Working Paper #08-29

October 2008

Estimating Search with Learning

Sergei Koulayev
Columbia University

Estimating search with learning

Sergei Koulayev*

Job Market Paper
PRELIMINARY VERSION

November 13, 2008

Abstract

In this paper we estimate a structural model of search for differentiated products, using a unique dataset of consumer online search for hotels. We propose and implement an identification strategy that allows us to separately estimate consumer's beliefs, search costs and preferences. Learning plays an essential role in this strategy: it creates variation of posterior beliefs across consumers that is orthogonal to the variation in search costs. We show that ignoring endogeneity of choice sets due to search may lead to significant biases in estimates of consumer demand: from 50 to more than 200 percent depending on informational assumptions. Second, the median search cost is about 25 dollars per 15 hotels; there is also a significant heterogeneity of search costs among the population. We perform a statistical test between models of search from known (Stigler 1967) and from unknown (Rothschild 1974) distribution and find that our data favors the latter: we find a statistically significant amount of Bayesian learning.

*Department of Economics, Columbia University. I'm indebted to my advisors, Michael Riordan, Bernard Salanié and Kate Ho for their continual advice and encouragment. I'm grateful to participants of Applied Micro Theory colloquium and Industrial Organization colloquium, members of Friday talks for their helpful suggestions. Special thanks for Ting Wu for his support at early stages of this research. Financial support from Program of Economic Research at Columbia University, from the NET institute (www.netinst.org) and Kauffman Foundation is gratefully acknowledged.

1 Introduction

In markets with multiple sellers and frequently changing prices, consumers often have to engage in costly search in order to collect information necessary for making a purchase. A number of theoretical models, from Stigler (1961) and Rothschild (1974) to Talmain (1992) characterized the behavior of a rational consumer in such situations: she would make a sequence of search efforts, stopping at the point where the expected benefit from another attempt falls short of the search cost. This process results in a choice set from which a purchase is ultimately made.

Inference about the consumer demand in such a context is complicated by the fact that both her preferences and her search decisions affect the likelihood of purchasing a particular product. Indeed, search decisions determine whether the product is included in the choice set of a particular consumer, and what kind of alternatives are also available. A static demand estimation, which ignores the process of formation of choice sets, would generally give biased results. We see three reasons for this. First, since search is costly, the resulting choice sets are quite limited. In contrast, a typical demand estimation that uses data on purchases and market shares, implicitly makes an assumption that consumers possess full information about available goods. Second, since the optimal stopping rule depends on consumer preferences, the choice set will be endogenous to these peferences as well. For example, if we observe a consumer who only sampled one or two goods and stopped searching, this is an indicator that she really likes what she has found. A static demand model does not account for this information and therefore makes a biased conclusion about preferences. Finally, even if consistent estimates of preferences were available, computing any quantity of interest - such as the price elasticity - requires the knowledge of search behavior. Indeed, a price perturbation affects not only the decision to buy, but also the decision to search and ultimately the composition of the choice set.

In this paper, we estimate a structural model of search for heterogeneous products, using data on searches for hotels online. We find that a static model over-estimates price elasticity by more than 200% if the full information assumption is used, and by 60% if we use information on actual choice sets. To the best of our knowledge, this is the first paper to empirically evaluate the importance of accounting for endogeneity of choice sets due to search in estimating consumer demand. In doing so, we make another step in extending the analysis of search problem to the case of differentiated goods, in the spirit of multi-attributes approach by Anderson, de Palma and Thisse (1989). This emphasis on detailed modeling of consumer behavior sets us apart from much of the existing literature on product search, which focused on explaining observed price variation with search costs, e.g., the "Diamond paradox" (Diamond (1971)). Consequently, these papers have considered the case of homogeneous products, and whenever heterogeneity was introduced, it was quite limited. See the next section for a review of that research.

From a methodological point of view, our contribution to the empirical literature on search is the identification strategy that allows to separately estimate three main components of the search process: preferences, beliefs and search costs. The inherent difficulty in doing so is that a consumer can make a search effort either because she is optimistic or because she has low search cost, and neither of them is observed. In previous research, a partial identification was achieved: a search cost distribution was identified by assuming that consumer beliefs are driven by empirical distribution of prices and, in some studies, by imposing equilibrium restrictions. Implicitly, this approach assumes that: a) consumers know the true equilibrium price distribution from which they search, and b) the econometrician, too, knows this distribution since he observes empirical density of prices. These are strong assumptions, and in this paper we attempt to relax both of them by estimating a model where consumers are uncertain about the true distribution and learn while searching (see Rothschild (1974)). For the econometrician, consumer beliefs are unobserved and estimated together with other parameters of the model. We are able to do so because in our data we observe consumer-specific information sets and search decisions. Learning imposes a restriction on their joint variation, from which beliefs are identified separately from search costs. Put simply, the set of observed prices is gives an indication of consumer optimism, that affects the decision to search independently of search costs. From the estimation point of view, this is also the first paper to introduce Bayesian learning in a model of sequential search[1]. Since a model of search from known distribution (Stigler (1961)) can be viewed as a nested version of a model with learning, we can perform a statistical test between the two theories of search. We find that the data favors the hypothesis of search with learning, at a high level of significance.

Continuing with the methodological perspective, a particular interest in estimating a search model is to see to what extent this way of thinking about search can explain the actual search patterns. Previous studies have used only purchase data, which is the final outcome of search. Therefore, they were restricted to estimating highly stylized equilibrium search models that would predict observed distribution of prices and market shares, which is one step removed from the actual consumer behavir. In this paper our detailed data enables us to focus on explaining search patterns themselves, and within a model that explicitly accounts for specifics of the actual search environment. This gives us more confidence in our conclusions. At the same time, this approach is much more data-intensive and computationally costly, which limits the scope of search behavior we can explain.

The estimation is performed on a unique dataset of searches for hotels in Chicago, May 2007, at a popular website www.kayak.com. Although this website has many different search

[1] As distinguished from literature on the search for experience goods, such as Crawford and Shum (2006), or marketing literature on consideration set formation, see Roberts and Lattin (1997), where Bayesian learning has long been an integral part of the analysis. However, the idea of using learning for identification of beliefs seems new to those literatures as well.

tools, in this paper we focus on consumers who use a particular search strategy, both simple and one of the most popular: to sort hotels by decreasing price. A consumer visits the website, and fills out a request: city (Chicago), dates of stay, number of people, rooms, etc. Based on that request, the website gives her an assortment of currently available options. The user sorts them by increasing price and observes the first page of results that consists of 15 lowest priced hotels. At this point, she has three options: leave the website; choose a hotel on the first page; go to the next page. Since we allow her to turn only one page, on the second page the user has either to choose a hotel or leave the website. In our data, we observe everything that a consumer does and sees on website, in real time: request parameters, page contents, turning and clicking decisions.

To explain observed decisions, we assume that turning the page entails a search cost, with can be interpreted as a cost of processing information on that page. Every consumer is endowed with a search cost that is independently drawn from a common distribution. At the same time, there is a possibility of finding a better deal among the 15 hotels on the second page. In response to the uncertainty about the content of the next page, the consumer formulates a prior belief about the joint distribution of hotel characteristics (price, star rating, etc), and updates her belief from the information on the first page. The decision to turn the page is then based on a comparison of search cost to the expected benefit, evaluated from the posterior beliefs. Further, the decision to click on a hotel (on not click at all) is explain by a model of utility, together with composition of the choice set. We formulate parametric models of utility, prior beliefs (common for all consumers), the search cost distribution and perform the estimation by maximum likelihood. The median search cost is around \$25 for a collection of 15 hotels, or about \$1.7 per hotel; there also significant heterogeneity of search costs among the population. The estimated prior belief about conditional price distribution, although has theoretically correct signs, shows less sensitivity to hotel's characteristics than price regression. The estimated amount of prior uncertainty is small, but statistically significant; at the same time, the degree of Bayesian updating is economically meaningful, because the first page of hotels is very informative about the underlying price-quality relationship. We also find that travelers who stay over a weekend are more price sensitive, while those traveling in a group and/or searching in a shorter time are more likely to click on a hotel now than leave website without clicking. These findings seem to be intuitive from an economic point of view.

This paper is organized as follows. In the next section, we provide a review of the existing literature on empirical models of product search. Section (3) describes the data; in Section (4) we present our search model and then discuss issues of identification and estimation in Sections (5) and (6). Results are discussed in Section (7) and in Section (8) we compute and compare price elasticities of demand. Section (9) concludes. Appendix contains all tables and figures.

2 Literature

By this time there exists a large empirical literature[2] examining consumer search, which varies widely in methods applied and questions addressed. A comprehensive review of this literature is beyond the scope of this paper. Here we only review a small part of it, which is directly related to our research. These are structural models of sequential product search, e.g. models that explicitly incorporate optimal search decisions. We also distinguish ourselves from papers on search for experience goods, such as Crawford and Shum (2006), Ackerberg (2003), Erdem, Keane and Strebel (2005).

Hong and Shum (2003) use online price data for consumer electronics products and books to estimate a search model, separately for each product. This paper proposes a method to estimate a parametric form of search cost distribution from price data alone, without information on purchases or market shares. The idea is to use equilibrium indifference conditions of the type suggested by Burdette and Judd (1983), as identifying restrictions. These restrictions postulate the distribution of reservation prices in the population, which implies a functional relationship between search cost distribution and price beliefs. Every observed price is then represented as a reservation price for a group of consumers, and parameters of search cost distribution are estimated by maximizing the likelihood of observing these prices. In this way, it is implicitly assumed that price beliefs are given ML approximation to the empirical density of prices. Hong and Shum find that sequential models give rather large search cost estimates, which seems unrealistic given the existence of price comparison websites. This is a negative conclusion, but it is also not surprising: in their model, search costs alone should explain the variation of prices in equilibrium. Also, this is a highly stylized model, which does not account for seller's heterogeneity, in terms of market shares or marginal costs, as well as product heterogeneity. In a follow-up paper, Moraga-Gonzalez and Wildenbeest (2008) propose an alternative estimation method for the non-sequential search model, which is shown to perform better in their Monte-Carlo experiment.

Babur de los Santos (2008) on search with unequal sampling

Hortacsu and Syverson (2003) study search for mutual finds, whose value depends not only on price, but also on other characteristics, including unobserved taste shock. However, all consumers are assumed to agree on their valuation of a particular fund, in other words, goods are vertically differentiated. Although this is still a strong assumption, authors show that it fits the data better than the homogeneous goods one. Similarly to Hong and Shum, consumers are assumed to know the empirical distribution of utilities of available funds, hence the search is motivated by uncertainty of what location offers what value. The methodological contribution of this paper is to show that if price data is supplemented by market shares data, then CDF of search cost distribution is non-parametrically identified at a number of cutoff

[2] There is also a well developed literature on labor search, but it addresses quite different set of questions (such as recovering equilibrium distribution of wages) - see van der Berg and Ridder (1999) for a review.

points equal to the number of firms. This is a remarkable result, given that no data on the search process itself is available.

Sorensen (2001) is perhaps the closest paper to what we do. she estimates a model where consumers go around local farmacies looking for the best price of a particular drug. Although drugs themselves are the same, some degree of differentiation is brought by characteristics of a farmacy (such as location), as well as consumer specific taste shock. Therefore, both horizontal and vertical product differentiation is introduced, albeit to a limited extent, both in terms of product attributes and the fact that an outside option is omitted, which makes it difficult to interpret price elasticities. Prior to search, consumers know the non-price characteristics of available goods (including taste shock, which is observed to consumers, but not to the econometrician), as well as the empirical distribution of prices on the market. However, they are uncertain about drug's price at every location, which motivates costly search. Search decision in this model has binary nature: either examine all locations or only one (the the closest). Estimation is performed on data on retail farmacy transactions. Since search decisions are unobserved, they are integrated out together with search costs and taste shocks. Despite substantial price variation, the predicted search intensity is low, about 10%, and such effort costs on average $15. However, search intensity varies across different drugs, it is lower for one-time purchases and higher for maintenance medications. It is important to note that, contrary to the above cited papers, the search cost distribution is identified without equilibrium restrictions: this is a single agent decision problem, which takes existing prices as given. We follow this approach in our paper as well.

Sailer (2006) on search for auctions

Conlon and Mortimer (2008) on demand estimation under incomplete availability, and Fox (2007), related.

3 Data

3.1 Description

Consumers are searching for a hotel in Chicago on the website www.kayak.com. To begin search, the user submits a search request, which includes city (Chicago), dates of stay, number of guests and number of rooms. On average, a search request results in more than 140 available hotels, which makes it a non-trivial search problem. To navigate among search results, users can just flip through pages, or employ various sorting and filtering tools, such as sorting by price or filtering by neighborhood. Each search action (flipping, sorting, filtering) results in a display of at most 15 hotel options. As soon as the user finds a preferred hotel, she can click on it: this website does not sell hotels itself, so the click redirects the user to another website where a booking can be made. This sequence - request, search actions, displays, clicks - comprises what we call a "search history". In total, we have 24 321 of unique search histories, and Table (3) represents the most popular ones. Unfortunately, we cannot tell whether any two histories were made by the same person; therefore, in this paper we treat each history independently.

Fully structural approach to model the optimal search behavior in this strategy space seems unfeasible. Instead, we focus on a subset of population who employed a particularly simple search strategy: those who started their search by sorting by price, then flipped at most one page, and finished their process by booking a hotel or leaving the website. The position of this subset in the total population is described in Table (3). To summarize:

(1) Number of consumers engaged in active search: 24321-7029 = 17 292

(2) Those who started search by price sorting: 2664

(3) Those who started search by price sorting and then only flipped: 1436

(4) Those who started search by price sorting and then only flipped at most one page: 1123.

In what follows we refer to group (1) is "general population", while group (4) is our estimation sample. It consists of 1123 consumers, of whom 848 never turned a page, and 275 turned one page. This represents only 6.5% of general population, and 78% of searchers who employed "sort by price and flip" strategy. Such low number is mostly explained by the large number of other possible strategies: 10 240 searchers used relatively rare ones. Plus, we are not allowing to continue searching on the website using other strategies, such as sorting by distance or by filtering by neighborhood. Nevertheless, as we can see from Table (3), such strategy as price sorting is the most popular way to start search, after simple page flipping.

For every search history of this type, we observe almost everything, that is:

- parameters of initial request: date of search, dates of stay, number of people, number of rooms
- number of pages flipped
- contents of each flipped page: 15 hotel options with prices and other characteristics
- clicked hotels

Since a click redirects the user to another website, we do not observe whether the actual booking was made. See Section (6) for a discussion of this issue. If we observe more than one click, then the last click is used.

Now we discribe the available data fields in more detail; while the main focus is on the estimation sample, we also present summary statistics for groups (1)-(3) to, to see the effect of selection on the distribution of variables.

3.2 Chicago hotels

A search request for Chicago hotels typically returns 130-140 hotel options, depending on availability. During May 2007, the maximal number[3] of returned hotels was 148; these are Chicago hotels with online pricing. As we can see from the Figure (2), there is wide variation in geographical position of hotels. These are not only hotels located in the city of Chicago itself, but also in the satellite towns (Evanston, Skokie, etc), as well as in the proximity of airports (O'Hare, Midway). Table (4) summarizes these hotels in terms of their quality and location. Hotels in "Gold Coast", "Loop" and "West side" are close to the city center - in the estimation, they are combined under "Center" neighborhood. Hotels in "SW (South west)" and "Midway" are relatively far from the center, grouped them under "South" dummy; hotels to the north side are "North" and a special category is "O'Hare" - those close to the airport. On Figure (1) we have a distribution of hotels by their distance to the city center. There are two well defined clusters, those that are located within 5 miles from the city center, and those far from the city, between 10 and 20 miles. These clusters are largely accounted for by neighborhood dummies; also, we try to suppress the effect of outliers by including distance in the log form.

Unobserved quality is an important issue in such good as hotel accomodation. Part of it comes from quality standards among different hotel chains, which we control for by including a set of brand dummies in the estimation. There is a large number of hotel brands present in Chicago market, but for most of them the estimation sample has very little or no data on clicks. Therefore, we include only top 5 hotel brands: Null, Rodeway Inns, Econo Lodge, Days Inn, Best Western - together, they attract 28% of impressions and 56% of clicks. The Null brand stands for hotels that don't belong to any chain; all other hotels are grouped under a default category.

Since we do not observe the total availability for each request, we assume that all N=148 hotels are available at the time of request. This assumption is needed only for specification of consumer's beliefs, as it will be clear below.

3.3 Request types

To start search, the user has to enter a request. Its parameters include: date of search; dates of stay; number of people; number of rooms. From the dates of search and stay, we can derive advance purchase, length of stay, and whether the Saturday night is included. From Table (5) we can see that, relative to the general population (group 1), consumers in our sample search on average 2.8 days less in advance, and are a bit less likely to stay over weekend (only by 3%, significant at 10% level). There are no statistically significant differences in terms of

[3] If one includes hotels without online pricing (i.e. those who advertise themselves on the internet but give price quotes only by phone), their number raises to around 220. However, by default the website shows only hotels with online pricing, and only these I use for estimation.

length of stay or other parameters.

Here we aggregate this request parameters into a number of types, based on advance purchase (within 7 days, between 7 and 14, more than 14), whether Saturday night stay is included, and how many people are traveling (one or more).

	one person		few people		
advance	nw	w	nw	w	total
a<7	121	92	90	106	409
7<a<14	45	36	30	47	158
a>14	90	140	114	212	556
total	256	268	234	365	1123

To summarize:

- people in our dataset either book shortly (within one week of travel) or well in advance (more than 2 weeks).

- more people stay over Saturday night than not: 633 obs. against 490

- there is almost equal number of searches where the traveler is alone and where two or more people are traveling

The most popular types are:

- a group of people books in advance and stays over weekend

- one person books in advance and stays over weekend

- one person books shortly and does not stay over weekend

These variables reflect the booking behavior of consumer, and as such can be used as a measure of her "type", in terms of intent of travel, or price sensitivity. For example, one could argue that people who stay over the weekend are more likely to belong to leisure travelers; the same could be thought of those who search well in advance. In this paper I test the empirical validity of this argument, by including request parameters into the utility of outside option.

3.4 Searching and clicking activity

In this model, we assume that consumer can turn at most one page, after she sorts results by price. As a result, the average length of search in the estimation sample is 1.3 pages. The table below compares this number to other groups:

	group1	group2	group3	group4	T1-4
mean	3.91	3.52	1.87	1.30	87.76
sd	3.36	3.18	1.34	0.51	

Among the general population, the search intensity is significantly higher than in the estimation sample. Note, however, that most of the difference is not the result of our restriction on the number of pages: compare the mean in group 3 that has no such restriction to the mean in group 1, which is a general population. The limited search is probably the result of a reliance on a price sorting as a search strategy; indeed, people who search for 3 pages or more usually employ several strategies. The table below presents distribution of search intensity among group 3:

pages turned	0	1	2	3	total
N people	848	275	166	92	1,436
Cumul. %	59.05	78.2	89.76	96.17	100

In fact, 78.2% of people who used price sorting as their search strategy, flipped no more than 1 page and thus belong to our estimation sample; at the same time, the number of people who flipped 2 pages is of comparable magnitude to the number of people who flipped only 1 page; as a result, extension of the model to the 3rd page seems to be desirable and is subject of future research.

In terms of raw click rates, there is no statistically significant difference between groups, so we do not report these results here. On average, the click rate is 0.36-0.38, with standard deviation of 0.48. However, if we break click rates across different consumers with different parameter requests, then some differences appear, as shown in Table (6).Contrary to the general population, the click rate of people who use only price sorting: a) is positively and significantly affected by length of stay; b) no affected by weekend stay; c) not affected by advance purchase. In all groups an increase in number of travelers has strong effect on click rate. This is a preliminary evidence that parameter of request may be relevant for consumer type; to test this idea more formally, we include request variables in the mean of outside option.

Combining clicking and turning activity, one can distinguish between various types of demand that hotels on the first page receive: "fresh" demand (those people who don't go to the second page) and "returning" demand (those who went to the second page and returned); hotels on the second page receive "residual" demand. The joint distribution of clicking and turning is:

	no turn	turn	Total
no click	546	194	740
click	302	81(19)	383
Total	848	275	1,123

Most of the clicks for hotels on the first page belong to "fresh" demand, 302 out of 383, while the "returning" demand is negligible (19 clicks only). The rest of the clicks belongs to "residual" demand for hotels on the second page. However, the amount of that demand is small relative to the number of people who turn the page: 62 out of 275. These facts will be helpful in understanding some of the estimation results, as we'll see below.

3.5 First page variation

Sufficient amount of variation in prices of hotels, observed on the first page, is very important from the perspective of identification of the model: holding the quality of a hotel constant, variation in its price helps identify the consumer's preference for quality; holding the set of hotels on the first page constant, joint variation of their prices leads to different posterior beliefs among consumers, which in its turn helps identify learning aspects of the search model.

Luckily for us, hotel market is characterized by fluctuating demand and price-discrimination strategies (otherwise called revenue management) employed by hotels. This is reflected in the ample variation of prices of hotels observed on the first page:

var	min	max	mean	std
price	32	567	97.44	46.55
nobs	16845			

To offer some evidence of price variation on a hotel level, on Figure (3), we plot 10% and 90% quantiles of price distribution (from the first page data), for each hotel separately. Although not all hotels (118 out of 148) were observed on the first page, most of the observed ones offer significant price variation.

A special role is played by maximal prices on the first page. According to the search model, these prices serve as truncation points for the distribution of prices on the second page. This source of variation in posterior beliefs is necessary for identification of model without learning, as we argue below. As common intuition suggests, consumers whose maximal prices on the first page are already high, should turn less frequently, expecting even higher prices on the second page. Table (10) presents a summary statistics of maximal prices, separately for turners and no turners. For those most part, there is no difference between truncation prices seen by these types; it appears only for very high prices - 90%, 95% quantiles, and the largest 4 price observations.

Finally, do people see structurally different first pages? On Figure (5), for every hotel we plot the share of first pages on which this hotel has appeared. Most hotels, appear on the first page only from time to time (in less than 40% of cases), and only 15 of them appear on every second page or more often. These are mainly 2* hotels and a couple of cheap 1* hotels. In other words, there some diversity in the content of the first page, but it is not as substantial as price variation.

3.6 Search induced demand

Consumer's preferences, as revealed from joint variation of clicks and choice sets, are one of the central elements of the search model. Here we present some empirical facts that characterize consumer's demand and its relationship to the search process.

To compare demand across selection groups (1)-(4), in Table (7) we present means of various characteristics of clicked hotels. As expected, people in group 4 are clicking on hotels that belong to the lower tail of price distribution: on average, these hotels have lower star rating and are located further from the city center (most notably, close to O'Hare airport). Further, Table (8) gives more detail about configuration of choice sets of consumers, and about hotels that received most of the clicks. It seems that people who sort by price are mostly looking to stay closer to airports or places in the South direction. The presence of airports as strong points of attraction suggests that there is probably a caterogy of travelers who don't care about proximity to the city center. Also, most observed hotels are concentrated in the middle of quality range, 2-3 stars.

Table (9) sheds some light on the dynamics of demand, by breaking down clicks for the first and the second pages. The question is, when deciding to turn the page, what one should expect? We can see that on the second page, very cheap 1* hotels disappear, cheaper 2* hotels lose their weight in favor of 3* and 4* hotels. Shares of clicks change accordingly, especially notable is the increase in demand for 3* hotels (also some demand for 4* hotels). In terms of neighborhoods, centrally located hotels (Gold Coast, Loop) gain both in the impressions and in the demand. An interesting case is the Southwest (SW) neighborhood: although its share of impressions increases by 5%, the demand drops significantly. This probably means that people willing to stay in that neighborhood are quite price sensitive. In terms of airport neighborhoods, O'Hare is prevalent on the first page and Midway on the second, both in terms of impressions and clicks. A similar story goes for distances: hotels that are located far from the center (>10 miles) lose in both terms, and they are replaced by centrally located hotels. In fact, a striking 92% of observations on the first page are hotels located far from city center, while demand on the second page belongs in large part to hotels centrally located.

In sum, by turning the page a consumer should expect to see more of higher quality (3-4 stars) hotels, located closer to the city center (by neighborhood or by distance). Sometimes these kinds of hotels are also present on the first page; in such cases, their prices can be used to evaluate the benefit of turning the page. Intuitively, consumer should be more willing make an effort if she sees lower prices of this type of hotels on the first page.

Figure (6) presents some supporting evidence: it compares empirical CDF of prices of hotels close to city center (<5 miles), observed on first pages by turners and no-turners, separately. One can see that distribution of prices seen by no-turners clearly dominates, by first order stochastic dominance. For turners, the mean price of centrally located hotels is $176, median is $159; for no-turners, these numbers are 198 and 202 dollars, respectively. A similar picture holds for higher quality (>=3 stars) hotels, see Figure (7). For turners, the mean price of hotels of better quality is $138; for no-turners, it is 153 dollars. At the same time, there is no such difference for hotels of lower quality and/or located further from the city center, even though they occupy most of the space on the first page. Such dependence between prices on the first page and page turning can be explained by learning: seeing low prices for good hotels, consumers become more optimistic about the benefit of going to the next page; at the same time, low prices for bad hotels do not affect beliefs because this is not the kind of hotels a consumer expects to see on the second page.

4 Model

In this search environment, every consumer starts by observing the first page of 15 hotel options. At this point, she has three alternatives: a) leave the website without clicking; b) click on a hotel on the first page; c) go to the next page of results, which will give her another 15 hotels. In fact, we can merge a) and b) by including an outside option as a "null" hotel, that is always implicitly present on every page. Therefore, she first makes a search decision (turn the page), to gather information about available hotels. Although by turning the page she may find a better hotel, it is also costly: we assume that every consumer is endowed with a non-zero cost of processing information on the second page (the first page is given for free). After that, the decision to click is going to be based on comparing the values of hotel accomodations, including the outside option.

To explain these joint decisions, we need a statistical model that consists of three main components: first, a utility model that determines the value of a hotel as a function of its observed and unobserved characteristics; second, a model of consumer beliefs about benefits of turning the page; third, a distribution of search costs among population. We start with the model of utility, as it defines the relevant search space.

4.1 Utility

The information about every hotel that is displayed to the consumer includes name of the hotel, brand, price, geographical location, start rating and amenities. The mean utility from a particular hotel is a linear function of price, star rating, and geographical position of a hotel (distance to the city center, neighborhood). That is, the utility of a hotel j for consumer i:

$$\begin{aligned} u(p_j, q_j, \varepsilon_{ij}) &= \alpha_d d_j + \alpha_s s_j + \overrightarrow{\alpha}_n \overrightarrow{n}_j + \overrightarrow{\alpha}_b \overrightarrow{b}_j + \alpha_p^i P_j + \varepsilon_{ij} \\ \alpha_p^i &= \alpha_p + \alpha_{pwd} W_i \end{aligned} \tag{1}$$

- where P_j is hotel's price (in hundreds of dollars); $q_j = \left(d_j, s_j, \overrightarrow{n}_j, \overrightarrow{b}_j\right)$ is a vector of non-price characteristics of hotel j: distance to the city center, star rating, and a set of neighborhood and chain dummies. We take $d_j = \log(1 + D_j)$ - logarithm of distance (in miles), in order to smooth the outliers, see Figure (1). To capture a possible heterogeneity between business and leisure travelers, we allow the price sensitivity to depend on W_i - a dummy variable which is equal to one if a person stays over a weekend, and zero if not. There is also an additive term, a "taste shock", or "match value" that determines idiosyncratic taste of a given consumer for a given hotel. It is observable for consumer but not for the econometrician and its distribution is Type 1 EV, i.i.d across hotels and consumers.

As we have noted above, in the data we observe only clicks, but not the actual bookings. Even though a click may not result in a booking (for example, a consumer might learn something new about the hotel and change her mind), if we see that consumer clicks on a hotel j instead of hotel k, we interpret it as a revealed preference action: $u_j > u_k$ (see Section (6) for a discussion). Leaving the website without any click is interpreted as a preference for the outside option, whose utility is:

$$u_{i0} = \mu_{out} + \overrightarrow{\mu}_o \overrightarrow{R}_i + \varepsilon_{i0}$$

- where $\overrightarrow{R}_i$ is a vector of request parameters by consumer i: advance purchase, number of travelers, weekend stay. The utility model gives us a vector of unknown parameters: $(\alpha_d, \alpha_s, \alpha_p, \alpha_{pwd}, \overrightarrow{\alpha}_n, \overrightarrow{\alpha}_b, \mu_{out,} \overrightarrow{\mu}_o)$. Note that the utility specification (1) does not include a constant term. This exclusion restriction is necessary to identify μ_{out}; alternatively, we could identify a constant term in (1) and normalize μ_{out} to zero.

4.2 Page turning decision

A model of rational search implies that when making a search decision, the consumer takes into account the information she has collected so far. In our case, the relevant information set consists of 15 hotel options observed on the first page of results. Since prices are sorted in increasing order, these are the 15 lowest priced hotels among those available. Let $u_{ir} = u(p_{ir}, q_{ir}, \varepsilon_{ir})$ - utility of a hotel ranked r, for consumer i; also, let $r = 0$ correspond to the outside option. From the first page of results, the consumer receives the current best utility $U^*_{1i} = \max\{u_{ir}\}_{r=0}^{15}$, and the information set, $\Omega_i = \{p_{ir}, q_{ir}\}_{r=1}^{15}$. Note that the information set does not include taste shocks: they are independent of hotel's observable characteristics, and therefore uninformative about the posterior distribution of prices and qualities.

Going to the next page will reveal the next 15 hotels, which will be more expensive, but potentially of better quality. These hotels can be summarized by $U^*_2 = \max\{u(p_{ir}, q_{ir}, \varepsilon_{ir})\}_{r=16}^{30}$ - utility of the best one among them. Define $F_u(U^*_2|\Omega_i)$ as consumer's i posterior belief[4] about the distribution of U^*_2, conditional on her information set, Ω_i. Then, a rational consumer will

[4]Let $F^0(t)$ be the true, but unknown distrubution of best utility among hotels on the first page. For the consumer, F^0 is a random varible in an infinite-dimensional space of probability measures, and G is consumer's belief about distribution of F^0. Using the information set Ω_i, the consumer i computes her posterior belief, $G|\Omega_i$. However, since we are going to take expectations of gains of search, which is a linear operator, we only need to know one moment of the posterior belief, which is $F_u(t|\Omega_i) = E_G\left[F^0|\Omega_i\right](t)$. In the text, we refer to F_u as "posterior belief", keeping in mind that it is actually the mean of posterior $G|\Omega_i$.

turn the page if and only if the expected benefit of doing so exceeds the search cost:

$$\int_{U_{1i}^*}^{+\infty} (U_2^* - U_{1i}^*) \, dF_u(U_2^*|\Omega_i) > c_i \tag{2}$$

- where c_i is a search cost of consumer i. We assume that the logarithm of search costs follows normal distribution with mean and standard deviation (c_0, c_1), and every consumer receives an i.i.d draw from it. Unknown parameters (c_0, c_1) are going to be estimated with other parameters of the model. Note that the lower limit of integration is U_{1i}^*, since assume search with recall, i.e. the consumer can costlessly go back to the first page. We now discuss our assumptions on consumer's beliefs.

4.3 Beliefs and learning

A common idea behind various models of search is the extistence of some true data generating process that delivers goods available on the market and hence the outcomes of search. Part of that literature, beginning with Stigler (1967) assumes that the consumer knows this process; the other part, beginning with Rothschild (1974) assumes that she has only vague idea about it and learns while searching. In the former case, it is said that "consumer is searching from unknown distribution". In our case, let $\vartheta\left(p_j, q_j, \varepsilon_{ij}|a\right)$ be the true distribution from which characteristics of available hotels are generated. We assume that consumer knows the parametric family, but not the true parameter $a = a_0$. In this sense, consumers in our model are searching from unknown distribution. This formulation stands in between the two modeling traditions mentioned above, because papers based on Rothschild (1974) typically take non-parametric approach, such as Dirichlet process for the uknown distribution function. While such approach is convenient from analytical standpoint, it is less so in empirical implementation, as it requires a lot of data.

The learning process has a number of steps. Prior to search, the consumer holds a belief $G\left(a\right)$ about the unknown parameter a_0 of the joint distribution of hotel's characteristics; we assume that the prior belief is common across consumers. After observing her information set, Ω_i, which is a collection of prices and qualities of hotels on the first page, she updates this prior in the Bayesian fashion, to arrive at $G(a|\Omega_i)$ as her posterior belief about unknown parameter a_0. This, together with the knowledge of $\vartheta\left(p_j, q_j, \varepsilon_{ij}|a\right)$ allows her to find the distribution of about prices, qualities and match values of hotels on the second page. In doing so, she takes into account price sorting: prices on the second page are truncated order statistics. Then, using the utility model (1), she can transform her posterior belief from the multi-dimensional space of hotel's attributes into the space of scalar utilities, to obtain $F_u(U_2^*|\Omega_i)$. Below specify the parametric families of (ϑ, G) and the process of updating, but now a couple of comments are in order.

The assumption of common prior is rather strong; however, allowing for unobserved posteriors would require additional integration over the likelihood function, which is not feasible with the current method. One way to introduce some heterogeneity in priors would be to include parameters of request into the prior beliefs, which is the subject of future research.

Another issue is the rationality of beliefs. In an equilibrium model, the rationality of beliefs and the first order conditions of firm's profit maximization provide natural restrictions on the sort of price beliefs that may result from the estimation. For example, equilibrium model cannot produce beliefs where price decreases with star rating, because that would be incompatible with firm's behavior. In our model, which is a single agent decision problem, additional constraints may be necessary to obtain estimates of beliefs that are economically sensible. For example, here we require that the average hotel price predicted by consumer's prior should be correct (see Section (7) for details).

4.3.1 Belief structure

Using the chain rule and the assumption of independence of taste shocks, we can re-write $\vartheta\left(p_j, q_j, \varepsilon_{ij}|a\right)$ as a product of conditionals:

$$\vartheta\left(p_j, q_j, \varepsilon_{ij}|a\right) = f_p(p_j|q_j, a)H(q_j|a)f_\varepsilon(\varepsilon_{ij})$$

- where the distribution of match values, $f_\varepsilon(\varepsilon_{ij})$, is Type 1 EV, i.i.d across hotels. The independence assumption implies that taste shocks observed on the first page are not going to affect posterior beliefs; this is why information set Ω_i, as defined above, includes only observable characteristics of hotels. Note that both consumer and econometrician are uncertain about match values of hotels that may appear on the second page: the motivation is that consumer i learns about ε_{ij} only when she observes hotel j. On the practical side, this assumption[5] greatly simplifies computations, to the extent that without it our current approach to estimation would be unfeasible.

We also assume that consumer knows the empirical distribution of non-price characteristics of existing hotels $X = \{q_j\}_{j=1}^{N}$:

$$H(q_j) = \frac{1}{N}\sum_{q_j \in X} I(q_j = q) \tag{3}$$

- where the equality $q_j = q$ is satisfied if all components of vector q_j are equal to the corresponding components of a vector q. Here we do not assume that consumer knows the identities of all Chicago hotels, because of the uncertainty about unobserved taste shock, that we discussed above. Instead, she knows the support of distribution of observable qualities and

[5] An alternative assumption could be that the consumer i actually knows the vector of match values of all available hotels, $(\varepsilon_{i1}, .., \varepsilon_{iN})$, in which case second-page match values will be drawn from a discrete distribution. The support of this distribution is unknown to the econometrician, so an additional integration step is required. Our current assumption requires only one integration step, when computing expected benefit of search.

perceives every observed hotel as a random draw from $H(q)$. Also, since we do not observe the actual availability of hotels for each search request, we assume that $N = 148$ hotels are available for every consumer. These are Chicago hotels that had online pricing in May 2007.

Finally, the distribution of logarithm of price $p_j = \ln(P_j)$, conditional on hotel's quality, belongs to normal family:

$$f_p(p_j|q_j, a) = \phi(p_j; a_0 + a_1 s_j + a_2 d_j, \sigma_0^2) \tag{4}$$

- where $a = (a_0, a_1, a_2)$ is unknown parameter. However, the variance log-prices is known and remains fixed during estimation, at $\sigma_0^2 = 0.35$, an estimate from a large dataset of hotel prices. To assess the empirical validity of this assumption, we ran a regression of logarithm of price on distance and star rating, on a large dataset of hotel prices[6], with error terms clustered on the hotel level. Results are presented in the table below:

Table 1: Results of price regression. Dependent variable is logarithm of price

	Coef.	Std. Err.	t	P>|t|	95 Conf	Interval
stars	0.23	0.02	9.53	0.00	0.18	0.28
dist	-0.26	0.02	-13.71	0.00	-0.30	-0.23
cons	0.35	0.09	3.99	0.00	0.18	0.53
R-squared	0.60					
N obs	137000					

To test the normality of residuals from this regression, we tried various tests, such as Shapiro-Wilk, Shapiro-Francia, skewness-curtosis, and found that all of them strongly support the null hypothesis. Which is perhaps not surprising, for a dataset of this size. As an illustration, we include normal quantile plot, see Figure (4) in the Appendix. We also tested normality of log-prices for every hotel in particular, with the same result.

Using her information set, Ω_i, consumer performs Bayesian updating of her beliefs about unknown parameter a and the first step is to find the likelihood of Ω_i, i.e. the joint density of prices and qualities of hotels on the first page.

4.3.2 Price densities

From the statistical point of view, prices on the first page, p_{ir}, $r = 1...15$, are r-th order statistics from the unconditional price distribution, whose density and CDF can be found

[6] To avoid biases due to unequal popularity of hotels, we make a random draw of 1000 price observations for every hotel. This eliminates 11 out of 148 hotels for which we have less than a thousand observations. This is a small portion of the dataset, so results almost do not change.

from (3) and (4):

$$f_p(p_j|a) = \frac{1}{N}\sum_{j=1}^{N}\phi\left(p_j; a_0 + a_1 s_j + a_2 d_j, \sigma_0^2\right) \tag{5}$$
$$F_p(p_j|a) = \frac{1}{N}\sum_{j=1}^{N}\Phi\left(p_j; a_0 + a_1 s_j + a_2, \sigma_0^2\right)$$

- where ϕ, Φ are normal p.d.f and c.d.f. The joint density of order statistics is a well-known result (see, for example, David and Nagaraja (2003), pp12-13):

$$f(p_{i1}, .., p_{i15}|a) = \frac{N!}{(N-15)!}\left(\prod_{k=1}^{15} f_p(p_{(k)}|a)\right)(1 - F_p(p_{(15)}|a))^{N-15} \tag{6}$$

Our problem, however, is a bit more difficult, because we need to find the joint distribution of price order statistics and corresponding qualities, (p_{ir}, q_{ir}), $r = 1...15$. The following result, which is an adaptation of Lemma 1 from Bhattacharya (1974) is going to be helpful.

Lemma 1 *Hotel qualities $q_{i1}, .., q_{i15}$ are conditionally independent given realized prices, $p_{i1}, .., p_{i15}$, with densities $H(q_{ir}|p_{ir})$, $r = 1, .., 15$.*

Proof. *From our assumption on price distribution in (4), the relationship between price and quality is $p_j = aq_j + \eta_j$, $\eta_i \sim N(0, \sigma_0^2)$. We can present it as $aq_j = p_j + \eta_j$, and the same relationship holds for ranked results, $aq_r = p_r + \eta_r$. Since η_j are independent of p_k, $k \neq j$, and mutually independent, the same is true about η_r. Therefore, when prices are fixed, the joint density of hotel's qualities, $aq_r = p_r + \eta_r$, is $\prod H(q_r|p_r)$.* ■

From this lemma, we get joint conditional density of hotel's qualities:

$$\begin{aligned} H(q_{i1}, .., q_{i15}|p_{i1}, .., p_{i15}, a) &= \prod_{r=1}^{15} H(q_{ir}|p_{ir}, a) \\ &= \prod_{r=1}^{15} \frac{f_p(p_{ir}|q_{ir}, a)}{f_p(p_{ir}|a)} H(q_{ir}) \end{aligned} \tag{7}$$

Multiplying (6) and (7), we obtain the likelihood of the first page, given parameter a:

$$L(\Omega_i|a) = \frac{N!}{(N-15)!}\prod_{r=1}^{15} f_p(p_{ir}|q_{ir}, a)H(q_{ir})(1 - F_p(p_{i15}|a))^{N-15}$$

For what follows, we need only the part of the likelihood that varies with a and across consumers:

$$L(\Omega_i|a) \propto \prod_{r=1}^{15} f_p(p_{ir}|q_{ir}, a)(1 - F_p(p_{i15}|a))^{N-15}$$

We are now ready to describe the process of Bayesian updating.

4.3.3 Learning

The learning process in our model is not limited to Bayesian updating. In addition to that, we have to account for price sorting, whereby the price distribution should be appropriately truncated; also, our assumption of recall implies that the consumer should not expect to see the same hotel on the first and the second page.

The prior belief $G(a)$ about the distribution of unknown vector parameter a is assumed to be normal:

$$G(a) = N(\beta_0, \Sigma_0)$$

- where covariance matrix Σ_0 is diagonal. Given information set Ω_i, the unnormalized posterior is:

$$G(a|\Omega_i) \propto L(\Omega_i|a)\phi(\beta_0, \Sigma_0)$$

- which is a non-standard density. For the estimation algorithm to be feasible, however, we should be able to draw samples from posterior quickly and easily. Even though methods of drawing from non-standard densities are readily available (e.g., Metropolis Hastings algorithm), for simplicity we approximate the posterior with, again, normal distribution.

There are many possible ways to construct such approximation, and one of them is Laplace's method (see Tierney and Kadane (1986)). Laplace's method provides an approximation to the normalizing constant of posterior density, using second order Taylor expansion of the natural logarithm of the density around its mode. This involves finding the mode and computing hessian at that point; as such, it is a much less expensive task than numerical integration. According to this method, we approximate density $G(a|\Omega_i)$ with a Gaussian with the mean and covariance matrix β_{1i}, H_{1i}^{-1}:

$$\begin{aligned} G(a|\Omega_i) &\simeq N(\beta_i, H_i^{-1}) \qquad (8)\\ \beta_{1i} &= \arg\max(\log(G(a|\Omega_i))) \\ H_{1i} &= -\frac{\partial^2 \log(G(a|\Omega_i))}{\partial a^2}|_{a=\beta_{1i}} \end{aligned}$$

- in other words, the mean of the approximating Gaussian coincides with the mode of true posterior (it is unimodal), and covariance matrix is an inverse of Hessian of that posterior at the mode. Contrary to many other methods, e.g. based on minimizing Kullback-Leibler or Kolmogorov distance, this method is computationally inexpensive and in our case gives good results: we computed Kullback-Leibler distance between exact and approximated posteriors for all consumers in our sample, and found it to be within 1% of own entropy of the exact posterior[7].

Due to price sorting, updating of beliefs involves the second step: conditional price distribution in (4) has to be truncated from below by the maximal price on the first page:

$$f_p(p_j|p_j > p_{i15}, q_j, a) = \frac{\phi(a_0 + a_1 s_j + a_2 d_j, \sigma_0^2)}{1 - \Phi(p_{i15}, \sigma_0^2)} \tag{9}$$

Finally, the distribution of hotel's qualities, $H(q)$, is also modified due to the fact that consumer should not expect to see the same hotel on the first and the second page:

$$H(q|\Omega_i) = \frac{1}{N-15} \sum_{q_j \in X/\Omega_i} I(q_j = q) \tag{10}$$

Therefore, the system of posterior beliefs about prices, qualities and match values of hotels on the second page is described by (8)-(10) and the assumption of EV Type 1 distribution of match values. Together with the model of utility, this implies a distribution of best utility from the second page, $F_u(U_2^*|\Omega_i)$.

4.4 Reservation property

Let us return now to the decision rule (2). For the consumer, the inequality is a deterministic statement: it is either satisfied or not. For the econometrician, who observes neither first-page taste shocks[8], $\{\varepsilon_1, .., \varepsilon_{15}\}$ (and hence U_{1i}^*), nor search cost, this is a probabilistic statement. For a given search cost, this inequality defines a set in the space of first-page taste shocks; we now show that this set has a particularly simple structure. We first notice that the integral in (2) depends on $\{\varepsilon_1, .., \varepsilon_{15}\}$ only through U_{1i}^*. Then we have the following lemma.

Lemma 2 *Suppose $F_u(U_2^*|\Omega_i)$ is a continuous distribution function. Then, the inequality (2) as a condition on unobservables $\{\varepsilon_1, .., \varepsilon_{15}, c\}$ can be equivalently written as:*

$$\{\varepsilon_1, .., \varepsilon_{15}, c\} \quad : \quad U_1^*(\varepsilon_1, .., \varepsilon_{15}) < \bar{u}(c) \tag{11}$$

$$where \quad \bar{u}(c) \quad : \quad \int_{\bar{u}}^{+\infty} (U_2^* - \bar{u})\, dF_u(U_2^*|\Omega_i) = c \tag{12}$$

Proof. *Consider the left side of the inequality (2). From our assumption, it is a continuous function, can it can be re-written as: $\int_{U_1^*}^{+\infty} (U_2^* - U_1^*)\, dF_u(U_2^*) = \int_{U_1^*}^{+\infty} U_2^* dF_u(U_2^*) - U_1^*(1 - F_u(U_1^*))$. Here we omit conditioning on Ω_1 for brevity. We are going to vary U_1^* by*

[7] The Kullback Leibler distance is a mesure of difference between two distributions, $P(x)$ and $Q(x)$. In the case of continuous distributions, $D_{KL} = \int P(x)\log(P(x)/Q(x))dx$. For comparison, we've chosen own entropy of the true distribution, $P(x)$, which is $E(P) = \int P(x)\log(P(x))dx$. In this sense, we are comparing to an approximation $Q(x) = 1$.

[8] In fact, since by assumption the concomitants of price order statistics $\varepsilon_{(j)}$ have the same distribution as ε_j, we adopt the latter notation for clarity of exposition.

changing only $\{\varepsilon_1, .., \varepsilon_{15}\}$*: in this way,* $F_u(U_2^*)$ *will not be affected. Taking the derivative with respect to* U_1^**, we obtain:* $-U_1^* f_u(U_1^*) - 1 + F_u(U_1^*) + U_1^* f_u(U_1^*) = F_u(U_1^*) - 1$*, which is less than zero provided* $U_1^* < +\infty$*. That is, the left side of (2) is a decreasing function of* U_1^**. At* $U_1^* = -\infty$*, its limit is* $+\infty$*, and at* $U_1^* = +\infty$ *it is equal to zero; hence, there exists a single crossing point where it is equal to search cost (which is strictly positive).* ■

Remark 1 *It would be incorrect to interpret this result as a reservation property of the decision rule, because the content of information set* Ω_i *is fixed. Once we allow it to vary, the monotonicity of the expected benefit of search with respect to* U_{1i}^* *will generally not hold. In fact, such general result is not needed for estimation purposes, as long as econometrician also knows* Ω_i*.*

4.5 Likelihood of clicking and turning decisions

For every consumer, we observe two kinds of decisions: first, whether or not she turned the page; second, what hotel was booked (including the null hotel). For example, if we observe a consumer who has turned the page and booked a hotel $r \in \{0, 1, .., 30\}$, then in terms of unobservables this implies two kinds of inequalities:

$$\begin{array}{rcl} U_{1i}^* & < & \bar{u}(c_i) \\ \mu_{ir} + \varepsilon_{ir} & \geq & \mu_{ir} + \varepsilon_{ir},\ r = 0, 1, .., 30 \end{array}$$

- where μ_{ir} is the mean utility of hotel ranked r on the first page, or $r - 15$ on the second page; μ_{i0} is mean utility of outside option. Integrating these inequalities with respect to variables unobserved by econometrician gives us the joint probability of two decisions. These variables are match values (or taste shocks), associated with every observed hotel and the search cost parameter. At this point, our assumption about Type 1 EV distribution of taste shocks becomes very helpful, and for a given reservation utility, analytic solution exists to the integration problem (see Appendix A).

Before presenting the likelihood function, let us summarize what is observed on a consumer level. The exogenous variables are $\Omega_i = \{p_{ir}, q_{ir}\}_{r=1}^{15}$ - characteristics of observed hotels on the first page, here r - position of a hotel on that page; also, define $\Omega_{2,i} = \{p_{ir}, q_{ir}\}_{r=16}^{30}$ be the contents of the second page. Taking the two sets together, define $S_i = \Omega_i \cup \Omega_{2,i}$. Part of this data is missing, because we don't observe $\Omega_{2,i}$ for consumers who didn't turn the page; however, this information is irrelevant for explaining their joint decisions[9]. Clearly, the choice set is defined as $CS_i = \Omega_i \cup \Omega_{2,i}$ for turners and $CS_i = \Omega_i$ for no-turners. Finally, we have R_i - parameters of request, that includes dates of search, dates of stay, number of people and other variables derived from this data.

[9] This is not the case if we only had to explain the clicking decisions, where we have to integrate over page turning decisions.

The endogenous variables are (T_i, C_i), where $T_i = 0, 1$ - page turning decision and $C_i = 0, 1, .., \#CS_i$ - the position of the clicked option in the choice set CS_i, with $C_i = 0$ for no click.

From the definition of reservation utility, we obtain: $\bar{u}_i = \bar{u}(\beta_1(\Omega_i, \beta_0), R_i, c_i)$, where the (vector-valued) function $\beta_1(\Omega_i, \beta_0)$ reflects the learning process, that relates parameters of prior beliefs β_0 and data Ω_i to parameters of posterior beliefs, β_1.

Proposition 1 *Conditional on exogenous variables $X_i = (S_i, R_i)$ and search costs, the probability of observing consumer not turning the page and clicking on the result ranked $r = h$ is:*

$$P(T_i = 0, C_i = h | X_i, c_i) = \frac{\exp(\mu_{ih})}{\sum_{r=0}^{15} \exp(\mu_{ir})} \left(1 - \prod_{r=0}^{15} F_\varepsilon(\bar{u}_i - \mu_{ir})\right) \tag{13}$$

- where $F_\varepsilon()$ is cdf of Type 1 EV distribution. For a consumer who turned the page and then went back to book something from the first page (including null hotel):

$$P(T_i = 1, C_i = h | X_i, c_i) = \frac{\exp(\mu_{ih})}{\sum_{r=0}^{30} \exp(\mu_{ir})} \prod_{r=0}^{30} F_\varepsilon(\bar{u}_i - \mu_{ir}), \quad if \;\; h \leq 15 \tag{14}$$

Similarly, for a consumer who turned the page and booked something from the second page:

$$\begin{aligned} P(T_i = 1, C_i = h | X_i, c_i) &= \frac{\exp(\mu_{ih})}{\sum_{r=0}^{30} \exp(\mu_{ir})} \prod_{r=0}^{30} F_\varepsilon(\bar{u}_i - \mu_{ir}) \\ &+ \frac{\exp(\mu_{ih})}{\sum_{r=16}^{30} \exp(\mu_{ir})} \prod_{r=0}^{15} F_\varepsilon(\bar{u}_i - \mu_{ir}) \left(1 - \prod_{r=16}^{30} F_\varepsilon(\bar{u}_i - \mu_{ir})\right), \quad if \;\; h \geq 16 \end{aligned} \tag{15}$$

Proof. *See Appendix.* ■

The results of this proposition provide some insight why static demand estimates are inconsistent, if choice sets are generated by search. Observe that for those consumers who clicked on the first page (including outside option), the likelihood contributions have a multiplicative form: $P(T_i, C_i | X_i, c_i) = P(C_i | X_i) P(T_i | X_i, c_i)$. This means that the likelihood of observed clicks, conditionally on turning decisions, is $P(C_i | T_i, X_i, c_i) = P(T_i, C_i | X_i, c_i) / P(T_i | X_i, c_i) = P(C_i | X_i)$. In other words, we can consistently estimate consumer's preferences from the first-page clicks. At the same time, there is no such multiplicative form for consumers who clicked on the second page. Therefore, if we include both types of consumers in the estimation, the results will be inconsistent, because the likelihood function (e.g., the logit form) is misspecified for the part of the observations. This result is a property of the particular distribution of taste shocks that we use, and is not likely to hold in the general case.

Finally, every likelihood contribution has to be integrated with respect to the unobserved search cost. We assume that logarithm of search cost follows normal distribution, with mean and standard deviation (c_0, c_1). The method of estimation is by maximum likelihood; more precisely, by simulated maximum likelihood, because much of the integration in computing probabilities (13)-(15) is done using simulations, as described below, in Section (6.2).

5 Identification

Non-parametric identification. We start our discussion with identification of a mean utility function, for hotels and for the outside option. Consistently with the remark we made above, we find that mean utilities of hotels and of the outside option are non-parametrically identified from first-page clicks.

Consider all consumers who entered the same request R_i, observed the same first page, Ω_i. Note that these could also be consumers who turned the page but then went back. Let P_{ih} be the proportion of those who on hotel on the first page, ranked $r = h$, with characteristics (p_{ih}, q_{ih}); also P_{i0} is the proportion of those who chose outside option. We can compute $P_{ih0} = P_{ih}/P_{i0}$ - the ratio of the two. From (13) and (14), the model predicts that the ratio is equal to:

$$P_{ih0} = \exp(\mu(p_{ih}, q_{ih}) - \mu_0(R_i))$$

By inverting this equation, we obtain the value of the function $\delta(p, q, R) = \mu(p, q) - \mu_0(R)$ at the point (p_{ih}, q_{ih}, R_i). The function $\mu_0(R)$ is identified from $-\delta(p_0, q_0, R)$ up to a constant, $c = -\mu(p_0, q_0)$; conversely, the function $\mu(p, q)$ is identified from $\delta(p, q, R_0) - \delta(p_0, q_0, R_0)$ up to the same constant.

This also gives a non-parametric test of the model: the derivative of $\delta(p, q, R)$ w.r.t R should not depend on (p, q). Note that we have not used observations for consumers who clicked on the second page, because their likelihood contributions, (15), do not have the necessary multiplicative forms. However, thanks to the IIA property of logit demand, we still obtain consistent estimates of mean utility functions.

Turning to the identification of search cost distribution, define $P(X_i) = P(T_i = 1|X_i)$ - share of page turners among those with X_i. The model predicts that:

$$P(T_i = 1|X_i) = \int_0^{+\infty} \prod_{ir=0}^{15} F_\varepsilon(\bar{u}(\beta_0, X_i, c_i) - \mu_{ir}) f(c_i) dc_i \tag{16}$$

- where μ_{ir} are mean utility functions, identified as above, and β_0 is unknown parameter of prior belief. Given β_0, equation (12) defines $\bar{u}(\beta_0, X_i, c_i)$ is a known function of data, X_i, and search cost, c. The product of CDF's is then integrated over unobserved search cost, where we explicitly use an assumption that search cost is independent of consumer's characteristics, and, most importantly, of information set Ω_i. Using definition of F_ε, this can be written as:

$$\begin{aligned}\prod_{k=0}^{15} F_\varepsilon(\bar{u}(\beta_0, X_i, c) - \mu_{ir}) &= \exp\left(-\sum \exp\left[-\bar{u}(\beta_0, X_i, c) + \mu_{ir}\right]\right) \\ &= \exp\left(-\exp\left(-\bar{u}(\beta_0, X_i, c)\right)\sum e^{\mu_{ir}}\right) \\ &= \exp\left(M(X_i)\exp\left(-\bar{u}(\beta_0, X_i, c)\right)\right)\end{aligned}$$

Therefore, $f(c)$ is a solution[10] to the integral equation of the first kind:

$$\begin{aligned}P(T_i &= 1|X_i) = \int_0^{+\infty} K(X_i, c) f(c) dc \\ K(X_i, c) &= \exp\left(M(X_i)\exp\left(-\bar{u}(\beta_0, X_i, c)\right)\right)\end{aligned}$$

For a given value of β_0, the kernel is a known function of (X, c). Although the existence of exact solution to this equation is hard to check, a minimum distance solution almost certainly exists. For example, we can use a rather flexible parametric model for $f(c)$, with number of parameters up to dimensionality of X. A collection of prices and qualities on the first page, plus request parameters, X consists of about 50 variables. Then, parameters of search cost distribution and prior beliefs are recovered as a minimum distance solution to the system of non-linear equations.

Identification of a parametric model. With a limited size of the estimation sample, - for example, a typical number of observations per X_i is just one, - it is necessary to adopt a fully parametric version of the model. Arguments for identification in such models are of somewhat different nature.

As has become standard, coefficients in the mean utility specification are identified from joint variation of clicks and the choice sets from which they are made. Another, less obvious, source of identification of demand parameters comes from joint variation of content of the first page and page turning decisions. As we know from the model, it is optimal to search if the best utility from the first page is lower than the reservation utility: $U_{1i}^* < \bar{u}_i$. An increase in first page prices makes the consumer less satisfied and hence more willing to search: the

[10] As a reference. All functions in this relationship, P, K, f are continuous and square integrable. As such, the kernel gives rise to a compact operator $K : H_1 \to H_2$ from real Hilbert space (of Lebesgue measurable functions) H_1 to another space H_2. Picard has established a necessary and sufficient condition for the existence of a solution to this equation:

$$\sum_{j=1}^{+\infty} \mu_j^{-2} |\langle P, u_j \rangle| < \infty$$

- where positive numbers μ_j are singular values of K, and u_j are singular vectors for the self-adjoint operator KK^*. This condition seems tough to check.

predicted probability of page turning increases in a way proportional to price coefficient.

The mean of outside option is identified from the proportion of people who didn't click on anything, and from the restriction that the linear mean utility specification (1) does not have an intercept. The coefficients on parameters of requests are identified from the variation of the shares of no-clickers across consumers with different combinations of these parameters.

Now consider the problem of separation of beliefs and search cost distribution. The extent to which it is possible depends both on data and structural assumptions on distribution of beliefs and search cost. Without the latter, we can rationalize any observed page turning decisions by choosing appropriate cominations of prior beliefs and search costs[11]. Therefore, it is necessary to impose some structure: here we assume common prior belief and common search cost distribution, from which every consumer receives an i.i.d draw.

In general, parameters of prior beliefs and search costs are identified from joint variation of content on the first page and turning decisions. The model presents a restriction on this variation, see (16) above. From this equation, we see that the effect of beliefs and search costs on consumer's decisions[12] is summarized by reservation utility: $\bar{u}(\beta_1(\Omega_i, \beta_0), R_i, c)$. The "learning" function $\beta_1(\Omega_i, \beta_0)$ emphasizes that the effect of information set Ω_i on page turning depends exclusively on β_0, and not the search cost. This function is a weighted "average" of the prior, β_0 and the data, where weights depend on the variance of prior, Σ_0. Therefore, changes in β_0 have differential impact on predicted page turning probabilities of consumers, depending on the position of their first page data relative to the prior. For example, an increase in β_0 makes everybody more pessimistic, but those who observed high prices become relatively more so than those with low prices. In its turn, the variance of prior beliefs determines to extent to which posterior can be affected by data.

On a more intuitive level, one can make the following argument. Suppose there can be only two kinds of pages: with high prices and with low prices. Then take consumers who all have seen the low price page. Given the assumption of common prior, they must have similar posteriors as well. According to the model, they can make different search decisions if and only if they have different search costs. In other words, the parameter of search cost distribution is identified from the share of page turners among people who received similar information. Further, observe that if search cost were the only factor, then share of page turners should be the same among those who saw low prices and those who saw high prices. In reality, they are not equal, which can only be explained by the fact that those consumers have different posteriors. Therefore, parameters of prior beliefs are identified from the variation of page turning activity across different types of first pages.

[11] Suppose for simplicity that we observe utilities. Then, for given beliefs, we can make interval estimates of consumer's search cost: $c_i \in [0, \bar{c}_i]$ if she turned the page, and $c_i \in [\bar{c}_i, +\infty)$ if she didn't. I conjecture that with longer panel, these intervals can be made smaller, but point identification is not possible.

[12] Note that the reservation utility only enters expressions like $F_\varepsilon(\bar{u} - \mu_{(k)})$, which means that assumptions on distribution of taste shocks are very important for estimates of beliefs and search costs.

Note that Bayesian learning is the key mechanism that helps disentangle the effects of beliefs and search cost distribution. It creates variation in posterior beliefs which serves as another source of variation in expected benefit of search, orthogonal[13] to variation in search costs. Another source of such variation is the non-stationarity of the search problem, where the sampling distribution is changing over time. Here there are two reasons for this: truncation of price distribution due to price sorting; assumption of recall (e.g., I do not expect to see the same hotels on the second page as I saw on the first one). See Section (4.3.3) for more details. Thanks to these additional factors, the connection between page turning and content of the first page remains. Yet the effect of the recall property is arguably small, considering the large number of hotels; therefore, in the model of search from known distribution the main identifying variable is the truncation price. Table (10) suggests that there is a substantial variation in truncation prices, so parameters of beliefs are still identified, but for slightly different reasons.

6 Estimation

6.1 Assumptions for consistency

The estimation exercise of this paper gives consistent and unbiased results under certain assumptions on unobservables, which we discuss here.

The hotel accomodation is a complex good: it has many dimensions that matter for the consumer's utility, but not observed by the econometrician. It is plausible that such unobservable characteristics of quality - reputation, for example - will be correlated with price, creating a problem of endogeneity. In some models of discrete choice, like in Berry, Levinsohn, and Pakes (1993), it is possible to use instrumental variables approach. In this paper, where error terms enter the model in a non-linear way, it is much harder; therefore, we need to assume that the unobservable component of utility (1) is uncorrelated with observable hotel's characteristics.

This assumption is related to the issue of clicks versus bookings. In the data we observe clicks but not bookings, and in fact only a proportion of clicks result in a booking. As such, a click is a noisy indicator of booking, which brings us to the problem of measurement error in the dependent variable. In the discrete choice framework, such problem is called "misclassification", and it is known that it makes MLE estimates inconsistent (see Abrevaya, Hausman, Scott-Morton (1998)). In our model, when we observe that a consumer clicks on hotel A when hotel B is also available, we interpret it as a preference of A over B, $u_A > u_B$. A misclassification occurs when this relationship breaks, for example, when the click is made for reasons other than utility maximization (information gathering, for example). In such

[13] Conditionally of the search length, of course.

case, the direction of the bias will depend on what type of hotels receive most of the "wrong" clicks: if these are higher quality hotels, which is a plausible assumption, then we are going to under estimate price sensitivity. An explicit modeling of various motives for clicking would take care of it, but it is outside scope of this paper; thoughout, we assume that a click means revealed preference.

Note that the fact that a consumer may click but not book does not necessarily mean misclassification. Consider an example. When the user clicks on a hotel, she is redirected to another website where she can make a booking but also she can get more information about the hotel. It is possible that after learning more she changes her mind and does not book. To formalize this situation, suppose that in the utility model of a sort $u_{ij} = \alpha p_j + q_j + \varepsilon_{ij}$ - the error term has two compoments: $\varepsilon_{ij} = \eta_{ij} + \zeta_{ij}$. The first component is an idiosyncratic taste by consumer i for hotel j, known to the consumer but not to the econometrician. The second component is consumer's residual uncertainty about the pleasure from staying at that particular hotel, either due to lack of information or experience. Ex-ante, the expected utility of hotel i is larger than that of hotel k: $\alpha p_j + q_j + \eta_{ij} + E_i\left[\zeta_{ij}\right] > \alpha p_k + q_k + \eta_{ik} + E_i\left[\zeta_{ik}\right]$, so she clicks. Ex-post, when she learns ζ_{ij} from the booking website, this inequality may reverse, and the purchase is not made. However, this does not represent a problem for estimation as long as a click remains a preference indicator, and both $\eta_{ij}, E_i\left[\zeta_{ij}\right]$ are uncorrelated with (p_j, q_j).

Further, the "no click" action also represents a certain ambiguity. Indeed, there can be a number of possibilities, some of which can be correlated with options observed before. The consumer may decide to call the desired hotel directly; continue searching at another time; abandon the idea about the trip; etc. In an attempt to control for different reasons for "no click", we include parameters of request in the mean value of outside option. Otherwise, we are assuming that the unobserved component in the value of outside option is independent from taste shocks in hotel's utilities.

Finally, the fact that the estimation sample is highly selected on the basis of search strategy (which is endogenous), the results should be interpreted with caution. The fact that someone preferred price sorting (and then leaving website) to everything else means that the distributions of her unobservables are truncated, if viewed from the perspective of a more general model. What is needed is a model of choice of a search strategy, but computational difficulties prevent from constructing such a model in a fully rational framework. A possible solution is to approximate the continuation value of search by a flexible function of components of the state vector (beliefs, current best utility).

6.2 Simulations

To compute the expected utility increase in (12), we simulate the contents of the second page from the posterior distribution. Every simulated second page, indexed by s, consists of

15 vectors of utility components, $\{p^s_{(k)}, q^s_{(k)}, \varepsilon^s_k\}$, $k = 16..30$, $s = 1..MS$. The simulation is performed in three steps. First, I make a random draw of parameter a^s from the posterior belief, $G^1(a|\Omega_1)$. Second, given a^s, we make a price draw for each of the remaining (N-15) hotels, conditional on their qualities (see (4)). The hotels with the lowest 15 prices are then chosen as $\{p^s_{(k)}, q^s_{(k)}\}$, $k = 16..30$, as corresponding order statistics and their concomitants. Finally, the set of ε^s_k are just random draws from EV Type 1 distribution, which is maintained fixed throughout the estimation. At every step, inversion method is used to make sure that simulated prices are smooth functions of parameters; however, due to sorting, simulated qualities behave in a non-smooth way, creating problems for optimization that I discuss later. Since integration is performed over multdimensional parameter, we choose rather large number of simulations, $MS = 1000$. Given the content of every simulated second page, we can compute the implied best utility from that page:

$$
\begin{aligned}
U_2^{s*} &= \max\{u^s_{(16)}, .., u^s_{(30)}\} \\
u^s_{(k)} &= q^s_{(k)} + \alpha_p p^s_{(k)} + \varepsilon^s_k
\end{aligned}
$$

Having the set of simulated best utilities from the second page, $\{u^{s*}\}_{s=1}^{MS}$, we can numerically solve the equation:

$$
\int_{\bar{u}}^{+\infty} (U_2^* - \bar{u})\, dF_u(U_2^*) \approx \frac{1}{MS}\sum_{s=1}^{MS} \max\{U_2^{s*} - \bar{u}, 0\} = c
$$

- for every value of search cost, c, to obtain an approximation to the reservation value function, $\bar{u}(c)$. After substituting this function into likelihood contributions (13)-(15), we numerically integrate them with respect to search cost.

6.3 Optimization

To summarize, parameters to be estimated are: $(\alpha_d, \alpha_s, \alpha_p, \alpha_{pwd}, \overrightarrow{\alpha}_n, \overrightarrow{\alpha}_b, \mu_{out,} \overrightarrow{\mu}_o)$ from the utility model, $(\beta_{0,1}, .., \beta_{0,3}, \theta_1, .., \theta_3)$ from beliefs and (c_0, c_1) from the search cost distribution.

There are several difficulties with optimizing the likelihood function: first, due to the employed simulation method, the objective function is discontinuous with respect to beliefs parameters; second, there are lots of local minima; finally, but probably most importantly, the computation is quite slow (250 sec per evaluation). Facing these difficulties, we settled on the following optimization procedure: first, make a uniform random draw of N starting values from the space of discontinuous parameters, $(\beta_{0,1}, .., \beta_{0,3}, \theta_1, .., \theta_3)$; second, for every starting value use Newton BFGS method to optimize over smooth parameters; third, take the best point from the results of the previous step and use Nelder-Mead algorithm to optimize over discontinuous parameters. Then repeat Newton/symplex step over the best point until convergence is reached. This procedure exploits the fact that belief parameters take the bulk of the computation: if you change them, it takes 250 sec to evaluate the likelihood; if you fix them, it takes 30 sec.

One also needs to decide on the cases when belief parameters go too far from the data, so that the Bayesian posterior is numerically undefined. In these cases, we revert to the prior and keep track of the proportion of observations that experience this problem.

7 Estimation results

In this section we report estimation results from a succession of models, adding more structural elements at each step.

Models D1 and D2 in Table (11) present estimates from logit demand models. These are models of multinomial choice that try to explain observed clicking decisions, taking choice set as given. In D1, the choice set includes all available hotels, i.e. as if consumer possessed full information; these are estimates one would obtain in a more common situation when the actual choice set is unobserved[14]. In its turn, the model D2 brings in variation in the actual choice sets to help identify demand parameters. The difference in estimates between the two models reflects the importance of controlling for the incomplete information about available goods, that consumers often have. In this particular context, the incompleteness is coming from costly search; in other contexts, such as ticket sales, it could depletion of stock.

Models S_L and S_{NL} take another step and try to explain the formation of choice sets through a search process. In S_{NL}, consumers are assumed to know the true distribution of prices: their mean belief is obtained by regressing log-price on constant, star rating and distance to the city center, on a large dataset of prices. See Table (1) for regression results. In terms of the model outlined in Section 4, we are estimating a restricted version with $\beta_0 = (0.35, 0.23, -0.26)$, $\theta = 0$. The last restriction is due to the fact that consumers in S_{NL} do not perform Bayesian updating during search. This approach follows the tradition of the preceding literature on product search, where the beliefs structure was fixed at some data-driven level. In this way, we ask a question: given that consumers are extremely rational (e.g., they know true equilibrium), what can we say about their preferences and search costs?

The model S_L relaxes this requirement and allows consumers to have some uncertainty about price distribution and learn about it while searching. At the same time, in this model the econometrician does not observe these beliefs, contrary to what we have implicitly assumed in S_{NL}; therefore, the parameter β_0 is now estimated together with other parameters of the model. During the estimation, we impose a certain rationality constraint on beliefs, namely that the predicted average hotel price should be equal to its true value. This implies that

[14] Since we observe only actual choice sets, prices of other hotels are imputed based on observations by other consumers, with similar dates of search and parameters of request.

$\beta_0 \in S_\beta$ where the constrained set is:

$$S_\beta = \left\{ (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}) : \frac{1}{N} \sum \exp(\beta_{0,1} + \beta_{0,2} s_j + \beta_{0,3} d_j + \sigma_0^2/2) = \bar{p} \right\}$$

- where $\bar{p} = 228.39$, an estimate from a large dataset of prices. Such restriction is needed to obtain economically sensible estimates of beliefs.

Table (12) presents estimation results from search models, together with standard errors, numerically computed using information matrix. Since these models are nested versions of each other, a likelihood ratio test can be employed to compare their performance. Specifically, we test a hypothesis H0: $\beta_0 = (0.35, 0.23, -0.26), \theta = 0$, against Ha: $\beta_0 \in S_\beta, \theta > 0$, using likelihood ratio test. Its statistic is computed as LR=-2*(Lur - Lr), where Lur and Lr are log-likelihoods of unrestricted and restricted versions, respectively. In our data, LR=29.77, while 99% quantile of Chi-4 distribution is $q_{99} = 13.28$, which means that the null hypothesis is strongly rejected. In other words, we find that our data favors the model of search from unknown distribution.

To illustrate the quality of fit, in this table we have average deviances of page turning decisions, $-2\log(p(T_i|X_i))$, separately for page turners and no-turners:

	SNL	SL
turn	2.8079	2.8379
no turn	0.5663	0.5584

As we can see, all search models do much better job at explaining why people don't turn the page than otherwise. This is due to the fact that no-turners take about 75% of the sample. At the same time, the contribution of S_L over S_{NL} is mainly about why people do turn the page.

Now we turn to the discussion of estimates in Table (11) and (12). In doing so, we would like to focus on three quantities of interest: price coefficient, because it is the main factor of price elasticity (see the next section); combinations of beliefs and search costs, because of they illustrate the ability of this model to predict search decisions.

7.1 Discussion

Changes in price coefficient across models. One way to interpret the value of price coefficient is in relation to other characteristics of hotel, for example star rating. As the basic version of D2 suggests, an additional star brings about 0.84 utils. An equivalent increase in prices is then (0.84/2.85)*100=29.50 dollars, since prices in the utility model are measured in hundreds of dollars (see equation (1)). This seems to be a reasonable value: from the regression estimates in Table (1), the semi-elasticity of price w.r.t stars is 0.23; with an average price of

2-star hotel being $110, an additional star brings (exp(0.23)-1)*110 = 28.45 dollars of price increase.

The difference between estimated price coefficients in D1 and D2 is empirically explained by the availability of higher quality hotels. In the actual choice sets from which consumers make their purchases we have only the left tail of price distribution: hotels of low star rating, located far from the city center. In order to prefer the same hotels when better quality ones are available (in model D1), consumer must demonstrate much higher price sensitivity, which is what we observe. As a result, D1 over-estimates price elasticity relative to D2 by a significant amount, about 130%. More generally, this finding highlights the importance of accounting for actual availability of goods when estimating demand. The same conclusion is obtained in the recent paper Conlon and Mortimer (2008), which use availability data from vending machines.

Even though D2 performs better by incorporating information on actual choice sets, it ignores their endogeneity and therefore its results are biased, too. Indeed, comparing D2 to the search models, we observe a significant drop (in magnitude) of the price coefficient. In a search model, the role of the price coefficient becomes more involved, because together with search costs it serves to match page turning decisions[15]. Holding everything else constant, a smaller price coefficient produces two opposite effects on the probability of turning. On the one hand, it increases utility from the best hotel on the first page, making the consumer more satisfied and thus less willing to search; on the other hand, it also makes second page hotels more attractive, thus increasing benefit of search. In an environment where prices are sorted in the increasing order, one would expect the second effect to dominate: second page prices are higher, which means changes in price coefficient are applied to a larger base, which eventually leads to a relatively larger increase in utilities. So we are going to focus on the second effect, on benefits of search.

As we know from the model, the benefit of search is summarized by a reservation utility. To explain page turning, a model must predict a relatively high reservation utility for those consumers who turned the page, and a low one for those who didn't. In S_{NL}, where there is no Bayesian learning, price sorting produces variation of reservation utilities across consumers: those who saw higher maximal price on the first page will have relatively higher belief about prices on the second page. Changes in search cost parameters have uniform effect on reservation utilities of all consumers; therefore, they simply match the average propensity to search. In contrast, changes in price coefficient produce differential impact: from the above argument, a lower price coefficient induces more search, but relatively more so for people who observed higher truncation prices. In sum, price coefficient and search costs attempt to solve problem of matching between reservation utilities and observed page turning across consumers with

[15] Of course, all other coefficients of utilty play a role. However, the role of price coefficient is far more pronounced.

different truncation prices. The result is shown on Figure (9): the upper part are reservation utilities as functions of truncation prices, for various search models; the lower part are page turning probabilities induced by these reservation utilities, plotted against actual shares of page turners. Search cost is fixed at $c = 0.2$ throughout, and quantities are normalized for illustration purposes. Here, the model denoted by $S_{NL}(D2)$ is a search model with beliefs as in S_{NL} but with preferences kept at the level of D2; therefore, the difference between two curves on both plots is only due to demand parameters, and price coefficient in particular. We can see that for $S_{NL}(D2)$ the reservation utility is a quite steep function of truncation price; as a result, we see on the lower plot that according to $S_{NL}(D2)$ people who saw prices higher than 200 dollars should not turn the page. However, this prediction is inconsistent with the data, as shown by this figure and by Table (10). In reponse to this, we have an increase in the price coefficient, as in S_{NL}, which makes the reservation utility curve much flatter.

Empirically, there are two factors behind the fact that D2's estimates fails to predict enough search. First, 848 out of 1123 consumers don't turn the page, so D2's estimates are driven primarily by first-page choice sets, where a lot of these consumers choose outside option. Second, from 275 consumers who turn the page, only 62 do it successfully, i.e. they click on something at the second page where higher quality hotels are availabe. If, on the other hand, it were the case that a lot of consumers went to the second page and clicked there, we would see that D2 predicts too much search. Therefore, the sign of the bias due to endogeneity of choice sets cannot be predicted, it depends on the pattern of demand across search types.

From the perspective of a static demand model, these facts can be explained by concluding that consumers are quite price sensitive. From the point of the search model, clicking decisions should be explained conditionally on page turning: for example, if we observe someone who turned the page, she must be relatively less price sensitive, and vice versa. Therefore, explanations of the fact that people rarely click on second page have to put less weight on price sensitivity, and more on other factors, such as taste for quality.

While the price coefficient is an important determinant of benefit of search, it is a costly instrument for explaining search decisions, because of its other role as predictor of clicking decisions. In fact, any distortion relative to D2's estimates is costly, and as soon as belief parameters are free to vary - as in the search model S_L - we should expect some reversal of the price coefficient to the level of D2. And indeed, we see some of it in the estimates; as a result, the explanatory power of S_L is better than that of S_{NL} both in terms of clicking and page turning decisions. While higher search cost decreases reservation utilities in a relatively uniform fashion, the impact of price beliefs depends on the truncation price, which has ample variation across consumers, as documented by Table (10). If the truncation price is low, an increase in the mean belief has stronger negative impact on search than if the truncation price is high (to see why, think of two extreme cases). This is similar to the effect of price coefficient, with a difference that beliefs only affect search decisions, not purchasing.

Search costs, beliefs, learning. The estimates of search costs are quite similar across models, most likely due to the rationality constraint we imposed. In the basic specification, the mean search cost in S_{NL} is $E(c) = \exp(c_0 + c_1^2/2) = 0.38$, which can be interpreted in the following way: an average consumer is not going to turn the page if the expected increase in utility is lower than 0.35. In dollar terms, this change in utility can by caused by (0.38/1.00)*100 = 38 dollars of increase in the price of the best hotel from the first page. At the same time, there is a high variation of search costs among population.

Here we interpet this number as the cost of processing information on a page of results.

To compare these numbers with the actual benefits of search predicted by the model, we perform a small Monte Carlo experiment. Note that expected gain of search, $E\max(U_2 - U_1, 0)$ is unobserved for us, because of U_1 - best utility from the first page depends on taste shocks. Taking a consumer from our sample with a typical[16] vector of $\{\mu_0, .., \mu_{15}\}$, we make a draw of her taste shocks, $\{\varepsilon_0, .., \varepsilon_{15}\}$ from EV Type 1 distribution, which gives a value of U_1. Repeating this exercise many times, we obtain a distribution of U_1 for this consumer, and, as a result the distribution of $E\max$. Figure (8) plots this distribution against the estimated search cost distribution (ignore the dashed line for now). From this picture, we can see that the mean of search costs largerly overstates the actual gains of search predicted by the model: for the chosen consumer, the median gain is only 11 dollars, and in fact the mean search cost of 38 dollars corresponds to a 90% quantiles in the distribution of benefits. In other words, a single statistic of search cost distribution can be a misleading indicator of gains of search, and simply because only 25% of consumers actually turn the page.

Nevertheless, even the average expected gain of 11 dollars seems to be a rather large quantity, considering how little people search. Why does the model predict such high benefit of search? A small example may help understand this.

Suppose there is one good per page, with utilities $u_1 = \mu_1 + \sigma\varepsilon_1$ and $u_2 = \mu_2 + \sigma\varepsilon_2$, where ε_i are independent, Type 1 EV; consumer believes that $\mu_2 = \mu_1$, which roughly corresponds to beliefs in SNL. Now, if we want this consumer to turn the page with 25% probability, what does it imply about her search cost? First, we find her reservation utility, $\bar{u}$, from the equation: $F_{EV}(\frac{\bar{u}-\mu_1}{\sigma}) = 0.25$, to be $\bar{u} = \mu_1 - 0.33\sigma$; second, we substitute it into (12) to find $c = 1.02\sigma$. This result is quite intuitive: the value of an option to turn the page is positively related to the variance of search results. In our model, where $\sigma = 1.28$, this implies a cutoff value of search cost as $c = 1.30$. Although the dollar value of search cost will depend on the price coefficient, this example shows that high values of search cost are not surprising given our assumption on the variance of match value. As an illustration, consider again Figure (8),

[16] Typical in the sense of predicted distribution of gains of search. The average gain of search from this consumer is equal to its average level in the sample.

curve "Benefits 80%": this is CDF of expected gains from search where standard deviation of match value is reduced by 20% (keeping other parameters constant). As a result, benefits from search drop sharply: the median benefit is about 5 dollars now, and almost no one will turn the page if the search cost is more than 20 dollars.

A natural question is, can we identify σ together with other parameters of the model? The answer is no, because utilities are identified up to a shift and scale parameters. At the same time, if we divide equation $u = \alpha_p p + \alpha_q q + \varepsilon$ by price coefficient: $\tilde{u} = p + \tilde{\alpha}_q q + 1/\alpha_p \varepsilon$, then σ is identified as inverse of the price coefficient. This gives another way to look at the role of this coefficient we've discussed above: lower price coefficient is equivalent to higher variance of match value, which increases benefit of search.

Turning to the learning component of the model, the variance of the prior in S_L is small, but precisely estimated: as we've shown above, the model with learning is stronly preferred by the data. Also, small variance of the prior doesn't mean little learning. Even though the value of prior variance indicates a high confidence that consumers have in their beliefs, the informativeness of the first page - which contains the lowest 15 order statistics from the sample of 148 normal random variables - is also very high. Together with ample price variation on the first page, as documented in Section (3), this leads to a meaningful variation in posterior beliefs. On Figure (10), we illustrate an empirical distribution of average hotel price, predicted by posterior beliefs across consumers. While most of posterior means are between 200 and 250 dollars, some consumers update to as high as 350 dollars. This means that updating of beliefs plays an active role in the identification of the model: we see that the estimated prior has has shifted away from the price regression, its much flatter now; also, search cost estimates have somewhat decreased, but perhaps more importantly, their variance decreased as well.

8 Price elasticity of demand

In the hotel market, the relevant definition of the good is a stay at a given hotel at given dates. For a consumer, the decision to purchase this good will be determined both by her preferences and by her choice set, generated during the search process. Because of unobserved heterogeneity in preferences and in search costs, the individual demand for good X is defined as a probability that a search by a randomly picked consumer will result in the purchase of this good. This probability is conditional on extisting prices and hotel availability on the market, both of which determine the content of the first and the second price sorted pages observed the a searcher. In the estimation sample, such data is available only for consumers who turned the page, which means that we can compute individual demand elasticities only for 275 out of 1123 of search histories in the sample.

Then we define two supplementary notions of demand, depending on the level of aggregation. One the one hand, one can break down the search induced demand into a fresh, returning and residual demand. Fresh demand is a probability that a consumer will click on a given hotel on the first page and doesn't turn the page; the returning demand is a probability of turning the page, and then going back to click on the first page; the residual demand is a probability to click on a given hotel on the second page. Distinction between these types of demand is helpful in evaluating the role of the position of hotel in search rankings on elasticity of its demand. On the other hand, one can aggregate demand for all hotels into a market demand, which is the probability of clicking on any hotel in the choice set, or one minus the probability of choosing the outside option.

The individual demand elasticities are obtained by increasing the price of a hotel on the first page by 1% and the relevant probabilities before and after the disturbance. The identity of the disturbed hotel is fixed both across the models, and across consumers. For that purpose, I have chosen hotel "Extended stay America", 2 stars, 15 miles from the city center, neighborhood of O'Hare - a typical hotel that often appears on the first page. Table (2) presents elasticities, averaged over consumers in the sample. These are estimates of unconditional demand elasticities; in other words, they represent an expected change in demand of a random consumer in response to price increase by 1%, where the expectation is taken over possible search histories.

From Table (2) we see that new structural elements added to every model beginning D1, lead very different demand elasticities, both in material and statistical sense. Starting from D1, the model D2 adds information about actual (and endogenous) choice sets; in S_{NL}, we add the search component where beliefs are fixed in ad-hoc fashion; in S_L, we free up belief parameters and estimate them, allowing consumers to learn while searching.

Table 2: Average elasticities of demand for a hotel on the first page of results, nobs=275. The second row are standard errors, computed by delta method. In D2, the choice set consists only of the first page.

Mdl	ind	out	fresh	return
D1	-3.0387	0.0341		
	0.1516	0.0029		
D2	-2.2012	0.0399		
	0.1789	0.0040		
SNL	-0.7199	0.0110	-0.7277	-0.7128
	0.0345	0.0036	0.1842	0.2124
SL	-0.7498	0.0140	-0.7501	-0.7478
	0.0832	0.0037	0.2323	0.2041

On one end of the distribution, we have D1-D2 with very high elasticites, between 2 and 3, which is probably not very realistic. The elastictiy of demand in D1 is much higher than in D2. While most of the difference is explained by the price coefficient, another argument is the limited nature of choice sets in D2. It is straightforward to show that, with logit demand, more alternatives make consumers more price elastic, holding taste parameters constant. Indeed, the own price elasticity is found as (assuming $\alpha_p = -1$):

$$\frac{dq_i}{dp_i}\frac{p_i}{q_i} = -p_i \frac{\sum_{j \neq i} \exp(\mu_j)}{\sum_{j=1}^{n} \exp(\mu_j)}$$

- which is increasing in absolute value with n. As the number of firms increases, this has two effects on elasticity. On the one hand, there is a "price-sensitivity effect", when the demand curve $q_i(p_i)$ becomes steeper because more alternatives are available; on the other hand, there is a "market share effect" due to lower quantity of sales per firm. With EV Type 1 distribution of error terms, the second effect dominates; however, this may not be true for in other cases, as discussed in Chen and Riordan (2008). On the other end, we have the model S_{NL} with an elasticity about 60 percent lower than in D2 in the expanded specification, and about 200 percent lower in the basic specification.

9 Conclusions

In this paper we have estimated a structural model of sequential search for hotels online. Using its results, we have attempted to show that accounting for search frictions matters for estimation of consumer demand. One implication of costly search is that it results in limited choice sets. If we ignore this and assume that consumers had full information about available options when making a purchase, our estimates may be significantly biased. In our model, we find that the full information assumption leads to over-estimation of price elasticity by more than 200%, relative to results from a search model. This results confirms a general intuition that consumers are less price sensitive when choices are more limited. Further, from the perspective of consumer decisions, incomplete information about available goods is equivalent to incomplete availability. From the perspective of estimation, however, there is an important difference: while it may be argued that incomplete availability is exogenously given, as in Conlon and Mortimer (2008), the incomplete information is not. According to the search model, a consumer stops collecting information if she is satisfied enough with what was found so far, which is a statement about her preferences. Therefore, choice sets are not only limited, but also endogenous to preferences. In our model, we find that if we account for actual choice sets but ignore their endogeneity (e.g., the search process), we over-estimate price elasticity by more than 50%. Unfortunately, the sign of the bias is specific to our dataset and cannot be determined apriori.

Besides their implications for estimation of demand, our results from the search model are of separate interest. In this paper, we proposed and implement a way to identify consumer's prior beliefs, separately from search costs and preferences. Bayesian learning is the key mechanism that helps achieve such identification: it creates variation in posterior beliefs that's orthogonal to variation in search costs. Although in the current specification freeing up beliefs does not change demand estimates very much, we believe that in general it is important to do so, because according to the search model beliefs are one of the key components explaining search. If they are fixed at some arbitrary level, as it is done in the previous papers, the extra burden of explaining search decisions lies on preferences and not search costs. In our model, we find that consumers have relatively flat priors about conditional price distribution, as compared to regression results. The estimated amount of prior uncertainty is small, but since the informativeness of the first page of hotels is very high, we find a meaningful amount of Bayesian learning. The median search cost is around 25$ per 15 hotels, or 1.7 dollars per hotel. This is much smaller than what was found in some previous studies, such as Hong and Shum (2003) and Hortacsu and Syverson (2004).

On the point of model's fit, we find that it does a good job at predicting average search intensity, but a very modest job at picking heterogeneous incentives to search among consumers. For example, for a range of plausible search cost values, the means predicted page turning probability, computed separately for page turners and no-turners, are within 2% from the average search intensity. This result points to some limitations of our model, which provide directions for future research. First, extend the length of the search process to three pages and introduce other search strategies, such as sorting by distance and star rating. This woud allows to include many more observations in the sample and alleviate the selection problem. Second, increase the amount of consumer heterogeneity, both in priors and in the search cost distribution. Currently, there are two sources of heterogeneity: parameters of request and contents of the first page.

Despite these limitations, our paper makes the first step in estimating structural models of search with learning, when detailed data on search histories is available.

References

[1] Abrevaya, J., Hausman, J., Scott-Morton, F. (1998), "Misclassification of the dependent variable in a discrete-response setting", Journal of Econometrics 87 (1998)

[2] Ackerberg, D. (2003): "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Examination," International Economic Review, 44, 1007–1040.

[3] Anderson, S.P., DePalma, A., and Thisse, J.-F (1989), "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach." Review of Economic Studies, Vol. 56 (1989), pp. 21-35.

[4] Anderson, S. and Renault, R., (1999). "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," RAND Journal of Economics, vol. 30(4)

[5] Berry, S, Levinsohn, J., and Pakes (1993). "Automobile Prices in Market Equilibrium." NBER Working Paper no. 4264, 1993.

[6] Bhattacharya, P.K. (1974), "Convergence of Sample Paths of Normalized Sums of Induced Order Statistics", The Annals of Statistics, Vol. 2, No. 5. (Sep., 1974), pp. 1034-1039

[7] Bhattacharya, P.K (1984), "Induced order statistics: theory and applications", in Handbook of Statistics, vol 4

[8] Bikhchandani, Sushil and Sharma, Sunil, (1996), "Optimal search with learning", Journal of Economic Dynamics and Control, vol 20, 1996

[9] Chen, Yongmin, and Michael Riordan (2008), "Price-increasing competition", discussion paper 0506-26, Columbia University

[10] Conlon, Christopher and Julie Mortimer (2008), "Demand estimation under incomplete product availability", working paper

[11] Crawford, Gregory and Matthew Shum (2006), "Uncertainty and learning in pharmaceutical demand", Econometrica, Vol. 73, No. 4 (July, 2005), 1137–1173

[12] David, H.A and Nagaraja, H.N (2003), "Order Statistics", 3-d edition

[13] de los Santos, Babur (2008), "Consumer search on the internet", PhD dissertation, Indiana University

[14] Diamond, P., (1971), "A model of price adjustment", Journal of Economic Theory, 3, (1971)

[15] Diamond, P.A., Maskin, E. (1979), "An equilibrium analysis of search and breach of contracts, I: steady states", Bell Journal of Economics, vol. 10

[16] Fox, Jeremy (2007), "Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices", RAND journal of economics, vol 38 (4),

[17] Hong, H. and Shum, M. (2003), "Can search cost rationalize equilibrium price dispersion in online markets?", Rand Journal of Economics, Summer 2006, vol. 37, no. 2, pp 258–276

[18] Hortacsu, 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", Quarterly Journal of Economics , v.119, May 2004, p.403-456

[19] Johnson, Eric, Wendy W. Moe, Peter S. Fader, Steven Bellman and Gerald L. Lohse (2004) "On the Depth and Dynamics of On-line Search Behavior," Management Science 50-3, 299-308, 2004.

[20] Kohn, M., Shavell,S., (1974), "The Theory of Search," Journal of Economic Theory, vol 9, 1974

[21] Moraga-Gonzalez, Jose and Matthijs Wildenbeest, (2008), "Maximum likelihood estimation of search costs", European Economic Review, 52

[22] Morgan, Peter (1985), "Distributions of the Duration and Value of Job Search with Learning", Econometrica, Vol. 53, No. 5. (Sep., 1985)

[23] Roberts, John H and James Lattin (1997), "Consideration: Review of research and prospects for future insights", Journal of Marketing Research, Vol. 34, Issue 3

[24] Rosenfield, D, and Shapiro, R., (1981),"Optimal adaptive price search", Journal of Economic Theory. Vol. 25, no. 1, pp. 1-20. 1981

[25] Rothschild, Michael (1974), "Searching for the Lowest Price When the Distribution of Prices Is Unknown". Journal of Political Economy, Vol. 82, No. 4. (Jul. - Aug., 1974), pp. 689-711

[26] Sailer, Katharina (2006), "Searching and Learning in Internet Auctions The eBay Example", PhD dissertation, University of Munchen

[27] Sorensen, Alan (2001), "An Empirical Model of Heterogeneous Consumer Search for Retail Prescription Drugs", NBER Working Paper No. 8548

[28] Talmain, Gabriel (1992), "Search from an unknown distribution: an explicit solution", Journal of Economic Theory, vol 57, 1992

[29] Tierney, L. and Kadane, J. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities" Journal of the American Statistical Association, Vol. 81, No. 393. (Mar., 1986), pp. 82-86.

[30] van den Berg, G. and Ridder, G. (1999), "An Empirical Equilibrium Search Model of the Labor Market," Econometrica 66, 1183-221, 1998.

[31] Wolinsky, A. (1987), "Matching, search and bargaining", Journal of Economic Theory, vol. 42

10 Appendix. Tables and figures.

Table 3: Most popular search strategies on kayak.com. The first column is number of searches that used a strategy; the second is click rate; then used actions, in preserved order. Note: 7029 people did nothing, 3499 only flipped pages

N	CTR	cause1	cause2	cause3	cause4	cause5	cause6	cause7
7029	0.250							
1169	0.290	page						
899	0.344	page	page					
869	0.367	sortprice						
708	0.422	landmark						
427	0.389	page	page	page				
407	0.435	page	page	page	page			
289	0.374	page	page	page	page	page		
271	0.295	filter_price						
256	0.297	sortprice	page					
242	0.248	filter_dist						
230	0.317	sortdist						
188	0.431	landmark	filter_dist					
176	0.415	page	page	page	page	page	page	
174	0.293	nbhd						
150	0.340	sortprice	page	page				
132	0.417	page	page	page	page	page	page	page
132	0.402	hotelname						
117	0.342	filter_price	filter_price					
114	0.360	landmark	page					
102	0.284	sortrating						
14081		subtotal						
10240	0.376	less than 100 searches						
24321	0.377	total						

Table 4: Distribution of non-price characteristics of Chicago hotels, by the number of establishments (stimation sample)

c	n	nbhd	n	stars	n
null	34	Chinatown	3	1	9
best western	7	Gold Coast, Old Town	51	2	40
hampton inns	6	Loop	22	3	55
holiday inn hotels	6	SW	15	4	42
marriott (all)	6	midway	12	5	2
hilton (all)	5	north side	21		
super 8 motels	5	ohare	20		
comfort inns	4	west side	3		
hyatt (all)	4				

Table 5: Comparing samples by parameters of request. First row is the mean, second row is standard deviation of the variable. The right column is t-test comparing means of group1 and group4

mean/sd	group1	group2	group3	group4	T1-4
obs	17292	2664	1436	1123	
advance	18.86	17.56	17.03	16.02	7.64
	11.79	12.06	12.14	12.12	
weekend	0.59	0.58	0.58	0.56	1.94
	0.49	0.49	0.49	0.50	
N days	2.41	2.41	2.37	2.32	1.66
	1.61	1.74	1.67	1.61	
N rooms	1.07	1.07	1.07	1.07	0.06
	0.25	0.25	0.25	0.25	
N guests	1.84	1.81	1.81	1.82	0.70
	0.98	0.99	1.00	1.02	

Table 6: Logit estimates of click rates, depending on type of request. The first row is the coefficient, the second is t-statistic.

coef/t	group1	group2	group3	group4
advance	0.0066	-0.0001	-0.0041	-0.0079
	4.73	-0.04	-0.85	-1.43
N days	-0.0070	0.0189	0.0763	0.1219
	-0.67	0.77	2.17	2.87
weekend	-0.0786	-0.0019	-0.0421	-0.1190
	-2.33	-0.02	-0.36	-0.90
N rooms	-0.0820	-0.2410	-0.0534	-0.1818
	-1.22	-1.35	-0.23	-0.67
N guests	0.1240	0.1900	0.1674	0.1538
	7.04	4.29	2.78	2.30
N pages	0.0297	0.0482	0.0341	-0.1112
	6.41	3.84	0.82	-0.89
const	-0.8706	-0.8003	-0.9668	-0.6292
	-11.17	-4.04	-3.65	-1.87

Table 7: Characteristics of clicked hotels in groups1-4: mean values and standard deviations on stars and distance, and percentage shares for neighborhoods.

	group1	group2	group3	group4
stars	3.07	2.74	2.59	2.45
	0.88	0.89	0.85	0.80
dist	5.00	7.60	9.26	10.95
	6.12	6.40	6.10	5.29
Chinatown	4.02	4.17	2.31	3.00
Gold Coast	35.58	21.23	14.81	8.25
Loop	22.99	17.76	14.00	8.25
SW	7.57	16.27	22.31	27.75
midway	5.28	4.56	3.65	3.25
north	6.02	7.34	7.69	7.75
ohare	17.21	27.18	34.00	41.75
west side	1.33	1.49	1.15	0.00

Table 8: Observed and clicked hotels in the choice sets, % of total

stars	obs	click	nbhd	obs	click	dist	obs	click
0	0.00	0.00	Chinatown	2.02	3.00	0-5	10.09	20.25
1	10.85	4.50	Gold Coast	4.97	8.25	6-10	2.04	2.50
2	48.91	61.50	Loop	2.44	8.25	10-15	79.23	73.50
3	27.56	19.00	SW	20.59	27.75	16+	8.65	3.75
4	12.65	15.00	midway	10.56	3.25			
5	0.02	0.00	north side	16.10	7.75			
			ohare	43.16	41.75			
			west side	36.00	0.00			

Table 9: Observed and clicked hotels on the first and the second page of results

	page1		page2	
stars	obs	click	obs	click
1	13.29	5.61	0.55	0.00
2	53.78	69.47	28.35	30.65
3	22.66	13.40	48.29	50.00
4	10.28	11.53	22.80	19.35
nbhd				
Chinatown	1.8	2.8	2.97	1.61
Gold Coast	3.41	4.67	11.55	19.35
Loop	1.51	2.18	6.38	27.42
SW	19.69	32.71	24.4	9.68
midway	9.05	1.87	16.94	11.29
north side	16.54	7.17	14.19	12.9
ohare	47.96	48.6	22.81	17.74
west side	0.04	0	0.76	0
dist				
0-5	6.95	9.66	23.32	56.46
6-10	0.77	0.62	7.43	9.68
11-20	92.27	89.72	69.24	33.87

Table 10: Summary statistics of maximal prices on the first page, observed by turners and no turners

	percentiles		max 4 prices	
%	no turn	turn	no turn	turn
1	90	90	469	396
5	92	92	479	419
10	95	95	529	421
25	98	98	567	429
50	103	104		
75	125	130		
90	199	179		
95	265	209		
99	409	419		
Mean	127.34	124.69		
Std. Dev.	62.7	50.82		
Skewness	3.24	3.41		
Kurtosis	15.31	17.75		
Obs	848	275		

Table 11: Estimation of logit models of demand.

	basic				with brands and requests			
var	D1	std	D2	std	D1	std	D2	std
dist	-0.55	0.15	-0.74	0.18	-0.66	0.19	-0.71	0.19
stars	1.06	0.09	0.84	0.09	1.01	0.11	0.70	0.10
price	-3.91	0.20	-2.85	0.23	-3.14	0.25	-1.55	0.27
n1	0.82	0.31	-0.41	0.43	-0.09	0.44	-0.51	0.44
n2	-0.58	0.29	-1.00	0.48	-0.36	0.44	-0.62	0.44
n3	1.15	0.32	-0.61	0.46	0.69	0.44	-0.05	0.44
out	2.75	0.23	0.41	0.29	2.90	0.48	2.21	0.44
b1					0.74	0.26	0.74	0.28
b2					-1.22	0.30	0.40	0.26
b3					0.86	0.23	0.62	0.23
b4					1.23	0.19	1.37	0.18
b5					2.08	0.24	2.09	0.22
out_nppl					-0.23	0.13	-0.26	0.13
out_wnd					-0.07	0.24	0.24	0.25
out_adv					0.09	0.13	0.17	0.13
price_wnd					0.08	0.07	-0.13	0.14
logL	2115.84		1711.86		1997.19		1645.64	
star	27.17		29.50		32.26		45.27	

Table 12: Estimation results from search models. Results reported only for the final specification. The last three rows are dollar values of additional star, of median search cost and its sdandard deviation.

Var	SNL	sd	SL	sd
dist	-0.63	0.18	-0.60	0.18
stars	0.47	0.08	0.50	0.08
price	-1.00	0.16	-1.05	0.17
n1	-0.42	0.41	-0.48	0.42
n2	-0.50	0.41	-0.58	0.42
n3	0.19	0.41	0.07	0.42
out	2.53	0.31	2.53	0.30
b1	0.68	0.26	0.67	0.26
b2	0.49	0.24	0.47	0.24
b3	0.62	0.22	0.63	0.22
b4	1.45	0.17	1.43	0.17
b5	2.40	0.18	2.34	0.18
out_nppl	-0.13	0.16	-0.17	0.15
out_wnd	0.31	0.25	0.30	0.25
out_adv	0.10	0.13	0.12	0.13
price_wnd	-0.16	0.16	-0.16	0.16
c0	-1.31	2.24	-1.38	1.87
c1	0.82	2.13	0.81	1.52
beta0	0.31		0.46	
beta1	0.23		0.12	0.06
beta2	-0.26		-0.02	0.05
theta	0.00		1.23E-04	7.80E-06
logL	2273.01		2258.13	0.00
star	47.20		48.08	
cost_med	26.89	60.16	23.94	44.67
cost_sd	32.11		28.02	

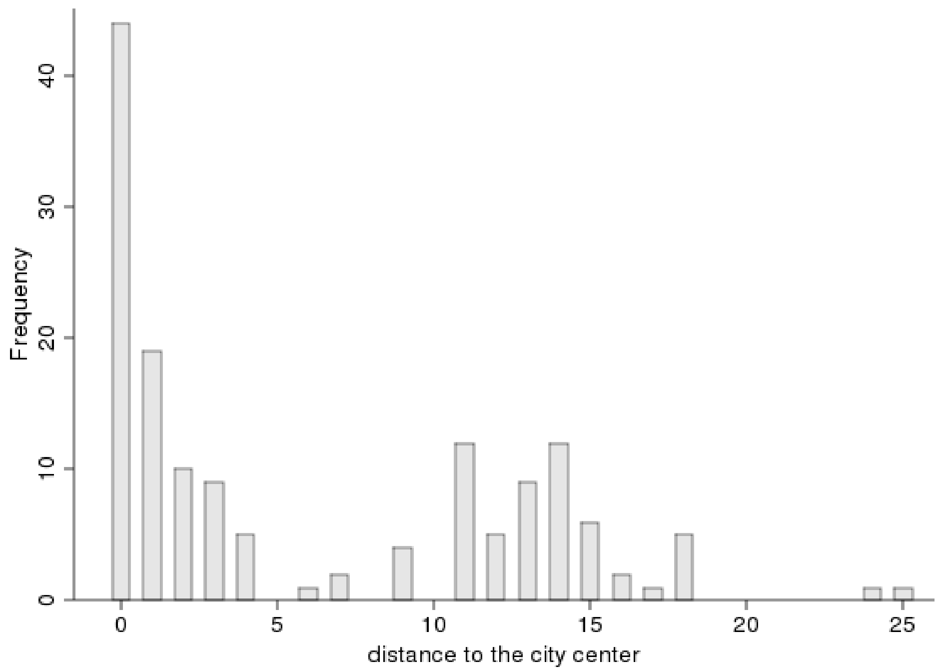

Figure 1: Distribution of distances to the city center (in miles) of Chicago hotels (estimation sample)

Figure 2: Geographical position of Chicago hotels

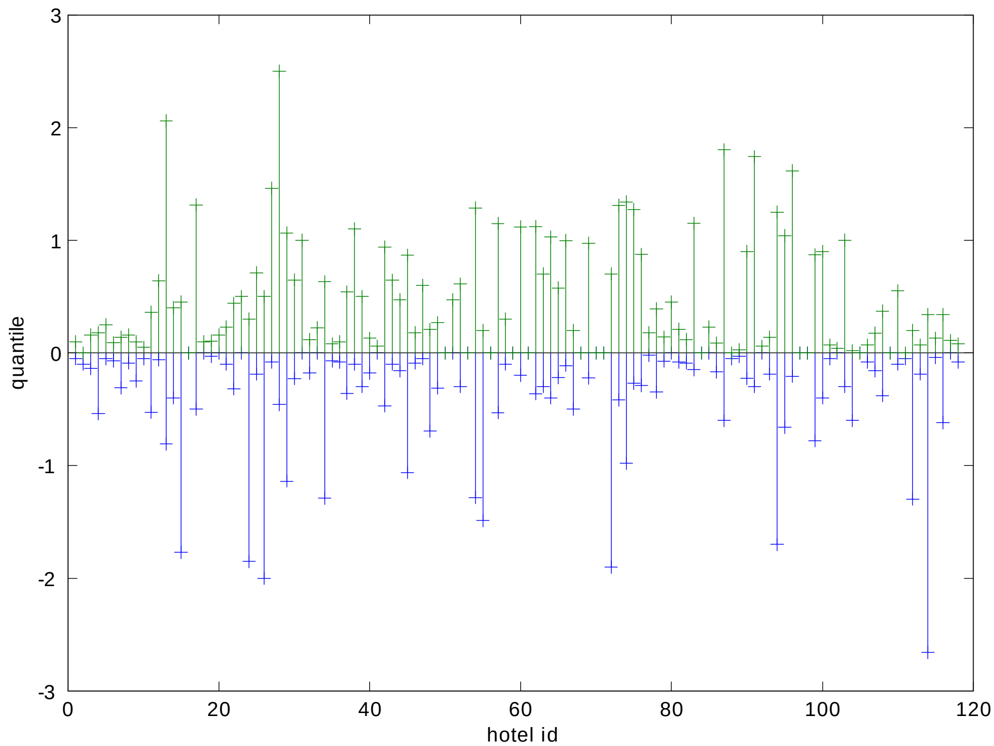

Figure 3: Quantiles (10%, 90%) of hotel prices, observed on the first page (for each hotel separately). The y-axis is measured in hundreds of dollars. Centered around median.

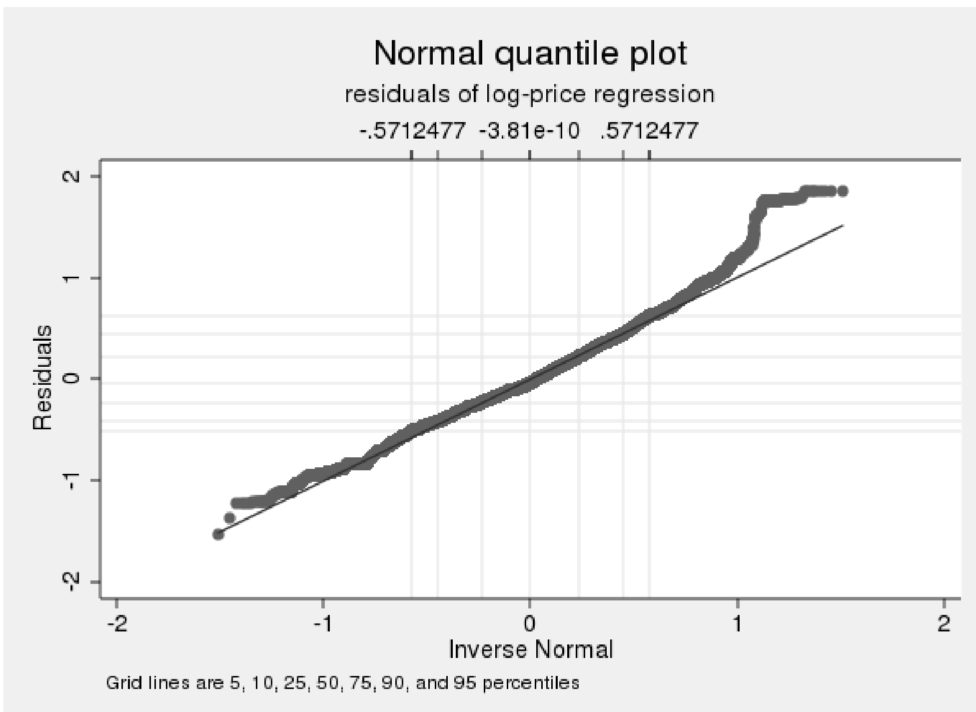

Figure 4: Normal quantile plot, residuals from regression of logarithm of price on distance and star rating.

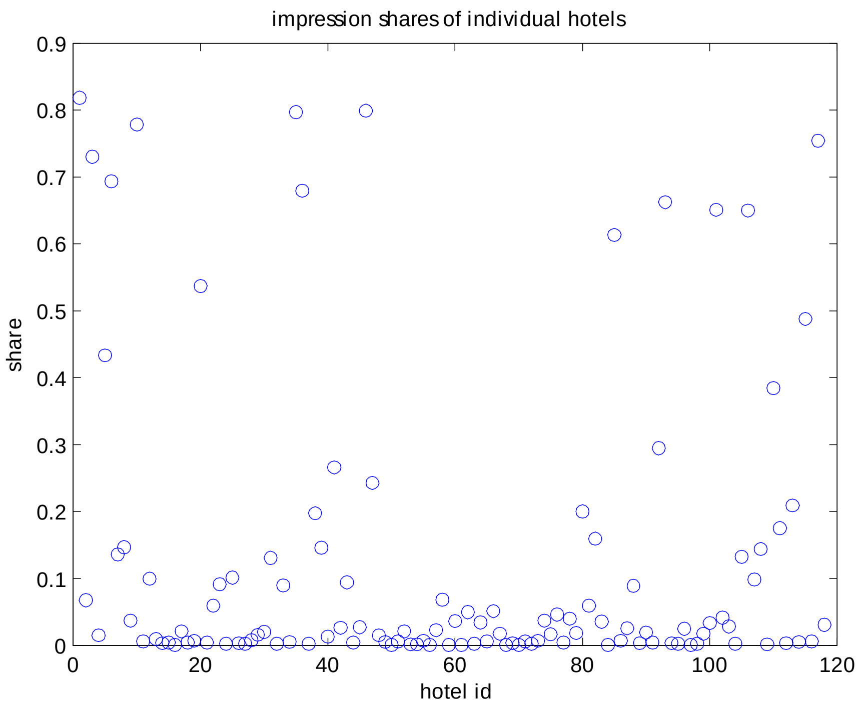

Figure 5: Proportions of first pages at which individual hotels were displayed, estimation sample. In total, 118 hotels on 1123 observed first pages.

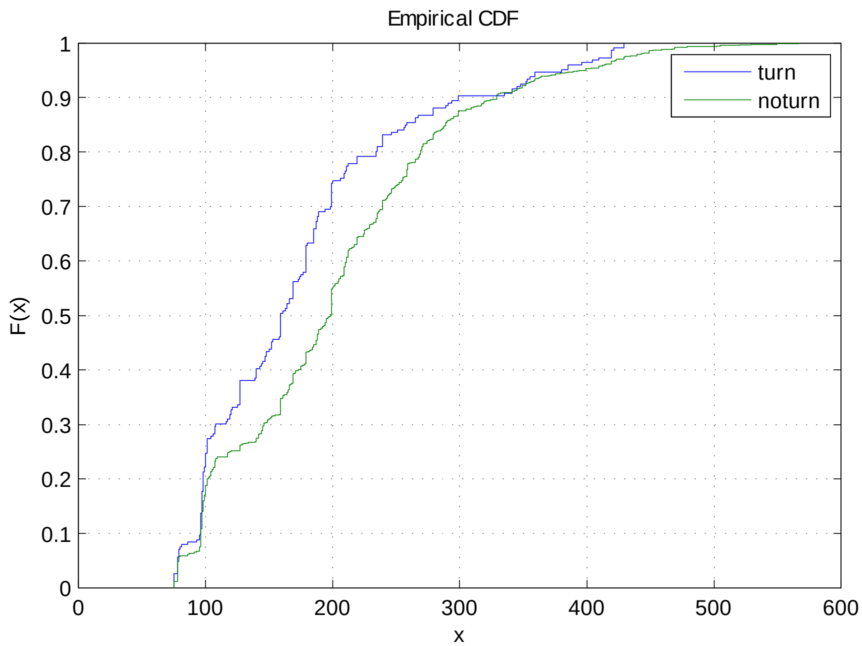

Figure 6: Empirical CDF of prices of hotels <5 miles from city center, seen on the first page by turners and no-turners

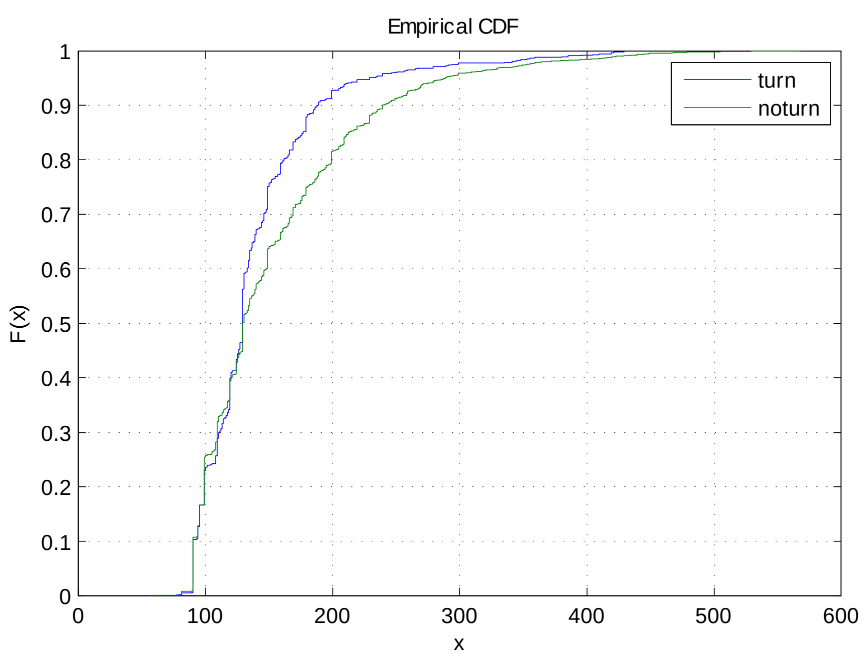

Figure 7: Empirical CDF of prices of hotels with 3 stars and more, seen on the first page by turners and no-turners

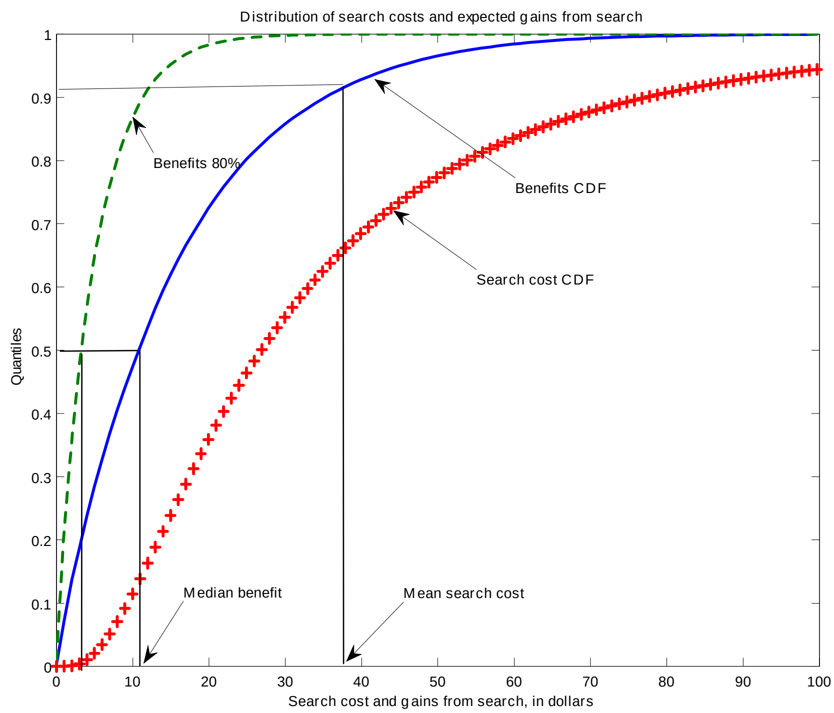

Figure 8: Distribution of search costs and gains from search for a representative consumer, model SNL (basic specification)

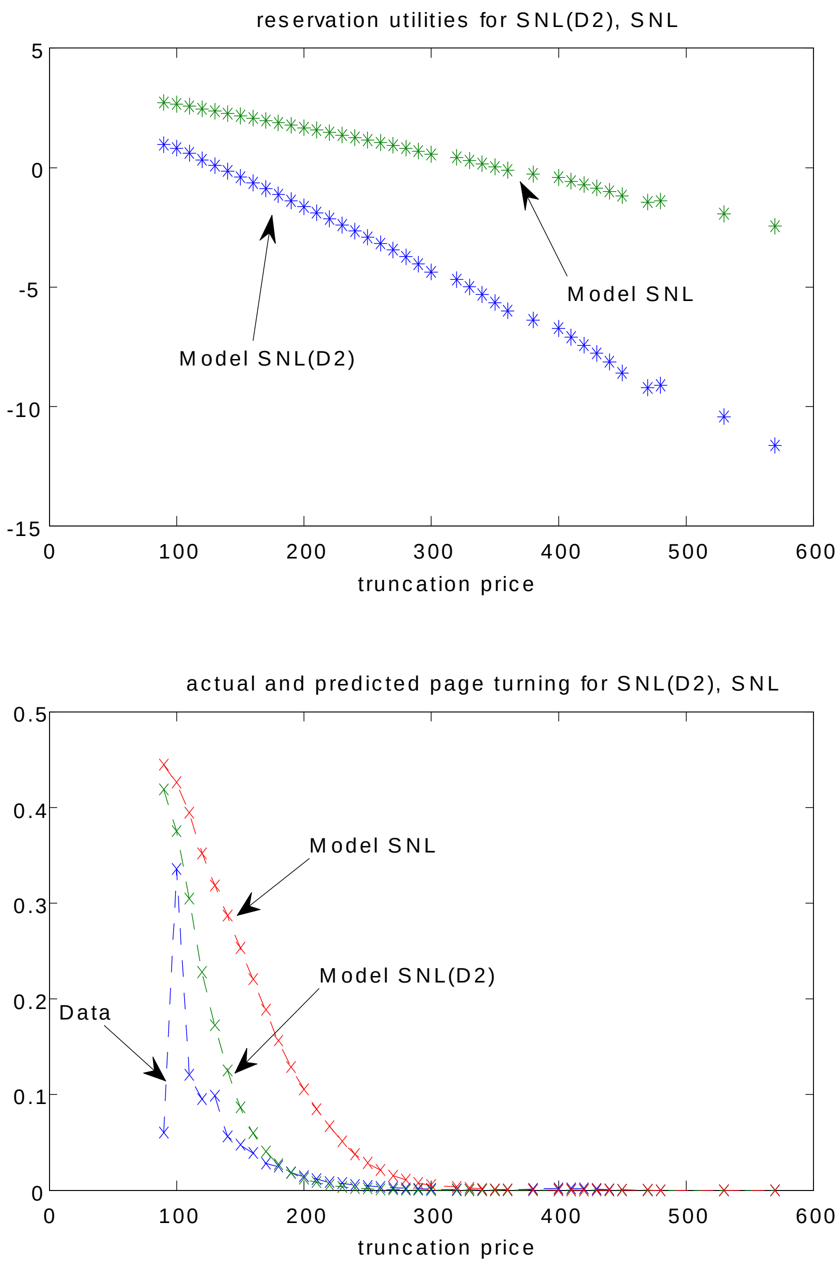

Figure 9: Reservation utitlities and predicted probabilities of page turning for various models.

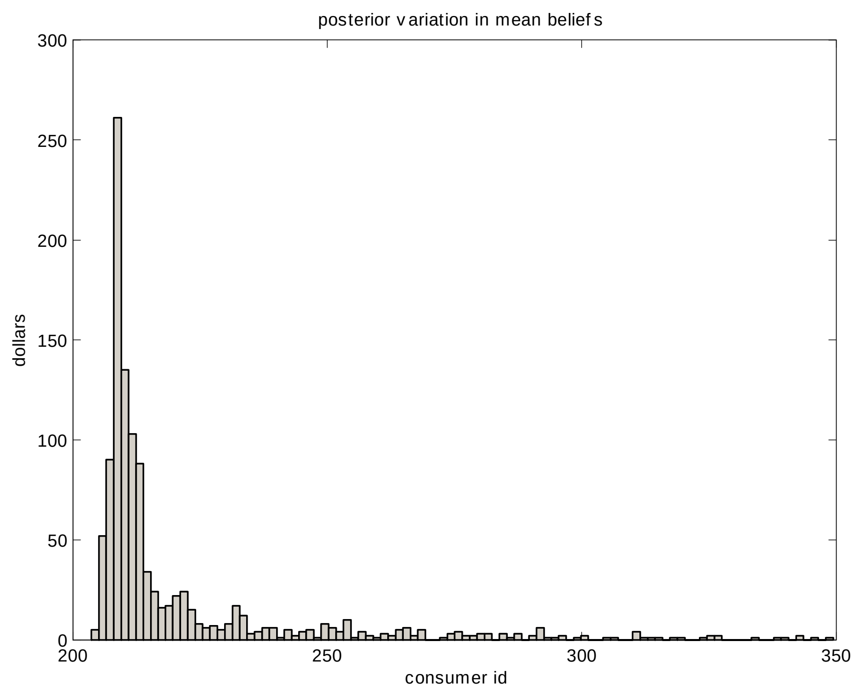

Figure 10: Histogram of average hotel prices, predicted by posterior beliefs across consumers (after they observe the first page).