Hyper Media Search and Consumption

Jason M.T. Roos[*] Carl F. Mela[†] Roni Shachar[‡]

March 15, 2013

[*]Assistant Professor, Rotterdam School of Management, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, Netherlands; email: roos@rsm.nl; phone: +1 206 317 1713.

[†]T. Austin Finch Foundation Professor of Business Administration, Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, North Carolina, 27708; email: mela@duke.edu; phone: +1 919 660 7767.

[‡]Dean, Arison School of Business, Interdisciplinary Center (IDC) Herzliya, Herzliya 46150, Israel; email: ronshachar@idc.ac.il; phone +972 09 960 2408.

The authors would like to thank comScore for the data used in this study as well as Peter Arcidiacono, Andrew Ching, Andres Musalem, Ken Wilbur, seminar participants at Duke (Fuqua, Dept. of Economics), Erasmus (RSM), Ohio State, Yale, Wash. U. in St. Louis, Georgia Tech, the 34th ISMS Marketing Science Conference, the 2011 Tilburg Christmas Camp, UNC Chapel Hill, U. of San Diego, U. of Michigan, 2012 HEC Marketing Camp, U. of Frankfurt, and U. of Houston for their thoughtful comments. This paper stems from the first author's dissertation.

Abstract: Hyper Media Search and Consumption

In the past five years, the number of Americans using the Internet as their main source of news has doubled to more than 40%, while those choosing newspapers has dropped to just over 30%. Although this trend signals a shift in the consumption of information in favor of hyper-media, wherein excerpting and linking is commonplace, research regarding how consumers acquire news in online environments has not proceeded apace. We develop a model of forward-looking consumers who gain knowledge and utility through the search and consumption of hyper-media, and explore its implications for welfare and site policy.

The model is estimated using comScore browsing data for five celebrity news and gossip sites, supplemented with link data scraped from site archives. The model is estimated by coupling Imai, Jain, and Ching's (2009; *Econometrica* 77(6):1865–1899) method with the Riemann manifold sampling algorithm of Girolami and Calderhead (2011; *JRSS-B* 73(2):123–214). The pairing of these recent advances enables researchers to estimate dynamic models with many more variables and states than previously considered.

Results indicate consumers are heterogeneous in their preferences for certain types of celebrity sites (e.g., females prefer sites that emphasize gossip over pictorials of female models and entertainers). Results also indicate that links to other sites are informative about the target site's potential quality; after observing one link to another site, consumers' uncertainty about the linked site's quality is about 20% lower.

Counterfactual analyses show that network throttling, in which access to all sites is slowed down uniformly in an effort to reduce congestion, has an unequal effect on site traffic. For example, sites frequented by consumers with higher than average browsing costs lose a greater share of their traffic. A related anlysis shows when carriers take bandwidth from one site and reallocate it to others, some consumers who browse less are actually better off, and some who browse more are worse off. A third counterfactual analysis shows how a change in fair use law, in which excerpts and links become less informative of the linked sites' quality, causes consumers with more extreme tastes to curtail their searches, consequently decreasing traffic at sites with mainstream content and many inbound links.

1 Introduction

Over the past decade, the consumption of news and information has fundamentally and irrevocably changed, as millions of consumers have migrated from print, television, and radio to electronic media. This profound shift in media consumption continues unabated: in just the past five years, the number of Americans identifying the Internet as their main source of news doubled to more than 40%, while the number choosing newspapers dropped to just over 30%.[1] Perhaps as a consequence, 2012 marked the first year online advertising spend exceeded that of print, with online spending expected to be double that of print by 2016.[2] Yet in spite of the growing importance of hyper-media, there remains a dearth of empirical research seeking to explain and forecast consumer behavior in the context of networked media. A coherent characterization of consumers' browsing behaviors is critically important to firms seeking to optimize their content, linking strategies, and advertising policies. Accordingly, this paper considers the problem of how users consume information via networked media ("hyper-media") where information from one provider can be excerpted and linked to by another.

To illustrate hyper-media search and consumption, Figure 1 depicts the behavior of two individuals observed in the data (see §3 for a complete description). Panels 1 and 2 show histograms for the number of sites visited each day over a period of 14 months. Immediately apparent are the differences between these consumers: while both are equally likely to browse on a given day, Consumer 1 typically visits 1 site and rarely more than 2, whereas Consumer 2 usually visits 2 or more (and as many as 10). Among other factors, these differences could be attributed to differences in search costs (higher search costs offset the expected return from visiting the next site), the relevance of information available (more relevant information enhances consumption), and the degree of redundancy of information between sites (more redundancy increases substitution across sites).

There is also substantial variation in the order of site visits, as seen in Panels 3 and 4, which depict, for the 7 most popular sites with each consumer, the order of site visits. Each path represents a distinct consumption occasion. Considerable difference is evidenced in the search order between these users, with Consumer 1 typically starting a session at the most frequented site and Consumer 2 starting from a larger set. Moreover, Consumer 2's browsing order appears to depend critically on the first site visited. For example, a visit to the fifth most frequented site (site "5") initiated

[1]Source: Pew Research Center for the People & the Press, January 4, 2011.
[2]Source: Advertising Age, January 19, 2012.

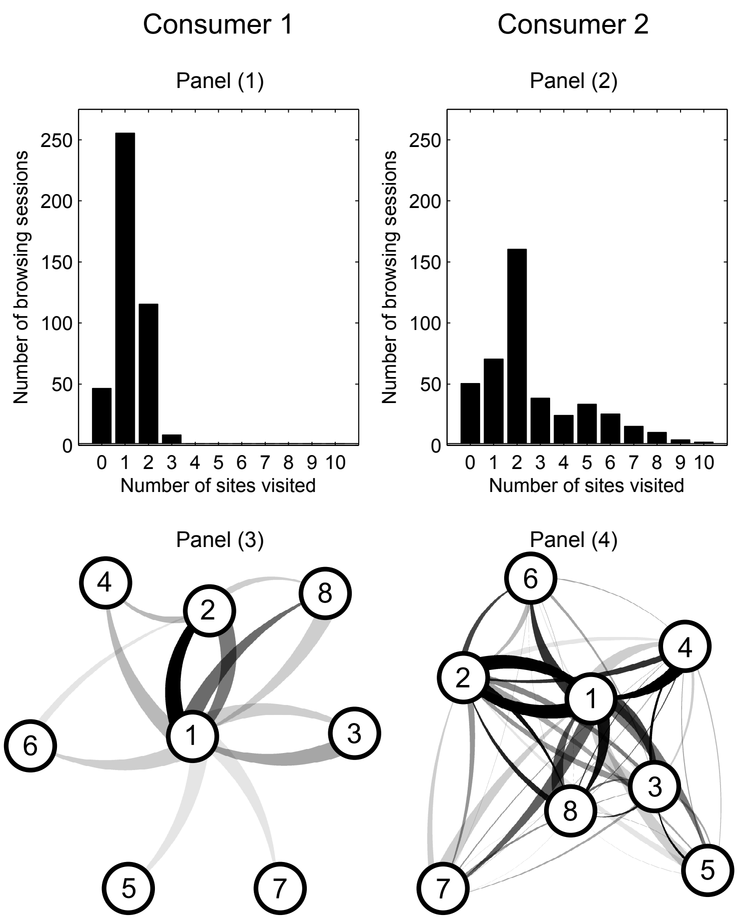

Figure 1: Comparison of search behavior for two consumers over 14 months. Panels (1) and (2) depict the frequency of search by the number of sites visited. Panels (3) and (4) depict search paths through the 7 most popular sites for each consumer. Circles represent sites and are numbered in the order of popularity with each consumer (thus, site (1) represents the site visited most frequently by each consumer). Site (8) represents all sites other than the seven most popular. Arcs between circles represent browsing sessions in which the site at the wider end of the arc directly preceded the site at the narrower end. Darker arcs indicate more frequent browsing.

a session more often than a visit to the third most frequented site ("site 3"). Moreover, a visit to site 5 preceded a visit to site 3 on more than 50 occasions, but the reverse (site 5 before site 3) occurred only four times. Such variation in choice paths could reflect differences in the match values of the various sites (more preferred sites are visited earlier), their subject matter, the anticipated amount of information, or the sites' excerpting behaviors (more links and excerpts provides better information about remaining sites). In light of the growing ubiquity of hyper-media as a source of information for consumers, this analysis can shed light on several questions of increasing relevance to hypermedia sites and regulators:

- How does the volume of information search and consumption vary by users and over time? For example, how does a change to the link structure or to the amount or relevance of information on sites affect whether and how many sites are visited? One finding suggests two contrasting effects of information; more information provides a greater incentive to begin searching, but also increases redundancy, resulting in shorter searches (i.e., fewer sites visited). Accordingly, an intermediate level of information leads to the longest sessions.

- How does the order of sites visited vary over time? For example, do consumers visit sites because of their innate preference for the sites' information, or because excerpts and outgoing links reduce uncertainty about the type of information at other sites? Can sites, via their outbound links, reduce uncertainty about the information at other sites, thereby changing the option value of visiting those linked sites?

- What are the demand and welfare implications of various Internet regulations? For example, net neutrality rules (e.g., regulating broadband carriers' ability to throttle bandwidth to high-traffic sites) and copyright laws (e.g., limitations on fair use exemptions that currently make it possible to excerpt large amounts of content from other sites) are likely to affect the consumption of news, and consequently, consumer welfare. Findings suggest that the welfare loss from network throttling is greatest among individuals with the highest browsing costs, and that the sites catering to this group's tastes suffer the greatest traffic losses. We also find that welfare effects from changes in fair use copyright laws are felt most acutely by individuals with strong preferences.

To answer these questions, we develop an empirical model of hyper-media search and consumption. In this setting, consumers surf for bits of information across linked sites as long as the likelihood of finding novel information exceeds the effort of the additional search. In this model, Internet users consume information—in other words, any utility consumers gain while reading hyper-media is the product of whatever new information they acquire along the way. Accordingly, we model information at the most fundamental level: that of individual bits of information. This approach allows the model to reflect the most essential features of hyper-media in a theoretically principled way. Of note, the demand system can also serve as a foundation to consider supply side issues such as whether to use a paywall and the effect of linking strategies on advertising revenue.

Because the set of information available at each site changes every day and thus is unknown prior to a visit, this model of search and consumption is grounded in the learning literature (Erdem and Keane, 1996). Before visiting a site, consumers form beliefs, based on past experience, about the amount and relevance of information, as well as the number and type of links they expect to find. Of note, the signaling value of excerpts and links, the rapid turnover in the information available each day, and the redundancy of information across sites lead to a number of extensions to the standard learning model. The complexity arising from these contextual factors in the face of forward-looking consumers also vastly expands the information states relevant to consumer choices. The approach we develop to estimate this problem could facilitate the estimation of dynamic models with a large number of states in other research contexts.

We present and estimate the model in the context of one of the most interesting and important types of hyper-media: blogs. Blogs have become an important source of information for consumers, with about a third of all online adults reading blogs on a regular basis.[3] Moreover, the practice of extensively excerpting and linking to outside sources is a widely-accepted norm among bloggers. Thus, blogs provide an ideal environment in which to study the consumption of news and information online.

Contribution and Literature Review. Few studies in the marketing and economics literature have addressed issues specifically related to hyper-media. Mayzlin and Yoganarasimhan (2008) and Dellarocas et al. (2010) present analytic, supply-side models aimed at understanding blogs' strategic decisions in the face of homogeneous consumers. In contrast, the centerpiece of this study is an empirical demand side model; hence we consider forward-looking agents with heterogeneous preferences. In contrast to the small number of marketing studies looking at blogs stands a larger body of research in the machine learning and computer science literatures. Work in this area has been concentrated primarily in two areas: understanding how information propagates through networks (Leskovec et al., 2007; Lloyd et al., 2006; Leskovec et al., 2009; Yang and Leskovec, 2011), and developing techniques to automatically classify digital content (Fujimura et al., 2005; Nakajima et al., 2005; Brooks and Montanez, 2006), both of which pertain to the generation of hyper-linked content and not its demand.

Because this paper focuses on issues related to the demand for online content, it is related to a set of studies in marketing that have modeled consumer demand for web sites. Some of these model *aggregate* demand across *many* web sites (e.g., Park and Fader, 2004; Danaher, 2007), whereas others look at *individual* demand for a *single* web site (e.g., Bucklin and Sismeiro, 2003; Montgomery et al., 2004; Ansari and Mela, 2003; Chatterjee et al., 2003). This study however is closest to a third set of papers modeling *individual* demand across *many* web sites. Johnson et al.

[3] Source: Pew Internet & American Life Project. December 16, 2010.

(2004) model the total number of sites visited during a single web session, as do we. However, they include the probability of ending a session as a model primitive, whereas this event arises endogenously in this model. Like this study, Goldfarb (2002) and Lee et al. (2003) model web site choice while allowing prior actions to affect current decisions, however they do not include forward-looking consumers or explicitly consider search.

This study is closely related to the information search literature in microeconomics and marketing, beginning with Stigler (1961), in which consumers are uncertain about future consumption experiences, and therefore seek information to reduce their uncertainty. Because the model presented in this study uses a dynamic discrete choice framework, involves quality signals, and represents consumers as Bayesian updaters, it is most similar in form to the class of learning models in the vein of Erdem and Keane's (1996) seminal paper.[4] However, the hyper-media context provides a unique opportunity to advance this literature along a number of dimensions. First, information turns over quickly in this environment, with a typical news item receiving almost no attention after about 40 hours (Yang and Leskovec, 2011). Thus, the quality of information (meaning its relevance to an individual consumer) and the quantity of information are non-stationary and must be learned at each browsing session. Second, there is a high degree of redundancy in the information available at various sites (Fujimura et al., 2005; Lloyd et al., 2006; Leskovec et al., 2007, 2009; Yang and Leskovec, 2011). These redundancies, such as two sites reporting the same story, are relevant to consumers because they imply diminishing marginal returns to visiting more sites. Fifth, excerpting (quoting extensively from and linking back to online sources; Blood, 2003) is pervasive. An excerpt embedded in a blog signals what type of information one might find at the excerpted site, thereby reducing some of the uncertainty surrounding the next decision. Table 1 links some of the methodological advances in this study to prior research, and is organized around aspects of the model that arise from the three unique aspects of blogs in the context of learning.[5]

[4]Like Erdem and Keane (1996), we consider forward looking consumers. Several factors in the considered context can induce forward looking behavior. First, their is an option value to search. Even if the first site has a likelihood of not having relevant information, and therefore a negative expected value of information, there is still the potential to visit that site because the consumer can continue searching in the event they do not find something relevant. Second, sites with more information (e.g., news or commentary) enable users to better assess how much information is available across all sites on a given day. Hence, consumers have an incentive to visit information intensive sites first—not only for the utility of their information—but also because such sites allow consumers to update their beliefs about the utility from content remaining at other sites. Third, sites with many links also provide information useful in making future browsing more efficient. Consumers might visit such sites simply to learn which other sites to visit later.

[5]Because consumers in this model decide what type of information they wish to obtain, this paper bears some resemblance to studies in this literature that model the process by which consumers obtain brand information (e.g., Punj and Staelin, 1983; Simonson et al., 1988; Ratchford and Srinivasan, 1993; Moorthy et al., 1997). These studies typically assume however that myopic consumers gather information about a small number of brands and end their

Table 1: Comparison of selected search models with the present study.

		Non-stationarity		*Signals*		*Temporality*
Study	Product(s)	Quality	Quantity	Extrapolate to other goods	Competitor-initiated	Forward looking
Roberts and Urban (1988)	Automobiles					
Erdem and Keane (1996)	Detergent					✓
Ackerberg (2003)	Yogurt					✓
Byzalov and Shachar (2004)	TV Shows	✓				
Hortaçsu and Syverson (2004)	Mut. Funds					✓
Mehta et al. (2004)	Detergent					
Erdem et al. (2005)	Computers					✓
Shin et al. (2007)	Toothpaste					
Erdem et al. (2008)	Ketchup					
Lovett (2008)	(Monte Carlo)	✓				
Kim et al. (2010)	Camcorders					✓
Zhao et al. (2010)	Books					✓
Anand and Shachar (2010)	TV Shows	✓				
Chintagunta et al. (2012)	Drugs					✓
This Study	Information	✓	✓	✓	✓	✓

It is worth noting that this study is the first in which the assortment of goods available from one provider provides information about the assortment available from others. Previous studies are varied in the types of signals they have modeled, with sources including consumption of the focal brand (e.g., Erdem and Keane, 1996; Ackerberg, 2003; Mehta et al., 2004; Shin et al., 2007), advertising (e.g., Byzalov and Shachar, 2004; Anand and Shachar, 2010), prices (e.g., Hortaçsu and Syverson, 2004; Erdem et al., 2008), and word-of-mouth recommendations / product reviews (e.g., Roberts and Urban, 1988; Byzalov and Shachar, 2004; Erdem et al., 2005; Anand and Shachar, 2009; Zhao et al., 2010; Kim et al., 2010). With hyper-media, the set of information at one site signals the relevance and availability of information at other sites. Moreover, excerpts signal consumers' match with the type of information at the excerpted site. The signals produced by excerpts are similar in some ways to comparative advertising, and we suspect aspects of the proposed model might be relevant in that context as well.

In addition to the substantive and theoretical considerations noted above, our paper builds on the computational literature pertaining to the estimation of single agent dynamic models in the presence of a large number of states (e.g., Imai et al. (2009); Nevo and Aguirregabiria (2011)). Due to the so-called "curse of dimensionality," the number of states considered in dynamic models is often quite limited (Bellman, 1957; Keane et al., 2011). Owing to the considerations noted above, however, the dimension of the state space in our model is in fact quite sizable (specifically, 17 state variables). Therefore, in §4.4 we develop an approach to estimating single agent dynamic discrete

searches by purchasing one of them. By contrast, we model forward-looking consumers gathering information because it provides utility at a fundamental level.

choice models with many states.

The remainder of this paper is structured as follows: In §2, we present a theoretical model of hyper-media search. We also simulate data from this model and discuss how changes to the model primitives affect consumers' search patterns. In §3, we describe the browsing and link data used in estimation, and in §4, we present the strategy used to estimate the model primitives. In §5, we discuss the results of estimation, and in §6, we conduct counterfactual simulations relevant to policy makers.

2 Theoretical Model

The model presented here describes a forward-looking consumer who gains new knowledge by reading blogs and other hyper-media. Although the consumer receives utility from any new information he may find, the process of visiting a site and extracting its information is costly. The consumer's goal, therefore, is 1) to visit whichever sites he thinks will provide the greatest amounts of relevant information, while 2) considering the benefits of doing so relative to the expected cost. To achieve this goal, however, he must consider the following: First, whereas some information might be available from more than one site, seeing that information twice does not double its usefulness. Hence, as the consumer gains more knowledge, he finds diminishing marginal returns to continuing the session and eventually stops browsing. Second, because the set of available information varies across sessions, the consumer faces a high degree of uncertainty about both the amount and type of information he might find.[6] To overcome both of these obstacles, the consumer must collect information about information—i.e., at each site he visits, the consumer must take note of its content, excerpts, and links and use them to predict the amount and type of information available elsewhere.

We develop this model formally in the sections that follow: First, in §2.1, we characterize what information entails in the context of hyper-media and define the concept of a bit of information. In §2.2, we articulate the search and consumption process—i.e., the consumer's objective function, what is being searched, learned, and consumed, and the stopping rule. With these basics out of the way, we then present the consumer's dynamic choice problem. To facilitate exposition, we start in §2.3 with the simplest context of choosing among web sites that are symmetric in quality and have no hypertext links; we refer to this as the "basic model." We then relax these

[6] Following the economics literature, we refer to the process of sequentially visiting sites as a "search." As the applied setting is blogs, we also refer this process as a "browsing session." Moreover, in this empirical setting, consumers engage in one browsing session per day, hence we speak of a "daily" search. These conventions are meant to be illustrative, as the model is in fact quite general with respect to media and timing.

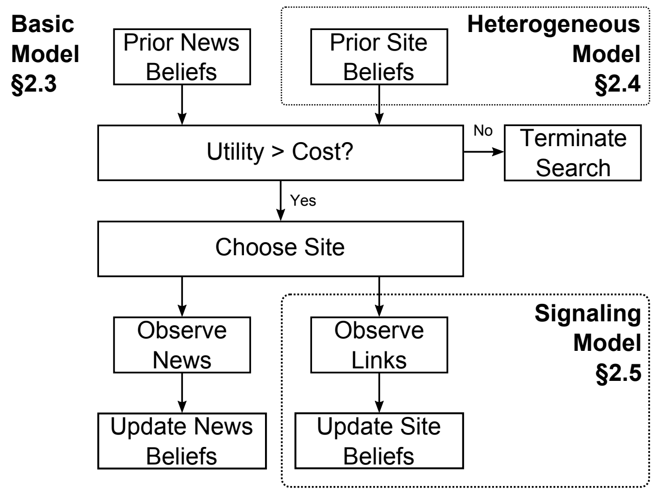

Figure 2: Structure of the choice problem. In the Basic Model (§2.3), the consumer chooses to browse to a site or end his search on the basis of his prior beliefs about news available at each site. Upon seeing the news at a particular site, the consumer updates his beliefs about the news at the remaining sites. In the Heterogeneous Model (§2.4), consumers have prior beliefs about the content at each site that also inform their choices. In the Signaling Model (§2.5), the consumer sees links to other sites and updates his beliefs about their content.

assumptions. In §2.4 (the "heterogeneous model") sites are asymmetric in both the amount and quality of information provided. In §2.5 (the "signaling model") these blogs excerpt content from each other, and consumers use these excerpts to infer the quality of information at the linked blogs. Table 2 relates each of these sub-models to the consumer's choice problem presented in §2.2. We close by generalizing the model to many consumers and searches (§2.6), and exploring the implications of the model primitives on search behavior (§2.7).

2.1 The Role of Information

Information plays a central and dual role in this model. First, it provides utility to the consumer. In this role, information behaves much like any other good. Second—and somewhat recursively—information helps the consumer gather information more efficiently. In this role, information behaves as it does in other learning models. We discuss each of these roles below.

2.1.1 Information as a Good (Consumption)

The consumer's goal is to gain new knowledge, which he accomplishes by visiting sites and reading their content. The information at a site provides utility because acquiring knowledge is intrinsically beneficial. For example, readers of sports blogs enjoy learning about their favorite athletes' performances, and readers of political blogs might be interested in how economic news affects government spending. Any *new* information the consumer finds provides him with utility and adds to his prior knowledge. Hence, subsequent encounters with the same information do not provide additional utility, and the likelihood of finding novel information generally decreases as the search progresses.

Information is represented in the following way. Within the context of a given search, there exists a finite amount of information that can be acquired. Following Allen (1986, 1983, 1990), we represent this information as a set of N unique and indivisible "bits." A bit represents the smallest amount of information that can be relevant (i.e., provide utility) to a consumer. Thus, a bit might reflect the most granular amount of information in a sentence or brief paragraph, and can represent either facts or opinions. Bits of information are distributed heterogeneously across blogs and search sessions. Thus, some bits, such as those representing basic facts about a topic (e.g., "Smith wins re-election!"), can be found at a large number of sites; whereas others, perhaps representing unusual or insightful commentary (e.g., "Smith's win is great news for local beekeepers!"), might be unique to a single blog.[7]

2.1.2 Information about Information (Search)

Information also plays the same role as in other learning models. Within each session, the consumer is uncertain about three aspects of the available information: 1) the average amount of information at each blog, 2) the average utility per bit of information, and 3) the consumer's match with the type of information at each blog (i.e., his predilection for the blog). Under such a high degree of uncertainty, the consumer initially relies on his prior beliefs regarding these three factors when choosing which blogs to visit. After visiting a blog, however, the consumer updates these beliefs. Specifically, he sees how many bits of information are at the site and updates his beliefs about the amount and relevance of any remaining information. He can also observe the average utility per bit of information and update his beliefs about how relevant the remaining information is. Furthermore, any excerpts he finds provide signals about the information at other sites, thereby reducing uncertainty regarding those sites. Thus, there are three learning processes in this model.

2.2 Structure of the Search and Browsing Session

The consumer engages in many browsing sessions over time, and over the course of each session, he learns about whatever information is available. However, by the time the next session begins (e.g., the next day), the information at sites (e.g., the news of the day) will have changed, and any previous learning about the quality or quantity of information will not be relevant for the current

[7] It is possible that some bits might not appear at any site. Thus, N can be thought of as an upper bound on the set of information that might possibly be reported each day.

session. Therefore, the consumer reverts to his prior beliefs at the start of each session. Hence, the model presented in this section focuses on the events occurring within one session (note that we estimate the model using data from many sessions for each consumer).

The search session for day $d = 1, \ldots, D$ comprises a sequence of steps, indexed $t = 1, \ldots, T_d$, corresponding with the site choices made over the course of the session (i.e., which site to visit next or whether to end the session).[8] At each step, the consumer chooses whichever action maximizes his total expected utility from that point forward, conditional on his current state of knowledge (e.g., his beliefs about the average utility per bit of information). The consumer's options are to visit a blog ($j \in \{1, \ldots, J\}$), or end the session ($j = 0$). The events at each step t are ordered as follows.

1. Observe relevant state variables affecting the next decision, including all information accumulated through step $t-1$ (step $t = 0$ indexes his prior beliefs).
2. Take the action $a_t \in \mathcal{A}_t$ maximizing total expected utility going forward.
3. End the session, or else visit a blog, observe its information, and (possibly) receive utility.
4. Update beliefs about the availability and relevance of information based on observations in step #3.

To simplify matters, we assume the consumer acquires all available information at each blog he visits, and thus there is little reason for him to return to a blog he has already visited. We therefore assume that he visits each blog no more than once per session, i.e., $\mathcal{A}_{t+1} = \mathcal{A}_t \backslash a_t$, with $\mathcal{A}_1 = \{0, \ldots J\}$.

2.3 The Basic Model

In the basic model, we assume blogs are homogeneous in the type and amount of information they typically provide, and that they do not link or post excerpts. We begin by presenting the fundamental parts of the dynamic discrete choice problem: the utility function, state variables, and value functions. We then present the consumer's prior and updated beliefs and develop the state transition density. A list of nomenclature appears in Appendix A.

2.3.1 Utility from Information

The net utility gained by visiting blog j at step t depends on three quantities: 1) the utility from any new bits of information found at blog j, 2) the cost of switching to blog j, and 3) an idiosyncratic shock. We discuss each of these in turn.

[8] To facilitate exposition, we suppress the d subscript when doing so does not lead to ambiguity.

First, the utility from learning new information is a function of: 1a) the consumer's prior state of knowledge (i.e., the subset of the N bits that have already been consumed during the session), $k_t \in \{0,1\}^N$; 1b) the set of bits at blog j, $\iota_j \in \{0,1\}^N$; and 1c) the utility from each bit of information, $u \in \Re^{+N}$. The utility from new knowledge is always greater than the utility from remaining uninformed (which is set to zero), hence the utility from each bit is positive, i.e., $u_b > 0$ for all bits b. The utility from information at step t is the sum of the bit utilities for all bits at blog j that the consumer hasn't already learned.

$$\beta_{jt} \equiv \beta\left(u, \iota_j | k_{t-1}\right) = \textstyle\sum_{b=1}^{N} u_b \iota_{jb} \left(1 - k_{t-1,b}\right) \tag{1}$$

The consumer makes choices without knowing which bits are at each blog j.[9] His choices therefore depend on his beliefs about the distribution of β_{jt}, which he updates after each visit to a new blog.

The utility from visiting blog j also depends on 2) a switching cost, γ, and 3) an idiosyncratic shock to utility, ϵ_{jt}, which is private information learned just prior to the decision at step t, but not observed by the researcher.

$$U(u, \iota, k_{t-1}, \gamma, \epsilon_t, a_t = j) = \beta(u, \iota_j | k_{t-1}) - \gamma + \epsilon_{jt} \tag{2}$$

Ending the session yields net utility of $U(u, \iota, k_{t-1}, \gamma, \epsilon_t, a_t = 0) = \epsilon_{0t}$.

2.3.2 State Variables

The state variables affecting choices at step t are 1) the total number of bits learned during the session, $K_{t-1} = \sum_{b=1}^{N} k_{t-1,b}$; 2) the average utility from those bits, $\overline{u}_{t-1} = \sum_{b=1}^{N} u_b k_{t-1,b} / K_{t-1}$; and 3) the set of blogs that were visited, $h_{t-1} \in \{0,1\}^J$ (the number of blogs visited is denoted $H_{t-1} = \sum_j h_{t-1,j}$). The set of state variables at step t is thus $\mathcal{S}_t = (\mathcal{X}_t, \epsilon_t)$, with $\mathcal{X}_t = (K_{t-1}, \overline{u}_{t-1}, h_{t-1})$. The utility from information at step t (Equation 1) can be rewritten as a function of the transition from state $\mathcal{X}_t$ to state $\mathcal{X}_{t+1}$. Hence, $\beta_t \equiv \beta\left(\mathcal{X}_{t+1}, \mathcal{X}_t\right) = \overline{u}_t K_t - \overline{u}_{t-1} K_{t-1}$.[10] We drop the j subscript in this expression because information utility does not depend on the blog providing the information. The consumer's beliefs about the process by which $\mathcal{S}_t$ transitions to $\mathcal{S}_{t+1}$ conditional on his choice a_t are reflected in the probability distribution $p\left(\mathcal{S}_{t+1} | \mathcal{S}_t, a_t\right)$ (discussed in §2.3.5).

[9] We assume the consumer knows the maximum number of bits available, N.

[10] Please see the Technical Appendix for proofs of this (Claim 1) and other results.

2.3.3 Value Function

Conditional on the current state $\mathcal{S}$, the expected present discounted value of current and future utilities (within the current session) is given by the value function $V(\mathcal{S})$.[11] We make a number of assumptions regarding the value function, $V(\mathcal{S})$, that are standard to the literature on single-agent dynamic discrete choice problems (Aguirregabiria and Mira, 2010). In the Technical Appendix, we enumerate these assumptions and explain how they lead to the state transition density, $p(\mathcal{S}'|\mathcal{S}, a = j)$, being factored as $f(\mathcal{X}'|\mathcal{X}, a = j)\, g(\epsilon)$, where $g(\cdot)$ denotes the $EV(0,1)$ distribution. Hence, the value function is defined as

$$V(\mathcal{X},\epsilon) \;=\; \max\left(\epsilon_0, \max_{\substack{j\in A(\mathcal{X})\\ j\neq 0}}\{\mathbb{E}[\beta(\mathcal{X}',\mathcal{X})\,|\mathcal{X}] - \gamma + \epsilon_j + \int V(\mathcal{X}',\epsilon')\, f(\mathcal{X}'|\mathcal{X},j)\, g(\epsilon')\, d\mathcal{X}' d\epsilon'\}\right) \tag{3}$$

(we drop the index t, as its value does not affect the consumer's choices). Equation (3) indicates the consumer chooses the greater of the value of 1) ending the session (the outer max function) or 2) the highest expected utility among sites yet to be visited (the inner max function). The latter option contains two terms: 2a) the period utility of information consumption at each site and 2b) the value of future site visits—i.e., the value of future knowledge in the face of lower uncertainty. The value of the action j maximizing Equation (3) is $V_j(\mathcal{X}) + \epsilon_j$. The choice-specific value function, $V_j(\mathcal{X})$, depends on the expected information utility and switching cost, as well as the expected value of all subsequent (optimal) decisions in the current session (i.e., the expected value function, denoted $W(\mathcal{X})$). The choice-specific value function for option $j > 0$ is

$$V_j(\mathcal{X}) = \mathbb{E}\left[\beta(\mathcal{X}',\mathcal{X})\,|\mathcal{X}\right] - \gamma + \int W(\mathcal{X}')\, f(\mathcal{X}'|\mathcal{X},j)\, d\mathcal{X}', \tag{4}$$

whereas the expected value function is

$$W(\mathcal{X}) = \int V(\mathcal{X},\epsilon)\, g(\epsilon)\, d\epsilon = \log \textstyle\sum_{j\in\mathcal{A}(\mathcal{X})} \exp\{V_j(\mathcal{X})\}\,. \tag{5}$$

To complete the specification of the value function, we need to compute 1) the expected utility from information, $\mathbb{E}[\beta(\mathcal{X}',\mathcal{X})]$, and 2) the state transition density, $f(\mathcal{X}'|\mathcal{X},j)$. These expressions

[11] We assume that the present discounted value of future browsing sessions is unaffected by decisions made in the current browsing session. Accordingly, we ignore the present value of future sessions in the equations that follow, as this quantity is the same for each option under consideration (visit a site or end the current browsing session). The discount rate for future utility therefore equals 1, given the short search duration in the empirical application.

are developed in the remainder of this section.

2.3.4 Expected Utility ($\mathbb{E}\left[\beta\left(\mathcal{X}',\mathcal{X}\right)|\mathcal{X}\right]$)

In the Technical Appendix (Claim 2), we show that the expected utility from the information gained in step t is simply the product of expected knowledge, K_t, and expected average utility per bit, $\overline{u}_t$, minus the total utility from information from the previous step, $K_{t-1}\overline{u}_{t-1}$.

$$\mathbb{E}\left[\beta_t|K_{t-1},\overline{u}_{t-1},h_{t-1}\right]=\mathbb{E}\left[K_t|K_{t-1},h_{t-1}\right]\mathbb{E}\left[\overline{u}_t|\mathbb{E}\left(K_t|K_{t-1},h_{t-1}\right),\overline{u}_{t-1},K_{t-1}\right]-K_{t-1}\overline{u}_{t-1} \quad (6)$$

As Equation (6) depends on the distributions of $\overline{u}_t|K_t$ and K_t, we explain these beliefs next.

Expected Number of Bits of Information ($\mathbb{E}\left[K_t|K_{t-1},h_{t-1}\right)$]. To derive the expected number of bits of information seen at step t (K_t) we need to first specify the process that governs information availability, and then derive how this process is related to beliefs about the information remaining as the session proceeds. We discuss each next.

Prior and Updated Beliefs. Recall that within a given session, the set of bits at blog j is denoted ι_j. As blogs are undifferentiated in the basic model, we assume the probability of bit b appearing at any site that day is π_b (π_b is therefore proportional to the number of blogs expected to have bit b that day). The true distribution for the availability of information is therefore $\iota_{jb}|\pi_b\sim Bern\left(\pi_b\right)$. The consumer's prior beliefs about the π_b's and ι_{jb}'s are[12]

$$\iota_{jb}|\tilde{\pi}_b\overset{\text{iid}}{\sim}Bern\left(\tilde{\pi}_b\right),\qquad\tilde{\pi}_b|\tilde{\alpha}_0\overset{\text{iid}}{\sim}Beta\left(\tilde{\alpha}_0,1\right),\quad\tilde{\alpha}_0>0, \quad (7)$$

where $\tilde{\alpha}_0$ governs the ex ante (i.e., prior to search) expected likelihood a bit will appear on a blog. We explain in §1 of the Technical Appendix why the i.i.d assumptions are not restrictive.

After visiting $t-1$ blogs, the consumer updates his beliefs about the $\tilde{\pi}_b$'s for each of the $N-K_{t-1}$ bits he hasn't seen yet. In the Technical Appendix (Claim 3), we show these updated beliefs are

$$\tilde{\pi}_{bt}|k_{t-1,b}=0,H_{t-1}\sim Beta\left(\tilde{\alpha}_0,H_{t-1}+1\right), \quad (8)$$

where H_t reflects the number of sites observed as of step t than have not contained bit b. The intuition for this result can be seen by considering the expectation of this distribution, $\mathbb{E}\left[\tilde{\pi}_{bt}|H_{t-1},k_{t-1,b}=0\right]=$

[12]Throughout the remainder of this paper, we will always mark variables representing the consumer's beliefs with a tilde (~) in order to differentiate them from the "true" variables of interest. Thus, the consumer's beliefs about π_b are denoted $\tilde{\pi}_b$.

$\tilde{\alpha}_0 / (\tilde{\alpha}_0 + H_{t-1} + 1)$. As the number of sites at which bit b has not appeared (H_{t-1}) increases, the likelihood of finding that bit at a heretofore unvisited site decreases.

Expected Number of New Bits. The expected change in the consumer's knowledge, $K_t - K_{t-1}$ can be thought of as the expected number of successes in $N - K_{t-1}$ Bernoulli draws taken with probabilities $\tilde{\pi}_{b,t}$. Because 1) the consumer integrates over the unobserved bit probabilities, $\tilde{\pi}_{b,t}$, and 2) each Bernoulli draw has the same marginal probability, the change in the consumer's knowledge, $K_t - K_{t-1}$, is a binomial random variable with sample size $N - K_{t-1}$, and probability $\tilde{\alpha}_0 / (\tilde{\alpha}_0 + H_{t-1} + 1)$ (see Claim 4 in the Technical Appendix). Hence the consumer's expected change in total knowledge after step t is equal to the number of bits remaining times the marginal probability of finding any one of them at an unvisited site. Accordingly, expected total knowledge is

$$\mathbb{E}\left[K_t | K_{t-1}, H_{t-1}\right] = \underbrace{K_{t-1}}_{\text{Prior knowledge}} + \underbrace{(N - K_{t-1})}_{\text{Number of bits left}} \times \underbrace{\left(\frac{\tilde{\alpha}_0}{\tilde{\alpha}_0 + H_{t-1} + 1}\right)}_{\text{Prob. of finding a new bit}} \tag{9}$$

With every site visit, the consumer expects to gain less new knowledge because 1) by acquiring new information, he depletes the stock of remaining bits (K_{t-1} increases), and 2) the more times he fails to find bit b, the less he expects it to appear (H_{t-1} increases).

Expected Utility per Bit of Information ($\mathbb{E}\left[\overline{u}_t | \mathbb{E}\left(K_t | K_{t-1}, h_{t-1}\right), \overline{u}_{t-1}, K_{t-1}\right]$). Before deriving the expected utility per bit of information, we first need to discuss the distribution of bit quality in the context of the consumer's prior and updated beliefs.

Prior and Updated Beliefs. We assume that consumers know their utility for various bits of information (u_b). However, because consumers do not know which bits will be available on a given news day, they have uncertainty about the average utility from each *available* bit. We assume, therefore, that the true distribution of the u_b's available within the current session d is exponential with scale σ_d; thus $u_b | \sigma_d \sim Expo\left(\sigma_d^{-1}\right)$, where σ_d is the unobserved (by the consumer) expected utility per bit on day d.[13] The consumer's prior beliefs about the distribution of bit utility available in a given session are represented by the variables $\tilde{\sigma}$, $\tilde{\kappa}_0$, and $\tilde{\lambda}_0$, and the following distributions:

$$u_b | \tilde{\sigma} \overset{\text{iid}}{\sim} Expo\left(\tilde{\sigma}^{-1}\right), \quad \tilde{\sigma} | \tilde{\kappa}_0, \tilde{\lambda}_0 \sim Inv\text{-}Ga\left(\tilde{\kappa}_0 + 1, \tilde{\kappa}_0 \tilde{\lambda}_0\right), \quad \tilde{\kappa}_0 > 0,\ \tilde{\lambda}_0 > 0, \tag{10}$$

[13] We use the exponential distribution for two reasons. First, because irrelevant information is more prevalent than relevant information, the p.d.f. of u_b should be a decreasing function of u_b (recall we impose the restriction $u_b > 0$). Second, the exponential distribution permits a conjugate distribution for σ_d.

where $\tilde{\lambda}_0$ represents the consumer's ex ante expected utility per bit (the expected value of $\tilde{\sigma}$ is $\tilde{\lambda}_0$), and $\tilde{\kappa}_0$ reflects the strength of this belief.[14] Recall that the utility from new information, β_t, is the sum of the u_b's for $K_t - K_{t-1}$ new bits. Due to the i.i.d. assumption, this sum follows a gamma distribution: $\beta_t|\tilde{\sigma}_t, K_t, K_{t-1} \sim Ga\left(K_t - K_{t-1}, \tilde{\sigma}_t\right)$. Hence, at step $t-1$ (i.e., after observing $K_{t-1} - K_{t-2}$ bits providing utility equal to β_{t-1}), the consumer updates his beliefs about $\tilde{\sigma}_t$ to be (see Claim 5 in the Technical Appendix)

$$\tilde{\sigma}_t|K_{t-1}, \overline{u}_{t-1} \sim Inv\text{-}Ga\left(\tilde{\kappa}_0 + K_{t-1} + 1, \kappa_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}\right). \tag{11}$$

Given the expectation of the gamma distribution, this equation suggests the perceived mean and variation of bit quality approaches its true value, and that the rate of this convergence depends upon the strength of prior beliefs, the availability of information, and the number of sites visited.

Expected Utility per Bit. The consumer's predictions for $\overline{u}_t$ are derived by integrating the posterior distribution of β_t over the consumers' uncertainty about the current session's average utility per bit, $\tilde{\sigma}_t$; making the appropriate change of variables from β_t to $\overline{u}_t K_t - \overline{u}_{t-1}K_{t-1}$; and noting that when no new bits are found, $\overline{u}_t = \overline{u}_{t-1}$. In the Technical Appendix, we derive this p.d.f. (Claim 6), as well as the expected average utility per bit at step t (Claim 7). The latter is shown to be

$$\mathbb{E}\left[\overline{u}_t|\overline{u}_{t-1}, K_{t-1}, K_t\right] = \frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\left(1 - \frac{K_{t-1}}{K_t}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{K_t}\right). \tag{12}$$

Thus the expected value of $\overline{u}_t$ is a weighted average of K_{t-1} bits with average utility $\overline{u}_{t-1}$ and $K_t - K_{t-1}$ bits with *expected* utility of $\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}$. Therefore, as more bits are accumulated, the expectation moves away from the consumer's prior belief, $\tilde{\lambda}_0$, and closer to the observed average utility per bit $\overline{u}_t$.

Expected Information Utility ($\mathbb{E}\left[\beta_t|K_{t-1}, \overline{u}_{t-1}, h_{t-1}\right]$). The total expected utility from information, $\mathbb{E}\left[\beta_t|—\right]$, is derived in the Technical Appendix (Claim 8) and, as noted in (6), is simply the expected number of bits ($\mathbb{E}\left[K_t|K_{t-1}, h_{t-1}\right]$) times the expected utility per bit ($\mathbb{E}\left[\overline{u}_t|K_t, \overline{u}_{t-1}, K_{t-1}\right]$).

$$\mathbb{E}\left[\beta_t|K_{t-1}, \overline{u}_{t-1}, H_{t-1}\right] = \underbrace{\left((N - K_{t-1})\,\frac{\tilde{\alpha}_0}{\tilde{\alpha}_0 + H_{t-1} + 1}\right)}_{\text{Expected number of new bits}} \times \underbrace{\left(\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)}_{\text{Expected utility per new bit}} \tag{13}$$

[14] In the Technical Appendix we again show that the i.i.d. assumptions are not restrictive.

Thus the consumer's predicted utility from the information at the next site is the product of two point estimates: 1) the predicted amount of information, which depends on the number of bits left and the number blogs visited; and 2) the predicted quality of that information, which depends on the average quality for all the information he's already encountered. The simplicity of the expression in (13) is reassuring given the strong assumptions about Bayesian updating that provide its foundations.

2.3.5 State Transition Density

Per §2.3.2, the state transition density $f(\mathcal{X}'|\mathcal{X},j)$ is the product of the state transition densities for K_t and $\overline{u}_t|K_t$. Hence, $f(\mathcal{X}'|\mathcal{X},j) = p(K_t|K_{t-1},h_{t-1})\, p(\overline{u}_t|K_t,K_{t-1},\overline{u}_{t-1})$.

2.3.6 Choice-Specific Value Function

Substituting the expressions for $\mathbb{E}[K'|\mathcal{X},j]$, $\mathbb{E}[\overline{u}'|K',\mathcal{X},j]$, and $f(\mathcal{X}'|\mathcal{X},j)$ developed in the previous sections into the choice-specific value function (4) yields the following expected value for option $j > 0$.[15]

$$V_j(\overline{u}_{t-1},K_{t-1},h_{t-1}) = \underbrace{(N-K_{t-1})}_{\#\text{ bits left}} \underbrace{\left(\tfrac{\tilde{\alpha}_0}{\tilde{\alpha}_0+H_{t-1}+1}\right)}_{\text{Pr. of a new bit}} \underbrace{\left(\tfrac{\tilde{\kappa}_0\tilde{\lambda}_0+K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0+K_{t-1}}\right)}_{\text{Exp. util per bit}} \underbrace{-\ \gamma}_{\text{Switch. cost}}$$
$$+\int \underbrace{W(\overline{u}_t,K_t,h_t)}_{\text{Emax function}} \times \underbrace{p(K_t|K_{t-1},h_{t-1})\, p(\overline{u}_t|K_t,K_{t-1},\overline{u}_{t-1})\, d\overline{u}_t\, dK_t}_{\text{State trans. for bits}} \quad (14)$$

It is evident from (14) that the expected value of choice $j > 0$ depends on the current stage of the search, but not on the attributes of blog j (i.e., blogs are in expectation undifferentiated).

2.4 Extension to Asymmetric Blog Content ("Heterogeneous Model")

In this section, we relax the assumption that blogs are undifferentiated in the information they provide. Instead, we assume that blogs differ in their respective information sets, leading the consumer to prefer some sites other others. Differentiation occurs along two dimensions. The first dimension is the amount of information at each blog, owing to the reality that some blogs publish more content than others. The second dimension is the editorial position of each blog, i.e., whether the tonal quality of information is congruent with a reader's tastes. Below, we consider the ramifications of quality and quantity differentiation on blog consumption.

[15] We simplify notation in Equation (14) and throughout the remainder of this paper by writing $\int p(K_t)\, dK_t$ instead of $\sum_k \Pr(K_t = k)$, even though K_t is a discrete random variable.

2.4.1 Quality of Information and Editorial Match with Consumer Tastes

The consumer prefers some types of information over others. For example, a consumer who prefers sports-related news might prefer (and thus have a higher match with) a site focused on sports, whereas a consumer who prefers celebrity gossip might not. The match between the consumer's tastes and a blog's editorial focus however varies from day to day, depending in part on whatever happens to be newsworthy, and in part on the blog's coverage of the day's events. In the basic model, sites are chosen with equal probability at the start of the session. Now that sites are differentiated, the consumer's knowledge of his average match with each blog informs this initial choice.

Heterogeneous Information Quality. Recall that by Equation (2) in the basic model, the period t utility from visiting blog j is $\beta_{jt} - \gamma + \epsilon_{jt}$. Here we replace this expression with one taking into account differences in the match between the consumer's tastes and a blog's editorial focus. As noted above, the degree of match varies from day to day. For example, the *Boston Globe*'s coverage of the Red Sox beating the Yankees might be better suited to the consumer's taste than the same coverage from the *New York Times*. Hence, match utility depends on: 1) the long-run average match utility from blog j, δ_j, reflecting a long-term preference for some sites over others; and 2) a session-specific, idiosyncratic deviation from this average, ν_j^d, arising due to daily variation in available information and execution.[16] Hence, the consumer's actual match on day d is $\delta_j + \nu_j^d$. Thus, the utility from blog j is now

$$U\Big(\mathcal{X}_{t+1}, \mathcal{X}_t, \gamma, \delta_j, \nu_j^d, \epsilon_{jt}, a_t = j\Big) = \beta_{jt} + \Big(\delta_j + \nu_j^d\Big) - \gamma + \epsilon_{jt}. \tag{15}$$

Beliefs about Quality Heterogeneity. We assume ν_j^d is a mean-zero deviation from average match, hence $\nu_j^d \sim N\left(0, \tau_\delta^{-1}\right)$. Based on a potentially long history of browsing, the consumer knows his average match utility, δ_j, and the typical amount of match variation across searches, τ_δ^{-1} based on his prior experiences browsing blogs; δ_j and τ_δ^{-1} are therefore constant across searches. The consumer's beliefs about the unknown part of his match within a given session are represented by $\tilde{\nu}_j \sim N\left(0, \tau_\delta^{-1}\right)$.

Because the consumer does not observe any signals for $\tilde{\nu}_j$, he cannot meaningfully update his beliefs about the daily component of match until *after* he visits blog j. The *expected* utility from information is therefore unaffected by $\tilde{\nu}_j$, whose expected value is 0. The value of future

[16] Here we use the notation ν_j^d to emphasize this variable changes with each session whereas δ_j does not.

browsing decisions, on the other hand, depends on the level of variation in $\tilde{\nu}_j$, as can be seen in the updated choice-specific value function (16). When uncertainty about $\tilde{\nu}_j$ is great, the value of future browsing is higher due to the possibility of finding a site providing unusually high match utility. Hence, uncertainty about $\tilde{\nu}_j$ increases the chances of continued search; this is commonly called an option value (Ho et al., 1998).

$$V_j\left(\overline{u}_{t-1}, K_{t-1}, h_{t-1}\right) = \underbrace{\left(N - K_{t-1}\right)}_{\text{\# bits left}} \underbrace{\left(\frac{\tilde{\alpha}_0}{\tilde{\alpha}_0 + H_{t-1} + 1}\right)}_{\text{Pr. of a new bit}} \underbrace{\left(\frac{\tilde{\kappa}_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)}_{\text{Exp. util per bit}} \underbrace{+\delta_j}_{\text{Avg. match util.}} \underbrace{-\gamma}_{\text{Switch. cost}}$$
$$+ \int \underbrace{W\left(\overline{u}_t, K_t, h_t, \{\tilde{\nu}_k\}\right)}_{\text{Emax function}} \underbrace{p\left(K_t | K_{t-1}, h_{t-1}\right) p\left(\overline{u}_t | K_t, K_{t-1}, \overline{u}_{t-1}\right) d\overline{u}_t dK_t}_{\text{State transition for bits}}$$
$$\prod_{k \in \mathcal{A}(\mathcal{X})} \underbrace{p\left(\tilde{\nu}_k | \tau_\delta\right) d\tilde{\nu}_k}_{\text{Idiosyncratic match}} \tag{16}$$

The updated expected value, or emax, function (c.f. Equation 5) is

$$W\left(\overline{u}_t, K_t, h_t, \{\tilde{\nu}_j\}\right) = \log \textstyle\sum_{k \in \mathcal{A}(\mathcal{X})} \exp\left\{V_k\left(\overline{u}_{t-1}, K_{t-1}, h_{t-1}\right) + \tilde{\nu}_k\right\}. \tag{17}$$

One computational cost of incorporating heterogeneous match into the model is that the dimension of the integral in (17) expands by the J quality match terms.

2.4.2 Quantity of Information

Some blogs provide information that is readily available elsewhere, whereas others provide more original content. The consumer might receive greater utility from visiting the former, due to greater amounts of new information, but he might also gain high utility from the latter if his match utility is particularly high.

Heterogeneous Information Quantities. The amount of information typically found at each blog j is represented by the parameter, $\alpha_j \in (0, 1)$. Lower values of α_j indicate less information on average, whereas higher values indicate more. We consider contexts where the consumer has learned the average amount of information at each blog based on his prior browsing experience (i.e., the α_j for each blog j is known).

In the basic model, the probability of observing bit b is the same at each blog. Here we assume this probability depends on two factors: the first, π_b, pertains to an individual bit b and is common across blogs (due to blogs facing a common information environment), the other, α_j, is particular to blog j and common across bits (due to differences in execution across blogs). The interplay between π_b and α_j is given in the following expression for the probability bit b appears at blog j, which we now denote $\rho_{jb} \equiv \Pr[\iota_{jb} = 1] = 1 - (1 - \pi_b)^{\alpha_j}$. This relationship is such that when the blog publishes more information on average ($\alpha_j \to 1$), then $\rho_{jb} \to \pi_b$. However, when the blog publishes less ($\alpha_j \to 0$), we find $\pi_b > \rho_{jb} \to 0$. The parameter α_j thus attenuates the probability of finding information at blog j. An implication of heterogeneity in quantity is that consumers are more likely to visit sites that embed a lot of information because 1) the chance of finding relevant information increases, and 2) such sites are more informative about the distribution of information available on a given day.

Beliefs about Quantity. As in the basic model, the consumer's prior beliefs about the availability of news are that $\tilde{\pi}_b$ follows a beta distribution: $\tilde{\pi}_b \sim Beta(\tilde{\alpha}_0, 1)$. However, because the probability of seeing bit b at blog j depends on the blog's content strategy (via α_j), the updating equations are different. We show in the Technical Appendix (Claim 9) that the posterior distribution of $\tilde{\pi}_b$, for any unseen bit b, is Beta with parameters $\tilde{\alpha}_0$ and $1 + \alpha_{t-1}$, where α_{t-1} is the sum of the α_j's for any blogs that have been visited as of step $t-1$: $\alpha_{t-1} = \sum_{j=1}^{J} \alpha_j h_{t-1,j}$. The term α_{t-1} plays the same role as H_{t-1} in in the basic model: thus rather than increasing by 1 with each blog visited, the second parameter in the beta distribution now increases by α_j. As noted above, the implication is that blogs with higher quantities of information provide better information about the remaining $\tilde{\pi}_b$'s.

Because the expected number of bits at blog j now depends on that site's coverage of information, α_j, the probability of transitioning from K_{t-1} to K_t depends on the consumer's choice, $a_t = j$. We show in the Technical Appendix (Claim 10) that the conditional distribution of K_t is still binomial with sample size $N - K_{t-1}$, however the probability is now $1 - B(\tilde{\alpha}_0, 1 + \alpha_{t-1} + \alpha_j) / B(\tilde{\alpha}_0, 1 + \alpha_{t-1})$, where $B(\cdot, \cdot)$ denotes the Beta function.

Although the consumer's beliefs at step t now depend on α_{t-1}, the state space does not change because α_{t-1} is a function of the state variable h_{t-1} (the history vector) and the α_j's (which are known to the consumer): $\alpha_{t-1} = \sum_j h_{t-1,j} \alpha_j$. The updated choice-specific value function (for

$j > 0$) is now

$$V_j\left(\overline{u}_{t-1}, K_{t-1}, h_{t-1}\right) =$$
$$\underbrace{(N-K_{t-1})}_{\text{\# bits left}}\underbrace{\left(1-\frac{B(\tilde{\alpha}_0, 1+\alpha_{t-1}+\alpha_j)}{B(\tilde{\alpha}_0, 1+\alpha_{t-1})}\right)}_{\text{Pr. of a new bit}}\underbrace{\left(\frac{\tilde{\kappa}_0\tilde{\lambda}_0+K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0+K_{t-1}}\right)}_{\text{Exp. util per bit}}\underbrace{+\delta_j}_{\text{Avg. match}}\underbrace{-\gamma}_{\text{Switch. cost}}$$
$$+\int\underbrace{W(\overline{u}_t, K_t, h_t, \{\tilde{\nu}_k\})}_{\text{Emax function}}\underbrace{p(K_t|K_{t-1}, h_{t-1})\, p\left(\overline{u}_t|K_t, K_{t-1}, \overline{u}_{t-1}\right) d\overline{u}_t dK_t}_{\text{State transition for bits}}$$
$$\prod_{k\in\mathcal{A}(\mathcal{X})}\underbrace{p(\tilde{\nu}_k|\tau_\delta)\, d\tilde{\nu}_k}_{\text{Idiosyncratic match}} \tag{18}$$

Equation (18) reveals three differences between the heterogeneous and basic models. First, the probability of finding a new bit now depends on whichever blog is chosen. Second, the utility from the visit now depends on the expected match with the chosen blog. And third, the value of future decisions depends on the variability of match utility across searches—i.e., the emax function depends on the variability of match utility in a way that increases the value of continued search when uncertainty is higher.

2.5 Extension to Linking and Excerpts ("Signaling Model")

We now consider the third and final step in the development of this model by relaxing the assumption that blogs do not link to each other. In the heterogeneous model just considered, the consumer cannot reduce uncertainty about his daily match with any blogs yet to be visited (the $\tilde{\nu}_j$'s) because he receives no information about these prior to his visit. In the signaling model, however, links and excerpts provide signals about the information at the linked sites. In this section, we first present this signaling mechanism, then consider the attendant implications for quality beliefs and the value function, and conclude with a discussion of the effect of links on switching costs.

2.5.1 Heterogeneous Signals

Blogs link to each other asymmetrically. Let ω_{kj} denote the probability that blog k excerpts blog j, noting that ω_{kj} and ω_{ki} are allowed to differ. Over the course of a session, the consumer might encounter many links to (and excerpts from) the same blog, each of which provides a signal about his match quality with the linked blog. For example, a link indicating a favorite author has posted on another site signals the consumer's match with the linked site that day. The number of signals observed for blog j accrues as the consumer observes more links, and at step t is denoted n_{jt}.

Signals are assumed to be noisy but unbiased reflections of the excerpted blog's true match value.

$$s_{jt}|\delta_j, \nu_j \sim N\left(\delta_j + \nu_j, \tau_s^{-1}\right) \tag{19}$$

The notation s_{jt} indicates that signal s was observed at blog $a_t = k$. The level of noise in signals, denoted τ_s^{-1}, is constant across blogs and is known by the consumer.

2.5.2 Beliefs about Quality

Recall that the consumer's prior beliefs about his match with blog j are $\delta_j + \tilde{\nu}_j$, where δ_j is known and represents long-term beliefs about average site match, and $\tilde{\nu}_j \sim N\left(0, \tau_\delta^{-1}\right)$ represents an unknown deviation from that average for the current session. As noted earlier, such deviations are influenced by the nature of information coverage on a given day (which might vary, as in the above example, with the set of authors on the site that day). We reiterate this point in order to emphasize that quality is not stationary across searches, and so the consumer is not just concerned with the average quality of a given blog—rather, he also needs to learn about that day's deviation from the average (Lovett, 2008).

The consumer knows that excerpts provide unbiased signals about his idiosyncratic match. After visiting $t-1$ blogs, the consumer will have seen $t > n_{j,t-1} \geq 0$ signals for blog j with an average quality of $\bar{s}_{j,t-1} = \sum_{r=1}^{t-1} s_{jr}/n_{j,t-1}$.[17] His updated beliefs about $\tilde{\nu}_j$ follow from the standard normal updating equations (West and Harrison, 1997). Hence the posterior mean of $\tilde{\nu}_j$ is $m_{jt} = \frac{\tau_s n_{j,t-1}(\bar{s}_{j,t-1} - \delta_j)}{n_{j,t-1}\tau_s + \tau_\delta}$ and the posterior precision (i.e., inverse variance) is $M_{jt} = n_{j,t-1}\tau_s + \tau_\delta$. As more signals for blog j are encountered, the consumer's beliefs about match grow increasingly concentrated around the average value of the observed signals (i.e., $\delta_j + m_{jt}$ converges to $\bar{s}_{j,t-1}$ and M_{jt}^{-1} shrinks to zero).

The consumer knows that seeing excerpts will reduce his uncertainty, but doesn't know a priori which blogs will contain links, nor what those links will signal. He consequently forms beliefs about both, and accordingly, we now account for the transitions from $\bar{s}_k$ to $\bar{s}'_k$ and from n_k to n'_k in the density $f(\mathcal{X}'|\mathcal{X}, j)$. We also redefine $\mathcal{X}$ so that it also includes the number and average value of signals for each blog j (the $\bar{s}_j$'s and n_j's). Hence, $\mathcal{X} = (K, \bar{u}, h, \{\bar{s}_j\}, \{n_j\})$.

Conditional on any observed excerpts, the predicted level of the next signal for blog j, s_{jt} is $\delta_j + m_{jt}$, with variance $\tau_s^{-1} + M_{jt}^{-1}$. The signal s_{jt}, however, is observed if and only if the next blog

[17] As most sites include at most one embedded link, we assume the consumer observes at most one signal for site j at any other site.

visited, $a_t = k$, links to blog j. As we stated earlier, this event occurs with probability $\omega_{k,j}$. Hence, with probability $\omega_{k,j}$, $n_{j,t-1}$ transitions to $n_{j,t-1}+1$ and a new $\overline{s}_j$ is observed with updated mean, $\left(\delta_j + m_{jt} + n_{j,t-1}\overline{s}_{j,t-1}\right)/\left(n_{jt-1}+1\right)$, and variance, $\left(\tau_s^{-1} + M_{jt}^{-1}\right)/\left(n_{jt-1}+1\right)^2$. With probability $1-\omega_{kj}$, no signal is observed, so $n_{j,t} = n_{j,t-1}$ and $\overline{s}_{j,t} = \overline{s}_{j,t-1}$.

2.5.3 Choice-Specific Value Function in the Signaling Model

The choice-specific value and emax functions now include the consumer's beliefs about his level of match (the $\tilde{\nu}_j$'s), as well as his beliefs about the number and value of signals (the n_j's and $\overline{s}_j$'s). The choice-specific value function for option $j > 0$ therefore becomes

$$
\begin{aligned}
&V_j\left(\overline{u}_{t-1}, K_{t-1}, h_{t-1}, \{\overline{s}_k\}, \{n_k\}\right) = \\
&\quad \underbrace{(N - K_{t-1})}_{\#\text{ bits left}} \underbrace{\left(1 - \tfrac{B(\tilde{\alpha}_0, 1+\alpha_{t-1}+\alpha_j)}{B(\tilde{\alpha}_0, 1+\alpha_{t-1})}\right)}_{\text{Pr. of a new bit}} \underbrace{\left(\tfrac{\tilde{\kappa}_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)}_{\text{Exp. util per bit}} \underbrace{+\delta_j + \tfrac{\tau_s n_{j,t-1}(\overline{s}_{j,t-1}-\delta_j)}{n_{j,t-1}\tau_s + \tau_\delta}}_{\text{Exp. match util.}} \underbrace{-\gamma}_{\text{Switch. cost}} \\
&\quad + \int \underbrace{W\left(\overline{u}_t, K_t, h_t, \{\tilde{\nu}_{jt}\}, \{\overline{s}_{jt}\}, \{n_{jt}\}\right)}_{\text{Emax function}} \underbrace{p\left(K_t | K_{t-1}, h_{t-1}\right) p\left(\overline{u}_t | K_t, K_{t-1}, \overline{u}_{t-1}\right) d\overline{u}_t\, dK_t}_{\text{State trans. for bits}} \\
&\qquad \prod_{k \in \mathcal{A}(\mathcal{X}_t)} \underbrace{p\left(\overline{s}_{kt} | \overline{s}_{k,t-1}, n_{k,t-1}\right) d\overline{s}_{kt}}_{\text{State trans. for signals}} \underbrace{p\left(\tilde{\nu}_{k,t} | \overline{s}_{k,t-1}, n_{k,t-1}\right) d\tilde{\nu}_{k,t}}_{\text{Idiosync. match}},
\end{aligned} \tag{20}
$$

and the emax function is now

$$
W\left(\overline{u}_t, K_t, h_t, \{\tilde{\nu}_{jt}\}, \{\overline{s}_{jt}\}, \{n_{jt}\}\right) = \log \sum_{k \in \mathcal{A}(\mathcal{X}_t)} \exp\left\{V_k\left(\overline{u}_t, K_t, h_t, \{\overline{s}_{jt}\}, \{n_{jt}\}\right) + \tilde{\nu}_{kt}\right\}. \tag{21}
$$

The differences between these expressions and their counterparts in the model with heterogeneity are 1) the expected match utility at blog j now depends on the values of any signals received for blog j, and 2) the variance of the consumer's beliefs about this match depends on the number of signals observed. Regarding the first point, the presence of excerpts leads to heterogeneity in expected match quality across blogs. Regarding the second point (variance), recall that in the model with heterogeneity but no signals, greater variance in match utility provides greater incentives to continue searching (due to the higher option value of future search). When links signal match, this uncertainty is lower. Thus a positive signal about match both increases the likelihood of search, by making the signaled blog more attractive, and decreases its likelihood, by lowering its future value. Which effect dominates is an empirical question.

2.5.4 Extension to Switching Costs

Although it is conceivable that links decrease switching costs, the current specification of the model does not accommodate this possibility. As clicking on a link is typically easier than typing a URL, it is possible that links, by lowering switching costs, have an additional, positive effect on the likelihood of visiting linked sites. Specifically, the cost of switching from blog j to blog k would be $\gamma - \zeta \ell_{j,k}$, where γ is the baseline switching cost, $\zeta > 0$ is the reduction in switching costs due to linking, and $\ell_{j,k} = 1$ when j links to k (and is 0 otherwise). The expected switching cost is therefore $\mathbb{E}\left(\gamma - \zeta \ell_{j,k}\right) = \gamma - \zeta \omega_{j,k}$.

We shall consider these potential switching cost-reducing effects as a robustness check; however, it is worth noting these cost effects, to the extent they are benign, 1) may be difficult to identify, and 2) may have little impact on estimates of the signaling effect. The latter point is amplified by the observation that switching costs operate independently of the depth of the search session (i.e. they are constant across steps), whereas link informativeness effects are moderated by the amount of search. Hence, if the effect of links on switching costs is small, the likelihood that their omission substantively alters inferences about link informativeness is also small.

To ascertain the potential size of these effects, §3.3 presents model-free evidence from the data to suggest any switching cost-reducing effects of links may be negligible in comparison to their informative effects. In over half of all instances, the presence of a link does not increase the average likelihood of a person visiting a particular blog. Were the reduction in switching costs of note, then one would generally expect links to increase visits to linked sites. Thus, estimating a model in which links have both informative and cost-reducing effects is left as a future robustness check.

2.5.5 Search Model Summary

This completes the specification of the model. There are four dynamic forces that affect the consumer's decision to continue search. Two of these relate to the quantity of information: the number of bits remaining (K_{t-1}), and the likelihood that the consumer will encounter any of those bits at the next blog (via the α_j's). The other two forces relate to the quality of information: the expected utility per bit ($\overline{u}_{t-1}$), and the expected match utility from the information at each blog (via the $\overline{s}_j$'s and n_j's). In the remaining sections, we extend the model to many consumers and searches, and show how the structural parameters, via their effects on these forces, control the length of the consumer's session and the blogs he chooses to visit.

2.6 Heterogeneity

So far, the model describes the actions taken by one consumer over the course of a single session. We now extend the model to multiple consumers and searches. First, we index consumers by $i = 1, \ldots, I$. Second, we assume that the consumer's prior beliefs about the average amount of information available ($\tilde{\alpha}_0$) and the average quality of each bit ($\tilde{\lambda}_0$ and $\tilde{\kappa}_0$) are correct (i.e., we assume unbiased expectations). Thus, the average information available, α_0, becomes common knowledge among all consumers. However, because consumers are heterogeneous with respect to the relevance of a piece of information, each consumer i's bit utilities are drawn from a distribution conditioned on their own $\lambda_{0,i}$'s and $\kappa_{0,i}$'s. Third, consumers are heterogeneous in the costs incurred by switching to another blog, which are now denoted γ_i. Likewise, consumers' average and daily match utilities are also heterogeneous. Accordingly, the average match between consumer i and blog j is now denoted δ_{ij}, and the daily deviation is ν_{ij}^d. Both the level of variation in match utility each day, τ_δ^{-1}, and the variation in signals, τ_s^{-1}, vary by individual, as described below in §2.6.1. Note that all of the preceding parameters except the ν_{ij}^d's are constant across searches.

Consumers engage in many searches, which are indexed by $d = 1, \ldots, D_i$ (in the empirical application, there is one search occasion each day). Furthermore the t^{th} action taken by consumer i on day d is denoted $a_{idt} \in \mathcal{A}(\mathcal{X}_{idt})$, where $\mathcal{X}_{idt}$ is the state vector at step t, and $\mathcal{A}(\mathcal{X}_{idt})$ includes the set of blogs that have not yet been visited (as described in §2.2).

2.6.1 Consumer-specific Match Utility

We use a "structural MDS" approach (Goettler and Shachar, 2001; Roos and Shachar, 2011) in formulating the match parameters (the δ_{ij}'s in 2.6). This approach has several advantages, including parsimony and ease of interpretation, as it can produce a joint map of consumer preferences and blog locations. Under this approach, each blog j occupies a point in a Q dimensional space, $z_j \in \Re^Q$, which represents its editorial position. Similarly, each consumer i occupies a point in the same space, $v_i \in \Re^Q$, which represents her preferences. Daily match utility is formulated as $(z_j + \nu_{jd})\, v_i$.[18] Thus, match value is decomposed into a blog position and a consumer position. This implies, for example, that a right-leaning political blog is likely to be preferred by a right-leaning consumer (and vice-versa). The ν_{jd}'s capture daily variation in blogs' positions around their long

[18] Hence, average match utility is $\delta_{ij} = z_j v_i$, daily match is $\nu_{ij}^d = \nu_{jd} v_i$, and the variances of the ν_{ij}^d's and link signals (the τ_δ^{-1}'s and τ_s^{-1}'s) are implicitly scaled by the level of consumers' preferences (the v_i's). Consumers therefore prefer sites satisfying $\operatorname{sign} z_j = \operatorname{sign} v_i$. For example, a conservative consumer with tastes $v_i = -1$ prefers a conservative blog with $z_j = -1$ (and thus $\delta_{ij} = 1$) over a liberal blog with $z_j = 1$ (and $\delta_{ij} = -1$).

term averages (right-leaning readers might be especially likely to read-right leaning blogs the day after an election). It is this variation that consumers learn anew each day (and which is signaled through links and excerpts).

The prior distribution of the site locations is $z_j \sim N(0, I_Q)$ (the unit variance is necessary for identification, which we discuss in §4.2), whereas the prior distribution of match preferences is $v_i \sim N(x_i\phi_v, \varsigma_v^2)$, with x_i a $1 \times P$ vector of demographic variables for consumer i, and ϕ a $P \times Q$ matrix of taste parameters. Under this approach, consumers with similar demographic characteristics should have ideal points that are closer together (assuming $|\phi_{v,pq}| > 0$). Consequently, for any blog j, the δ_{ij}'s for consumers with similar characteristics will be correlated. Moreover, the location of the z_j's will be such that, for any consumer i, blogs providing similar match utility will tend to be closer together. Hence we can produce a joint "map" of sites and consumers capable of generating insights into the way blogs are differentiated from one another and consumers are similar to each other (segmented). We estimate the model with a single latent attribute dimension ($Q = 1$).

We assume the the $\lambda_{0,i}$'s and γ_i are drawn from log-normal distributions. For each of these parameters $\vartheta \in \{\lambda_{0,i}, \gamma_i\}$, the prior distribution is defined as $\log\vartheta \sim N(\eta_\vartheta + x_i\phi_\vartheta, \varsigma_\vartheta^2)$. Thus we account for both observed heterogeneity (from the consumer characteristics, x_i) and unobserved heterogeneity (through ς_ϑ^2).

In §4.1, we show how blog visitation decisions across consumers and searches can be used to construct a likelihood function for estimation. Before proceeding, however, we explore the theoretical properties and empirical implications of the model.

2.7 Simulations

To understand how the structural primitives of the model affect search and information consumption behavior, we simulate data from the basic model under a variety of parameter settings. Simulating 10,000 browsing sessions for a single consumer, 15 blogs, and 100 bits; we set the base model's four parameters at the following "baseline" values: $\alpha_0 = 0.33$, $\lambda_0 = 0.08$, $\gamma = 3.75$, and $\kappa_0 = 10$. We then perturb each parameter, simulate additional browsing sessions, and compare these to the baseline. Because blogs are undifferentiated in the basic model, we compare just the number of sites visited in each condition. Figure 3 shows the proportion of searches during which the consumer visited some number of blogs.

From Figure 3, we see that as the amount of information at each site grows (i.e., as α_0 grows larger), consumers are more willing to initiate a session. However, when the amount of information

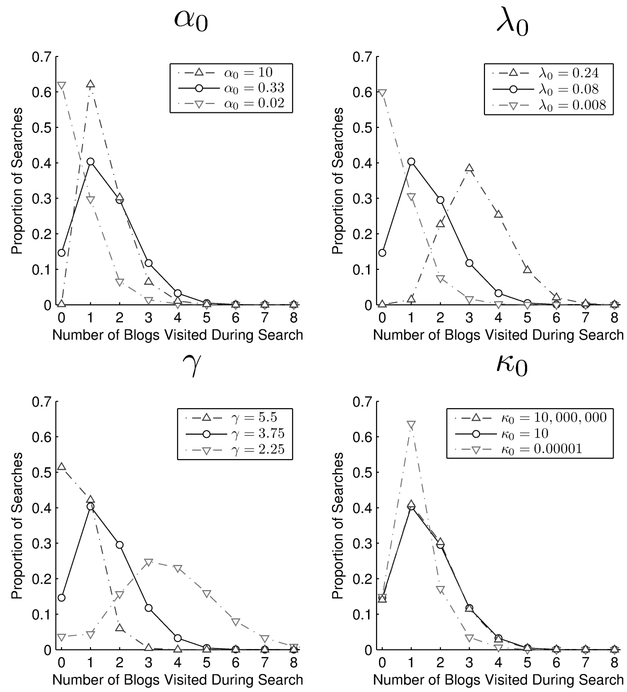

Figure 3: Simulated information searches in the basic model. Plots the proportion of searches (y-axis) for which the consumer visited $0, 1, 2, \ldots$ blogs. Changes to the structural primitives change search and information consumption behavior relative to the baseline $\alpha_0 = 0.33$, $\lambda_0 = 0.08$, $\gamma = 3.75$, $\kappa_0 = 10$.

at each site becomes especially large, consumers curtail session length. This attenuation in the length of sessions arises from the high degree of overlap in sites' information sets—consumers collect almost all of the available information at the first site, and hence have little reason to visit the next. Moreover, the effect of α_0 on search and consumption behavior is not symmetric: when information is scarce, searches are both less likely and shorter. Figure 3 also shows increasing the utility per bit (λ_0) and decreasing switching costs (γ) have similar, but not identical, effects due to differences in how these parameters affect period utility. Specifically, a decrease in γ affects utility by the same amount in every state, whereas an increase in λ_0 is attenuated by the number of new bits at each site. Also evident is the longer right tail for λ_0 relative to γ because the effect of utility per bit attenuates as search proceeds and consumer information satiates, but switching costs remain constant for every step. Finally, because expected utility does not depend on κ_0, changes to this parameter have little effect on the likelihood of initiating a session. A lower value of κ_0, however, increases the variability of information quality, and because the true quality level is not encountered until the first site is visited, increases the number of searches ending after step one.

The simulation suggests that even the most basic hyper-media search and consumption model can capture an array of different consumption patterns and explain some of the different search styles to which we alluded in Figure 1. Moreover, the differential effect of the various parameters on search and consumption behaviors also indicates that these parameters can be identified, a point we shall discuss further in §4.

3 Data

The data used to estimate this model of hyper-media search and consumption come from two sources: the first is a set of Internet log files containing information about consumers and their browsing activities (§3.1); the second contains information about the link structure between blogs (§3.2). We discuss each of these respectively before presenting preliminary model-free analysis (§3.3).

3.1 Internet Browsing Data

Internet browsing data are provided by comScore. The raw panel data used in this analysis, comprising more than 113 million records, contain information about the url-level browsing activities of over 2.5 million U.S. panelists, at 3000 blogs, over a period of two years (1/09–12/10). As the model describes choices that are made at the level of blogs, not individual pages or urls, we aggregate panelists' page requests by blog. We define a search/browsing session as all browsing activity on a given day, for two reasons. First, blogs operate under the same 24-hour news cycle as other media. A typical news item receives the most attention from blogs a few hours after it first breaks, but from there, the level of attention rapidly declines over the next 12–18 hours (Leskovec et al., 2007, 2009; Yang and Leskovec, 2011). Thus, two browsing sessions executed on the same day might yield a high degree of redundant information, whereas sessions on adjoining days might yield very little redundant information. Second, panelist browsing activity is often concentrated on a single day followed by days, or even weeks, of inactivity.[19]

The data also include demographic information for the panelists, such as age, gender, race, income, and family composition. Many of the individuals appear in the browsing data too briefly for us to have corresponding demographic information. Thus, individuals with fewer than two months of browsing data are excluded, as are those who lack demographic data.

Half of the panelists never visit any of the 3,000 blogs in the data set. Of those who do, many engage in only one or two visits per month. We focus on active consumers of blog information who are likely to have some degree of knowledge about blogs. In this regard, the present study is

[19] Prior studies have aggregated data to "sessions" lasting as long as 20 minutes Montgomery et al. (2004), a day Park and Fader (2004), or even a month Johnson et al. (2004).

Table 2: Summary of web browsing data, after aggregating by day and filtering for (A) frequent readers of blogs and (B) frequent readers of the five celebrity news blogs used for estimation.

	(A) Frequent Readers ($n \approx 22,000$)						(B) Estimation Panel ($n = 127$)					
	Min	25%	Median	Mean	75%	Max	Min	25%	Median	Mean	75%	Max
Sessions Per Day	186	812	1,089	1,094	1,382	1,997	39	62	71	70	77	85
Sessions Per Consumer	8	14	24	37	45	506	17	34	50	50	64	92
Blogs Per Session	1	1	1	1.6	2	33	1	1	1	1.3	1	5
Months in Panel Per Consumer	2	2	4	4.6	6	24	3	5	7	7.6	9	18

consistent with that of others who have dealt with this issue in the context of consumer packaged goods by eliminating "light users" (Gupta, 1988). Specifically, we focus on panelists who visit an average of four or more blogs per month. This subset includes 22 thousand panelists engaging in 750,000 sessions. Table 2 (A) provides summary statistics for this subset. The median panelist appears in the data for 4 months and initiates a session on 24 different occasions. The data are clearly weighted toward infrequent sessions involving a small number of blogs. Searches of any length are initiated about 35% of the time, and searches that end after visiting one blog are initiated about 25% of the time. Finally, with respect to demographics, 51% of the panelists are male, 55% are between the ages of 18 and 34, and 61% have children. Household incomes in the range of \$60–\$75 thousand are the most prevalent.

Estimating the model against browsing data for all 3,000 blogs in the panel is not yet feasible (we discuss estimation in §4). We therefore limit the empirical application to a small group of closely related sites. Specifically, we use browsing data for five celebrity news blogs: *celebuzz.com, dlisted.com*, *egostastic.com*, *perezhilton.com*, and *thesuperficial.com*. Celebrity news sites exemplify many of the features of hyper-media, such as redundant coverage of news events and linking. The data used for estimation were taken from Q4 2009. Over this period, consumers in the estimation panel needed to have engaged in at least 16 sessions (with a minimum of 4 per month) and visited at least two of the sites. The former requirement ensures we have enough data to estimate consumers' preferences, whereas the latter limits estimation to consumers who are more likely to be familiar with all five sites. 127 consumers met these criteria. Table 2 (B) provides summary statistics for the browsing this group's browsing activity in comparison to the larger panel, and Figure 4 shows the distribution of the number of sites visited in each of their sessions.

Among the panelists used to estimate the model, we find the median panelist to be female, aged 26–54, with household income less than \$65k, and with one or more children living at home.[20]

[20]Income data are categorical, and due to the relatively small number of consumers in each group, we combine these into two categories (less than and greater than or equal to \$65k per year).

Table 3: Market share (%) by demographic group for five celebrity news blogs.

	Gender		Age			HH Income		Children		HH Size		Afr. Amer.		
Blog	Female	Male	≤25	26–54	≥55	<\$65k	≥\$65k	≥1	0	1	>1	Yes	No	Total
celebuzz.com	16	8	12	10	30	9	18	11	15	14	12	1	13	13
dlisted.com	17	12	16	21	1	16	14	17	13	12	17	54	14	15
egotastic.com	5	22	12	9	7	13	10	12	12	15	10	19	12	12
perezhilton.com	54	45	50	48	58	52	48	49	52	52	50	25	51	50
thesuperficial.com	8	13	10	13	4	10	10	12	8	7	11	0	10	10
	62	38	41	54	5	59	41	51	49	30	70	2	98	100

Only 2% of the panelists are African American. Table 3 gives the relative share of traffic for each of the five celebrity news sites, broken out by demographics. *Perezhilton.com* receives half of all site traffic. The second most popular site among female panelists is *dlisted.com*, but *egotastic.com* is more popular among male panelists.

3.2 Link Data

We augment the browsing data by recording the full text and outgoing hyperlinks from every post published at the five celebrity news blogs during Q4 2009. These text and link data were scraped from the blogs' archives by a "web crawler" (a.k.a. a "spider") that extracts text and outbound links from blog posts. The web crawler systematically requests web pages and follows any embedded links it finds; when it finds a blog post, it extracts the date of publication, full text, and any embedded links. Hence these data indicate which links a panelist might have seen on each day. We use the link data during estimation, and the text data in an exploratory analysis of the parameter estimates. Table 4 shows the number of links observed during Q4 2009.

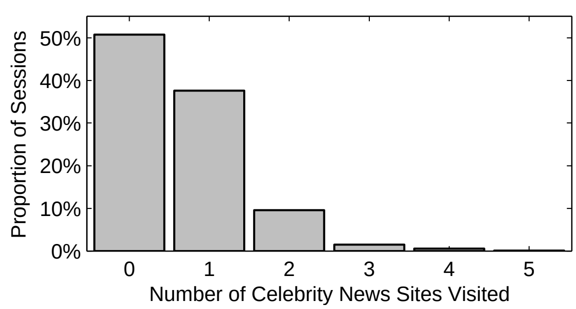

Figure 4: Distribution of searches by number of celebrity news sites visited (limited to the 5 used in estimation), aggregated across all panelists. Searches of any length are initiated about 50% of the time, and searches ending after one blog are initiated about 38% of the time.

Table 4: Total number of observed links between sites for Q4 2009. Links are made from sites listed in rows to sites listed in columns. Hence *dlisted.com* linked to *celebuzz.com* on 64 out of 92 days.

	Link Target				
Linking Site	*dlisted.com*	*perezhilton.com*	*thesuperficial.com*	*celebuzz.com*	*egotastic.com*
dlisted.com	–	2	2	64	63
perezhilton.com	0	–	0	7	0
thesuperficial.com	0	0	–	58	0
celebuzz.com	6	1	9	–	0
egotastic.com	60	0	0	0	–

3.3 Preliminary Analysis

Or model predicts that linking to another site might actually *discourage* consumers from visiting the linked site because 1) learning reduces uncertainty, and hence the option value of continuing a search, and 2) information can be redundant. This prediction runs counter to what previous studies have assumed—that after observing a link to another site, consumers are never less likely to visit the linked site (Mayzlin and Yoganarasimhan, 2008; Dellarocas et al., 2010). Hence, we consider model-free evidence regarding the relationship between linking and traffic flow.

3.3.1 Aggregate Analysis

Using the full two years of link and browsing data for the five celebrity news sites, we calculate for each (ordered) pair of sites (A, B) the correlation (across sessions) between 1) the number of links going from site A to site B and 2) the number of site A's visitors who subsequently visit site B during the same session. To account for differences in site popularity, we also calculate the correlation between 1) the proportion of A's links pointing to B, and 2) the proportion of site A's traffic that subsequently visits site B (i.e., the share of site A's traffic that is "sent" to site B). The correlation between the rank order of these variables provides a third robustness check.

The correlations are -0.41 between the number of links and visitors, -0.36 between the proportion of links and visitors, and -0.42 between the rank orders of links and visitors, all suggesting a negative association between linking and traffic flow.[21] Table 5 illustrates this negative association by showing linking data for *thesuperficial.com*. Even though *thesuperficial.com* linked to *celebuzz.com* more often than any other, it sent the least amount of traffic there—i.e., among people familiar with both sites, only 19% of those visiting *thesuperficial.com* went on to visit *celebuzz.com*. By contrast, even though *perezhilton.com* was linked to the least, 32% of those visiting *thesuperficial.com* subsequently visited *perezhilton.com*. Although limited to aggregate traffic patterns, this

[21]When we exclude the most popular site (*perezhilton.com*), the results are qualitatively similar, with correlations of -0.41 (no change), -0.42 (increase in magnitude), and -0.25 (decrease in magnitude) respectively.

Table 5: Number of links from *thesuperficial.com* to four other celebrity gossip sites, compared to the number of visitors subsequently visiting those sites. The correlation between links received and subsequent visits is estimated between −0.36 and −0.41.

	Links Received			Subsequent Visits		
Paired Site	Total	Share	Rank	Total	Share	Rank
perezhilton.com	3	<1%	4	2716	32%	1
egotastic.com	31	4%	3	1067	31%	2
dlisted.com	227	30%	2	1100	25%	3
celebuzz.com	501	66%	1	685	19%	4

analysis supports the idea that observing links might reduce the need for further search.

3.3.2 Individual-Level Analysis

We conduct a second analysis using the estimation panel in which we compare two conditional choice probabilities that are estimated directly from the data. The first is the empirical probability of visiting a site after having seen one or more links to it,

$$\widehat{\Pr}\left(a_{idt} = j | n_{ijdt} > 0\right) = \frac{\sum_d \sum_t \mathbf{1}\left(a_{idt} = j \wedge n_{ijdt} > 0\right)}{\sum_d \sum_t \mathbf{1}\left(n_{ijdt} > 0\right)}; \tag{22}$$

the second is the empirical probability of visiting a site when no links are observed,[22]

$$\widehat{\Pr}\left(a_{idt} = j | n_{ijdt} = 0\right) = \frac{\sum_d \sum_t \mathbf{1}\left(a_{idt} = j \wedge n_{ijdt} = 0\right)}{\sum_d \sum_t \mathbf{1}\left(n_{ijdt} = 0\right)}. \tag{23}$$

If links have no effect on choices, then we expect these conditional probabilities to be equal. If instead links encourage site visits, then we expect $\widehat{\Pr}\left(a_{idt} = j | n_{ijdt} > 0\right) > \widehat{\Pr}\left(a_{idt} = j | n_{ijdt} = 0\right)$, and if they discourage site visits, we expect the opposite.

In fact, we find instances of all three in the data. Figure 5 shows that about 30% of the time, observing a link decreases the chance of a subsequent visit to the linked site, about 20% of the time, the link increases it, and about 50% of the time, there is no difference (as indicated by the point mass over 0). The magnitude of the difference in choice probability varies. The average decrease in choice probability is about 5%, whereas the average increase is about 23%. About 70% of consumers saw links to more than one site; among these individuals, about 25% showed an increase at one site, and a decrease at another. While these results provide further model-free evidence that links may reduce the likelihood of visiting linked sites, they also indicate that any

[22] If consumer i never observed a link to site j (or only observed links *after* visiting site j) then we exclude the consumer-site pair (i, j) from the analysis because the condition $\sum_{dt} \mathbf{1}\left(n_{ijdt} > 0\right) = 0$ leaves the conditional probability $\widehat{\Pr}\left(a_{idt} = j | n_{ijdt} > 0\right)$ undefined. After eliminating these incomplete observations, we are left with 42% of the original $I \times J$ pairs.

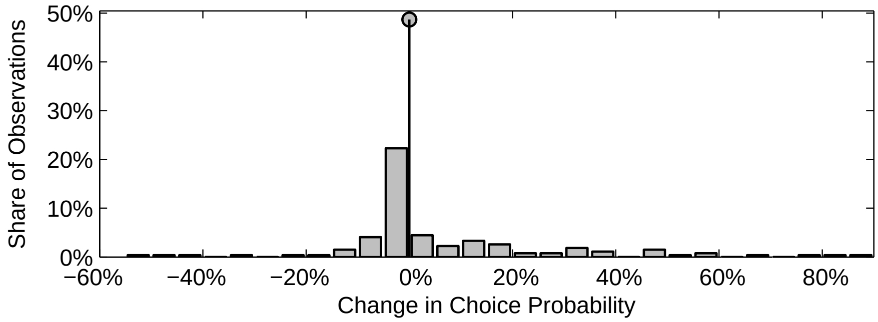

Figure 5: Empirical effect of one or more links on site choice probability. Observations are at the individual-site level. About 50% of cases show no change; these cases are represented by the point mass over 0%. In about 30% of cases, the probability of visiting a site was lower after seeing a link to it.

such effect will likely vary across individuals and sites.

4 Estimation Issues

We now discuss estimation of the model. We begin with the model likelihood function (§4.1). We then discuss identification of the model parameters (§4.2), their prior distributions (§4.3), and the estimation strategy (§4.4).

4.1 Likelihood Function

As noted earlier (§2.3.3), the unobserved (by the researcher) idiosyncratic utilities (the ϵ_{jt}'s) follow an extreme value distribution and are i.i.d. across options and choice occasions. Accordingly, we can integrate over the distribution of these unobserved shocks (McFadden, 1973) and express the probability of observing choice a_{idt}, conditional on state $\mathcal{X}_{idt}$ and the parameters, as a function of the choice-specific value functions (20) in the familiar multinomial logit form:

$$\Pr\left[a_{idt} = j|\mathcal{X}_{idt}, \theta, \theta_i\right] = \frac{\exp\left[V_j\left(\mathcal{X}_{idt}; \theta, \theta_i\right)\right]}{1 + \sum_{j' \in \mathcal{A}(\mathcal{X}_{idt})} \exp\left[V_{j'}\left(\mathcal{X}_{idt}; \theta, \theta_i\right)\right]}, \quad j \in \mathcal{A}\left(\mathcal{X}_{idt}\right), \tag{24}$$

where the parameter vectors θ and θ_i contain the model primitives: $\theta = \left(\alpha_0, \{\alpha_j\}, \tau_\delta, \tau_s, \{\omega_{j,k}\}, \{z_j\}\right)$, $\theta_i = (\lambda_{0,i}, \kappa_{0,i}, \gamma_i, v_i)$. The conditional probability of choice a_{idt} follows a multinomial distribution, hence the likelihood of the parameters (θ and $\{\theta_i\}$), conditional on the observed browsing choices (a) and unobserved states ($\mathcal{X}$), is the product of the choice probabilities (24) across all consumers and sessions:

$$L\left(\theta, \{\theta_i\}|a, \mathcal{X}\right) \propto \prod_{i=1}^{I} \prod_{d=1}^{D_i} \prod_{t=1}^{T_{id}} \prod_{j \in \mathcal{A}(\mathcal{X}_{idt})} \Big(\Pr\left[a_{idt} = j|\mathcal{X}_{idt}, \theta, \theta_i\right]\Big)^{a_{idt,j}}, \tag{25}$$

Table 6: Key model primitives estimated in the full model.

Parameter(s)	Description
λ_i	Average utility per bit of information
γ_i	Browsing cost
v_i	Preference for site position
α_j	Average bit availability
z_j	Average site position
τ_s	Relative precision of link signals

where $a_{idt,j} = 1$ if $a_{idt} = j$ (and 0 otherwise). We denote the step t at which consumer i's session ended on day d by $T_{id} = \{t|a_{idt} = 0\}$. Equation (25) depends on the unobserved states $\mathcal{X}$, hence the marginal likelihood $L(\theta, \{\theta_i\}|a)$ is found by integrating over the distribution of these unobserved states. We use the standard Bayesian approach of treating $\mathcal{X}$ as a latent model parameter vector, estimating the joint distribution of $(\theta, \{\theta_i\}, \mathcal{X})$ and then numerically integrating over $\mathcal{X}$. A detailed description of this procedure can be found in the Technical Appendix

4.2 Identification

We now discuss how variation in browsing and linking each day identifies the model primitives (Table 6 lists the key model primitives to be estimated in the full model). We begin with identification of the consumer and site parameters (§4.2.1), followed by the parameters and state variables associated with linking (§4.2.2). We then describe the parameter normalizations needed for estimation (§4.2.3).

4.2.1 Identification of Individual and Site-Specific Parameters

Identification of the latent site attributes (the α_j's and z_j's) and consumer preferences (the $\lambda_{0,i}$'s and v_i's) comes directly from consumers' choices, which we observe over the span of many browsing sessions. These choice data reveal each consumer's average expected utility from each site. When the expected utilities for two individuals are positively correlated (across sites and sessions), we expect those individuals to have similar preferences. By the same token, when expected utilities at two sites are positively correlated (across consumers and sessions), we expect those sites to have similar characteristics. In essence, individuals' preferences and sites' characteristics are identified by the covariance of expected utilities over many sessions (Goettler and Shachar, 2001).[23]

Each individual, however, has *two* parameters that characterize his preferences: one representing

[23]Note that because all of these parameters represent latent characteristics and preferences, their scale might not be identified. We discuss this point below.

his utility from information, $\lambda_{0,i}$, the other his match utility from site positions, v_i. Likewise, each site is characterized in two ways: by its expected amount of information, α_j, and by its average position, z_j. This raises the issue of whether all four consumer parameters can be identified from the data. In fact, identification of all four parameters is possible if we observe variation in consumers' choice probabilities at each step of a typical browsing session.

Consider the problem of separately identifying average site positions (z_j) and information levels (α_j), and assume for the moment that all sites provide the same amount of information on average ($\alpha_j = \alpha_k \forall\, j, k$). If 20% of all readers typically visit *perezhilton.com* at step 1, then another 16% (i.e., the 20% of the 80% of visitors who did not visit at step 1) is expected do so at step 2, since average site positions are constant throughout the session. Identification of the α_j's, therefore, depends on systematic deviations from each site's long-run share of traffic, at each step of the search. Sites with higher than average shares at step 1 are expected to have more information, whereas sites with higher shares at later steps should have less. In contrast, the effect of the average site positions (z_j) is not moderated by the step in the search process, but reflects the average blog shares over time.

By a similar argument, consumers' utilities from information (the $\lambda_{0,i}$'s) are identified by systematic deviations from their long-run average site choice probabilities, at each step. An individual whose choice probabilities do not vary at each step is expected to receive less utility from information than an individual who visits sites with higher α_j's early on. Thus the order of site visits plays an important role in identifying $\lambda_{0,i}$. In contrast, the consumer match values, (v_i's), reflect long-term choice shares that are not directly moderated by the step in the search.

Variation in the number of sites visited in a typical session allows us to separately identify consumers' browsing costs (γ_i's) from their utility from information ($\lambda_{0,i}$). If two individuals with identical preferences visit at least one site each day, their behavior might be due either to lower than average browsing costs or higher than average information utility. But if one of these individuals typically visits twice as many sites as the other, then he must have lower browsing costs, since the expected utility from bits of information decreases with each step.[24]

[24]Browsing costs are separately identified from match preferences owing to the fact that match preferences affect the utility of each site by an amount that depends on the site's z_j. Hence a change in γ_i affects the net utility from every site by the same amount, whereas a change in v_i does not.

4.2.2 Identification of Link Parameters

Data describing links between sites help identify the parameters and state variables related to links. If most consumers are systematically more or less likely to visit a site after seeing a link, then links must be informative in some way. Hence differences in sites' shares of traffic among individuals who saw a link and those who did not identify the overall informativeness of links, τ_s.

Differences in how individuals with different preferences respond to a particular link identifies the value of that link's signal, $s_{jk,d}$. When consumers who prefer one set of blogs react in the opposite way to consumers who prefer a different set, then we expect the signal $s_{jk,d}$ to be non-zero (the sign of the signal will depend on whichever group's choice probabilities increase and the layout of the site locations, z_j); if all consumers have the same reaction to a link, then the signal value of that link is expected to be zero. Differences in choice probabilities among individuals who saw different links to the same site identify the daily deviations in site positions, ν_{jd} (these state variables are not identified when fewer than two links are observed, however). While the data are expected to provide sharp identification of the informativeness of links, τ_s, given the sample size, identification of daily deviations in site position and signal values (ν_{jd} and $s_{jk,d}$) should be weak. Accordingly, we treat these variables as missing data and do not discuss them when presenting results.

4.2.3 Parameter Normalizations

The parameters describing the information environment (κ_0, N, and α_0) cannot be separately identified from the individual taste parameters. We could, for example, double the number of bits, while dividing κ_0 and the λ_i's by two, and describe the observed choices equally well. We therefore normalize $N = 30$ and $\kappa_0 = 4$ during estimation. The α_j's cannot be separately identified from α_0 because a change in α_0 can be directly offset through an adjustment of each α_j (we demonstrate this non-identification numerically in the Technical Appendix); thus we set $\alpha_0 = 1$ during estimation. As discussed earlier, the link data identify the informativeness of links, τ_s, as well as the relative value of each signal, $s_{jk,d}$, but their scales are not separately identified. Moreover, deviations in site position, ν_{jd}, and signal values, $s_{jk,d}$, are both latent constructs, thus we cannot identify both τ_s and τ_δ. We can, however, estimate their ratio, and set $\tau_\delta = 1$ accordingly. Finally, as we alluded earlier, the scale of consumer preferences (v_i) and site locations (z_j) cannot be identified separately from the model primitives and unobserved states (namely, $\overline{u}$, K, and s). We therefore enforce the constraints $\mathbb{E}(z_j) = 0$ and $\mathbb{V}(z_j) = 1$ during estimation.

Finally, even though the link probabilities ($\omega_{j,k}$'s) are identified from the number of links observed between pairs of sites, estimating these probabilities greatly increases the complexity of the value function calculation while adding little to our understanding of browsing behavior. Hence we set the link probabilities to be equal to their observed frequencies: $\omega_{j,k} = D^{-1} \sum_d \ell_{j,k,d}$, where $\ell_{j,k,d} = 1$ if site j linked to site k on day d.

4.3 Prior Distributions

The prior distribution for link informativeness, τ_s^{-1}, is $Inv\text{-}Ga\,(10, 9)$ (hence $\mathbb{E}\left(\tau_s^{-1}\right) = 1$ and $\mathbb{V}\left(\tau_s^{-1}\right) = 1/8$). The inverse-gamma distribution ensures that the precision of link signals is positive. The prior distribution of the site-specific information parameters (the α_j's) is logit-normal: $\operatorname{logit} \alpha_j \sim N\,(0, 1)$.

As we discussed earlier in §2.6.1, the prior distributions of the v_i's is standard normal, and the prior distribution of the $\lambda_{0,i}$'s and γ_i is log-normal: $\log \vartheta \sim N\left(\eta_\vartheta + x_i \phi_\vartheta, \varsigma_\vartheta^2\right)$, $\vartheta \in \{\lambda_{0,i}, \gamma_i\}$. Hyper-prior distributions for the η_ϑ's, ϕ_ϑ's and the ς_ϑ^2's are standard, conjugate normal-scaled-inverse-χ^2 distributions, leading to prior expectations of λ_i and γ_i's equal to $\exp(-1)$ and $\exp(1.5)$, respectively.

4.4 Estimation Strategy

In this section we outline the approach to estimation (details can be found in the Technical Appendix). The goal is to sample from the posterior distribution of the model parameters, hence we define an MCMC sampling routine based on the Metropolis-Hastings algorithm (Hastings, 1970, M-H hereafter). In order to execute this sampling routine, however, we must repeatedly estimate the value of the emax function (equation 21), a computational challenge that grows exponentially more complex with each additional state variable (the so-called "curse of dimensionality"). This complexity limits most single-agent DDC models to no more than six or seven state variables.

This model, however, has 17 state variables. Estimating the emax function in a model with so many state variables is made possible through the integration of three recent technical advances: (1) IJC's Bayesian dynamic programming algorithm, (2) Girolami and Calderhead's (2011) Riemann manifold Metropolis adjusted Langevin algorithm (mMALA), and (3) automatic differentiation (Griewank et al., 2010). We provide an overview of these advances in the context of the model below.

IJC reduces the computational burden of estimating the emax function in two ways. First, as in any DDC estimation, the emax function must be numerically integrated over all state variables.

In IJC, however, this averaging occurs over the states with the highest probability, and hence does not waste resources on states that almost never arise (and therefore have little influence over the estimate). Second, emax approximation is an iterative procedure in which the above integration is typically repeated many times; moreover, in a Bayesian setting, a new approximation is needed for every step of the MCMC sampler. But rather than fully iterate the expected value function at each step, IJC reuses previous approximations in such a way that only one iteration is needed. We discuss this procedure in further detail in the Technical Appendix.

While the computational gains from IJC are great, they come at a cost: MCMC samples exhibit higher autocorrelation (due to reusing information from previous MCMC steps), thus reducing the overall efficiency of the MCMC sampler. We alleviate some of this cost by using Girolami and Calderhead's (2011) mMALA procedure, which uses the derivatives of the log-posterior density function to construct a high-quality proposal distribution with two important qualities. First, a deterministic component of the proposal distribution "aims" the sampler in the direction of higher density regions of the parameter space. Second, the covariance of the random component is adjusted at each step to approximate the target distribution. Together, these features improve the rate of convergence and reduce autocorrelation.

In order to generate an mMALA proposal distribution, we need to know the first, second, and third partial derivatives of the log-probability function, which for single-agent DDC's are not available in any known closed form. We therefore use a technique known as automatic differentiation (also referred to as AD; Griewank et al., 1996; Su and Judd, 2010) to obtain the values of these derivatives as a byproduct of calculating the log-posterior function. By repeatedly applying the chain rule to the basic operations, such as addition, multiplication, etc., that make the log-posterior function, AD provides a way to calculate the *exact* value of the required derivatives. Thus calculation of both the log-posterior function and its first through third derivatives takes place in a single procedure. (Further details regarding mMALA and AD are available in the Technical Appendix.)

Together, these three advances allow us to estimate a model with many more state variables than previously possible. Note that the approach taken here is not specific to hyper-media search and consumption, but rather can be used to estimate other single-agent DDC models.

5 Results

We now present results from estimation of the full model. We begin with a discussion of the estimates for site positions (z_j) and information quantities (α_j). We then discuss the consumer

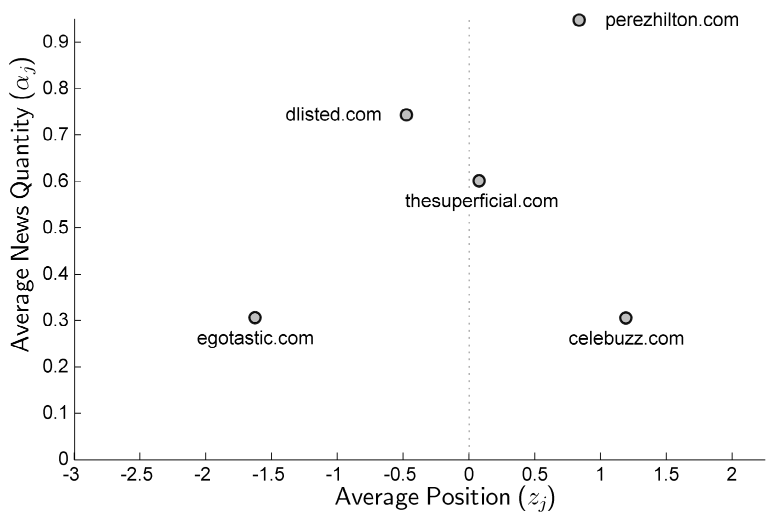

Figure 6: Estimated average site positions (z_j) and news quantities (α_j). Sites with negative z_j's emphasize salacious content, whereas those with positive z_j's emphasize gossipy content. Higher α_j's correspond with greater amounts of news.

parameters: their tastes for site positions (v_i), utility from information ($\lambda_{0,i}$), and browsing costs (γ_i). We conclude with a discussion of the parameter governing the informativeness of links (τ_s).

Figure 6 maps average site positions and information quantities, illustrating the strategies sites use to attract readers. For example, *egotastic.com* provides niche content containing relatively little news, whereas *thesuperficial.com* provides mainstream content with relatively more news.

The estimated level of news at each site varies considerably across sites: twice as many bits are expected at the site with the most information (*perezhilton.com*) than at the sites with the least information (*egotastic.com* and *celebuzz.com*). As we mentioned in §3.2, we collect word counts at each site. The left panel of figure 7 plots a regression of sites' estimated level of information (α_j) against their average (log) word counts, and shows that sites with higher expected levels of information also tend to have higher word counts.

The dimension along which sites differentiate from one another (via z_j) has a straightforward interpretation. Sites located at the left of Figure 6 (*thesuperficial.com*, *dlisted.com* and especially *egotastic.com*) regularly publish sexually-oriented content, for example pictorials of attractive female entertainers and models, and report on celebrities using a mix of humor and sarcasm. Sites at the right of Figure 6 (*celebuzz.com* and *perezhilton.com*) also rely on sexual content, although they do so less often, and sometimes feature male celebrities. Reporting at these sites is not without humor and sarcasm, but in contrast to the sites at the left, it is also at times positive and earnest. The most obvious differentiating factor, however, is the type and amount of sexual content pub-

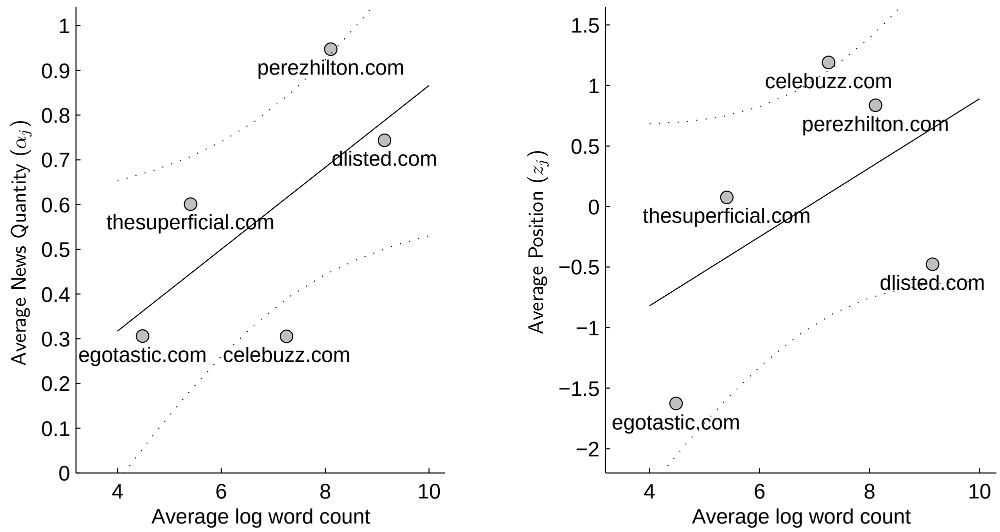

Figure 7: Regression of expected information (left) and site position (right) against average log word count. Sites with higher estimated information tend to also have higher word counts. Sites with z_j's greater than 0 also tend to have more words. Curved lines represent ± 2 SE around predictions.

lished. The right panel of Figure 7 shows a relationship between site positions and word counts that is consistent with the notion that sites with lower z_j's might place greater emphasis on publishing photos than text. Thus, we interpret the z_j's as points along a continuum between content that is primarily "sexy" at one end ($z_j < 0$) and "gossipy" at the other ($z_j > 0$).

Given this interpretation of the site positions, one might expect males, on average, to prefer sites on the left-hand side of Figure 6.[25] Figure 8 plots a histogram of the match preferences (posterior means of the v_i's) for males and females, and shows that men indeed prefer sites positioned to the left of 0 more strongly than females (median -0.02 versus 2.29, Mann-Whitney U-test $p = 0.002$).[26] With the exception of gender, there is no strong association between demographic variables and site position. Information utility and browsing costs generally do not vary significantly across demographics, again, with the exception of gender: the median female has 35% higher browsing costs compared to the median male (Mann-Whitney U-test $p = 0.02$ before controlling for multiple comparisons).

The estimated relative noisiness of links (the ratio τ_s/τ_δ) is about 2.20 (0.01). Figure 9 plots the posterior uncertainty about ν_{jd} after having seen one or more links to some unvisited site j. After observing one link, consumers' uncertainty is about 20% lower; after seeing a second link,

[25] Unfortunately, we do not have data regarding the sexual orientation of the panelists. We assume that the majority are heterosexual.

[26] To compensate for multiple comparisons, we apply a Bonferroni correction (equal to the number of demographic variables) to all p-values.

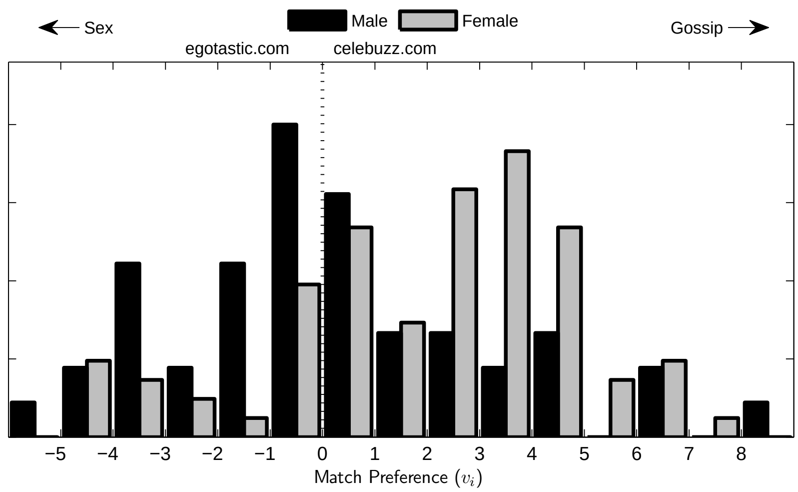

Figure 8: Distribution of match preferences (v_i) by gender. Males tend to prefer sites with negative z_j's (sexy) whereas females prefer sites with positive z_j's (gossipy).

uncertainty is reduced by another 10%. Links, therefore, provide useful information about site positions; at the same time, consumers recognize these are noisy signals and thus do not discount their prior beliefs completely.

6 Policy Experiments

A key goal of this study is to ascertain how various policies affect information consumption and consumer welfare, especially in the context of some recent public policy considerations. We present three counterfactual simulations, in which we examine the effects of net neutrality regulations and changes in fair use laws. The simulation procedure is discussed in the Technical Appendix.

6.1 Net Neutrality

On November 20, 2011, the FCC's first set of "net neutrality" rules went into effect. Although these rules prohibit broadband carriers from limiting consumers' access to high bandwidth sites, such as *netflix.com*, in order to reallocate bandwidth elsewhere, the new rules do not prevent carriers from "throttling" their networks—i.e., slowing down all traffic on the network in an effort to reduce congestion during periods of high demand.[27] We consider the implications of both types of intervention.

Because bandwidth throttling increases consumers' browsing costs, but does not affect their utility, we conduct these policy experiments by manipulating consumers' browsing costs in the

[27] See http://arstechnica.com/tech-policy/news/2011/09/us-net-neutrality-rules-finalized-in-effect-november-20.ars.

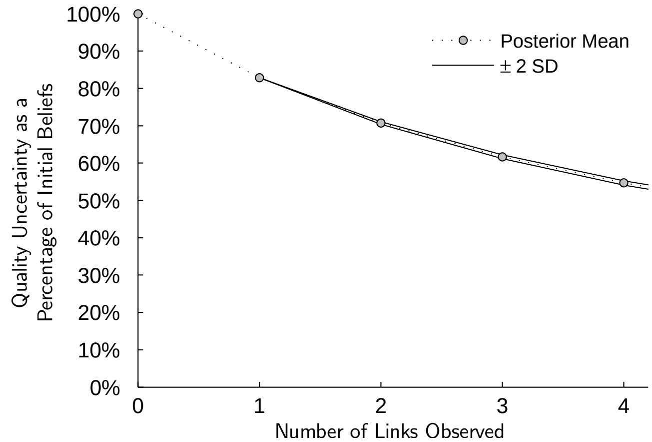

Figure 9: Estimated effect (±2SE) of links on beliefs about site positions. Uncertainty about site position is reduced by about 20% after seeing one link, and by 55% after seeing four.

counterfactual scenario. We simulate browsing under the counterfactual scenario and compare it to simulated browsing under the baseline condition (in which browsing costs are not manipulated). The effects of the counterfactual intervention are integrated over the posterior distribution of the parameters by simulating browsing (under both the baseline and counterfactual scenarios) once for each of 200 MCMC samples. We use the median outcome when calculating the change in traffic and welfare under the counterfactual scenario.

6.1.1 Bandwidth Throttling

We first explore a counterfactual scenario in which a broadband carrier throttles all traffic on its network. Specifically, we consider a scenario wherein every consumer's browsing costs are 1% higher. We find that for this 1% increase in browsing costs, the median consumer is expected to visit 4.4% fewer sites, causing site traffic to decline between 2.4% (*thesuperficial.com*) and 3.8% (*perezhilton.com* and *celebuzz.com*). Table 7 lists the expected change in traffic at each site, along with the probability that traffic will decrease rather than increase. The greatest traffic loss is found

Table 7: Simulated effect of various counterfactual interventions on site traffic. The expected change in traffic is calculated assuming a risk-neutral absolute loss function, and therefore equals the median across all simulations.

	Bandwidth Throttling		Bandwidth Reallocation		Fair Use	
Site	%Δhits	Pr [Δ < 0] (%)	%Δhits	Pr [Δ < 0] (%)	%Δhits	Pr [Δ < 0] (%)
dlisted.com	-3.2	81	18.2	0	-0.9	56
perezhilton.com	-3.8	97	-16.7	100	-0.1	54
thesuperficial.com	-2.4	66	15.5	0	-1.9	62
celebuzz.com	-3.8	86	18.1	0	-1.2	60
egotastic.com	-2.9	76	10.4	0	-0.4	53

at *perezhilton.com*, and this loss is realized primarily at the start of consumers' sessions. Consistent with this result, we find that browsing costs are relatively high among consumers who begin their sessions at these two sites (compared to those starting at other sites). In addition, *celebuzz.com*, which tends to be visited later in consumers' sessions, is disproportionately affected by the shorter sessions arising due to higher browsing costs.

The change in consumer welfare under higher browsing costs can be quantified via two measures: 1) the change in net utility, and 2) the change in the amount of information acquired. With increased costs, net utility for the median consumer decreases by 4.8% and the amount of information acquired is 4.4% lower. Moreover, because we consider income in the model, the estimates of ϕ_γ and ϕ_λ, which relate household income to search costs and bit utilities (see §4.3), can be used to translate the welfare loss into dollars (see the Technical Appendix for details). Following this approach, we estimate the welfare loss accrued to the median consumer at $-\$0.00002$ per year. This estimate is a consequence of consumers with higher incomes 1) receiving greater utility from information, and 2) having higher browsing costs. Hence when consumers curtail their browsing due to higher switching costs, the subsequent decreases in total utility and total cost, both due to fewer site visits, each correspond with an income loss. Although this estimate is small, note that it is based on a small number of sites. Should this result extrapolate to other consumers and other sites, the aggregate welfare loss is not likely to be insignificant.

6.1.2 Bandwidth Reallocation

At the center of the issue of bandwidth throttling is the idea that bandwidth intensive sites, such as *netflix.com*, use more than their fair share of available network resources, and in so doing slow down traffic to other sites (raising browsing costs). The argument in favor of limiting traffic to bandwidth-intensive rests on the assumption that higher utility outcomes obtain for those who do not use these sites. In the next counterfactual we test this conjecture. Specifically, we consider the case in which the broadband carrier places lower priority on traffic routed to the most popular site in our sample (*perezhilton.com*), and higher priority on traffic routed to the other four. Given a fixed level of bandwidth, we assume that the broadband carrier is constrained in such a way that the aggregate level of traffic across all sites does not change between the baseline and counterfactual conditions. That is, the carrier reallocates its available bandwidth in a way that benefits visitors to some sites at the expense of visitors to *perezhilton.com*. To simulate this policy outcome, we increase browsing costs by 4% whenever a consumer visits *perezhilton.com*, and lower them by

3% when browsing to any of the other sites. Table 7 shows that traffic is about 18% lower at *perezhilton.com*, and 10–15% higher at the other sites. Aggregate browsing, however, is effectively unchanged under this particular cost structure (expected traffic is about 1% lower, with $p = 0.25$).

Figure 10 depicts the change in utility for each consumer as a function of the change in his or her browsing behavior. Panel A plots the change in each consumers' utility as a function of the change in the number of site visits under the counterfactual policy. Under the counterfactual policy, we see some consumers browsing more and others less. These differences in browsing habits depend on two factors. First, because the cost of visiting *perezhilton.com* is higher under the couterfactual, those who visited *perezhilton.com* more often before the policy change (depicted by larger circles) browse less. Similarly, costs are lower at the other sites, so those who visited *perezhilton.com* less prior to the change now browse more. Second, the magnitude of the effect is strongest among consumers with the highest browsing costs (depicted with darker circles). Thus, at one extreme, we find consumers with high costs and a stronger preference for *perezhilton.com* browsing less (and receiving lower utility); and at the other extreme, we see consumers with high costs and a weaker preference for *perezhilton.com* browsing more (and receiving higher utility).

This outcome, however, focuses only on the utility gained from browsing, while ignoring the change in total cost associated with increased or decreased browsing. Hence, in Panel B, we consider the net effect of the policy, which is the change in the utility of information minus the change in browsing costs. Here we observe an interesting pattern: Among consumers with the lowest browsing costs (lighter circles), the impact of the bandwidth reallocation is as expected, depending mostly on how often one visits *perezhilton.com*. Among consumers with the highest browsing costs (darker circles), however, the results are mixed. Somes consumers cut back on their browsing and are better off for it, whereas others increase their browsing and are worse off. This result arises because the change in utility arising from a change in browsing is strongly linear (as shown in Panel A), but the relationship between changes in browsing and total costs is not. It is somewhat surprising that heavy users of *perezhilton.com* might be better off under a policy that restricts access to that site, and that increasing access to other sites could make some those sites' visitors worse off. Of note, the median female's browsing cost is 35% higher than the median male's, suggesting net neutrality regulations might have a greater effect on females' browsing behaviors than males'.

Finally, even though aggregate browsing is slightly lower (or at best, unchanged) under bandwidth reallocation, the median consumer is slightly better off (utility, net of browsing costs, is about 7% higher). This result demonstrates that the impact of bandwidth reallocation on welfare

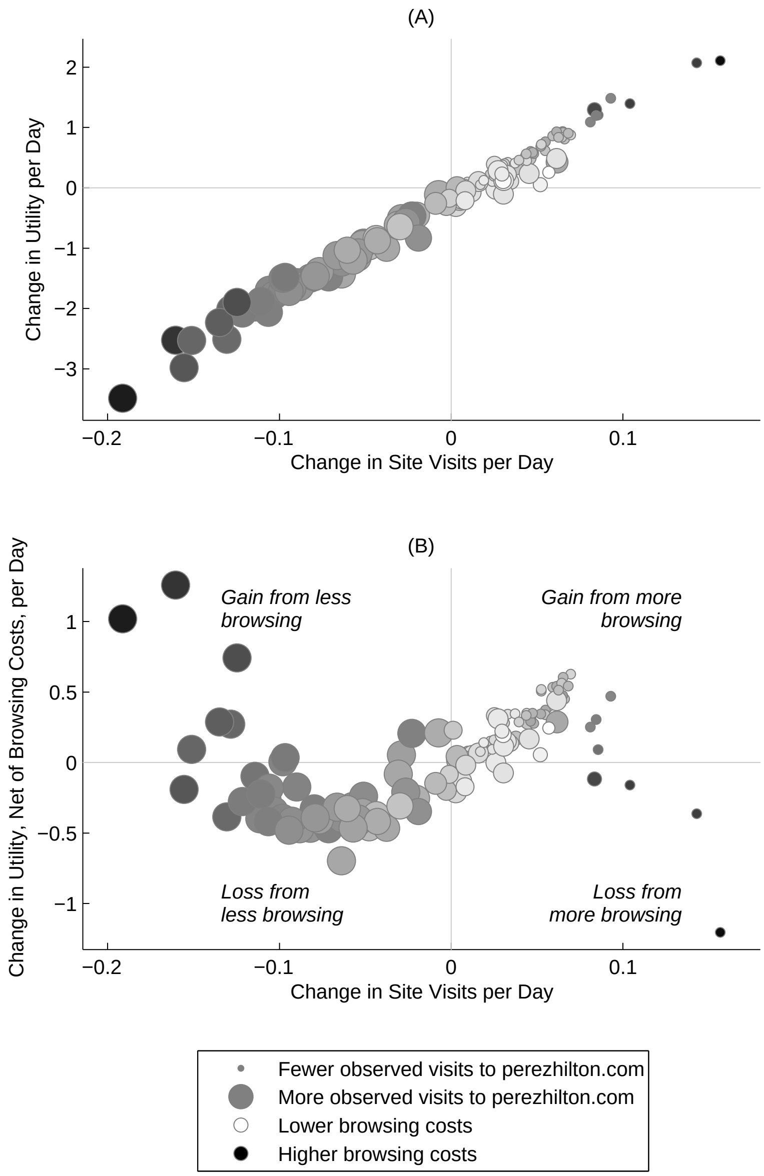

Figure 10: Effect of bandwidth reallocation on utility.

is not limited to simple first-order effects occurring at the affected sites. Rather, the way consumers substitute across sites, and the reasons for those substitutions (preferences v. costs) play important roles in determining the overall welfare implications of bandwidth reallocation. When considering the effect of net neutrality regulations, policy makers would be well advised to consider not only aggregate traffic changes, but also traffic changes among subpopulations with different browsing costs .

6.2 Fair Use/Copyright

Hyper-media sites frequently excerpt content from other sites under the auspices of fair use, however copyright law can change quickly due to legislative or judiciary action. If sites are restricted in their ability to excerpt from other sites (e.g., limiting excerpts posted at non-academic sites to a few sentences), then the informativeness of these signals would be diminished. The model can help us understand how such changes to copyright law impact browsing behavior. Specifically, by altering the value of the ratio τ_s/τ_δ, we can simulate the effect of an exogenous change to the informativeness of links.

We consider the effect of making links 20 times less informative, which might correspond with a policy under which linking is allowed, but excerpting content of any kind (including article or post titles) is not. Thus, links are allowed, but provide almost no information about linked sites (the first link reduces uncertainty about site location by just 1%). Under this scenario, the median consumer visits the same number of sites in the counterfactual as in the baseline. However, among the third of consumers with the most extreme tastes for site positions (i.e., with $|v_i| \gg 0$), about 0.9% fewer sites are visited under the counterfactual scenario. Knowledge and net utility decreased by 0.4% and 0.9% for the median consumer, with a net welfare loss occurring in 52% of simulations for both measures. These results suggest a small, but consistently negative effect of decreasing the informativeness of links. Hence, the aggregate welfare implications of link informativeness across sites are negligible.

There do exist notable differences at the site level, however. As shown in Table 7, changes in traffic volume are unequal across sites, ranging from -0.1% (*perezhilton.com*) to -1.9% (*thesuperficial.com*). These differences arise due to variation in these sites' linking habits and site locations. More specifically, a decrease in link informativeness can affect site traffic in two ways. For one, it can make visiting sites with many outbound links less appealing. Because links tell consumers which sites have higher than average match, visiting a site with many outbound links increases

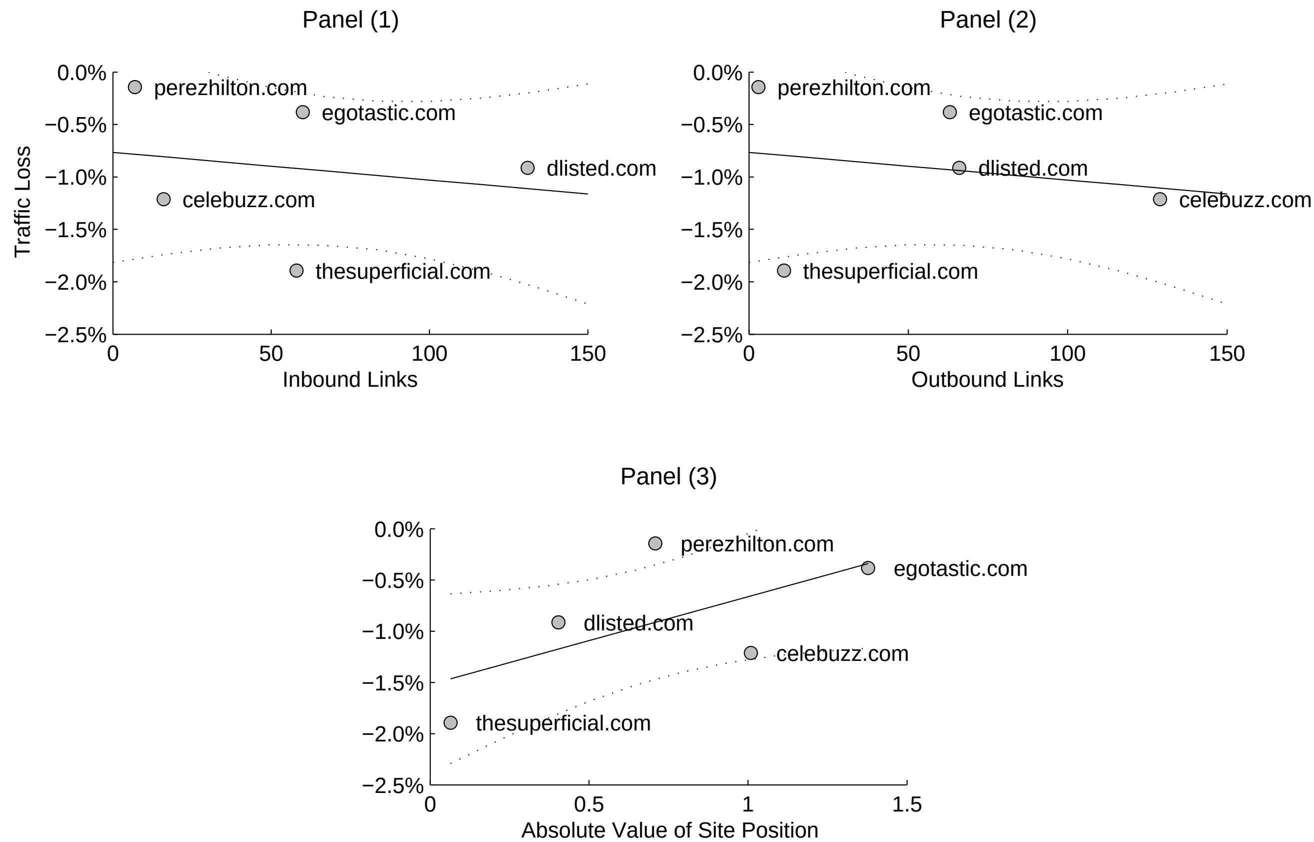

Figure 11: Simulated change in site traffic when links become 20 times less informative, plotted against the number of outbound links (1), inbound links (2), and absolute value of site position (3). Solid and dotted lines represent OLS predictions ±2SE

the expected value of subsequent browsing. This value is greatly diminished when links lose their signaling power, hence all else equal, we expect traffic to decrease at sites with many outbound links. Panel (1) of Figure 11, which plots the loss in traffic at each site against its number of outbound links, provides some evidence that sites with more outbound links are more negatively affected by decreased link informativeness.

A decrease in link informativeness can also affect traffic at sites with many inbound links. Consumers who only visit these sites when links indicate higher than average match will be less likely to visit them when links are less informative, because links signaling higher than average match will be more heavily discounted. Panel (2) of Figure 11 plots traffic loss against the number of inbound links, and shows that sites receiving a greater number of links tend to lose more traffic when links lose their informativeness. This effect, however, might be moderated by site location. Sites with niche content ($|z_j| \gg 0$) provide reliably high or low match utility according to individual preferences. For example, someone who finds content at *egotastic.com* to be extremely distasteful will not be influenced by a link indicating it is less distasteful than normal. This same individual, however, might find links to be useful when deciding whether to visit *thesuperficial.com*, whose z_j is close to zero and hence has $z_j > 0$ about half the time. Evidence in favor of this is provided in Panel (3) of Figure 11, which plots share loss against the absolute value of site location, $|z_j|$, and shows that sites with more mainstream content (i.e., closer to zero) lose a greater share of their traffic when links are less informative.

The various routes by which link informativeness affects site traffic help explain why traffic at *perezhilton.com* is almost unaffected by this counterfactual intervention: it gives and receives almost no links, and its content is somewhat niche-oriented. By contrast, *thesuperficial.com*, with many inbound links and a location close to zero, loses the greatest share of traffic. The effect of decreased link informativeness on sites with many outbound links appears somewhat weaker than the effect on sites with many inbound links. Thus, a change in fair use law prohibiting excerpts could have the most detrimental effect on traffic at sites with 1) mainstream content and 2) many inbound links.

7 Conclusion

This paper considers hyper-media search and consumption by reflecting essential features of this environment: 1) the rapid turnover of information (and consumers' concomitant uncertainty about its relevance and availability), 2) the redundancy of information (and the resulting diminishing marginal returns to continuing the session), and 3) the important signaling role played by linked excerpts (and its attendant effect on information search and consumption behavior). By capturing these aspects of information search and consumption, we explain a diverse assortment of consumers' search and consumption behaviors. As hyper-media become even more important to consumers, content producers, and advertisers, the need to explain this behavior is growing more acute.

The dynamic, structural approach to modeling hyper-media search and consumption taken here is based on how consumers use hyper-media to obtain information available from many linking and excerpting sites. We represent information at the most fundamental level—that of individual bits of information—and model links and excerpts as quality signals. As a result of the novel aspects of this context (rapid turnover, redundancy, and links), standard models of learning are ill equipped to capture these phenomena. Accordingly, we extend this literature to accommodate these phenomena, and in the process, develop an approach to facilitate the estimation of dynamic models with a large number of states.

The model we develop generates insights regarding the relationship between consumer behavior and the availability and relevance of information on a given day. For example, as the amount of information at each site grows either very large or very small, searches grows shorter—maximum search length is achieved at intermediate levels of information. We also find that, although links reduce uncertainty about the quality of linked sites, lower uncertainty can discourage consumers from continuing their sessions. The model also allows us to conduct policy experiments relevant to content producers, regulators, and researchers—such as how changes to copyright law (e.g., limitations on fair use) could affect demand for hyper-media, consumer learning, and welfare. Counterfactual analysis suggests that a change to fair use law prohibiting sites from excerpting each other could cause individuals with niche preferences to visit heavily linked, mainstream sites less often. In the context of net neutrality, counterfactual results also show that bandwidth throttling might have the most detrimental effect on sites serving consumer segments with higher than average browsing costs. Further, we find that limiting bandwidth intensive sites in order to free network resources for less intensive sites can actually harm some users of the less bandwidth-intensive sites. Findings suggest that fairness arguments in favor of throttling certain sites should consider not only the effect of this intervention on site visits, but also interactions between bandwidth allocation and users' browsing costs.

There are a number of extensions to the present study that offer interesting directions for future research. One is that we do not observe the total amount of information available each day, nor do we observe the amount of information at each site. However, by leveraging recent advances in machine learning (Leskovec et al., 2007, 2009; Yang and Leskovec, 2011), one may be able to infer a set of bits at each site. A second extension would be to consider potential dependencies between constructs that might be related. For example, daily match utilities might be correlated among sites providing similar types of information, or the number of excerpts and the amount of information at each site could be related. While estimating a model with such dependencies would be quite challenging, new methods and data sources should make such an effort feasible in the near future. A third extension would be to consider predictions regarding the amount of time spent on, or number of pages requested from each site. Better data would enable one to model the decision to leave a site, only to return shortly thereafter. A fourth extension would be to further extend the policy simulations along a number of dimensions. For example, one simulation involving paywalls would be to consider the myriad fee structures available to sites, such as multi-part tariffs or allowing linked traffic to avoid a paywall (as is currently practiced at the *New York Times*).

A fifth extension is to use the model to maximize an objective measure such as GRP's, reach, or frequency for the purposes of optimally placing display advertising; identifying the set of sites that achieves a particular objective can be challenging in an environment in which sites can direct traffic to each other.

A final route for future research is to leverage this demand model in a study aimed at understanding supply-side behavior. For example, we might look at the choices facing blogs, such as how much and what type of content to publish, or who to link to and when. Another promising area is to look at the decisions facing advertisers confronted with maximizing share vs. reach in a hyper-media setting. We have taken a step in this direction already by acquiring a data set from an advertising network specializing in blogs—hence for each of the blogs used in this study (and for the same period covered by the comScore panel) we have detailed information on the set of ads that ran, including the price paid for each run. The model of hyper-media search and consumption developed in this study is a necessary first step in developing a model of advertisers' behavior; we hope that our research can serve as a foundation for further inquiry regarding the role of hyper-media search and consumption on the Internet.

Appendix

A Nomenclature

The following table provides nomenclature for the model.

Symbol	Description	Role
Basic Model		
N	Total number of bits	Assumed
β_{jt}	Utility from information obtained from blog j at step t	Latent construct
ι_j	Set of information at blog j	Latent construct
k_t	Consumer's knowledge at step t	Latent construct
u_b	Utility from learning bit b	Latent construct
γ	Switching cost	Model primitive
ϵ_{jt}	Idiosyncratic taste shock at blog j, step t	Integrated out by the consumer
a_t	Action taken at step t	Data
$\mathcal{A}_t$	Set of actions that can be taken at step t	State variable
h_t	Set of blogs that have been visited at step t	Data
K_t	Number of bits that have been learned through step t	State variable
$\overline{u}_t$	Average utility per bit through step t	State variable
H_t	Number of blogs that have been visited through step t	State variable
$\mathcal{X}$	State variables that evolve according to a conditional distribution	Notational simplification
S	Set of all state variables, including $\mathcal{X}$ and per-period idiosyncratic shocks	Notational simplification
$V(\cdot)$	Value function (Bellman equation)	Latent construct
$V_j(\cdot)$	Choice-specific value function, or the expected value of choosing $a = j$ and making optimal decisions thereafter.	Latent construct
$W(\cdot)$	Emax function, or the expected value of optimal future choices prior to observing idiosyncratic shocks	Latent construct
π_b	Probability of observing bit b at any blog	Integrated out by the consumer
α_0	Ex ante number of bits expected at a blog for each bit that is missing	Model primitive
N_{t-1}	Maximum number of bits that might be found at step t	State variable
σ	Average utility per bit on a given day	Integrated out by the consumer
λ_0	Ex ante expected utility per bit	Model primitive
κ_0	Strength of ex ante beliefs about bit utility	Model primitive
Heterogeneous Model		
α_j	Bit availability at blog j	Model primitive
$\rho_{j,b}$	Probability of observing bit b at blog j	Integrated out by the consumer
δ_j	Average match utility with blog j	Model primitive
ν_j	Random deviation from match utility	Integrated out by the consumer
τ_δ^{-1}	Variance of match utility	Model primitive
Signaling Model		
$s_{j,t}$	Signal for blog j observed at step t	Integrated out by the consumer
$\overline{s}_{t,j}$	Average value of signals for blog j at step t	State variable
$n_{t,j}$	Number of signals observed for blog j at step t	Data
τ_s	Precision of signals	Model primitive
$\omega_{j,k}$	Empirical probability that blog j excerpts from blog k	Model primitive

References

Ackerberg, Daniel A. 2003. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review* 44(3):1007–1040.

Aguirregabiria, Victor, and Pedro Mira. 2010. Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1):38–67.

Allen, Beth. 1983. Neighboring information and distributions of agents' characteristics under uncertainty. *Journal of Mathematical Economics* 12(1):63–101.

———. 1986. The demand for (differentiated) information. *The Review of Economic Studies* 53(3):311.

———. 1990. Information as an economic commodity. In *Papers and proceedings of the hundred and second annual meeting of the american economic association*, ed. Ronald L. Oaxaca and Wilma St. John, 268–273. Pittsburgh: American Economic Association.

Anand, Bharat N., and Ron Shachar. 2009. Targeted advertising as a signal. *Quantitative Marketing and Economics* 7(3):237–266.

———. 2010. Advertising, the matchmaker. Working paper.

Ansari, Asim, and Carl F. Mela. 2003. E-customization. *Journal of Marketing Research* 40(2):131–145.

Bellman, R. 1957. Dynamic programming and its application to optimal control .

Blood, Rebecca. 2003. Weblogs and journalism: Do they connect? *Nieman Reports* 57(3):61–63.

Brooks, Christopher H., and Nancy Montanez. 2006. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *Proceedings of the 15th international conference on World Wide Web*, 625–632. New York: ACM.

Bucklin, Randolph E., and Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research* 40(3):249–267.

Byzalov, Dmitri, and Ron Shachar. 2004. The risk reduction role of advertising. *Quantitative Marketing and Economics* 2(4):283–320.

Chatterjee, Patrali, Donna L. Hoffman, and Thomas P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science* 520–541.

Chintagunta, Pradeep, Ronald Goettler, and Minki Kim. 2012. New drug diffusion when forward-looking physicians learn from patient feedback and detailing. *Chintagunta, Pradeep K., Ronald L. Goettler, and Minki Kim," New Drug Diffusion When Forward-Looking Physicians Learn from Patient Feedback and Detailing," Journal of Marketing Research, Forthcoming* .

Danaher, Peter J. 2007. Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Science* 26(3):422–437.

Dellarocas, Chrysanthos, Zsolt Katona, and William Rand. 2010. Media, aggregators and the link economy: Strategic hyperlink formation in content networks. Tech. Rep., The Networks, Electronic Commerce, and Telecommunications ("NET") Institute. Working paper 10-13.

Erdem, Tülin, and Michael P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* 15(1):1–20.

Erdem, Tülin, Michael P. Keane, T. Sabri Öncü, and Judi Strebel. 2005. Learning about computers: An analysis of information search and technology choice. *Quantitative Marketing and Economics* 3(3):207–247.

Erdem, Tülin, Michael P. Keane, and Baohong Sun. 2008. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science* 27(6):1111–1125.

Fujimura, Ko, Takafumi Inoue, and Mawayuki Sugisaki. 2005. The eigenrumor algorithm for ranking blogs. In *Proceedings of the www 2005 workshop on the weblogging ecosystem: Aggregation, analysis and dynamics*.

Girolami, Mark, and Ben Calderhead. 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(2):123–214.

Goettler, Ron L., and Ron Shachar. 2001. Spatial competition in the network television industry. *The RAND Journal of Economics* 32(4):624–656.

Goldfarb, Avi. 2002. Analyzing website choice using clickstream data. *Advances in Applied Microeconomics* 11:209–230.

Griewank, Andreas, David Juedes, and Jean Utke. 1996. Algorithm 755: ADOL-C: A package for the automatic differentiation of algorithms written in C/C++. *ACM Transactions on Mathematical Software* 22(2):131–167.

Griewank, Andreas, L. Lehmann, H. Leovey, and M. Zilberman. 2010. Automatic evaluations of cross-derivatives. Working paper.

Gupta, Sunil. 1988. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research* 25(4):342–355.

Hastings, W.K. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57(1):97–109.

Ho, Teck-Hua, Christopher S. Tang, and David R. Bell. 1998. Rational shopping behavior and the option value of variable pricing. *Management Science* 44(12):145–160.

Hortaçsu, Ali, and Chad Syverson. 2004. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics* 119(2):403–456.

Imai, Susumu, Neelam Jain, and Andrew Ching. 2009. Bayesian estimation of dynamic discrete choice models. *Econometrica* 77(6):1865–1899.

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse. 2004. On the depth and dynamics of online search behavior. *Management Science* 50(3):299–308.

Keane, Michael P., Petra E. Todd, and Kenneth I. Wolpin. 2011. Chapter 4 - the structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. vol. 4, Part A of *Handbook of Labor Economics*, 331 – 461. Elsevier.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg. 2010. Online demand under limited consumer search. *Marketing Science* 29(6):1001.

Lee, Sukekyu, Fred Zufryden, and Xavier Drèze. 2003. A study of consumer switching behavior across internet portal web sites. *International Journal of Electronic Commerce* 7(3):39–63.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, 497–506. New York: ACM.

Leskovec, Jure, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. 2007. Cascading behavior in large blog graphs. `http://arxiv.org/abs/0704.2803`.

Lloyd, Levon, Prachi Kaulgud, and Steven Skiena. 2006. Newspapers vs. blogs: Who gets the scoop. In *AAAI spring symposium on computational approaches to analyzing weblogs*. Menlo Park: AAAI.

Lovett, Mitchell J. 2008. Unstable consumer learning models: Structural estimation and experimental examination. Ph.D. thesis, Duke University.

Mayzlin, Dina, and Hema Yoganarasimhan. 2008. Link to success: How blogs build an audience by promoting rivals. Working paper, Yale School of Management.

McFadden, Daniel. 1973. Conditional logit analysis of qualitative behavior. In *Frontiers in econometrics*, ed. P. Zarembka, 105–142. New York: The Academic Press, Inc.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan. 2004. Role of forgetting in memory-based choice decisions: A structural model. *Quantitative Marketing and Economics* 2(2):107–140.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Science* 579–595.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar. 1997. Consumer information search revisited: Theory and empirical analysis. *The Journal of Consumer Research* 23(4):263–277.

Nakajima, Shinsuke, Junichi Tatemura, Yoichiro Hino, Yoshinori Hara, and Katsumi Tanaka. 2005. Discovering important bloggers based on analyzing blog threads. In *Workshop on the weblogging ecosystem, 14th international World Wide Web conference*.

Nevo, Aviv, and Victor Aguirregabiria. 2011. Recent developments in empirical io: Dynamic demand and dynamic games. *Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games (December 29, 2010). Northwestern University Center for the Study of Industrial Organization Working Paper* (0107).

Park, Young-Hoon, and Peter S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Science* 280–303.

Punj, Girish N., and Richard Staelin. 1983. A model of consumer information search behavior for new automobiles. *The Journal of Consumer Research* 9(4):366–380.

Ratchford, Brian T., and Narasimhan Srinivasan. 1993. An empirical investigation of returns to search. *Marketing Science* 12(1):73–87.

Roberts, John H., and Glen L. Urban. 1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science* 34(2):167–185.

Roos, Jason M.T., and Ron Shachar. 2011. When Kerry met Sally: Politics and perceptions in the demand for movies. Working paper.

Shin, Sangwoo, Sanjog Misra, and Dan Horsky. 2007. Disentangling preferences, inertia and learning in brand choice models. Working paper.

Simonson, Itamar, Joel Huber, and John Payne. 1988. The relationship between prior brand knowledge and information acquisition order. *The Journal of Consumer Research* 14(4):566–578.

Stigler, George J. 1961. The economics of information. *The Journal of Political Economy* 69(3):213–225.

Su, Che-Lin, and Kenneth L. Judd. 2010. Constrained optimization approaches to estimation of structural models. Working paper.

West, Mike, and Jeff Harrison. 1997. *Bayesian forecasting and dynamic models:*. 2nd ed. New York: Springer-Verlag.

Yang, Jaewon, and Jure Leskovec. 2011. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on web search and data mining*, 177–186. New York: ACM.

Zhao, Yi, Sha Yang, Vishal Narayan, and Ying Zhao. 2010. Modeling consumer learning from online product reviews. Working paper.

Hyper-Media Search and Consumption

Technical Appendix

Jason M.T. Roos[*] Carl F. Mela[†] Ron Shachar[‡]

June 13, 2012

1 Assumptions from the Basic Model

In this Appendix, we elaborate on key assumptions from the Basic Model.

1.1 Value Function Assumptions

We make a number of assumptions regarding the value function that are standard to the literature on single-agent dynamic discrete choice problems (Aguirregabiria and Mira, 2010). First, per Equation (2) in the main document, the idiosyncratic shock to utility, ϵ, is additively separable from the other components of utility—namely the utility from information, $\beta\left(\mathcal{X}', \mathcal{X}\right)$, and the switching cost, γ. We further assume that these random shocks are i.i.d over time, conditionally independent (i.e., conditional on the previous state) of the other state variables, $\mathcal{X}'$, and distributed $EV\left(0,1\right)$ independently across alternatives. Thus, the state transition density, $p\left(\mathcal{S}'|\mathcal{S}, a=j\right)$, can be factored as $f\left(\mathcal{X}'|\mathcal{X}, a=j\right) g\left(\epsilon\right)$, where $g\left(\cdot\right)$ denotes the $EV\left(0,1\right)$ distribution.

1.2 I.I.D. Bit Probabilities

The i.i.d. assumptions in Equation (7) are not restrictive. The intuition behind this claim is that the consumer only cares about the bits he hasn't seen yet. By definition, these bits have so far been absent from the blogs he has visited. Thus, even though the consumer can reasonably update his beliefs about dependencies among bits he has seen, these bits are empirically uncorrelated with the ones he hasn't seen. The consumer knows he will not be able to update his prior beliefs in a way that affects his decisions, thus the i.i.d. assumption is not restrictive.

[*]Assistant Professor, Rotterdam School of Management, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, Netherlands; email: roos@rsm.nl; phone: +1 206 317 1713.

[†]T. Austin Finch Foundation Professor of Business Administration, Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, North Carolina, 27708; email: mela@duke.edu; phone: +1 919 660 7767.

[‡]Dean, Arison School of Business, Interdisciplinary Center (IDC) Herzliya, Herzliya 46150, Israel; email: ronshachar@idc.ac.il; phone +972 09 960 2408.

1.3 I.I.D. Bit Utilities

Dependencies between u_b's, conditional on their common mean σ_d, do not affect the consumer's choices. The rationale for this claim is that the consumer cannot update his beliefs in any meaningful way. Within a single search, the consumer observes one draw of u—i.e., one draw from the joint distribution of the u_b's—and more often, he will observe just a subset of the available u_b's. In the absence of repeated observations from this distribution, it is difficult for the consumer to update his beliefs in a meaningful way, hence the i.i.d. assumption is not restrictive. Also, note that even though the u_b's are conditionally independent, σ_d is unobserved, making the u_b's *unconditionally* dependent. (Indeed, this unconditional dependency is what allows the consumer to update his beliefs about $\tilde{\sigma}_d$.)

2 Proofs of Claims from the Theoretical Model

Claim 1. We can express the utility from information at step t (Equation 1) as a function of the transition from state $\mathcal{X}_t$ to state $\mathcal{X}_{t+1}$

$$\beta_t = \beta_{jt} = \overline{u}_t K_t - \overline{u}_{t-1} K_{t-1}$$

Proof. The consumer knows bit b at step t because either 1) it was already prior knowledge, or 2) he learned b at blog $a_t = j$—but not both. Thus,

$$k_{tb} = k_{t-1,b} + \iota_{jb} - \iota_{jb} k_{t-1,b},$$

and solving for $\iota_{jb}\left(1 - k_{t-1,b}\right)$,

$$\iota_{jb}\left(1 - k_{t-1,b}\right) = k_{tb} - k_{t-1,b}. \tag{A1}$$

Recall Equation (1) from the main document:

$$\beta_{jt} = \sum_{b=1}^{N} u_b \iota_{jb}\left(1 - k_{t-1,b}\right).$$

Substituting (A1) into (1),

$$\begin{aligned}
\beta_{jt} &= \sum_{b=1}^{N} u_b \left(k_{tb} - k_{t-1,b}\right) \\
&= \sum_{b=1}^{N} u_b k_{t,b} - \sum_{b=1}^{N} u_b k_{t-1,b} \\
&= \frac{\sum_{b=1}^{N} u_b k_{t,b}}{K_t} K_t - \frac{\sum_{b=1}^{N} u_b k_{t-1,b}}{K_{t-1}} K_{t-1} \\
&= \overline{u}_t K_t - \overline{u}_{t-1} K_{t-1}
\end{aligned}$$

□

Claim 2. The expected utility from the information at a blog visited in step t is equal to the following:

$$\mathbb{E}\left[\beta_t | K_{t-1}, \overline{u}_{t-1}, N, h_{t-1}\right] = \mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right]$$
$$\times\, \mathbb{E}\left[\overline{u}_t | \mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right], \overline{u}_{t-1}, K_{t-1}\right] - K_{t-1}\overline{u}_{t-1}$$

Proof. This proof relies on material discussed in the main text after (6) is first presented (in particular, see the next two proofs). From the definition of β_t we know $\mathbb{E}\left[\beta_t | K_{t-1}, \overline{u}_{t-1}, N, h_{t-1}\right] = \mathbb{E}\left[\overline{u}_t K_t - \overline{u}_{t-1} K_{t-1} | K_{t-1}, \overline{u}_{t-1}, N,\right.$
Using the law of iterated expectations, the right-hand side expression can be written as follows.

$$\mathbb{E}\left[K_t \overline{u}_t - K_{t-1}\overline{u}_{t-1} | K_{t-1}, \overline{u}_{t-1}, N, h_{t-1}\right] = \mathbb{E}\left[K_t \mathbb{E}\left[\overline{u}_t | K_t, K_{t-1}, \overline{u}_{t-1}\right] | K_{t-1}, N, h_{t-1}\right]$$
$$- K_{t-1}\overline{u}_{t-1}$$

Substitution from (12) gives

$$= \mathbb{E}\left[K_t \left\{ \frac{\kappa_0 \tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}} \left(1 - \frac{K_{t-1}}{K_t}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{K_t}\right) \right\} \middle| K_{t-1}, N, h_{t-1}\right] - K_{t-1}\overline{u}_{t-1}$$

and after simplifying,

$$= \mathbb{E}\left[K_t - K_{t-1} | K_{t-1}, N, h_{t-1}\right] \left\{ \frac{\kappa_0 \tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}} \right\}$$

This is the expected number of new bits times the expected utility per new bit, which can be written as

$$= \left(\mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right] - K_{t-1}\right) \left\{ \frac{\kappa_0 \tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}} \right\} + K_{t-1}\overline{u}_{t-1} - K_{t-1}\overline{u}_{t-1}$$
$$= \mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right] \left\{ \frac{\kappa_0 \tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}} \left(1 - \frac{K_{t-1}}{\mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right]}\right) \right.$$
$$\left. + \overline{u}_{t-1}\left(\frac{K_{t-1}}{\mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right]}\right) \right\} - K_{t-1}\overline{u}_{t-1}$$

Finally, this expression is equal to the expected value of K_t times the expected value of $\overline{u}_t$ conditional on $\mathbb{E}\left[K_t\right]$ bits, minus the prior total utility from information, $K_{t-1}\overline{u}_{t-1}$.

$$= \mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right] \mathbb{E}\left[\overline{u}_t | \mathbb{E}\left[K_t | K_{t-1}, N, h_{t-1}\right], \overline{u}_{t-1}, K_{t-1}\right] - K_{t-1}\overline{u}_{t-1}$$

□

Claim 3. The posterior distribution of $\tilde{\pi}_{bt}$ for an unseen bit is:

$$\tilde{\pi}_{bt} | k_{t-1,b} = 0, H_{t-1} \sim Beta\left(\tilde{\alpha}_0, H_{t-1} + 1\right)$$

Proof. The prior distribution of $\tilde{\pi}_b$, the likelihood of bit b appearing at any blog, is $Beta\left(\tilde{\alpha}_0, 1\right)$. The likelihood of not finding bit b after visiting H_{t-1} blogs is $\left(1 - \tilde{\pi}_b\right)^{H_{t-1}}$. Applying Bayes' rule, the posterior distribution of $\tilde{\pi}_b$ is

$$\frac{\tilde{\pi}_b^0 \left(1-\tilde{\pi}_b\right)^{H_{t-1}} \times \tilde{\pi}_b^{\tilde{\alpha}_0-1} \left(1-\tilde{\pi}_b\right)^{1-1} \left[B\left(\tilde{\alpha}_0,1\right)\right]^{-1}}{\int_0^1 \tilde{\pi}_b^0 \left(1-\tilde{\pi}_b\right)^{H_{t-1}} \times \tilde{\pi}_b^{\tilde{\alpha}_0-1} \left(1-\tilde{\pi}_b\right)^{1-1} \left[B\left(\tilde{\alpha}_0,1\right)\right]^{-1} d\tilde{\pi}_b}$$

Reducing terms,

$$\frac{\left(1-\tilde{\pi}_b\right)^{H_{t-1}+1-1} \tilde{\pi}_b^{\tilde{\alpha}_0-1}}{\int_0^1 \left(1-\tilde{\pi}_b\right)^{H_{t-1}+1-1} \tilde{\pi}_b^{\tilde{\alpha}_0-1} d\tilde{\pi}_b}$$

The denominator is equal to the complete beta function, $B\left(\tilde{\alpha}_0, H_{t-1}+1\right)$, so the above expression is

$$\frac{\left(1-\tilde{\pi}_b\right)^{H_{t-1}+1-1} \tilde{\pi}_b^{\tilde{\alpha}_0-1}}{B\left(\tilde{\alpha}_0, H_{t-1}+1\right)}$$

which is the p.d.f of the $Beta\left(\tilde{\alpha}_0, H_{t-1}+1\right)$ distribution. □

Claim 4. The marginal posterior predictive distribution of $K_t - K_{t-1}$ is binomial.

Proof. Recall that conditional on $\tilde{\pi}_{bt}$, ι_{jb} is a Bernoulli random variable.

$$\Pr\left[\iota_{jb}=1|\tilde{\pi}_{bt}\right] = \tilde{\pi}_{bt}$$

By assumption, the ι_{jb}'s and $\tilde{\pi}_{bt}$'s are independent across bits. Thus the marginal probability of $\tilde{\pi}_{bt}$ is

$$\Pr\left[\iota_{jb}=1\right] = \int \tilde{\pi}_{bt} p\left(\tilde{\pi}_{bt}|h_{t-1}, k_{t-1}\right).$$

We are concerned only with the bits for which $k_{t-1,b}=0$. The posterior distribution of $\tilde{\pi}_{bt}$ for these bits was defined above as $\tilde{\pi}_{bt}|k_{t-1,b}=0, H_{t-1} \sim Beta\left(\tilde{\alpha}_0, H_{t-1}+1\right)$ main text. Thus,

$$\begin{aligned}
\Pr\left[\iota_{jb}=1|k_{t-1,b}=0\right] &= \int \tilde{\pi}_{bt} Beta\left(\tilde{\pi}_{bt}|\tilde{\alpha}_0, H_{t-1}+1\right) \\
&= \mathbb{E}\left[\tilde{\pi}_{bt}|\tilde{\alpha}_0, H_{t-1}+1\right] \\
&= \frac{\tilde{\alpha}_0}{\tilde{\alpha}_0+H_{t-1}+1},
\end{aligned}$$

where the last equality substitutes the mean of the Beta distribution. The marginal probability of all bits that haven't been seen is equal to $\tilde{\alpha}_0/\left(\tilde{\alpha}_0+H_{t-1}+1\right)$, thus the sum $\sum_{b=1}^{N} \iota_{jb}\left(1-k_{t-1,b}\right)$ by definition follows a binomial distribution. There are $N-K_{t-1}$ draws, each occurring with probability $\tilde{\alpha}_0/\left(\tilde{\alpha}_0+H_{t-1}+1\right)$. Thus,

$$K_t - K_{t-1} \sim Bin\left(N-K_{t-1}, \frac{\tilde{\alpha}_0}{\tilde{\alpha}_0+H_{t-1}+1}\right).$$

□

Claim 5. The posterior distribution of $\tilde{\sigma}_t$ is

$$\tilde{\sigma}_t | K_{t-1} \sim \mathit{Inv\text{-}Ga}\left(\tilde{\kappa}_{t-1} + 1, \tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)$$

Proof. The prior distribution of $\tilde{\sigma}$ is given in (10), and as discussed in the main paper, the probability of the $K_t\overline{u}_t - K_{t-1}\overline{u}_{t-1} = \beta_t$ bits follows a gamma distribution:

$$\beta_t | \tilde{\sigma}_t, K_t, K_{t-1} \sim Ga\left(K_t - K_{t-1}, \tilde{\sigma}_t\right).$$

The posterior distribution of $\tilde{\sigma}_t$ is found by applying Bayes' rule. Let $\nu_t = K_t - K_{t-1}$, $\tilde{\kappa}_{t-1} = \tilde{\kappa}_0 + K_{t-1}$, and $\tilde{\lambda}_{t-1} = \left(\tilde{\kappa}_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}\right) / \left(\tilde{\kappa}_0 + K_{t-1}\right)$ and consider the posterior distribution of $\tilde{\sigma}_t$ at $t = 1$.

$$\frac{\beta_t^{-\nu_t} \exp\left(-\frac{\beta_t}{\tilde{\sigma}_t}\right) \left(\Gamma\left(\nu_t\right)\right)^{-1} \tilde{\sigma}^{-\nu_t} \left(\tilde{\lambda}_0\tilde{\kappa}_0\right)^{\tilde{\kappa}_0+1} \left(\Gamma\left(\tilde{\kappa}_0+1\right)\right)^{-1} \tilde{\sigma}_t^{-\tilde{\kappa}_0-1-1} \exp\left(-\frac{\tilde{\lambda}_0\tilde{\kappa}_0}{\tilde{\sigma}_t}\right)}{\int_0^\infty \beta_t^{-\nu_t} \exp\left(-\frac{\beta_t}{\tilde{\sigma}_t}\right) \left(\Gamma\left(\nu_t\right)\right)^{-1} \tilde{\sigma}^{-\nu_t} \left(\tilde{\lambda}_0\tilde{\kappa}_0\right)^{\tilde{\kappa}_0+1} \left(\Gamma\left(\tilde{\kappa}_0+1\right)\right)^{-1} \tilde{\sigma}_t^{-\tilde{\kappa}_0-1-1} \exp\left(-\frac{\tilde{\lambda}_0\tilde{\kappa}_0}{\tilde{\sigma}_t}\right) d\tilde{\sigma}_t}$$

Simplifying,

$$= \frac{\tilde{\sigma}_t^{-\nu_t-\tilde{\kappa}_0-2} \exp\left(-\left(\beta_t + \tilde{\lambda}_0\tilde{\kappa}_0\right)/\tilde{\sigma}_t\right)}{\int_0^\infty \left(\tilde{\sigma}_t^{-1}\right)^{\nu_t+\tilde{\kappa}_0+2} \exp\left(-\left(\beta_t + \tilde{\lambda}_0\tilde{\kappa}_0\right)\left(\tilde{\sigma}_t\right)^{-1}\right) d\tilde{\sigma}_t}$$

Substituting $\tilde{\kappa}_t = \tilde{\kappa}_0 + K_t$ and $\tilde{\lambda}_t = \frac{\kappa_0\tilde{\lambda}_0 + K_t\bar{u}_t}{\tilde{\kappa}_t}$ and making the change of variables $b = \tilde{\lambda}_t\tilde{\kappa}_t/\tilde{\sigma}_t$ in the integral,

$$= \frac{\tilde{\sigma}_t^{-\tilde{\kappa}_t-1-1} \exp\left(-\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)/\tilde{\sigma}_t\right)}{\int_0^\infty \tilde{\sigma}_t^{-\tilde{\kappa}_t-1-1} \exp\left(-\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)/\tilde{\sigma}_t\right) d\tilde{\sigma}_t}$$

$$= \frac{\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)^{\tilde{\kappa}_t+1} \tilde{\sigma}_t^{-\tilde{\kappa}_t-1-1} \exp\left(-\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)/\tilde{\sigma}_t\right)}{\int_0^\infty b^{\tilde{\kappa}_t+1-1} \exp\left(-b\right) db}$$

The integral in the denominator is equal to the complete Gamma function $\Gamma\left(\tilde{\kappa}_t + 1\right)$,

$$= \frac{\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)^{\tilde{\kappa}_t+1} \tilde{\sigma}_t^{-\tilde{\kappa}_t-1-1} \exp\left(-\left(\tilde{\lambda}_t\tilde{\kappa}_t\right)/\tilde{\sigma}_t\right)}{\Gamma\left(\tilde{\kappa}_t+1\right)}$$

This is the p.d.f. for the $\mathit{Inv\text{-}Ga}\left(\tilde{\kappa}_t + 1, \tilde{\kappa}_t\tilde{\lambda}_t\right)$, or equivalently, $\mathit{Inv\text{-}Ga}(\tilde{\kappa}_0 + K_{t-1} + 1, \kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1})$ distribution. The result holds for future periods $t > 1$ with $\tilde{\kappa}_0 = \tilde{\kappa}_{t-1}$ and $\tilde{\lambda}_0 = \tilde{\lambda}_{t-1}$ in the equations above. $\square$

Claim 6. The marginal posterior predictive distribution of $\overline{u}_t$ conditional on K_t is

$$p\left(\overline{u}_t|K_t, K_{t-1}, \overline{u}_{t-1}, \tilde{\lambda}_0, \tilde{\kappa}_0\right) =$$

$$\begin{cases} \frac{K_t\left(\frac{K_t\overline{u}_t - K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0\tilde{\lambda}_0 + K_t\overline{u}_t}\right)^{K_t-K_{t-1}}\left(\frac{\tilde{\kappa}_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0\tilde{\lambda}_0 + K_t\overline{u}_t}\right)^{\tilde{\kappa}_0+K_{t-1}+1}}{(K_t\overline{u}_t - K_{t-1}\overline{u}_{t-1})\, B(\tilde{\kappa}_0 + K_{t-1}+1, K_t - K_{t-1})}, & K_t > K_{t-1} \\ \delta_{\overline{u}_{t-1}}\left(\overline{u}_t\right), & K_t = K_{t-1} \end{cases}$$

Proof. When $K_t = K_{t-1}$, no new utilities have been observed. Thus the average utility per bit received does not change. When $K_t > K_{t-1}$, the utility received is β_t, which has a $Ga\left(K_t - K_{t-1}, \tilde{\sigma}_t\right)$ distribution. $\tilde{\sigma}_t$ is a random variable with an $Inv\text{-}Ga\left(\tilde{\kappa}_t + 1, \tilde{\kappa}_t\tilde{\sigma}_t\right)$ distribution, and needs to be integrated over. Thus,

$$p\left(\beta_t|K_t, K_{t-1}, \overline{u}_{t-1}, \tilde{\lambda}_0, \tilde{\kappa}_0\right) =$$

$$= \int \frac{\beta_t^{K_t-K_{t-1}-1}\exp\left(-\beta_t/\tilde{\sigma}_t\right)}{\Gamma\left(K_t-K_{t-1}\right)\tilde{\sigma}_t^{K_t-K_{t-1}}} \cdot \frac{\left(\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{\tilde{\kappa}_{t-1}+1}\exp\left(-\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}/\tilde{\sigma}_t\right)}{\tilde{\sigma}^{\tilde{\kappa}_{t-1}+2}\Gamma\left(\tilde{\kappa}_{t-1}+1\right)} d\tilde{\sigma}_t$$

$$= \frac{\beta_t^{K_t-K_{t-1}-1}\left(\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{\tilde{\kappa}_0+K_{t-1}+1}}{\Gamma\left(K_t-K_{t-1}\right)\Gamma\left(\tilde{\kappa}_0+K_{t-1}+1\right)} \int \exp\left(-\left(\beta_t + \tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)/\tilde{\sigma}_t\right)\tilde{\sigma}_t^{-K_t-\tilde{\kappa}_0-2} d\tilde{\sigma}_t$$

Perform a change of variables inside the integral,

$$= \frac{\beta_t^{K_t-K_{t-1}-1}\left(\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{\tilde{\kappa}_0+K_{t-1}+1}\left(\beta_t+\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{-K_t-\tilde{\kappa}_0-1}}{\Gamma\left(K_t-K_{t-1}\right)\Gamma\left(\tilde{\kappa}_0+K_{t-1}+1\right)} \int \exp\left(-b\right) b^{K_t+\tilde{\kappa}_0} db$$

$$= \frac{\beta_t^{K_t-K_{t-1}-1}\left(\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{\tilde{\kappa}_0+K_{t-1}+1}\left(\beta_t+\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{-K_t-\tilde{\kappa}_0-1}}{\Gamma\left(K_t-K_{t-1}\right)\Gamma\left(\tilde{\kappa}_0+K_{t-1}+1\right)} \Gamma\left(K_t+\tilde{\kappa}_0+1\right)$$

$$= \frac{\beta_t^{K_t-K_{t-1}}\left(\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{\tilde{\kappa}_0+K_{t-1}+1}\left(\beta_t+\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}\right)^{-K_t+K_{t-1}-K_{t-1}-\tilde{\kappa}_0-1}}{\beta_t B\left(K_t-K_{t-1}, \tilde{\kappa}_0+K_{t-1}+1\right)}$$

$$= \frac{\left(\frac{\beta_t}{\beta_t+\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}}\right)^{K_t-K_{t-1}}\left(\frac{\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}}{\beta_t+\tilde{\kappa}_{t-1}\tilde{\lambda}_{t-1}}\right)^{\tilde{\kappa}_0+K_{t-1}+1}}{\beta_t B\left(K_t-K_{t-1}, \tilde{\kappa}_0+K_{t-1}+1\right)}$$

Now perform the change of variables $\beta_t = \overline{u}_t K_t - \overline{u}_{t-1}K_{t-1}$:

$$= \frac{K_t\left(\frac{\overline{u}_tK_t-\overline{u}_{t-1}K_{t-1}}{\overline{u}_tK_t+\tilde{\kappa}_0\tilde{\lambda}_0}\right)^{K_t-K_{t-1}}\left(\frac{\tilde{\kappa}_0\tilde{\lambda}_0+\overline{u}_{t-1}K_{t-1}}{\overline{u}_tK_t+\tilde{\kappa}_0\tilde{\lambda}_0}\right)^{\tilde{\kappa}_0+K_{t-1}+1}}{\left(\overline{u}_tK_t-\overline{u}_{t-1}K_{t-1}\right)B\left(K_t-K_{t-1}, \tilde{\kappa}_0+K_{t-1}+1\right)}$$

□

Claim 7. The expected value of $\overline{u}$, when $K_t > K_{t-1}$, is

$$\mathbb{E}\left[\overline{u}_t|\overline{u}_{t-1}, K_{t-1}, K_t > K_{t-1}, \tilde{\kappa}_0, \tilde{\lambda}_0\right] = \frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\overline{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}$$

Proof. Since $\beta_t = \overline{u}_tK_t - \overline{u}_{t-1}K_{t-1}$, we would like to find the expected value of $\overline{u}_t = \frac{\beta_t+\overline{u}_{t-1}K_{t-1}}{K_t}$. When $K_t > K_{t-1}$, the expected value of β_t is $\left(K_t-K_{t-1}\right)\tilde{\sigma}_t$. Therefore, conditional on $\tilde{\sigma}_t$, the expected value of

$\overline{u}_t$ is

$$\mathbb{E}\left[\overline{u}_t|\tilde{\sigma}_t, K_t, K_{t-1}, \overline{u}_{t-1}\right] = \frac{\tilde{\sigma}_t\left(K_t - K_{t-1}\right) + \overline{u}_{t-1}K_{t-1}}{K_t}$$
$$= \tilde{\sigma}_t\left(1 - \frac{K_{t-1}}{K_t}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{K_t}\right)$$

This is a weighted average of the average utility per bit for the first K_{t-1} bits and the expected utility per bit for the next $K_t - K_{t-1}$ bits. Integrating over $\tilde{\sigma}_t$, which has an expected value of $\tilde{\lambda}_{t-1}$, we get

$$\mathbb{E}\left[\overline{u}_t|\overline{u}_{t-1}, K_{t-1}, K_t > K_{t-1}, \tilde{\kappa}_0, \tilde{\lambda}_0\right] = \tilde{\lambda}_{t-1}\left(1 - \frac{K_{t-1}}{K_t}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{K_t}\right)$$
$$= \frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\left(1 - \frac{K_{t-1}}{K_t}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{K_t}\right)$$

□

Claim 8. The expected utility from information at step t is

$$\mathbb{E}\left[\beta_t|K_{t-1}, \overline{u}_{t-1}, H_{t-1}\right] = \left(\left(N - K_{t-1}\right)\frac{\tilde{\alpha}_0}{\tilde{\alpha}_0 + H_{t-1} + 1}\right)\left(\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)$$

Proof. We start by noting $\beta_t = K_t\overline{u}_t - K_{t-1}\overline{u}_{t-1}$. Thus,

$$\mathbb{E}\left[\beta_t|K_{t-1}, \overline{u}_{t-1}, H_{t-1}\right] = \mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right]\mathbb{E}\left[\overline{u}_t|\mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right], \overline{u}_{t-1}\right] - K_{t-1}\overline{u}_{t-1}$$

Expanding $\mathbb{E}\left[\overline{u}_t|\mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right], \overline{u}_{t-1}\right]$ from Equation (12), we get

$$= \mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right]\left\{\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\left(1 - \frac{K_{t-1}}{\mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right]}\right) + \overline{u}_{t-1}\left(\frac{K_{t-1}}{\mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right]}\right)\right\} - K_{t-1}\overline{u}_{t-1}$$
$$= \left(\mathbb{E}\left[K_t|K_{t-1}, H_{t-1}\right] - K_{t-1}\right)\left(\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)$$
$$= \left(\left(N - K_{t-1}\right)\frac{\tilde{\alpha}_0}{\tilde{\alpha}_0 + H_{t-1} + 1}\right)\left(\frac{\kappa_0\tilde{\lambda}_0 + K_{t-1}\bar{u}_{t-1}}{\tilde{\kappa}_0 + K_{t-1}}\right)$$

□

Claim 9. The posterior distribution of $\tilde{\pi}_b$ takes the following form:

$$\tilde{\pi}_b|k_{t-1,b} = 0, \alpha_{t-1} \sim Beta\left(\tilde{\alpha}_0, 1 + \alpha_{t-1}\right)$$

Proof. The prior distribution of $\tilde{\pi}_b$ is $Beta\left(\tilde{\alpha}_0, 1\right)$, and the likelihood of not observing bit b at the first $t-1$ blogs is

$$\left(1 - \tilde{\pi}_b\right)^{\alpha_{a_1}} \cdots \left(1 - \tilde{\pi}_b\right)^{\alpha_{a_{t-1}}} = \left(1 - \tilde{\pi}_b\right)^{\alpha_{t-1}}$$

The posterior distribution of $\tilde{\pi}_b$ at step t is therefore

$$\frac{(1-\tilde{\pi}_b)^{\alpha_{t-1}}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}\,(1-\tilde{\pi}_b)^{1-1}\,[B(\tilde{\alpha}_0,1)]^{-1}}{\int_0^1 (1-\tilde{\pi}_b)^{\alpha_{t-1}}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}\,(1-\tilde{\pi}_b)^{1-1}\,[B(\tilde{\alpha}_0,1)]^{-1}\,d\tilde{\pi}_b}$$
$$=\frac{(1-\tilde{\pi}_b)^{\alpha_{t-1}+1-1}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}}{\int_0^1 (1-\tilde{\pi}_b)^{\alpha_{t-1}+1-1}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}d\tilde{\pi}_b}$$

The denominator is equal to the complete Beta function $B(\tilde{\alpha}_0,\alpha_{t-1}+1)$, so the posterior distribution is

$$\frac{(1-\tilde{\pi}_b)^{\alpha_{t-1}+1-1}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}}{B(\tilde{\alpha}_0,\alpha_{t-1}+1)}$$

which is the p.d.f. of the $Beta(\tilde{\alpha}_0,\alpha_{t-1}+1)$ distribution. □

Claim 10. The conditional distribution of K_t is

$$p(K_t|K_{t-1},\alpha_{t-1},a_t=j)= Bin\left(K_t-K_{t-1}|N-K_{t-1},1-\frac{B(\tilde{\alpha}_0,1+\alpha_{t-1}+\alpha_j)}{B(\tilde{\alpha}_0,1+\alpha_{t-1})}\right).$$

Proof. The conditional probability of finding bit b at blog j is

$$(1-(1-\tilde{\pi}_b)^{\alpha_j})$$

To find the marginal probability of finding bit b, we integrate over the posterior distribution of $\tilde{\pi}_b$:

$$\begin{aligned}
\tilde{\rho}_{\alpha,jt} &= \int_0^1 (1-(1-\tilde{\pi}_b)^{\alpha_j})\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}\,(1-\tilde{\pi}_b)^{\alpha_{t-1}+1-1}\,[B(\tilde{\alpha}_0,\tilde{\alpha}_{t-1}+1)]^{-1}\,d\tilde{\pi}_b\\
&= 1-\int_0^1 (1-\tilde{\pi}_b)^{\alpha_j}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}\,(1-\tilde{\pi}_b)^{\alpha_{t-1}+1-1}\,[B(\tilde{\alpha}_0,\tilde{\alpha}_{t-1}+1)]^{-1}\,d\tilde{\pi}_b\\
&= 1-[B(\tilde{\alpha}_0,\tilde{\alpha}_{t-1}+1)]^{-1}\int_0^1 (1-\tilde{\pi}_b)^{\alpha_j+\alpha_{t-1}+1-1}\,\tilde{\pi}_b^{\tilde{\alpha}_0-1}d\tilde{\pi}_b
\end{aligned}$$

The integral is equal to the complete Beta function $B(\tilde{\alpha}_0,\alpha_j+\alpha_{t-1}+1)$, therefore

$$\tilde{\rho}_{\alpha,jt}=1-\frac{B(\tilde{\alpha}_0,\alpha_j+\alpha_{t-1}+1)}{B(\tilde{\alpha}_0,\tilde{\alpha}_{t-1}+1)}.$$

Summing across unobserved bits produces a binomial distributed variable due to the marginal probability $\tilde{\rho}_{\alpha,jt}$ being the same for all previously unobserved bits. □

3 Estimation Details

3.1 Non-Identification of α_j and α_0

The overall availability of information, α_0, cannot be separately identified from the amount of information at each site. Figure 1 demonstrates this lack of identification by plotting a continuum of α_0's and α_j's that

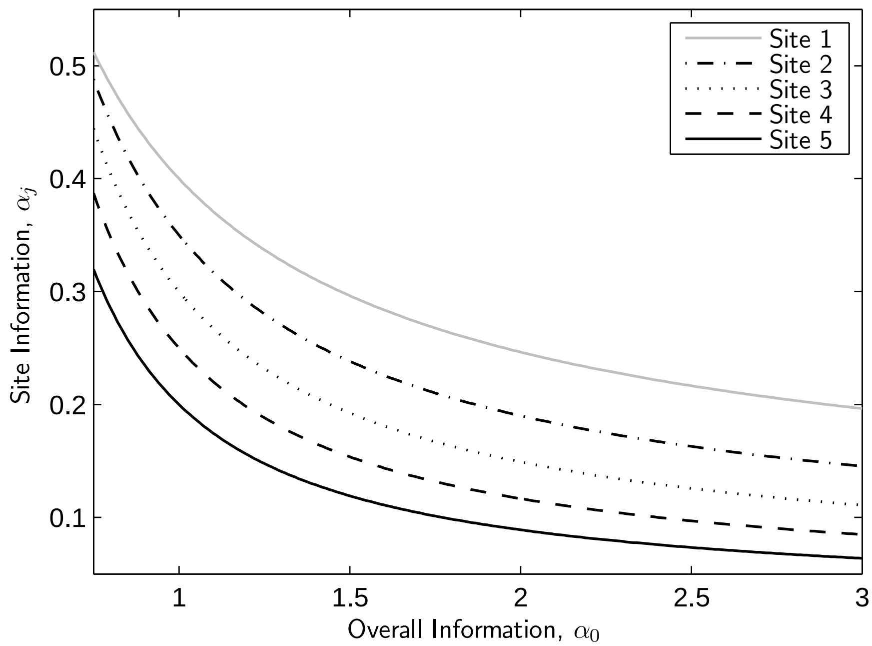

Figure 1: Non-Identification of α_0 and α_j. As α_0 varies, the α_j's adjust to produce the same browsing behavior.

produce identical amounts of information at each site. Each line represents the α_j for one of five sites; as α_0 varies along the x-axis, the α_j's can compensate for the change in α_0 in order to generate identical browsing behavior.

3.2 Sampling Algorithm

The general sampling procedure is outlined in Algorithm 1. In the sections that follow, we discuss the details of this procedure, as well as technical issues concerning reparametrization, unobserved state variables, value function iteration, and proposal distributions.

3.3 Reparametrization and Data Augmentation

The unobserved state variables at each choice occasion are defined in §2.5 as $\mathcal{X} = (K, \overline{u}, h, \{\overline{s}_j\}, \{n_j\})$. The amount of information, K, the average utility per bit, $\overline{u}$, and the average signal value for each site, $\overline{s}_j$ are unobserved (our data include the number of links to each site, n_j). We must integrate over these unobserved state variables, and do so via the standard approach of data augmentation (Tanner and Wong, 1987). As such, we sample these states along with the model primitives; we subsequently integrate over them numerically.

In order to improve the efficiency of our sampler, we employ two reparametrizations. First, we substitute σ with $\sigma^* \equiv \sigma\lambda_0^{-1}$, and $\overline{u}$ with $\overline{u}^* = \overline{u}\lambda_0^{-1}$. Accordingly, the prior distribution of σ^* is $\mathit{Inv\text{-}Ga}\,(\kappa_0 + 1, \kappa_0)$,

Algorithm 1: MCMC sampling procedure. At each iteration, parameters are sampled in blocks. The value function is then iterated and the result either replaces the oldest saved iteration or is appended to the set of saved iterations.

initialize saved MCMC samples: Θ
saved value function iterations: $\mathcal{W}$

foreach *MCMC iteration* t **do**

- **foreach** *parameter block* $\theta \equiv \theta_b^{(t-1)}$ **do**
  - Propose new θ using mMALA proposal distribution:
    - **calculate** marginal posterior probability: $p(\theta|\mathcal{W})$
      derivatives of log posterior probability: $\mathcal{D}_\theta$ (Algo. 3)
    - **set** $(\mu, \Sigma) \leftarrow f(\theta, \mathcal{D}_\theta)$
    - **draw** $\theta^c \leftarrow \mathcal{N}(\mu, \Sigma)$
  - Maintain detailed balance:
    - **calculate** $p(\theta^c|\mathcal{W})$ and $\mathcal{D}_{\theta^c}$ (Algorithm 3)
    - **set** $(\mu^\circ, \Sigma^\circ) \leftarrow f(\theta^c, \mathcal{D}_{\theta^c})$
  - Accept or reject proposal:
    - **set** $\alpha \leftarrow \frac{p(\theta^c)\mathcal{N}(\theta|\mu^\circ,\Sigma^\circ)}{p(\theta)\mathcal{N}(\theta^c|\mu,\Sigma)}$
    - **draw** $u \leftarrow Unif(0,1)$
    - **if** $u < \alpha$ **then set** $\theta_b^{(t)} \leftarrow \theta^c$ **else set** $\theta_b^{(t)} \leftarrow \theta$
- Iterate value function using IJC:
  - **draw** $\mathcal{X} \leftarrow p(\mathcal{X}|\theta^{(t)})$
  - **calculate** $\widehat{W} \leftarrow f(\mathcal{X}, \theta^{(t)})$ using IJC (Algorithm 2)
- Save parameters and value function:
  - **append** $\mathcal{W} \leftarrow \{\widehat{W}, \mathcal{X}, \theta^{(t)}\}$
  - **append** $\Theta \leftarrow \theta^{(t)}$

and it follows (c.f. §2, Claim 6) that the posterior distribution of $\overline{u}$ is

$$p\left(\overline{u}_t^* | K_t, K_{t-1}, \overline{u}_{t-1}^*, \kappa_0\right) =$$

$$\begin{cases} \dfrac{K_t \left(\frac{K_t \overline{u}_t^* - K_{t-1} \overline{u}_{t-1}^*}{\kappa_0 + K_t \overline{u}_t^*}\right)^{K_t - K_{t-1}} \left(\frac{\kappa_0 + K_{t-1} \overline{u}_{t-1}^*}{\tilde{\kappa}_0 + K_t \overline{u}_t^*}\right)^{\kappa_0 + K_{t-1} + 1}}{\left(K_t \overline{u}_t^* - K_{t-1} \overline{u}_{t-1}^*\right) B(\kappa_0 + K_{t-1} + 1, K_t - K_{t-1})}, & K_t > K_{t-1} \\ \delta_{\overline{u}_{t-1}^*}\left(\overline{u}_t^*\right), & K_t = K_{t-1}. \end{cases} \tag{1}$$

Note that this distribution depends only on the state variable K. Without the reparametrization, $\overline{u}$ also depends on λ_0, resulting in highly correlated MCMC draws. In the second reparametrization, we center the $\overline{s}$'s by defining $s_{jk,d}^* \equiv \left(s_{j,k,d} - z_j - \nu_{jd}\right) \tau_s^{-\frac{1}{2}}$; the distribution of $s_{jk,d}^*$ is therefore standard normal. We enforce the restrictions $\mathbb{E}\left(s_{jk,d}^*\right) = 0$ and $\mathbb{V}\left(s_{jk,d}^*\right) = 1$ (see §4.2 in) via pairwise sampling of the s^*'s, using the method of Musalem et al. (2009). A parallel strategy is used to sample the ν_{jd}'s.

3.4 Expected Value Function Calculation

Using IJC to estimate the expected value, or emax, function (Equation 21) greatly improves the computational efficiency of our MCMC sampler. In the standard Bayesian approach to estimating dynamic discrete choice models, the contraction mapping defined by the emax function (21) must be fully iterated (i.e., until it converges) at each iteration of the MCMC sampler. IJC's method uses calculations performed at prior MCMC iterations to mitigate the need to solve the expected value function in its entirety. Specifically, the emax function is iterated once at each MCMC iteration, and the result of this operation is saved alongside the MCMC chain. As sampling progresses, both the MCMC chain and the saved value function iterations converge to their invariant, joint distribution. In this way, the contraction mapping defined by the expected value function is iterated to convergence simultaneously with the model parameters.

An estimate of the expected value of future browsing, conditional on the current state and on choosing a particular option (the integral in Equation 20), is needed in order to calculate the model likelihood and iterate the emax function. This estimate is obtained as a weighted average of value function iterations generated at earlier points in the MCMC chain. (This procedure is detailed in Algorithm 2.) Given a state $\mathcal{X}$, parameter vector θ, and choice j, the weight assigned to a previous estimate of the expected value function, $\widehat{W}'$, is proportional to the probability density $p\left(\mathcal{X}' | \mathcal{X}, \theta, j\right) p(\theta' | \theta)$.

The first factor in this expression, $p\left(\mathcal{X}' | \mathcal{X}, \theta, j\right)$ represents the conditional probability of transitioning from state $\mathcal{X}$ (the one at which we wish to estimate the emax function) to the state $\mathcal{X}'$ from which the saved emax function iteration $\widehat{W}'$ was generated, assuming option j is chosen. In order to calculate this probability, we must have access not only past iterations of the emax function, but also the states that were used to generate them. Note that in practice, the states used to iterate the emax function are chosen at random. The second factor in the probability density, $p\left(\theta' | \theta\right)$, represents the "distance" between the current parameter vector, θ, and the parameter vector θ' under which the saved value function iteration was generated; this distance is typically expressed as a probability density function via Gaussian kernel smoothing. The distance between parameter vectors is included in the weighting density function because parameters are part of the choice state that determines the expected value of future browsing.[1] In total, the

[1] Hence, only choice-relevant parameters, and not, for example, variables controlling their prior distributions, need to be saved.

Algorithm 2: Value Function Approximation using IJC. The estimated value function given choice j is a weighted average of past iterations of the value function. The weight assigned to each iteration depends on 1) the probability of transitioning from the current state to the state used to generate the prior iteration; and 2) the kernel-smoothed distance between the current parameters and the parameters used to generate the prior iteration.

```
input: parameters: θ
       state: 𝒳
       saved iterations: 𝒲
output: estimated value function: Ŵ

initialize Ŵ_j ← 0 for all j
foreach unvisited site j do
    initialize (ŵ_j, Π_j) ← 0
    foreach saved value function iteration s do
        set (Ŵ_s, 𝒳_s, θ_s) ← 𝒲_s
        calculate state transition probability: p(𝒳_s|𝒳, θ, j)
                  kernel-smoothed distance: d(θ, θ_s)
        Build the weighted average:
            set π_js ← p(𝒳_s|𝒳, θ, j) × d(θ, θ_s)
            set ŵ_j ← ŵ_j + π_js Ŵ_s                    // numerator
            set Π_j ← Π_j + π_js                        // denominator
    set Ŵ_j ← ŵ_j / Π_j
return Ŵ
```

joint density function $p\left(\mathcal{X}'|\mathcal{X},\theta,j\right)p(\theta'|\theta)$ places the greatest weight on previous value function iterations that 1) were generated from parameters that are similar to the current parameters, θ; and 2) were generated from the states that are most likely to be reached when option j is chosen in state $\mathcal{X}$.

One computational advantage of using IJC is that value function approximation takes place in the regions of the state space that have the greatest posterior density. Contrast this approach with that of discretizing the state space into a grid of points, irrespective of the likelihood of the states they represent, which can waste computational resources calculating the value function in regions of the state space that will almost never be visited.

Another advantage of using IJC is that it greatly facilitates estimation of models that include 1) state variables that follow a continuous distribution, and 2) state variables that evolve deterministically. Indeed, our model has state variables with transition densities that are a mixture of the two (e.g., equation 1). When a continuous state variable $\mathcal{C}$ transitions deterministically to some new value $\mathcal{C}^*$, no weight would be placed on saved iterations that were not generated from state $\mathcal{C}^*$. The reason for this is that the unobserved states used for iteration are randomly sampled, and thus the probability of finding a saved iteration generated from state $\mathcal{C}^*$ is effectively zero. We avoid this problem by using a Gaussian kernel to approximate deterministic state transitions. For example, the weight assigned to a previous iteration generated in state $\mathcal{C}'$ will be proportional to $\exp\left(-\frac{1}{2}\left(\mathcal{C}'-\mathcal{C}^*\right)^2 h^{-2}\right)$, where h is a bandwidth parameter. Accordingly, expected value function iterations generated from states close to the new state $\mathcal{C}^*$ will receive the highest weight.

There is a connection between using kernel-smoothed densities to approximate deterministic state transitions on the one hand, and using them to account for the distance between parameter vectors on the other. A parameter is, in essence, an unobserved, continuous state variable that transitions deterministically to a single, constant value. The IJC weighting function can therefore be written as $p(\mathcal{X}'|\mathcal{X}, j)$, provided $\mathcal{X}$ is redefined to include θ with $p(\theta'|\theta, j) = 1$ if $\theta' = \theta$ (and 0 otherwise). Next, we explain how we leverage this property of IJC to reduce the effective size of our state space during estimation.

3.5 State Space

The state space for our model, $\mathcal{X} = (K, \overline{u}, h, \{\overline{s}_j\}, \{n_j\})$, has a relatively high dimension owing to the inclusion of site-specific state variables: the history vector, h, and the average level and number of signals, $\overline{s}$ and n, are each $J \times 1$ vectors. In a model with $J = 5$ blogs, there are 17 state variables to keep track of. Moreover, there are 3 parameters for each consumer (λ_i, γ_i, and v_i), 2 for each blog (α_j and z_j), and finally, τ_s (we ignore the $\omega_{j,k}$'s for the moment and discuss them below).[2] Thus, there are 31 variables to account for when computing the weights used to estimate the emax function: 17 state variables plus 14 parameters. We reparametrize the state space in such a way as to lower the dimension of the state points used to estimate the value function.

Our general approach is to embed the information in the history vector h in the site-level parameters and state variables. This approach exploits the fact that the *indexes* of the sites that have already been visited (and conversely, the indexes of those that have not) do not affect the value of the emax function. Thus, the vectors storing the site-specific state variables or parameters—e.g., average signal quality for each site, $\overline{s}$, or the average location, z—are defined at each step t as $(J - t + 1) \times 1$ vectors whose elements correspond to the $J - t + 1$ sites that have not been visited. Elements of these vectors are ordered by their original indexes. For example, if at step $t = 2$, site $j = 3$ has already been visited, then the vector representing z contains the elements (z_1, z_2, z_4, z_4), in precisely that order (and likewise for the other vectors). In this way, the number of parameters that need to be saved alongside each value function iteration is reduced: we only need to store 12 parameter values—3 consumer-level parameters (λ_i, γ_i, and v_i), $(J - 1) \times 2$ blog parameters (α_j and z_j), and τ_s—plus 11 state variables—K, $\overline{u}$, α_t, and the $(J - 1) \times 2$ link-related state variables ($\overline{s}$ and n)—23 variables in all. (In practice, we find also storing H improves the quality of the estimates.)

As mentioned earlier, we do not store the values of the site link probabilities ($\omega_{j,k}$'s), even though these variables affect the value of the emax function. We can avoid having to use these parameters to estimate the expected value function for two reasons. First, we do not estimate the $\omega_{j,k}$'s and hence there is no variation across MCMC iterations to account for via kernel-smoothing. Second, we can recover the identities of the sites that have not yet been visited, as long as we do not have $z_j = z_k$ and $\alpha_j = \alpha_k$ for any two sites j and k (in our MCMC sampler, this event occurs with probability 0). For these reasons, all information about link probabilities is already accounted for in the other parameters and states.

3.6 Efficient Proposal Distributions

Although reducing the size of the IJC state space improves the efficiency of our sampling procedure, we seek to further improve the quality of the candidate parameter draws in the Metropolis-Hastings accept/reject step of IJC (see Algorithm 1). We accomplish this by using the Riemann Manifold Metropolis Adjusted

[2] Imai et al. (2009) and Ching et al. (2010) discuss the need to save only one consumer's parameters.

Algorithm 3: Calculation of Posterior Probability. The state variable $\mathcal{X}$ is the same for all choices at step $t = 1$, hence the value function only needs to be evaluated once for each consumer at step $t = 1$. Derivatives of the log posterior density function are calculated via automatic differentiation.

input: parameters: θ
output: log posterior probability: $\mathcal{L}(\theta) = \log p(\theta)$
first three derivatives of log posterior: $\mathcal{D}_\theta = \{\nabla_\theta^d \mathcal{L}(\theta)\}_{d=1}^3$

initialize $L \leftarrow 0$

Browsing likelihood:
set initial state $\mathcal{X}_1 \leftarrow (\overline{u} = 0, K = 0, s = 0, n = 0, h = 0)$
foreach *consumer* i **do**
- Likelihood of choices at step $t = 1$:
- **set** $\widehat{W}_{i,1} \leftarrow f(\mathcal{X}_1, \theta)$ using IJC (Algorithm 2)
- **set** $V_{ij,1} \leftarrow \mathbb{E}\left[U_{ij}\left(\mathcal{X}_1, \theta\right)\right] + \widehat{W}_{ij,1}$ for each site j
- **set** $\ell_i \leftarrow -D_i \log\left(\sum_j \exp(V_{ij,1})\right)$ `// Logit denominator at step` $t = 1$ `on all days`
- **foreach** *day* d **do**
  - **set** $\ell_i \leftarrow \ell_i + V_{ij,1}$ for $j = a_{id,1}$ `// Numerator, step` $t = 1$`, day` d
- Likelihood of choices on step $t > 1$:
- **foreach** *day* d **do**
  - **for** $t = 2$ *to* $T_{i,d}$ **do**
    - **set** $\mathcal{X}_t$ from data-augmented θ, and the browsing and link data
    - **set** $\widehat{W}_t \leftarrow f(\mathcal{X}_t, \theta)$ using IJC (Algorithm 2)
    - **set** $V_{ij,t} \leftarrow \mathbb{E}\left[U_{ij}\left(\mathcal{X}_t, \theta\right)\right] + \widehat{W}_{ij,t}$ for each unvisited site j
    - **set** $\ell_i \leftarrow -\log\left(\sum_j \exp(V_{ij,t})\right)$ `// Denominator, step` t`, day` d
    - **set** $\ell_i \leftarrow V_{ij,t}$ for $j = a_{id,t}$ `// Numerator, step` t`, day` d
  - **set** $\ell_i \leftarrow \ell_i + p(\mathcal{X}_t|\theta)$
- Prior distribution of consumer i's parameters:
- **set** $\ell_i \leftarrow \ell_i + \log p(\lambda_{0,i}, \gamma_i, v_i)$
- **set** $L \leftarrow L + \ell_i$

Prior distribution:
set $\mathcal{L}(\theta) \leftarrow L + \log p(z, \alpha, \tau_s)$
set $\mathcal{D}_\theta$ using automatic differentiation
return $(\mathcal{L}(\theta), \mathcal{D}_\theta)$

Langevin Algorithm (mMALA) of Girolami and Calderhead (2011). mMALA uses the first, second, and third derivatives of the log-posterior density function to construct a proposal distribution. The gradient and higher-order tensors provide a deterministic component of the proposal distribution—candidate parameters are drawn from a normal distribution centered over regions of relatively higher density—whereas the second derivative provides a position-specific covariance matrix that is locally similar to that of the target distribution. mMALA thus generates proposal draws with relatively high acceptance rates and relatively low autocorrelation, both of which improve the efficiency of the sampler.

The gradient, Hessian, and derivative of the Hessian of the log-probability function are inputs to this procedure, but these tensors do not have closed analytical forms in a dynamic discrete choice model. Thus, we generate these derivatives through a numerical procedure known as automatic differentiation (also referred to as AD; Griewank et al., 1996; Su and Judd, 2010). Automatic differentiation provides exact values for the derivatives of a function (as opposed to the approximations provided by other numerical techniques, such as the method of finite differences). AD works by repeatedly applying the chain rule to the elementary operations (addition, multiplication, etc.) that comprise the function. One of the advantages of AD is that it requires very few changes be made to the original source code. AD libraries typically provide an object that replaces the standard single- or double-precision floating point type. While the function of interest is being executed using these replacement floating point types, the AD library records every operation into memory. Once the function has completed, the derivatives can be obtained from the saved information.

We approximate the derivatives of the log-posterior function, inasmuch as we do not use automatic differentiation when calculating the estimated value function via IJC. This approximation produces greater numerical stability (IJC estimates of the value function can be quite non-smooth in their arguments, especially earlier in the estimation procedure) and faster execution (the cost of calculating the derivatives increases with the number of operations performed).

4 Counterfactual Simulation under IJC

Our goal in running counterfactual analyses is to evaluate differences in browsing behavior and consumer welfare under two policy regimes. Our approach is to simulate browsing under two scenarios: In the baseline scenario, the model is parametrized using the results of our estimation procedure, whereas in the counterfactual scenario, these results are perturbed in some way. For example, in the fair use counterfactual scenario, the value of τ_s, representing link precision, is multiplied by a constant factor. We repeatedly simulate browsing under each scenario—once for every MCMC sample used in the counterfactual analysis.

We calculate the expected change in browsing activity and welfare from enacting the counterfactual policy. Expectations of these changes are formed assuming a risk neutral loss function, hence we report the median rather than the mean.

Before we can simulate browsing for a given parameter vector, we need to generate an estimate of the emax function. We cannot rely on the saved emax iterations used during estimation to approximate the emax function under the counterfactual scenarios for at least two reasons. First, the emax iterations generated during estimation come from the posterior distribution of the model primitives, not the distribution of the primitives under a counterfactual scenario. Second, the counterfactual scenario may require estimates of the emax function at parameters values falling well outside the range of values encountered during estimation. For example, in the fair use counterfactual analysis, the posterior distribution of τ_s is concentrated around a value somewhere in the range of 0.2; but in the counterfactual scenario, we divide this number by 20.

In order to generate an estimate of the emax function, then, we use IJC's method to generate a new value function approximation once for each parameter vector. That is, for each MCMC sample, we simulate a Markov chain in which the parameters evolve deterministically to a constant value; at each iteration of this chain, we iterate the emax function just as we do during estimation (this procedure is described in more detail in Algorithm 4). We repeat this step for both the baseline and counterfactual simulations to ensure the simulation error is approximately the same in both conditions.

Algorithm 4: Counterfactual simulation procedure.

foreach *MCMC sample* $\theta^{(t)}$ **do**
- **set** $\theta \leftarrow \theta^{(t)}$
- **simulate** browsing
- **set** $\mathcal{H}_t \leftarrow$ simulated browsing
- **set** $\mathcal{U}_t \leftarrow$ welfare measures
- **set** $\theta \leftarrow f(\theta^{(t)})$ `// Counterfactual intervention`
- **simulate** browsing
- **set** $\mathcal{H}_t^* \leftarrow$ simulated browsing
- **set** $\mathcal{U}_t^* \leftarrow$ welfare measures
- **set** $\mathcal{T}_t \leftarrow g(\mathcal{H}_t, \mathcal{H}_t^*, \mathcal{U}_t, \mathcal{U}_t^*)$ `// Test statistics`

calculate $\mathbb{E}(\mathcal{H},\mathcal{U})$ using the $\mathcal{H}_t$'s and $\mathcal{U}_t$'s
calculate $\mathbb{E}(\mathcal{H}^*,\mathcal{U}^*)$ using the $\mathcal{H}_t^*$'s and $\mathcal{U}_t^*$'s
calculate differences in expected browsing and welfare, $\mathbb{E}(\mathcal{H},\mathcal{U}) - \mathbb{E}(\mathcal{H}^*,\mathcal{U}^*)$
calculate expected value of test statistics, $\mathbb{E}(\mathcal{T})$ using the $\mathcal{T}_t$'s

4.1 Calculating Welfare Effects from Counterfactual Experiments

The change in welfare for each consumer equals to the change in total utility minus the change in total costs, $dU - dC$. Utility and costs can be equated to income through the prior distributions of λ_0 and γ, hence the change in income (denoted I) can be expressed in the following way.

$$dI \approx \frac{\partial I}{\partial U} dU + \frac{\partial I}{\partial C} dC$$

Expanding these expressions using the chain rule produces

$$dI \approx \frac{\partial I}{\partial \log \lambda_0} \frac{\partial \log \lambda_0}{\partial \lambda_0} \frac{\partial \lambda_0}{\partial U} dU + \frac{\partial I}{\partial \log \gamma} \frac{\partial \log \gamma}{\partial \gamma} \frac{\partial \gamma}{\partial C} dC.$$

The expression $\partial \log \lambda_0 / \partial I$ is defined in the prior distribution of λ_0, and is equal to $-\phi_{\lambda,I}$ when income changes from "less than \$60,000" to "more than \$60,000." We estimate the median income change in this range to be \$55,000, hence $\partial I/\partial \log \lambda_0 = -1/(\$55k\, \phi_{\lambda,I})$. The expression $\partial U/\partial \lambda_0$ is equal to the total number of bits accumulated in the baseline scenario, whereas the expression $\partial C/\partial\gamma$ is equal to the total number of sites visited in the baseline scenario. Making these substitutions yields the following expression.

$$dI \approx \frac{1}{\$55k\,\phi_{\lambda,I}\lambda_0 \sum_d K_{T_d}} dU + \frac{1}{\$55k\,\phi_{\gamma,I}\gamma \sum_d H_{T_d}} dC$$

References

Aguirregabiria, Victor, and Pedro Mira. 2010. Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1):38–67.

Ching, Andrew, Susumu Imai, Mazakazu Ishihara, and Neelam Jain. 2010. A practitioner's guide to bayesian estimation of discrete choice dynamic programming models. Working paper.

Girolami, Mark, and Ben Calderhead. 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(2):123–214.

Griewank, Andreas, David Juedes, and Jean Utke. 1996. Algorithm 755: ADOL-C: A package for the automatic differentiation of algorithms written in C/C++. *ACM Transactions on Mathematical Software* 22(2):131–167.

Imai, Susumu, Neelam Jain, and Andrew Ching. 2009. Bayesian estimation of dynamic discrete choice models. *Econometrica* 77(6):1865–1899.

Musalem, A., E.T. Bradlow, and J.S. Raju. 2009. Bayesian estimation of random-coefficients choice models using aggregate data. *Journal of Applied Econometrics* 24(3):490–516.

Su, Che-Lin, and Kenneth L. Judd. 2010. Constrained optimization approaches to estimation of structural models. Working paper.

Tanner, M.A., and W.H. Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association* 82(398):528–540.