom we do not observe right based on the same model.

The basic idea of our specification for browser demand is to model the consumer's choice of *NEWEST* as a function of *ADV_NEWEST*, of a list of consumer characteristics, *z*, and of interaction effects between them. Coefficients of the interaction effects are named *z_A*. Thus in column 1 of Table 5, which reflects the probit for *NEWEST*, the first coefficient is that of *ADV_NEWEST*, and the fourth coefficient, labeled *HOURS_A*, is of the interaction between *HOURS* (the number of hours per week the consumer spends on the Internet/100) and *ADV_NEWEST*. We also include a coefficient for *HOURS*, and for *DHOURS*, which is a dummy for consumers who weren't asked the question, didn't answer or made impossible answers such as 0 (the consumer is filling out an Internet survey.)[7] The idea of this specification is that it lets us determine (1) the relative importance of the default choice (measured by *ADV_NEWEST*) versus variations in consumer characteristics (*z*) and (2) the kinds of consumers who value new browsers if they must download them (coefficients of *z*) and (3) the changed impact of consumer type (*z*) if the consumer is less likely to have needed to opt-in.

Our descriptive model of consumer information has *RIGHT* as the dependent variable. The basic idea of this specification is to include all of the *z* variables we used in the model of browser demand (*HOURS*, etc.) but not their interactions with *ADV_NEWEST*. We also include some variables, particularly *SURVEY* and *Win98*, which might measure the value to the consumer of becoming well informed. The idea of this specification is to learn whether there is variation in the costs and benefits of a consumer becoming informed that predicts consumer informedness as an intermediate step before moving to a full structural model with a rational-ignorance flavor.

The table also reports some fit statistics as well as the marginal effects on the probability of the consumer having either the newest browser or correctly answering the question of each regressor.

To begin with the first column, in which we predict *NEWEST*, any examination of the coefficients must immediately note the importance of *ADV_NEWEST*. This coefficient is large, precisely estimated, and the increase in the probability of having the newest browser (ignoring the interaction effects) from an increase in *ADV_NEWEST* is 0.71. If we include the interaction

---

[7] We need not enter a *Dz* for all of the *z* variable since some of the missing values are perfectly correlated across uses (e.g., when the only reason the data are missing is because the questions were not asked in a survey wave.)

effects, i.e, calculate $\beta_{ADV\_NEWEST}+\sum_z\beta_{z\_A}\bar{z}$, this leads to an estimate of 1.13. That means that if the observable probability that a consumer got the newest browser with their computer rises by 1, the probability that they have the newest browser rises by somewhat more than one. An ML test of the hypothesis that *ADV_NEWEST* and all the interaction terms can be excluded rejects overwhelmingly. This is not surprising as the restricted model (whose coefficients are not shown in a table) predicts *NEWEST* far worse.[8]

In addition to the obvious conclusion that this is a market in which distribution is very important, we also conclude from this that our model of initial conditions is working well. In our structural model, we impose the restriction that a person whose initial conditions change such that the chance they got the newest version with their computer increases by some amount will increase the probability that they are using the newest version by that same amount. We cannot impose that restriction on the descriptive probit, but we note that it is not far wrong: our point estimate of it is not 1, but 1.13.

Second, these descriptive results also tell us some interesting things about the demand for the newest technology net of download costs and information. Looking at the variation in *z* at the point *ADV_NEWEST*=0 in the *NEWEST* equation and at the variation in *z* in the *RIGHT* equation tells us something about this. Focusing on reasonably precisely estimated coefficients, men are about 5% more likely to have the newest version of the browser (at the point *ADV_NEWEST*=0, a caveat I shall now stop repeating) than women, and older people are about 18% less likely to have the newest browser than younger people. These are unsurprising descriptive findings, but what do they tell us? If we look at the *RIGHT* equation, we see that men are more likely to be right about whether they have the newest browser, by 10% or so, and older people are more likely to be wrong about what their choice is, by about 17% or so. Based on these numbers, it is interesting to take up the question of whether information and fundamental underlying demand play different roles for different kinds of consumers. One might, for example, want to investigate the proposition that information processing costs play a different role in the age results (-0.173/-0.178) than they do in the gender results 0(.096/0.045).

---

8 In the restricted model, the average predicted probability of NEWEST=1 is .309 for users who are in fact running the newest version, and .258 for users who in fact are not.

A particularly interesting variable refers to who pays for the user's Internet access. We include a variable which is a dummy if the user does not pay herself, called *PAYWORK_DK*.[9] This gets a negative coefficient, reflecting the tendency of IT departments to slow down software upgrades in the interest of having an internal standard. That effect is more than reversed, however, if we increase the likelihood that the newest version came with the computer from zero to one.

Finally, note that we include in the *RIGHT* equation a regressor for the survey in which we observe the user (that is, a count variable which increases by 1 every six months since the surveys are twice a year) and a dummy for whether the user is running Windows 98. Both of these have negative coefficients, and the Windows 98 dummy is larger in absolute value. The marginal effects are not small: every six months, these Windows-IE users are 11% more likely to err in their report of their browser choice, and the Windows 98 users are 18% more likely to err. It is possible that some, but not all, of this effect is selection, as the composition of users of the Internet generally may be shifting over time to less sophisticated users, and Windows 98, a newer operating system, may have been bought by less sophisticated users.[10] Another interpretation of these effects is that the incentive of Windows-IE users to become informed about browser choice is declining. By the time of Windows 98, the newest version of the IE browser was tightly tied the operating system; the closeness of that tie had been growing over time within the period of our sample. The increasing strength of the default option implies a declining incentive to invest in information.

With that encouragement from the descriptive probits, we turn to estimation of a model which has a rational user investment in information.

4.2.1. Parameter Estimates

We report estimates of our structural model in Table 6. We report parameter vectors in which selected z predict (the logit of) the consumer information observation hazard ψ and the download cost φ. We also report the parameters of the equation for *RIGHT*. The sample used

---

[9] It has this odd name because we include the small group of users who don't know who pays for their Internet with the ones whose employer pays rather than with the ones who pay themselves.

[10] The coefficients are too large, quantitatively, to be explained by any such trend; newer users would have to be more than 100% more likely than older users of the Internet to be knolwedgeable. In our structural model, which conditions on the date the user was first on the Internet, any such selection / composition problem is far less likely to be a problem.

here is the same as in the descriptive bivariate probit analysis above and thus the descriptive statistics of the variables we are using are as above.

As a threshold matter, we note that even though our structural model imposes many limitations on the equation for *NEWEST*, the model predicts that variable not much worse than the unstructured bivariate probit did. The likelihood is lower in the structural model than in the unrestricted probit – not surprisingly, since we have imposed constraints – but the lower likelihood clearly arises only because we are predicting *RIGHT* less well than did the unstructured model. It is particularly notable in this context that our structural model sharply restricts the role of the distribution of browsers with new computers; *ADV_NEWEST*, the most important predictor of *NEWEST* in the descriptive probit, enters this model only through the (highly restricted) initial conditions.

Looking first at the coefficients in the net download cost (φ_i), which are located on the left of the table, the first thing to note is that the coefficients of very few regressors are estimated precisely enough to reject the hypothesis that they are zero. Statistically, only the consumer's modem speed variables (*DSPEED* and *LSPEED*) predict (net) download cost. A consumer who has a faster modem, or who does not know the modem speed (for example, because they get Internet connection services at work or at university) is estimated to have a lower download cost. These are the signs we would expect for these variables, of course.

In contrast, we can estimate the coefficients of most of the *z* variables that predict consumers' hazard for observing the newest product once it is in the marketplace, ψ. Further, confining attention to the variables which are precisely estimated, we have a rich set of estimates of consumer information-gathering capabilities and incentives. Consumers become informed about new versions if they use the Internet more, if they work in the computer industry, if they are men, if they are younger, or if they manage their own internet connection instead of having it done for them at work.

4.2.2. Probability Derivatives (Hazard Derivatives.)

The event, *NEWEST*, has a dynamic definition in our model, in which a consumer can become informed of the existence of the newest version of their brand of browser starting at the time it is introduced into the marketplace and continuing up to the time of observation. At each period, the event, "get newest," requires both observation and choice. Given this structure, we think the easiest way to understand the quantitative implications is by examining the two parts of the hazard function for getting the newest, that is, the hazard function for observation and the decision to install the newest version of the browser conditional on observing.

The hazard function for starting to use the newest browser has two parts. First, in each period, a user who has not yet observed the newest browser has a hazard for observing it of ψ_i (i.e., that is the period probability of observation occurring conditional on observation not having happened). The hazard for observing takes a comparatively simple form in our model, as ψ_i is a function of *z* and parameters.

If observation occurs, adoption of the new version occurs with probability:

$\Pr([u(x_t)+\beta\ W(x_t, X_t, s_t, 0; \varphi_i, \psi_i) < u(X_t)- \varphi_i+\varepsilon_t +\beta\ W(X_t, X_t, s_t, 0; \varphi_i, \psi_i)]$

We calculate this (and its derivatives) assuming that x_t, β and X_t, s_t are evaluated at central values in the dataset.[11] The probability of this event is called $H(\varphi_i, \psi_i)$:

$$H(\varphi_i, \psi_i)=\Pr([u(\underline{x}_t)+\beta\ W(\underline{x}_t, \underline{X}_t, 0.\ 0; \varphi_i, \psi_i) < u(\underline{X}_t)- \varphi_i+\varepsilon_t +\beta\ W(\underline{X}_t, \underline{X}_t, s_t, 0; \varphi_i, \psi_i)] \qquad (11)$$

We note that H() depends on the individual type, φ_i, ψ_i through the value function going forward and also depends on φ_i directly in the current period. As a result, H() also depends on z in <u>two ways,</u> since each z enters both φ_i, ψ_i

Table 7 reports the mean of H() and of ψ_i and probability derivatives for several parts of the hazard. At the bottom of the table, we provide a number of background statistics. These include the population mean predictions of the elements in the hazard. We also report the population mean of the time lag between the most recent browser introduction date and the survey date.

All of the probability derivatives reported in the table are evaluated at means of both the z variables (the regressors here) and other elements of our model which are not parameters. In particular, we report the average probability derivatives evaluated (1) at the time we observe each consumer, (2) averaging over the initial conditions in the way our probability model does, (3) holding all *z* other than the *z* of interest fixed at their means, (4) calculating a derivative for continuous valued *z* and a difference (from 0 to 1 for dummy *z*), and (5) holding all underlying parameters, such as the value to consumer of the newest browser, fixed at their means.

The first thing to note about Table 7 is the mean of H(), which is 0.96. Conditional on becoming informed, our estimates say that consumers overwhelmingly choose to download. In

---

[11] The values for x and X are given by prior estimates. We evaluate the functions after the introduction of a product so s_t=0.

short, the model locates the important economics of consumer acquisition of the latest version of the browser in information/observation, not in demand in the ordinary sense. This finding is, of course, entirely consistent with both our earlier finding that distribution of a browser with a new computer is very important and with the broader observation that there was tremendous strategic advantage to browser supplier Microsoft (whose products we are examining in these tables) to have their browser become the default choice for consumers. In that sense the finding is unsurprising. On the other hand, this finding is not compelled by our model or our specification. Indeed, our specification restricts the parameters of the hazard function for consumers becoming informed by linking it both to *NEWEST* and to *RIGHT*. That restricts the parameters in ψ_i, so on purely econometric grounds our model would appear likely to locate the "action" in the other parameters, in φ_i,. The opposite, however, has occurred, which leads us to conclude the model is working very well.

We conjectured above that is the role of the initial conditions that nails down the relative importance of information vs. (net) download costs, and this can be seen by looking at the Figure which immediately follows Table 7. The Figure shows the probability of *NEWEST* as a function of φ_i, ψ_i for a particular observable case. As you can see, for this case the floor on the probability of *NEWEST* is .4, driven by the date of OS introduction. For consumers in this class who nonetheless did not get the newest browser, the model could rationalize their choice either with a high download cost or a low probability of observing. But in the area of the Figure with high download cost, the probability surface has little slope in either direction. Thus, to rationalize the data, the model has chosen high information costs / low hazard for becoming informed.

Given this initial finding that most of the action is in the hazard for becoming informed, we do not discuss in depth the details of the separate impact of *z* on information and on H(). Table 7 reports all the relevant information, including probability derivatives of the hazard in two senses. The columns labeled "conditional" report derivatives of the hazard for acquiring *NEWEST* conditional on observing, $H(\varphi_i, \psi_i)$. For each regressor (z) , we report the impact that flows only through ψ_i, only through φ_i, and through both

We pay more attention to the overall hazard for adopting the newest browser. The columns labeled "unconditional" report derivatives of the unconditional hazard of acquiring

NEWEST, ψ_i H(φ_i, ψ_i). Here again we report, for each *z*, both flows of causation and the total (recognizing that most of the action flows through ψ_i.)

The mean of *NEWEST* is 0.40, and the mean number of months elapsed from the time the newest browser is released and the time the consumer is observed is 6.8. What determines this low level of use of the newest technology? Part of what is going on is the initial conditions: most consumers in this sample have to opt-in to get the newest browser.

The second part of what is going on is the low hazard rate for consumer information. The hazard for downloading the newest version can be written as the probability of getting informed about it times the conditional probability of downloading it if informed, i.e., as P=ψ_i H(φ_i, ψ_i). Evaluated at the means, this is 0.130. This is reflected in the simple table below as the first row. It is not instantly clear that this row should be compared to the mean of *NEWEST*, as the model is highly nonlinear, but it appears to predict a mean adoption rate of 0.61.

On the other hand, there is very considerable variation across *z* in the hazard for becoming informed, and this leads in turn to considerable variation in the hazard for getting the newest browser. The pattern of these estimates broadly follows that discussed above with regard to the structural estimates of the ψ_i function parameters. For example, while the mean unconditional hazard for getting the newest browser is 0.13, men and women differ in this hazard by 0.08.

5. Information is More Important

We reported two broad findings in the last section. Our model appears to have done a very good job on – perhaps not such a difficult problem -- the demand for browser novelty by those users whose preferred brand of browser is IE and whose preferred operating system is Windows. We say that this might not be such a difficult problem because it maps to our model in a very linear way; users have a default browser of their brand which came with their computer, and we need only model their decision to opt in to the newest version of that browser. That said, we subjected the model to a number of difficult comparisons and external validity checks, which it appears to have solved well. That leaves us with some confidence in the economic finding we have to report, which is that both the level and the variation in consumer information processing are the most important predictors of the demand for browsers. The speed of a consumer's modem comes in second, but a distant second. Other characteristics of consumers (*z*) appear to enter demand not through the ordinary channel of trading off the costs and hassles of downloading and installing the latest browser versus the benefits of having the newest technology, but rather through variation in the consumer's information gathering.

The empirical challenge of this paper was to begin to answer the question, "What is behind the transaction costs which make default browser installation so important?" We locate them in consumer information processing, at least for the decision we study (opt-in to the newest browser) and for the consumers we study. The ability to discern when information processing is central to the distinction between opt-in and opt-out is more generally important, and we will continue to investigate it.

Our approach is analytic; we anticipate that the finding of the importance of information costs is not universal and we would get somewhat different answers if we investigated different markets or products. We will turn next to investigating browser brand choice as well as version choice; consumers who choose Netscape later in our sample period need to opt-in to get that product, which we anticipate will let us tell a richer story of demand and information gathering costs. In other, similar, industries, the information held by consumers might tell a very different story. If we were to study PC music-playing software in the modern era, for example, we would expect music-focused consumers to become well informed about new versions of iTunes quickly, so that distribution of that product with new computers or with updates to the operating system might be far less influential than for less-popular music players. The ultimate message is that the distinction between opt-in and opt-out depends on deep parameters which can be measured as they have been measure here.

6. Tables

Table 1

Distribution of Responses across Survey Waves

Survey Number/Date	Number of Respondents
4 / October, 1995	8529
5 / April, 1996	4743
6 / October, 1996	8254
7 / April, 1997	13248
8 / October, 1997	4452
9 / April, 1998	6284
10 / October, 1998	2389

Table 2

Consumers' Answer to Questions about Their Choice

Possible response to "Do you think you are using the most up-to-date version of your browser?"	Frequency with which response was chosen (across surveys 8, 9, and 10)
Yes, I am quite certain	6280
Yes, and it is a pre-release/beta version	275
Yes, but I am not so certain	2238
No, I think I am using an older version	1119
No, I am definitely using an older version	2843
Don't know	416

Table 3

Initial Condition Cases (124 in total)

o	Brand/ platform	survey	t0_begin	t0_begin	t0_length	# obs
			year	month	(months)	
1	nswin98	7	1998	9	1	117
2	nswin95	2	1995	9	7	1007
3	nswin95	2	1995	10	6	247
4	nswin95	3	1995	9	13	1842
5	nswin95	3	1995	10	12	590
6	nswin95	3	1996	4	6	504
7	nswin95	4	1995	9	19	4273
8	nswin95	4	1996	4	12	1050
9	nswin95	4	1996	10	6	980
10	nswin95	5	1995	9	25	1929
11	nswin95	5	1996	10	12	292
12	nswin95	5	1997	4	6	249
13	nswin95	6	1995	9	31	2904
14	nswin95	6	1997	4	12	361
15	nswin95	6	1997	10	6	247
16	nswin95	7	1995	9	36	442
17	nswin95	7	1995	10	35	280
18	nswin95	7	1997	10	10	52
19	nswin95	7	1998	4	6	25
20	nswin31	1	1994	12	8	4084
21	nswin31	1	1995	4	6	1461
22	nswin31	2	1994	12	8	894
23	nswin31	2	1995	4	6	412
24	nswin31	2	1995	10	6	495
25	nswin31	3	1994	12	8	285
26	nswin31	3	1994	12	10	958
27	nswin31	3	1995	10	6	504
28	nswin31	3	1996	4	6	470
29	nswin31	4	1994	12	8	403
30	nswin31	4	1994	12	16	1050
31	nswin31	4	1996	4	6	444
32	nswin31	4	1996	10	6	360
33	nswin31	5	1994	12	8	119
34	nswin31	5	1994	12	22	235
35	nswin31	5	1996	10	6	50
36	nswin31	5	1997	4	6	67
37	nswin31	6	1994	12	8	104
38	nswin31	6	1995	4	24	186
39	nswin31	6	1997	4	6	29
40	nswin31	6	1997	10	6	24
41	nswin31	7	1994	12	8	16
42	nswin31	7	1995	10	24	8
43	nswin31	7	1998	4	6	1
44	nsmac	1	1994	12	0	1626
45	nsmac	1	1994	12	4	512

46	nsmac	1	1995	4	6	472
47	nsmac	2	1994	12	0	512
48	nsmac	2	1994	12	4	657
49	nsmac	2	1995	4	6	234
50	nsmac	2	1995	10	6	173
51	nsmac	3	1994	12	0	703
52	nsmac	3	1994	12	10	1005
53	nsmac	3	1995	10	6	282
54	nsmac	3	1996	4	6	167
55	nsmac	4	1994	12	0	1234
56	nsmac	4	1994	12	16	1539
57	nsmac	4	1996	4	6	277
58	nsmac	4	1996	10	6	172
59	nsmac	5	1994	12	0	403
60	nsmac	5	1994	12	22	385
61	nsmac	5	1996	10	6	55
62	nsmac	5	1997	4	6	35
63	nsmac	6	1994	12	0	408
64	nsmac	6	1995	4	24	350
65	nsmac	6	1997	4	6	47
66	nsmac	6	1997	10	6	19
67	nsmac	7	1994	12	0	375
68	nsmac	7	1995	10	24	132
69	nsmac	7	1997	10	6	6
70	nsmac	7	1998	4	6	7
71	ie win98	7	1998	9	1	292
72	ie win95	1	1995	9	1	372
73	ie win95	2	1995	9	7	80
74	ie win95	2	1995	10	6	13
75	ie win95	3	1995	9	13	532
76	ie win95	3	1995	10	12	170
77	ie win95	3	1996	4	6	148
78	ie win95	4	1995	9	19	896
79	ie win95	4	1996	4	12	158
80	ie win95	4	1996	10	6	137
81	ie win95	5	1995	9	25	410
82	ie win95	5	1996	10	12	65
83	ie win95	5	1997	4	6	42
84	ie win95	6	1995	9	31	1151
85	ie win95	6	1997	4	12	186
86	ie win95	6	1997	10	6	134
87	ie win95	7	1995	9	36	223
88	ie win95	7	1995	10	35	205
89	ie win95	7	1997	10	10	66
90	ie win95	7	1998	4	6	26
91	ie win31	2	1996	4	0	4
92	ie win31	3	1996	4	0	38
93	ie win31	3	1996	4	6	14
94	ie win31	4	1996	4	0	37
95	ie win31	4	1996	4	6	11
96	ie win31	4	1996	10	6	13
97	ie win31	5	1996	4	0	6
98	ie win31	5	1996	4	6	22
99	ie win31	5	1996	10	6	5
100	ie win31	5	1997	4	6	7
101	ie win31	6	1996	4	0	11
102	ie win31	6	1996	4	12	37
103	ie win31	6	1997	4	6	9
104	ie win31	6	1997	10	6	20

105	ie win31	7	1996	4	0	6
106	ie win31	7	1996	4	18	7
107	ie win31	7	1997	10	6	2
108	ie win31	7	1998	4	6	1
109	ie mac	2	1996	4	0	15
110	ie mac	3	1996	4	0	42
111	ie mac	4	1996	4	0	204
112	ie mac	4	1996	4	6	4
113	ie mac	4	1996	10	6	6
114	ie mac	5	1996	4	0	43
115	ie mac	5	1996	4	6	30
116	ie mac	5	1996	10	6	2
117	ie mac	6	1996	4	0	30
118	ie mac	6	1996	4	12	25
119	ie mac	6	1997	4	6	1
120	ie mac	6	1997	10	6	1
121	ie mac	7	1996	4	0	63
122	ie mac	7	1996	4	18	33
123	ie mac	7	1997	10	6	2
124	ie mac	7	1998	4	6	2

Table 4

Means of Data in Estimation sample n: 5556

	Mean	Std Dev	Minimum	Maximum
NEWEST[14]	0.40266	0.49052	0.00000	1.00000
RIGHT[15]	0.56154	0.49628	0.00000	1.00000
IE	1	0	0	1
AWIN98	0.052556	0.22317	0.00000	1.00000
AWIN95				
SURVEY	7.70644	1.70065	4.00000	10.00000
ADV_NEWEST	0.11009	0.16603	0.00000	0.62867
DHOURS[16]	0.081353	0.27340	0.00000	1.00000
I_HOURS	0.14147	0.12185	0.00000	0.50000
I_USE	1.14483	0.89117	0.00000	2.52000
I_PAYWORK_DK	0.25198	0.43419	0.00000	1.00000
DSPEED	0.95554	0.20613	0.00000	1.00000
LSPEED	3.76964	1.70072	0.00000	11.96582
OCCCOMP	0.27178	0.44492	0.00000	1.00000
MALE	0.71220	0.45278	0.00000	1.00000
DAGE	0.013499	0.11541	0.00000	1.00000
I_AGE	0.36144	0.13610	0.00000	0.83000
INCNS	0.13301	0.33962	0.00000	1.00000
I_INC	0.52582	0.42292	0.00000	5.00000

[14] NEWEST is 1 if the consumer is actually using the newest version of their browser.

[15] RIGHT is 1 if NEWEST=1 and the consumer gave any of the three "yes" answers listed in Table 2, or if NEWEST=0 and the consumer gave either of the two "no" answers listed there. Note this definition treats "don't know" as uninformed, and puts in the same class as an incorrect response.

[16] We include a dummy for each continuous regressor if the consumer does not answer the question, and give it a name Dz. Thus DHOURS is a dummy for no data on consumer hours on the internet per week.

Table 5

Descriptive Bivariate Probit Results

	NEWESTt			RIGHT		
	Est	SE	MFX	Est	SE	MFX
const	-1.434	0.162		4.583	0.438	
Win98				-1.004	0.107	-0.227
Survey				-0.516	0.047	-0.117
ADV_NEWEST	2.950	0.519	0.711			
DHOURS	0.122	0.082	0.029	-0.015	0.084	-0.003
HOURS	0.142	0.281	0.034	0.554	0.249	0.125
HOURS_A	3.448	1.234	0.831			
USE	0.092	0.039	0.022	0.116	0.040	0.026
USE_A	-0.139	0.176	-0.034			
PAYWORK_DK	-0.060	0.070	-0.014	-0.665	0.074	-0.150
PAYWORK_DK_A	3.342	2.270	0.806			
DSPEED	0.126	0.129	0.030	0.256	0.139	0.058
LSPEED	0.024	0.017	0.006	0.005	0.019	0.001
LSPEED_A	0.021	0.081	0.005			
OCCCOMP	0.170	0.059	0.041	0.173	0.064	0.039
OCCCOMP_A	0.138	0.295	0.033			
MALE	0.187	0.064	0.045	0.425	0.055	0.096
MALE_A	0.662	0.269	0.160			
DAGE	-0.016	0.190	-0.004	-0.500	0.247	-0.113
AGE	-0.737	0.230	-0.178	-0.765	0.201	-0.173
AGE_A	0.828	0.926	0.200			
DINC	-0.035	0.071	-0.008	0.053	0.090	0.012
INC	-0.090	0.068	-0.022	-0.066	0.077	-0.015
INC_A	0.212	0.297	0.051			
rho	0.633	0.028				

\# Obs=5556

Ln(likelihood)=-3708.12

For each dependent variable, *NEWEST* or *RIGHT,* the three columns are the estimate, the estimated standard error, and the estimated marginal effects ("MFX.")

Predictions of the model:

avg. Pr(newest)	0.494	for users who have newest
avg. Pr(newest)	0.178	for users who don't have newest
avg. Pr(right)	0.660	for users who are right
avg. Pr(right)	0.423	for users who are wrong

Table 6

Estimates of Structural Model

	φ_i		ψ_i			*RIGHT*	
	Est	SE	Est	SE		Est	SE
CONST	0.906	0.629	-1.855	0.402	CONST	7.074	0.723
DHOURS	-0.521	0.307	-0.142	0.236	AWIN98	-1.354	0.233
HOURS	-0.743	0.728	0.491	0.642	SURVEY	-0.811	0.078
USE	-0.120	0.106	0.225	0.109	α	0.669	0.192
PAYWORK	-0.097	0.199	-0.713	0.212	δ	2.719	0.124
DSPEED	-1.965	0.709	-1.242	0.500			
LSPEED	-0.459	0.190	-0.102	0.043			
OCCCOMP	-0.247	0.171	0.265	0.174			
MALE	-0.023	0.185	0.827	0.266			
DAGE	-1.613	1.217	-1.329	0.654			
AGE	0.360	0.541	-1.290	0.600			
DINC	0.031	0.217	0.122	0.230			
INC	-0.184	0.228	-0.206	0.201			

Obs=5556
Ln(likelihood)=- 5222.969

For the columns headed φ_i, and ψ_i, what is presented are the estimates of θ and their estimated standard errors. For RIGHT, which has a new set of row labels, the columns are the probit estimates and their standard errors.

Predictions:

avg. Pr(NEWEST)	0.543	for users who have NEWEST
avg. Pr(NEWEST)	0.329	for users who don't have NEWEST
avg. Pr(RIGHT)	0.753	for users who are right
avg. Pr(RIGHT)	0.681	for users who are wrong

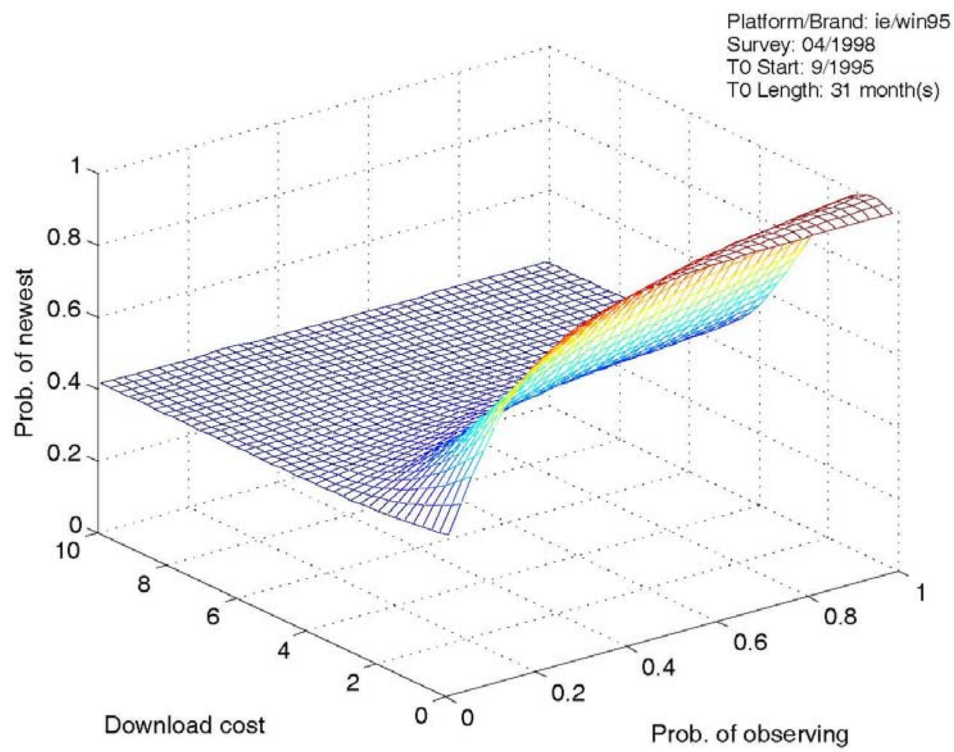
Table 7

Probability Derivatives for Hazard

Person	$P = \psi_i H(\varphi_i, \psi_i)$			$H(\varphi_i, \psi_i)$.		
	Unconditional Hazard for NEWEST			Conditional adoption hazard		
Mean						
Change in	ϕ_i only	ψ_i only	both	ϕ_i only	ψ_i only	both
DHOURS	0.015	0.009	0.021	0.002	-0.013	-0.011
HOURS	0.019	-0.050	-0.008	0.003	0.049	0.057
USE	0.005	-0.019	-0.012	0.001	0.021	0.023
PAYWORK	0.004	0.029	0.030	0.001	-0.054	-0.054
DSPEED	0.027	0.035	0.037	0.004	-0.079	-0.079
LSPEED	0.014	0.007	0.018	0.002	-0.009	-0.008
OCCCOMP	0.009	-0.023	-0.009	0.001	0.025	0.028
MALE	0.001	-0.102	-0.099	0.000	0.083	0.084
DAGE	0.025	0.035	0.037	0.003	-0.082	-0.082
AGE	-0.021	0.035	0.033	-0.003	-0.081	-0.081
DINC	-0.001	-0.009	-0.011	0.000	0.011	0.011
INC	0.007	0.012	0.017	0.001	-0.018	-0.017

Mean of H() .963

Mean of H()*ψ_i .130



Platform/Brand: ie/win95
Survey: 04/1998
T0 Start: 9/1995
T0 Length: 31 month(s)
Prob. of newest
1
0.8
0.6
0.4
0.2
0
10
8
6
4
2
0
Download cost
0
0.2
0.4
0.6
0.8
1
Prob. of observing

APPENDIX A:

Table 8

Variable Definitions and Sources

"this" refers to user/observation specific. "given" refers to subscripted variable.

Variable	Source, varies by	Definition
who	GVU (4-10)	unique identifier
IE	brand	dummy for choice of internet explorer
v	version	numerical value of version of browser used
newest	Browser, OS, survey	This browser is the newest within its brand on this OS (includes pre-release) in this survey
Adv_newest	Browser, survey, OS	Adv for newest IE browser on this OS in this survey
Disadv_newest	Same as above	disadv for newest NS browser on this OS in this survey
Survey	survey	Survey number
Surdate	survey	Date of survey
Amac	GVU (4-10) OS	Operating System Used, According to agent file Dummy for OS: Mac = 1
Awin31	GVU (4-10)	Operating System Used, According to agent file Dummy for OS: Windows 3.1 = 1
Awin95	GVU (4-10)	Operating System Used, According to agent file Dummy for OS: Windows 95 = 1
Awin98	GVU (4-10)	Operating System Used, According to agent file Dummy for OS: Windows 98 = 1
Age	GVU (4-8)	"What is your age?" Numerical value entered by respondent =0 for "Rather not say" (not an option in survey 4)
Age	GVU (9-10)	"What is your age?" Checkboxes for 5 year intervals Coded at interval midpoints =0 for "Rather not say"
Dage		Dummy=1 for "Rather not say" age
Male	GVU (4-10)	"What is your sex?" Dummy for sex: male = 1

White	GVU (4-10)	"How would you classify your race?" Dummy for race: white = 1
Asian	GVU (4-10)	"How would you classify your race?" Dummy for race: asian = 1
Black	GVU (4-10)	"How would you classify your race?" Dummy for race: black = 1
Raceot	GVU (4-10)	"How would you classify your race?" Dummy for race: hispanic, latino, indigenous, native, multi, spanish, other = 1
Racens	GVU (4-10)	"How would you classify your race?" Dummy for race: not say, na = 1
Educ11	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: grammar = 1
EducHS	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: HS, special, abitur, voctech = 1
EducSC	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: some college, some = 1
EducC	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: college = 1
EducPG	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: masters, professional, doctoral = 1
EducOT	GVU (4-10)	"Please indicate the highest level of education achieved." Dummy for education: other = 1
Educyr		Coded value of education in years Educ11=8, EducHS=12, EducSC=14, EducC=16, EducPG=19, EducOT=0
Inc	GVU (4-10)	"Please indicate your current household income." (in thousands of dollars) Checkboxes for $10-15 intervals Coded at interval midpoints Rather not say/na = -9
Dinc		Dummy for "Rather not say" income
Ynet	GVU (4-10)	"How long have you been using the internet?" Checkboxes for .5-2yr intervals Coded at interval midpoints
Ynet1		Dummy for has the respondent been on the web for a year or longer
NetStart		Date on which respondent started using the web Coded by subtracting ynet from survey date
USA	GVU (4-10)	"Where are you located?" Dummy for location: USA=1
EUR	GVU (4-10)	"Where are you located?" Dummy for location: Europe=1
LocOT	GVU (4-10)	"Where are you located?"

		Dummy for location: Africa, Antarctica, Asia, Canada, Central America, Mexico, Middle East, Oceania, South America, West Indies = 1
Dos	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: dos = 1
NT	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: nt = 1
Win	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: Windows = 1
Win95	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: Windows 95 = 1
Mac	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: Mac = 1
Mac8	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: Mac8 = 1 When Mac8 was not an option = -99 (survey 4-9)
Win98	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: Windows 98 = 1 When Windows 98 was not an option = -99 (survey 4-9)
PlatDK	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: don't know = 1
PlatOT	GVU (4-10)	"What is your primary computing platform?" Dummy for platform: unix, pc_unix, os2, vt100, next step, vms, t, webtv, other = 1
OSWrong		Dummy for match between agent coding of OS and self-reported platform when platform is specific (mac, mac8, win95, win98, ot) =1 when self-reporting is incorrect; or when platform is non-specific, but is clearly incorrect (e.g., amac = 1, but osns=1 and platdk=0)
OSNS		Dummy for non-specific self-reported OS (dos, nt, win, dk)
OccComp	GVU (10)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: support, consultant = 1
OccComp	GVU (4-9)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: computer = 1
OccProf	GVU (10)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: trained professional = 1
OccProf	GVU (4-9)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: professional = 1
OccMgmt	GVU (10)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: upper management, middle

		management, junior management = 1
OccMgmt	GVU (4-9)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: management = 1
OccOT	GVU(10)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: student, researcher, skilled labor, self employed, administrator, temporary, other = 1
OccOT	GVU (4-9)	"Which of the following categories best describes your primary occupation?" Dummy for occupation: education, other = 1
AccessW	GVU (10)	"What is the primary place you access the WWW from?" Coded as 1 when respondent indicated work as the most frequent place of access Ties were resolved in the following order: work, home, public, other
AccessW	GVU (4-9)	"What is the primary place you access the WWW from?" Dummy for place of access: work, primarily work = 1 When question was not asked = -99 (Survey 4)
AccessWna		Dummy for access question not asked
AccessH	GVU (10)	"What is the primary place you access the WWW from?" Coded as 1 when respondent indicated home as the most frequent place of access Ties were resolved in the following order: work, home, public, other
AccessH	GVU (4-9)	"What is the primary place you access the WWW from?" Dummy for place of access: home, primarily home, friend = 1 When question was not asked = -99 (Survey 4)
AccessP	GVU (10)	"What is the primary place you access the WWW from?" Coded as 1 when respondent indicated public as the most frequent place of access Ties were resolved in the following order: work, home, public, other
AccessP	GVU (4-9)	"What is the primary place you access the WWW from?" Dummy for place of access: public = 1 When public was not an option = -99 (Survey 4-7)
AccessOT	GVU (10)	"What is the primary place you access the WWW from?" Coded as 1 when respondent indicated other as the most frequent place of access Ties were resolved in the following order: work, home, public, other
AccessOT	GVU (4-9)	"What is the primary place you access the WWW from?" Dummy for place of access: distributed, school, other, na = 1 When question was not asked = -99 (Survey 4)
PaySelf	GVU (4-10)	"Who pays for your internet access?"

		Dummy for payer: self, parents = 1 When question was not asked = -99 (Survey 9)
PayWork	GVU (4-10)	"Who pays for your internet access?" Dummy for payer: work, school, other = 1 When question was not asked = -99 (Survey 9)
PayDK	GVU (4-10)	"Who pays for your internet access?" Dummy for payer: don't know = 1 When question was not asked = -99 (Survey 9)
Eng	GVU (4-10)	"What is your native/first language?" Dummy for language: english = 1 When question was not asked = -99 (Survey 4)
LangOT	GVU (4-10)	"What is your native/first language?" Dummy for language: all other languages = 1 When question was not asked = -99 (Survey 4)
Marr	GVU (4-10)	"What is your current marital status?" Dummy for marital status: married = 1
Single	GVU (4-10)	"What is your current marital status?" Dummy for marital status: single = 1
Marrot	GVU (4-10)	"What is your current marital status?" Dummy for marital status: divorced, separated, widowed, other, not say = 1
IEPref	GVU (4-10)	"What online service do you currently subscribe to?" Dummy for online services: aol, att, compuserve, delphi, ibm, mindspring, msn, netcom, prodigy = 1 When question was not asked = -99 (Survey 7)
OnlineOT	GVU (4-10)	"What online service do you currently subscribe to?" Dummy for online services: ambert, europeonline, genie, pipeline, t-online, other, other_local, other_national, web-based e-mail = 1 When question was not asked = -99 (Survey 7)
OnlineDK	GVU (4-10)	"What online service do you currently subscribe to?" Dummy for online services: don't know, none = 1 When question was not asked = -99 (Survey 7)
Netsc	GVU (4-10)	"Which browser do you expect to be your primary browser in 12 months?" Dummy for browser: netscape communicator, netscape navigator = 1 When question was not asked = -99 (Survey 4)
Micro	GVU (4-10)	"Which browser do you expect to be your primary browser in 12 months?" Dummy for browser: aol, microsoft = 1 When question was not asked = -99 (Survey 4)
BrowseOT	GVU (4-10)	"Which browser do you expect to be your primary browser in 12 months?" Dummy for browser: hotjava, lotus, lynx, netcom,

		netcruiser, psi, spry, other When question was not asked = -99 (Survey 4)
Speed	GVU (4-10)	"Which of the following connection speeds do you primarily use to connect to the internet?" Checkboxes for finite speeds Coded in x,000 Unsure = 0
Dspeed	GVU (4-10)	Dummy for speed unsure
Lspeed		Log(speed) When speed = -9, lspeed = 0
Use	GVU (4-10)	"On average, how often do you use your WWW browser?" Checkboxes for varying time intervals Coded as midpoints of intervals and converted to times per month When respondent did not answer the question = 0
Duse		Dummy for missing answer to use
Hours	GVU (4-10)	"On average, how many hours a week do you use your WWW browser?" Checkboxes for varying time intervals Coded as midpoints of intervals When respondent did not answer the question = 0
Dhours		Dummy for missing answer to hours

Table 9

Summary Statistics for Broad Sample

Variable	Mean	Std. Dev.	Min	Max	Variable	Mean	Std. Dev.	Min	Max
i1	0.007	0.084	0	1	payself	-12.381	33.662	-99	1
i2	0.009	0.094	0	1	paywork	-12.586	33.583	-99	1
i3	0.067	0.250	0	1	paydk	-12.983	33.426	-99	1
i4	0.043	0.203	0	1	eng	-16.863	38.231	-99	1
i5	0.000	0.021	0	1	langot	-17.571	37.902	-99	1
n1	0.175	0.380	0	1	marr	0.451	0.498	0	1
n2	0.200	0.400	0	1	single	0.359	0.480	0	1
n3	0.366	0.482	0	1	marrot	0.191	0.393	0	1
n4	0.132	0.339	0	1	iepref	-27.167	44.418	-99	1
other browser	0.000	0.000	0	0	onlineot	-27.042	44.496	-99	1
survey	6.600	1.754	4	10	onlinedk	-27.133	44.440	-99	1
amac	0.257	0.437	0	1	micro	-17.492	37.939	-99	1
awin95	0.465	0.499	0	1	netsc	-16.971	38.182	-99	1
awin31	0.270	0.444	0	1	browseot	-17.600	37.888	-99	1
awin98	0.009	0.092	0	1	speed	2033.745	14265.810	-9	57286
age	34.770	13.654	-9	90	speedus	0.107	0.309	0	1
agens	-17.613	37.881	-99	1	use	112.788	87.983	-9	252
male	0.668	0.471	0	1	hours	12.705	11.403	-9	50
white	0.882	0.323	0	1	dagens	0.015	0.121	0	1
asian	0.029	0.169	0	1	dhours	0.040	0.195	0	1
black	0.012	0.108	0	1	dspeed	0.107	0.309	0	1
raceot	0.051	0.219	0	1	lspeed	3.654	2.221	0	11.966
racens	0.026	0.160	0	1	accesswna	0.178	0.383	0	1
educ11	0.020	0.140	0	1	IE	0.126	0.332	0	1
educhs	0.125	0.331	0	1	Netscape	0.874	0.332	0	1
educsc	0.297	0.457	0	1	newest	0.300	0.458	0	1
educc	0.313	0.464	0	1	adv_newest	0.013	0.067	0	0.629
educpg	0.230	0.421	0	1	disadv_new~t	0.073	0.146	0	0.615
educot	0.015	0.121	0	1	ds4	0.178	0.383	0	1
educyr	15.199	3.114	0	19	ds5	0.099	0.299	0	1
inc	49.020	46.428	-9	500	ds6	0.172	0.378	0	1
incns	0.147	0.354	0	1	ds7	0.277	0.447	0	1
ynet	2.532	2.040	.25	7	ds8	0.093	0.290	0	1
ynet1	0.698	0.459	0	1	ds9	0.131	0.338	0	1
usa	0.812	0.391	0	1	ds10	0.050	0.218	0	1
eur	0.062	0.241	0	1	dynet1	0.143	0.351	0	1
locot	0.126	0.332	0	1	dynet2	0.159	0.365	0	1
oswrong	0.119	0.324	0	1	dynet3	0.425	0.494	0	1
osns	0.234	0.424	0	1	dynet4	0.192	0.394	0	1
dos	0.021	0.142	0	1	dynet5	0.081	0.273	0	1
mac	0.244	0.430	0	1	plat_brand	3.280	1.360	1	8
mac8	-94.051	21.602	-99	1	d_knowcurr~t	0.274	0.446	0	1

Variable	Mean	Std. Dev.	Min	Max	Variable	Mean	Std. Dev.	Min	Max
nt	0.017	0.130	0	1	right	0.189	0.392	0	1
win95	0.472	0.499	0	1	i_hours	0.131	0.108	0	0.500
win98	-94.052	21.594	-99	1	i_use	1.131	0.875	0	2.520
win	0.191	0.393	0	1	i_inc	0.503	0.449	0	5
platot	0.028	0.164	0	1	dage	0.015	0.121	0	1
platdk	0.006	0.075	0	1	i_age	0.349	0.133	0	0.900
occcomp	0.257	0.437	0	1	i_paywork	0.402	0.490	0	1
occmgmt	0.117	0.321	0	1	i_paydk	0.005	0.068	0	1
occprof	0.219	0.414	0	1	i_paywork_dk	0.407	0.493	0	2
occot	0.407	0.491	0	1	i_accessp	0.001	0.032	0	1
accessh	-17.110	38.118	-99	1	i_accessot	0.008	0.091	0	1
accessp	-71.872	44.158	-99	1	i_accessw	0.101	0.301	0	1
accessw	-17.371	37.996	-99	1	i_langot	0.057	0.232	0	1
accessot	-17.582	37.896	-99	1	i_micro	0.136	0.343	0	1
					i_browseot	0.028	0.166	0	1
					non_us	0.188	0.391	0	1