Search Duration*

Raluca M. Ursu[†] Qingliang Wang[‡] Pradeep K. Chintagunta[§]

Current version: June 20, 2017

PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CITE OR CIRCULATE WITHOUT THE AUTHORS' PERMISSION

Abstract

In studying consumer search behavior, researchers typically focus on which products consumers add to their consideration set (the extensive margin of search). In this article, we attempt to additionally study how much consumers search individual products (the intensive margin of search), by analyzing the time they spend searching (search duration). Using data on consumers searching for restaurants on an Asian review website, we document that search duration is considerable: most consumers search few restaurants, but the average time spent searching is 3.47 minutes. We also find that restaurants that are searched longer are more likely to be purchased and the more restaurants searched in a session, lesser is the time spent on any given one; together suggesting that the time spent is both a benefit and a cost to consumers. We develop a sequential search model in which consumers who are uncertain about the quality of a restaurant on the list page generated by the search, click on that restaurant to learn about the restaurant's quality. The restaurant's page then provides (noisy) signals about the restaurant's quality that consumers use to update their beliefs in a Bayesian fashion. The more time spent on the restaurant's page, the more information consumers can gather. However, at each time point, consumers need to decide whether to continue on that restaurant's page; return to the list page and click a different restaurant; or make a purchase decision. The model provides optimal search rules for the full set of decisions made by consumers during search: which products to search, how much time to search each product and whether to purchase. Two features distinguish this model from other search models: (i) ability to estimate not only consumer expected utility before search, but also estimate preferences for features of the information discovered through search, and (ii) quantify search costs in terms of consumers' opportunity cost of time. Our approach provides a general framework to study consumer engagement with a product

*We are thankful for comments from Daria Dzyabura, Tulin Erdem, Pedro Gardete, Wesley Hartmann, Xiao Liu, Harikesh Nair, Sridhar Narayanan, Navdeep Sahni, Stephan Seiler, Russell Winer, and seminar participants at Stanford University and attendees of the 2017 Marketing Science/INFORMS conference for their thoughtful comments. The usual disclaimer applies.

†Stern School of Business, New York University, E-mail: rursu@stern.nyu.edu.
‡Xi'an Jiaotong University, E-mail: wangqliang123@xjtu.edu.cn. Corresponding author.
§Booth School of Business, University of Chicago, E-mail: pradeep.chintagunta@chicagobooth.edu

through search, and is also able to capture decisions such as revisits to a previously searched product to resolve further uncertainty.

Keywords: online consumer search, search duration, revisits, sequential sampling, optimal search rules, online reviews.

1 Introduction

Nowadays, consumers have access to a plethora of information, especially online. This additional information allows consumers to make better or more informed choices. At the same time, paying attention to this information is costly. To understand how consumers make choices in such an environment, previous work has focused on which products consumers add to their consideration set before making a purchase decision, which we will refer to as the extensive margin of search. However, relatively little is known about how much consumers choose to search individual products, or what we will refer to as the intensive margin of search. Examples of such search decisions include the decision to spend time searching a product, the decision to revisit a previously searched option to resolve further uncertainty about it, etc.

In this article we attempt to fill this gap, by accomplishing two goals. First, we provide data patterns on consumer search on both the intensive and the extensive margin. We do so by analyzing consumer search for restaurants on an Asian review website where we observe which products consumers clicked (extensive margin), how much time they spent searching each product (intensive margin), and which products they purchased, if any. We find that even when consumers search very few products, they spend a considerable amount of time searching each product (on average 3.47 minutes). In addition, we find that search duration is related to purchases, so that restaurants that consumers spend more time searching are also the ones they purchase.

Second, we develop a sequential search model in which consumers search for the unobserved quality of a restaurant on the list page by clicking to observe additional information on the restaurant page. Search provides (noisy) signals about the restaurant's quality that consumers use to update their beliefs in a Bayesian fashion. The more time spent on the restaurant's page, the more information consumers can gather. However, at each stage, consumers decide whether to continue spending time on that restaurant's page; return to the list page and click a different restaurant; or make a purchase decision. This model captures both the intensive and the extensive margins of search and provides optimal search rules for the full set of decisions made by consumers during search: which products to search, how much time to search each product and whether to purchase. To model search duration, we interpret time spent searching a product as multiple searches of the same product. This approach allows us to use the sequential sampling theory developed by Chick and Frazier (2012) to characterize consumers' optimal search rules in this more general search framework. In addition, our model has two features that distinguish it from other search models: (i) ability to estimate not only consumer expected utility before search, but also estimate preferences for information discovered through search, and (ii) quantify search costs in terms of consumers' opportunity cost of time. To the best of our knowledge, we are the first to estimate a model of sequential search for costly information with imperfect observation of signals observed through search. Our approach provides a general framework to study consumer engagement with a product through search, being able to also capture decisions such as revisits to a previously searched product to resolve further uncertainty.

This paper contributes to an emerging literature studying consumer search and quantifying the impact of search frictions on search and purchase decisions. Papers such as Kim et al. (2010, 2016), Ghose et al. (2012), Chen & Yao (2016), Honka & Chintagunta (2016) quantify search costs in a sequential search model following Weitzman (1979), while papers such as Hong & Shum (2006), Moraga-Gonzalez & Wildenbeest (2008), Moraga-Gonzalez et al. (2015), Koulayev (2014), Seiler (2013), Honka (2014) use either a variation of sequential search or estimate search costs under an assumption of simultaneous search. In general, few product searches imply high search cost estimates (e.g. 2.3 hotel searches on an online travel agent's website lead to search cost estimates of $21.54 in Chen & Yao, 2016). In addition, papers such as Koulayev (2008) and De los Santos et al. (2016) relax the assumption that consumers search from known product distributions and introduce learning into the empirical search model. Ours is also a model of search with learning, as we model search duration as multiple searches of the same option. However, we base our estimation on the optimal search rules for sequential search models with multiple searches, as derived by Chick & Frazier (2012). Our model is also related to work on multiarmed bandit problems (Gittins 1979; Brezzi & Lai 2002). In these models, consumers look to maximize the sum of rewards from sequentially observing rewards from a set of options, including observing rewards from the same option multiple times. Such models are suitable to study repeat purchase occasions, as it is done in Lin, Zhang & Hauser (2014). In contrast, in our model consumers maximize the rewards from a single option chosen after sequentially searching for information about available options, including sampling the same option multiple times. As a result, this model captures repeated search decisions (e.g. time spent searching, revisits) as well as a (single) purchase decision. Similarly, papers such as Ke and Villas-Boas (2017), Ke et al. (2016), and Branco et al. (2012, 2016) also consider the problem of costly search for information before realizing rewards from a purchase decision. More precisely, Ke and Villas-Boas (2017), Ke et al. (2016) consider this problem in the case of multiple products, while Branco et al. (2012, 2016) consider the case of one product and search for information about its attributes. Most closely related to our paper, Ke & Villas-Boas (2017) focus on the case of two products (which they later extend to three products) and derive the optimal sequential search strategy when rewards are drawn from a two point distribution. They show that this depends on the level of outside option. In contrast, we follow Chick & Frazier (2012) who derive optimal search rules for the case of any number of alternatives, with rewards drawn from the normal distribution, and they show that this does not depend on parameters of the model. In addition, we take this model to data and describe the estimation strategy of this model. In terms of estimation strategy, this paper is most closely related to Honka (2014) and Honka & Chintagunta (2016) in using a logit-smoothed Simulated Maximum Likelihood technique. Finally, in terms of quantifying the benefits from searching on the intensive margin, our paper is related to Seiler & Pinna (2017). They measure the change in price paid from spending an additional minute searching in a super market setting and find a benefit of $2.10 per minute, similar to our estimate of consumers' opportunity cost of time.

Understanding how consumers spend time searching in addition to how many products they search, has implications for both managers and policy makers. More precisely, from a managerial perspective, search duration is a measure of engagement that may be related to higher conversions or repeat purchases. As such, having a model of search duration can be a useful tool to measure engagement. In addition, from a policy perspective, time spent searching is directly related to the economic construct of consumers' opportunity cost of time. To the extent that the opportunity cost of time affects how much information consumers choose to gather about companies, studying search duration becomes crucial to understanding firm competition.

The rest of the paper is organized as follows. The next section introduces the specific context for our analysis, the data we employ, and provides descriptive statistics on search duration. Section 3 describes the model, while sections 4 and 5 describe the estimation procedure and the identification of the model. In section 6 we describe our results. The last section concludes.

2 Data

2.1 Search Process

The specific context of our analysis is consumer search for restaurants on an Asian review website. At the time of our data collection, this website provided review information for many products and services, but mainly focused on restaurants (similar to Yelp). We start by describing consumers' three step search process on this website to introduce the observables in our data set. Figure 1 provides an illustration of these steps.

Figure 1: Search process illustration

Consumers start their search by visiting the homepage of the website. Here they can find a restaurant either by typing in a keyword in the search bar at the top, or by searching by cuisine type, location, dish tags, or another menu option (step 1). We refer to these actions as "specifying a query". In response to a query, consumers see an ordered list of restaurants,[1] typically divided into pages with 10 results per page. The consumer can make multiple queries. If these queries are less than one hour apart, we will interpret them as belonging to the same session, consistent with previous work (Wu et al. 2015). The list page contains some information about the displayed restaurants, such as the name, location, average quality information on three dimensions (taste, ambience and service) and average price. The consumer can then click on a restaurant to obtain additional information (step 2). In this case, they navigate to a second screen reserved to that restaurant, which we refer to as the restaurant page. Here they can see photos, restaurant and dish

[1]Restaurants are ordered by default according to the proprietary ranking algorithm used by the website. However, consumers can further sort or filter search results. Modeling such decisions is beyond the scope of this paper. The interested reader should refer to Chen and Yao (2016).

tags, a brief description of the restaurant, as well as previous consumer reviews ordered by posting date. Given the amount of information on this page, consumers decide how much time to spend on a restaurant page, whether to return to the list and make another click, or whether to purchase (step 3). Note that making clicks on the list page is equivalent to searching on the extensive margin, while spending time on the restaurant page is equivalent to searching on the intensive margin.

2.2 Data Sources

The data are comprised of two main sources. One data source is obtained from the Asian review website. This has three components. First is a click stream data set containing clicks consumers made on the site from December 2007 to March 2008. This includes not only clicks made on restaurants, but also clicks to the homepage of the website, clicks on consumer's personal homepage on the site, other member's homepage clicks, chat pages, etc. Importantly, these data contain information on the date and time of the click, which allows us to compute the duration of a click, using differences in time stamps. In general, having time stamp information would allow us to obtain duration information for all clicks but the last click made by the consumer (duration would be truncated). However, since we observe not only clicks made to restaurant pages, but all clicks made by the consumer, we are able to directly observe duration information for 79% of clicks and 40% of last clicks. One concern with using time stamps to measure duration is measurement error. More precisely, we observe when consumers clicked on the restaurant page and when they clicked to go back to the list page or another page. However, we do not observe exactly what they did in this time interval, that is whether they spent time reading about the restaurant or whether they were engaged in another activity. Although we cannot fully alleviate this concern, we do two things to partially address it. First, we collapse duration above 10 minutes since this is more likely to include activities not related to restaurant viewing. Second, we use the duration variable measured by comScore to cross check the time spent on a click on a similar website (Yelp). As we show in the section below, we find very similar duration measures in the two data sets.

Second data component describes restaurant page characteristics of the clicked restaurants for the period April 2003 to March 2008. This includes information on consumers' ratings of restaurants in terms of overall taste, ambience, service and price. We also collect restaurants' attributes such as their location, number of photos, number of reviews, promotions, length of introduction about the restaurant, number of restaurant and dish tags, average review length and whether restaurants bought search keywords.

Third, we have information on purchases for the period May 2005 to March 2008. The website uses a loyalty program. By using this card at the restaurant, consumers obtain 10-30% discount at collaborating restaurants. Note that consumers' use of this loyalty card allows us to link queries to transactions for (possibly) only a subset of consumers, and thus our transaction data is truncated. However, given the significant discount provided by the loyalty program, we anticipate this truncation to only have a minor impact on our data collection efforts. To further minimize the impact of truncation, we will focus our analysis on consumers who make a purchase. Although we limit the analysis to converting consumers, we observe both converting and non-converting sessions, where we call a non-converting session one in which more than 75% of the clicked restaurants participate in the loyalty program.

Since we are interested in modeling consumer search, we need to observe not only which restaurants consumers clicked, but also those they did not search, information which is not included in the first data source. Thus, to augment the data on the restaurants clicked, we use a second data source, which comes from an Internet archiving website called "Wayback Machine" (WBM).[2] Using the keywords that consumers searched and the time of search, we retrieve from the WBM the list of restaurants that consumers likely saw as a response to their query. We require that the keywords consumers searched should be exactly matched with the one save in WBM. However, because the time of search usually cannot be exactly matched, we retrieve the closest time that the keyword search was saved. Given that data on WBM becomes more sparse going further back in history, we are able to match 68% of queries, which we will use in the analysis.

2.3 Final Data Sample

In our final data sample, there are 343,270 observations, on 5,465 consumers searching across a total of 17,852 sessions and 34,912 queries and making 50,439 clicks and 7,538 transactions. In Table 1, we summarize restaurant and query characteristics we observe.

Each observation in the data is a restaurant shown on the list page displayed to consumers in response to their query. This includes information on the quality of the restaurant (weighted

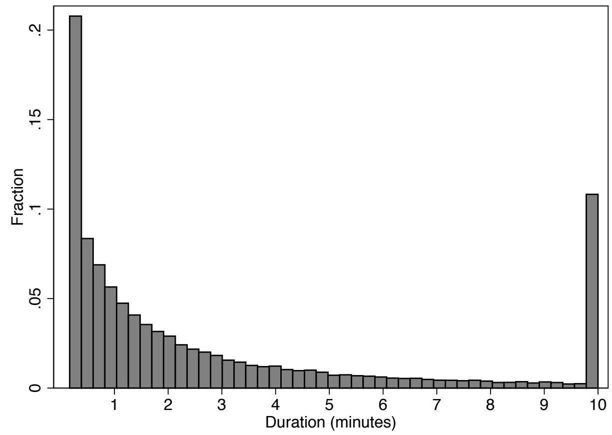
[2]The Wayback Machine website can be found at `https://archive.org/web/`.

Table 1: Restaurant characteristics

	All		Clicked		Duration>median		Purchased		Query	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List page information										
Rating (0-4)	2.71	0.32	2.72	0.31	2.74	0.29	2.78	0.25		
Price (100 RMB)	0.91	0.92	0.88	0.76	0.87	0.68	0.81	0.37		
Number of reviews (1000)	0.47	0.62	0.49	0.61	0.53	0.60	0.61	0.55		
Promotion	0.20	0.40	0.24	0.43	0.27	0.45	0.30	0.46		
Search ad	0.08	0.28	0.06	0.23	0.06	0.25	0.06	0.23		
Card	0.30	0.46	0.59	0.49	0.70	0.46	0.98	0.15		
Position	5.59	3.03	5.10	3.04	5.15	3.07	4.98	3.08		
Restaurant page information										
Number of photos			85.01	100.37	94.23	101.37	104.18	97.69		
Length of introduction (words)			98.54	49.44	103.26	45.33	112.54	32.25		
Number of restaurant tags			5.94	2.78	6.20	2.70	6.65	2.48		
Number of dish tags			13.42	7.18	14.20	6.85	15.62	6.12		
Average review length (1000 words)			0.37	0.18	0.38	0.19	0.41	0.19		
S.D. Rating			0.71	0.17	0.72	0.15	0.73	0.13		
S.D. Price			0.27	0.42	0.26	0.36	0.23	0.21		
Query information										
Days since registered on website									681.62	418.24
Weekend									0.21	0.41
Office hour									0.66	0.47
Time before first click (minutes)									0.72	2.73
Days between session and transaction									5.99	12.33
Number of search results									541.32	1588.18
Observations	343270		50439		25226		7538		34912	

sum of taste, ambience and service measures), average price and number of reviews. As can be seen, clicked, purchased or restaurants on which consumers spent more time generally have higher quality and lower prices. If the consumer clicked on a restaurant on the list page, then we observe additional information as contained on the restaurant page, such as the number of photos, description of the restaurant and several measures of the customer reviews posted. In general, the more information is displayed on the restaurant page (e.g. in terms of the number of photos, length of the introduction or review length), the longer consumers spend on the restaurant page. Finally, we also have information on several query observables. For instance, we observe that on average queries are made by consumers who registered with the website approximately two years in advance and that on average a transaction happens less than one week after the query. Note that we can also compute the time that consumers spent on the list page before clicking a restaurant, since we have both the time stamp of the query and of the click. However, as we show next, this duration is relatively small compared to the time consumers spend on a click on the restaurant page. Thus, we choose not to model this decision in our theory.

Figure 2: Extent of Search on the Intensive Margin: Search Duration



Notes: Histogram of duration (minutes) for observations with full duration information (no imputed values). The spike at the right tail is due to truncation and collapse of duration larger than 10 minutes (the 90th percentile).

2.4 Data Patterns on Search Duration

In this section, we show how consumers search on the intensive margin using data on the time they spend searching restaurants. More precisely, we provide evidence on how much time consumers search, on what affects search duration and on the effect of duration on purchase decisions.

2.4.1 How much time do consumers spend searching?

We find that consumer search on the intensive margin is considerable: the average (median) consumer spends 3.47 (2.45) minutes on a click, with a standard deviation of 3.07 minutes.[3] This large variation in search duration can also be seen in Figure 2, which presents the distribution of search duration for clicked restaurants. To support this result, we use comScore to check the time spent on a similar website, and we find that click duration on Yelp was 3.55 minutes (January, 2013), which is very similar to the result in our data. In contrast, although consumers spend a relatively long time searching each restaurant, their search on the extensive margin (that is, the number of restaurants clicked) is small. More precisely, we find that 60% (71%) of sessions (queries) have only one click, with an average (median) click number by session of 2.83 (1). Correspondingly, at the query level, the average (median) click number is 1.44 (1). Thus, even when consumers search very few restaurants, they search each option extensively.

[3]Considering only the observations for which we observe full duration information (79% of clicks), the average (median) consumer spends 2.88 (1.45) minutes on a click, with a standard deviation of 3.21 minutes.

2.4.2 What influences search duration?

We showed before that search duration is considerable. Next, we ask what influences the consumer's search duration decision, by considering three related questions. First, is duration considerable because certain consumers spend a long time searching or because certain restaurants are searched longer by all consumers. To answer this question, we divide consumers into types by tenure, that is the days since they registered with the website. A high (low) type consumer is one with tenure higher (lower) than the 75th (25th) percentile of the tenure distribution. Consumers with higher tenure should be more experienced with the website and thus might differ in terms of how much time they require to process restaurant information. We also divide restaurants into types such that a high (low) type restaurant is one with quality and price above (below) the 75th (25th) percentile of the respective distributions. In Table 2, we compute the average search duration by consumer and restaurant type. A $*$ identifies significant differences between two entries by means of a t-test. We find that low type restaurants are searched less than high types by all consumers. This is to be expected since low type restaurants present relatively less interest to most consumers. In contrast, high type consumers spend less time on both types of restaurants than a low type. In sum, we find that both consumer and restaurant characteristics affect search duration.[4]

Table 2: Duration by consumer and restaurant types

		Restaurant type		
		High		**Low**
Consumer type	**High**	3.55	*	2.93
				*
	Low	3.72	*	3.19

Second, in Table 3 we consider a larger set of factors influencing consumers' decision to spend time on a restaurant. In particular, we divide restaurant characteristics by whether they are displayed on the list or the restaurant pages and we also consider the effect of query characteristics on search duration. Focusing on the list page information, we find that higher quality restaurants that are promoted and displayed lower in the ranking lead to higher search duration. Also, the more information is displayed on the restaurant page, for example in terms of a larger number of photos, longer description of the restaurant or longer reviews, the more time consumers spend reading about

[4]Our results hold also when considering the median duration by type. The analysis is available upon request.

the restaurant. Figure 3 shows the same relation between the amount of information displayed on the restaurant page and search duration. Finally, more experienced consumers searching on a week day also spend longer on a click. In sum, we find that higher quality restaurants displaying more information are searched longer.

Table 3: Factors influencing search duration (OLS)

	Coefficient	
List page information		
Rating (0-4)	0.2284***	(0.0551)
Price (100 RMB)	−0.0267	(0.0288)
Number of reviews (1000)	0.0329	(0.0294)
Promotion	0.3385***	(0.0375)
Search ad	0.1971**	(0.0630)
Position 1	−0.3588***	(0.0567)
Position 2	−0.2730***	(0.0582)
Position 3	−0.3113***	(0.0601)
Position 4	−0.1789**	(0.0611)
Position 5	−0.0994	(0.0613)
Position 6	−0.1195	(0.0623)
Position 7	−0.0163	(0.0628)
Position 8	−0.1561*	(0.0636)
Position 9	−0.0914	(0.0660)
Restaurant page information		
Number of photos	0.0006**	(0.0002)
Length of introduction (words)	0.0024***	(0.0003)
Number of restaurant tags	0.0200*	(0.0084)
Number of dish tags	0.0179***	(0.0034)
Average review length (1000 words)	1.2172***	(0.0818)
S.D. Rating	0.1396	(0.0870)
S.D. Price	−0.1631**	(0.0506)
Query information		
Days since registered on website	−0.0001***	(0.0000)
Weekend	−0.2017***	(0.0328)
Office hour	0.3651***	(0.0286)
Number of search results	0.0000	(0.0000)
Constant	1.6900***	(0.1434)
R^2	0.0311	
Observations	50439	

Standard errors in parentheses
Notes: All estimates are conditional on a click.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Third, we investigate the relation between search duration and clicks. In particular, we ask whether how much time do consumers spend on a click depends on click order. To this end, we restrict our attention to sessions (or queries) with at least three clicks, and compute the average click duration for the first, last, and middle clicks. Figure 4 shows our results. We find that consumers spend more time on the first and last clicked restaurant in a session (or query) than on middle clicks. Consistent with a sequential model of search, click order captures consumers' expected utility from the considered alternatives (net of search costs). Our finding suggests that search duration captures consumers' revealed preference for restaurant characteristics, which is not

Figure 3: Amount of information and relation to duration

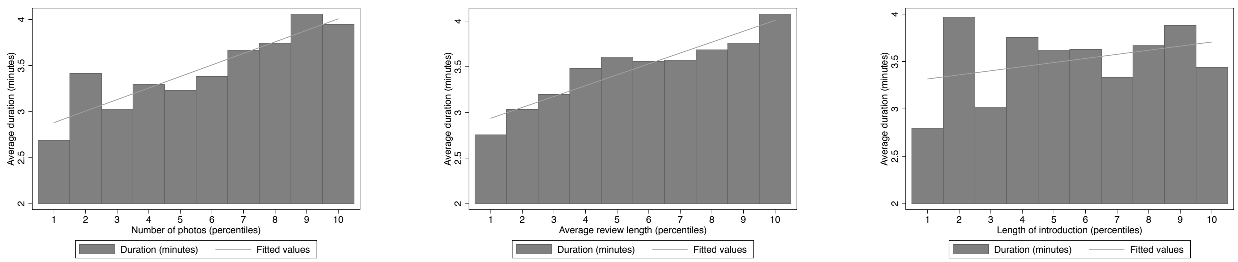
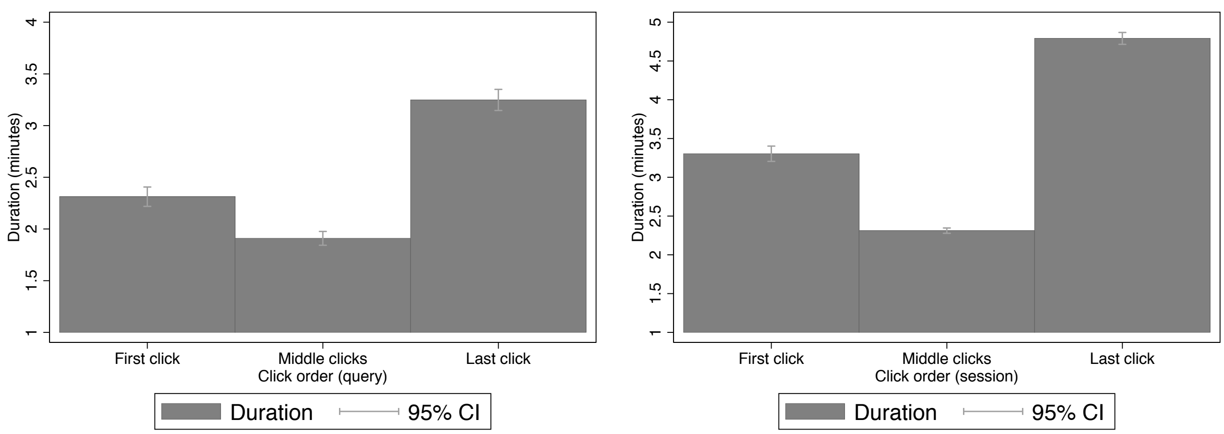


Figure 4: Click order and duration



Query level

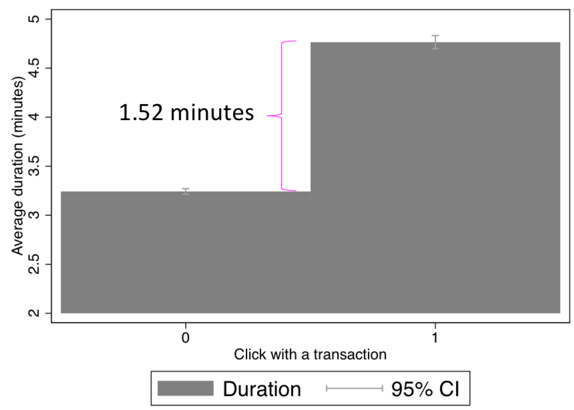
Session level

captures by click order.

2.4.3 Relation between search duration and purchases

Finally, we consider the relation between the time that consumers spend on a click and the probability of purchasing the clicked restaurant. As can be seen in Figure 5, clicked restaurants with that were purchased have a higher search duration (1.52 minutes difference) than those that were not purchased. To further decompose this effect, we condition on clicks, and model the purchase decision in a session as being influenced by duration, list page and restaurant page information. Our results can be found in Table 4. We find that restaurants of higher quality that are cheaper and ranked closer to the top of the ranking are more likely to be purchased. Similarly, generally restaurants with more information displayed on the restaurant page are also more likely to be purchased. Most interesting for this paper, we find that even after accounting for restaurant characteristics, observing consumers' search duration decision helps predict purchases, as restaurants with higher

Figure 5: Search duration for purchased versus not purchased restaurants



search duration are more likely to be purchased.

In sum, in this section we have shown that consumers spend a considerable amount of time searching, even when they search few products. Search duration increases with consumer inexperience, higher quality of the restaurants and the amount of information displayed on the restaurant pages. Finally, we have shown that search duration is higher for purchased restaurants. These results demonstrate that search duration represents both a benefit for consumers (in terms of the amount of information gained through time spent) and a cost for accumulating information. In other words, our results imply that search duration is a choice, made separately from the other two choices they have: which restaurants to click and whether or not to purchase. This implies the need to incorporate the intensive margin decision into a consumer search model, which is what we do in the next section.

3 Model

3.1 Consumer problem

Consider a consumer who seeks to purchase an alternative $j \in \{1, \ldots, J\}$ or choose the outside option (denoted by $j = 0$). The expected utility of the outside option is known, but the consumer faces uncertainty about the J options and can search for information before making a purchase decision. On the one hand, searching an option reveals some information without resolving all the consumer's uncertainty. Thus, the consumer can search the same option multiple times. In our empirical specification, we will interpret time spent searching an option as searching the same

Table 4: The effect of search duration on transactions (Conditional logit at session level)

	Coefficient	
Duration (minutes)	0.0704***	(0.0045)
List page information		
Rating (0-4)	0.6063***	(0.0657)
Price (100 RMB)	−0.4472***	(0.0459)
Number of reviews (1000)	0.1326***	(0.0267)
Promotion	0.1296***	(0.0354)
Search ad	−0.2629***	(0.0624)
Position 1	0.0877*	(0.0443)
Position 2	0.0332	(0.0452)
Position 3	0.0795	(0.0459)
Position 4	0.1748***	(0.0468)
Position 5	0.1677***	(0.0467)
Position 6	0.0796	(0.0473)
Position 7	0.0724	(0.0473)
Position 8	−0.0377	(0.0489)
Position 9	0.0901	(0.0491)
Restaurant page information		
Number of photos	−0.0017***	(0.0002)
Length of introduction (words)	0.0046***	(0.0004)
Number of restaurant tags	0.0114	(0.0084)
Number of dish tags	0.0301***	(0.0037)
Average review length (1000 words)	0.9124***	(0.0739)
S.D. Rating	0.1617	(0.0994)
S.D. Price	−0.1647*	(0.0773)
Outside option	3.9007***	(0.1681)
LL	-17008	
Observations	68291	

Standard errors in parentheses
Notes: All estimates are conditional on a click.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

option multiple times (e.g. spending 5 minutes on a restaurant will be equivalent to searching it 5 times). However, this same model can be used to model revisits of previously searched (e.g. clicked) options. On the other hand, searching an option, involves paying a cost per time, c. To ensure that search costs are positive, we assume c is an exponential function. The consumer's goal is to maximize her expected utility from the best option she will choose to purchase when search ceases net of total search costs.

The consumer's utility from choosing product j in period t has three components. First, the consumer values the quality of the product, q_{jt}, which can include product characteristics, match with needs, etc. The consumer is uncertainty about the quality of the product which is drawn independently from $N(\mu_j, \sigma_j^2)$, with unknown μ_j and known σ_j^2. Because draws are independent, the consumer does not learn about the quality of one product by searching another. Second, the consumer values the characteristics of each restaurant revealed on the list page, X_j^{list}. Third, the consumer observes an idiosyncratic standard normal shock, ϵ_{jt}, hidden from the researcher, which also affects her utility. In sum, the consumers utility from choosing product j in period t

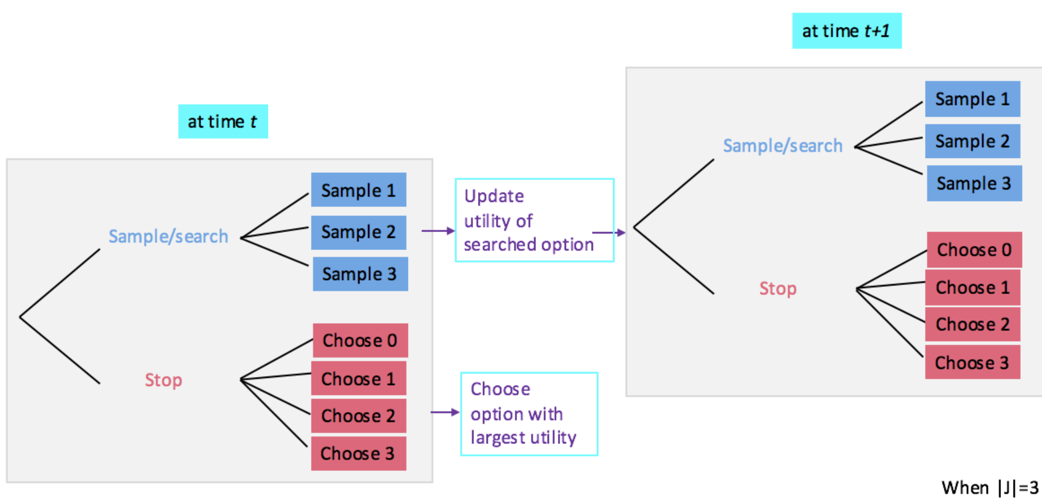
$$u_{jt} = q_{jt} + X_j^{\text{list}} \beta + \epsilon_{jt} \tag{1}$$

We model the expected utility of not purchasing, i.e. the outside option, as known, so that $u_0 = \epsilon_0$. The interpretation of this assumption is that the consumer chooses between one of the J alternatives or rejects all of them, obtaining zero mean utility.

Although the quality of the product is unknown, the consumer begins with a prior probability distribution describing her belief about quality, which she updates after each search decision. More precisely, at $t = 0$, let beliefs be summarized by $N(\mu_{j0}, \sigma_j^2/n_{j0})$, where n_{j0} gives the implied number of samples drawn to form the prior belief. If consumers want to obtain further information about an option, they click and spend time on the restaurant page. We model each minute spent on the restaurant page as another signal observed about the quality of that restaurant. More precisely, sampling option j reveals signal $s_{jt} \sim N(\mu_j, \sigma_j^2)$. We model signals as a function of restaurant page information, $\mu_j = \mu + X_j^{\text{rest}} \alpha$. The consumer then uses Bayes Theorem to update beliefs about quality. More precisely, sampling j at t implies posterior belief $N(\mu_{jt+1}, \sigma_j^2/n_{jt+1})$, with $\mu_{jt+1} = \frac{n_{jt}\mu_{jt} + s_{jt}}{n_{jt+1}}$ and $n_{jt+1} = n_{jt} + 1$, while for $k \neq j$, $\mu_{kt+1} = \mu_{kt}$ and $n_{jt+1} = n_{jt}$, that is, options that are not searched are not updated.

To illustrate the consumer's sequential search procedure, consider Figure 6. Suppose the consumer has three options to choose from. In a given time period t, the consumer can either continue searching or stop. If she continues searching, she then must choose the option to search next. If instead she decides to stop searching, then she chooses whether to purchase one of the three options or choose the outside option (option 0). At any stage in the search process, the consumer uses all data observed thus far to make a decision. If for example, she chooses to search option 2, then she observes a signal about this product, which she uses to update her belief about quality and thus utility of 2. Her beliefs about the other two options stay the same since the consumer did not observe any new information about these. Then in the next period, she again can choose to continue searching or to stop. Importantly, she can choose to search any of the three options available, include the previously search option 2. The possibility of searching the same option as before is what distinguishes this model from previous search models in the literature that assume all consumers' uncertainty about a product is resolved with a single search.

Figure 6: Sequential search illustration



3.2 Optimal search

At a given point during search, the consumer must decide whether to continue searching, and if so, which alternative to sample next. Upon stopping her search, she must decide whether to purchase. To model consumer's optimal search decision, we follow Chick and Frazier (2012). They provide an optimal policy of choosing at each time period whether to continue searching, if so which alternative, and upon termination, which alternative to choose. To characterize such a policy, let the state of information about option j at t be given by $\Theta_{jt} = (\mu_{jt}, n_{jt})$ and the state of the system at t be $\vec{\Theta}_{jt} = (\Theta_{0t}, \Theta_{1t}, \ldots, \Theta_{Jt})$. The optimal policy is one that determines which j to search/purchase at each t given $\vec{\Theta}_{jt}$ in order to maximize the expected utility from the outcome chosen once search terminates net of total search costs. Chick and Frazier (2012) frame this problem using dynamic programming and show that the optimal policy is one that attains the maximum of the following Bellman recursion problem

$$V(\vec{\Theta}_t) = \max\left\{\max_{j=1,\ldots,J} E(-c + V(\vec{\Theta}_{t+1}|\vec{\Theta}_t), \max_{j=0,1,\ldots,J} E(u_{jt}|\vec{\Theta}_t)\right\}.$$

Chick and Frazier (2012) solve for the optimal policy in two steps. First, after proving the existence of an upper bound on the total number of searches a consumer will make, they consider the case with one alternative ($J = 1$), one outside option, and normally distributed rewards with unknown means and known variances. This problem could be solved using the Bellman recursion above given the upper bound on the total number of searches. However, the solution depends on parameters of the problem and the solution would have to be recomputed when these change. Thus, instead of this approach, they choose to transform the discrete-time problem to continuous-time and use diffusion approximation to describe the solution (similar to approaches used for multiarmed bandit problems, e.g. Chernoff and Ray 1965, Lai 1987, Brezzi and Lai 2002, Chick and Gans 2009). This approach leads to a solution that is independent of parameters. Second, they use results from the case of one alternative to provide approximations to the solution for the case of $J > 1$.

The optimal policy for the case of $J > 1$ is characterized by three search rules. We follow the search rules based on the stopping boundary that Chick and Frazier (2012) derive:[5]

1. **Stopping Rule**: Continue to search if and only if $\exists\ j \in (1, \ldots, J)$ such that its posterior mean utility u_{jt} lies within the *continuation set*, that is, u_{jt} must be in the range $\max_{k\neq j} u_{kt} \pm M_{jt}(c, \sigma_j, n_{jt})$, where M_{jt} is the boundary of search.

   In other words, the consumer will continue searching if comparing the posterior mean of at least one alternative and the maximum posterior mean of all other alternatives (and the out-

[5] This is the approach recommended by Chick and Frazier (2012) because of its ease of implementation

side option) is smaller than the boundary of search. Thus, all alternatives that fall within the continuation set are potential candidates for further search, while those outside the continuation set will not be searched at t (although they might be searched in different periods). This condition can also be rewritten as follows: search will continue if $\exists\ j \in (1, \ldots, J)$ such that

$$M_{jt}(c, \sigma_j, n_{jt}) > \Delta_{jt},$$

where $\Delta_{jt} = |u_{jt} - \max_{k \neq j} u_{kt}|$ for $k \in (0, 1, \ldots, J)$.

2. **Selection Rule**: While the stopping rule is not satisfied, choose to sample the alternative $j \in (1, \ldots, J)$ that is furthest inside the continuation set as measured in standardized coordinates, that is, the alternative j such that

$$\arg\max_j \frac{M_{jt}(c, \sigma_j, n_{jt}) - \Delta_{jt}}{c^{1/3}\sigma_j^{2/3}}.$$

3. **Choice Rule**: Conditional on stopping, choose the alternative $j \in (0, 1, \ldots, J)$ with the largest posterior mean utility.

It remains to specify the functional form for the boundary of search. Chick and Frazier (2012) show that it is given by $M_{jt}(c, \sigma_j, n_{jt}) = c^{1/3}\sigma_j^{2/3} b(\sigma_j^{2/3}/(c^{2/3} n_{jt}))$, where $b(h)$ can be approximated by[6]

$$\hat{b}(h) = \begin{cases} 0.233h^2, & \text{if } h \leq 1 \\ 0.00537h^4 - 0.06906h^3 + 0.3167h^2 - 0.02326h, & \text{if } 1 < h \leq 3 \\ 0.705h^{1/2} ln(h), & \text{if } 3 < h \leq 40 \\ 0.642[h(2ln(h))^{1.4} - ln(32\pi)]^{1/2}, & \text{if } h > 40 \end{cases}$$

We note that the stopping and the selection rules may not be optimal when $J > 1$, because they are derived from an approximation to the dynamic programming problem. However, Chick and Frazier (2012) show that these perform very well when using numerical results and they are easier to implement than solving the dynamic programming problem using Bellman recursion.

[6] This approximation to the optimal stopping boundary $b(\cdot)$ is similar to results in the related problem of multi-armed bandits (Gittins, 1989; Brezzi & Lai, 2002).

3.3 Relation to other sequential search models

It is important to describe how these search rules differ from those in two similar problems: Weitzman's (1979) sequential search problem and the multiarmed bandit problem.

In Weitzman's (1979) sequential search problem, the consumer faces a set of options and can sequential sample each. Searching an option reveals all uncertainty about it and the consumer focuses on deciding whether to continue searching any of the unsearched options or stop and choose one of the searched options. Thus, this model cannot be used to study consumers' decision to search the same option multiple times (e.g. spend time searching, revisit a previously searched option). The optimal policy involves three search rules. The stopping rule dictates that the consumer will terminate search when the maximum utility observed u_j exceeds the reservation utility z_j of any unsearched option, where the reservation utility is defined by the solution to $c_j = \int_{z_j}^{\infty}(u_j - z_j)f(u_j)du_j$, where c_j gives the search cost and $f(\cdot)$ is the distribution of rewards. The reservation utility serves as the boundary of search: an option with expected utility higher than the reservation utility will be a candidate for search. The selection rule says that if a search is to be made (if the stopping rule is not satisfied), the option with the highest reservation utility z_j should be searched next. Finally, the choice rule says that once the consumer stops searching, she will choose to purchase the option with the highest utility u_j of those searched (including the outside option).

In contrast to Weitzman's (1979) sequential search problem which describes the case where consumers cannot search the same option multiple times, multiarmed bandit problems deal with the case where the same option is sampled multiple times. In such problems, the consumer has the option to observe rewards from a number of alternatives, described by different reward distributions. By sampling an option, she learns about the distribution of that option and has the option to continue sampling from all options. There is an implicit tradeoff the consumer is facing between exploiting her knowledge from the sampled options or exploring potentially less appealing options currently in order to learn about their reward distribution and make better choices in the future. The goal of the consumer is to maximize the (discounted) sum of rewards. The optimal policy is characterized in Gittins and Jones (1974) and Gittins (1979) in terms of an index rule that dictates the consumer should choose in each time period the option with the largest index. The Gittins index resembles the role of the reservation utility in the Weitzman's (1979) sequential search model. This model is well suited to study repeated purchase decisions (as is done in Lin, Zhang and Hauser, 2014). However, since in this model the consumer accumulates rewards after each period, it is not well suited for the problem in this paper where the consumer's goal is to maximize the (single) utility net of search costs from the option chosen after search ceases. In addition, adding a cost to sample, leads in most cases to the breakdown of the optimality of the index (see Bank and Sundaram, 1994), making the use of the multiarmed bandit framework in our case less desirable.

In sum, in order to model the time spent searching an option (or to revisit a previously searched option) a model that allows consumers to search the same option multiple times with each search costly is needed. The model presented in this paper provides exactly such a framework. In the next section, we discuss how this model can be estimated.

4 Estimation

The model contains four types of parameters to be estimated: mean utility parameters $(\mu_0, \mu, \beta, \alpha)$, belief uncertainty (σ_j), number of samples implied by the prior distribution of each option (n_{j0}), and mean search costs (γ). The optimal search rules presented in the previous section translate into the following restrictions on the parameters of interest. Suppose the consumer searched an option in each of $t \leq T$ periods, with period T denoting the final period in which search occurs. Then, in period $T+1$ we observe the consumer making a purchase decision. In this case, the stopping rule imposes two types of restrictions on parameters. First, in all periods $t \leq T$ when the consumer searched an option, for the searched option j it must be that

$$M_{jt}(c, \sigma_j, n_{jt}) - \Delta_{jt} > 0,$$

Second, in period $T+1$ when the consumer does not search, it must be that

$$M_{jT+1}(c, \sigma_j, n_{jT+1}) - \Delta_{jT+1} < 0, \ \ \forall j \in \{1, \ldots, J\}.$$

The selection rule requires that, if the consumer searched j at t, then

$$\frac{M_{jt}(c, \sigma_j, n_{jt}) - \Delta_{jt}}{c^{1/3}\sigma_j^{2/3}} > \max_{k \neq j} \frac{M_{kt}(c, \sigma_k, n_{kt}) - \Delta_{kt}}{c^{1/3}\sigma_k^{2/3}},$$

Finally, consistent with the choice rule, if the consumer chooses j (including the outside option) after terminating search, her utility from this choice must exceed the utilities of all other options and the outside option. Formally,

$$u_{jT+1} \geq \max_{k \neq j, k \in \{0,1,\ldots,J\}} u_{kT+1}, \;\; \forall j \in \{0, 1, \ldots, J\}.$$

If consumers search sequentially, they make search, search duration and purchase decisions jointly. Thus, the probability of observing a certain outcome in the data in period t is characterized by the joint probability of the stopping, selection and choice rules holding, as given by

$$L_t = Pr(\text{Stopping rule, Selection rule, Choice rule}).$$

Unfortunately, the likelihood function does not have a closed form solution. As a result, we use a simulated maximum likelihood (SMLE) approach to estimate the parameters of the model. In choosing the simulation method, we follow McFadden (1989), Honka (2014), Honka and Chintagunta (2017) and use the logit-smoothed AR simulator.

Simulation using the logit-smoothed AR simulator involves the following steps:

1. Make $d = \{1, \ldots, D\}$ draws of $(\epsilon^d_{jt},\ s^d_{jt})$ for each consumer-product-time period, so if the consumer spends a total of T time periods, then make DT draws. Note that T will differ by consumer.
2. Compute search costs and use draws to form M^d_{jt}, u^d_{jt}, Δ^d_{jt}.
3. Define the following expressions for each draw d
   (a) $\nu^d_{t1} = M^d_{jt} - \Delta^d_{jt}$ for the searched j.
   (b) $\nu^d_{t2} = \Delta^d_{jT+1} - M^d_{jT+1} \;\; \forall j \in \{1, \ldots, J\}$.
   (c) $\nu^d_{t3} = \frac{M^d_{jt} - \Delta^d_{jt}}{c^{1/3} \sigma_j^{2/3}} - \max_{k \neq j, k \in C_t} \frac{M^d_{kt} - \Delta^d_{kt}}{c^{1/3} \sigma_k^{2/3}}$ for the searched j.
   (d) $\nu^d_{t4} = u^d_{jT+1} - \max_{k \neq j, k \in \{0,1,\ldots,J\}} u^d_{kT+1}, \;\; \forall j \in \{0, 1, \ldots, J\}$.
4. Compute expression $R^d_t = \sum_{n=1}^{4} e^{-\frac{\nu^d_{tn}}{\lambda}}$, where $\lambda > 0$ is a scaling parameter.[7]
5. Obtain the sum of R^d_t over time and compute S^d for each draw d

$$S^d = \frac{1}{1 + R^d}.$$

[7] Little guidance is available on choosing the scaling parameter λ. We will determine the appropriate scaling parameter using Monte Carlo simulations, which are described in section 5.2.

6. The average of S^d over D draws of the error terms gives the simulated likelihood function.

In this section, we described the estimation strategy of the sequential search model. Next, we discuss identification, followed by estimation results using data on time spent searching.

5 Identification

5.1 Model Parameters

The parameter set includes mean utility parameters $(\mu_0, \mu, \beta, \alpha)$, belief uncertainty (σ_j), number of samples implied by the prior distribution (n_{j0}), and mean search costs (γ). For identification, we can only recover the ratio of the prior and the signal variance, which implies that we will not be able to estimate σ_j, but can estimate n_{j0}.

Mean utility parameters that vary by product (restaurant characteristics) are identified from the stopping, selection and choice rules. More precisely, the correlation between product characteristics and the frequency with which consumers click, spend time and purchase identify the mean utility parameters. Characteristics in the utility function that do not vary by product cannot be identified, because they shift the mean posterior utility of the focal product and the maximum of the posterior utilities of all other products by the same amount. Search costs do not enter the choice rule, and thus can serve as exclusion restrictions allowing us to identify preferences and search cost parameters separately (Chen and Yao, 2016). Finally, stopping, selection and choice rules identify n_{j0} from variation in consumers who continue searching the same product and those who switch, conditional on product characteristics revealed upon search.

5.2 Monte Carlo Simulation

In what follows, we show that Simulated Maximum Likelihood using the logit-smoothed AR simulator can recover utility and search cost parameters in this model. We do so using Monte Carlo simulation. More precisely, we generate a data set of 5,000 consumers, who have five options to choose from (four restaurants and one outside option). Restaurants have both list and restaurant page characteristics, which we assume are drawn from a normal distribution with mean and standard deviation equal to those found in the data. The true values of the parameters are chosen to be consistent with those from a preliminary estimation of the model. For estimation, we follow the steps described in Section 4 and use 100 draws from the distribution of the utility error terms and signals for each consumer-restaurant-time period combination and we repeat the estimation 50 times. Our results can be found in Table 5.[8] The first column shows the true parameters and the second column shows the estimated parameters. We find that our method recovers the parameters of the model well.

Table 5: Monte Carlo Simulation Results

	True values	Estimated values
Learning		
μ_0	1.5	1.3132***
		(0.0031)
μ	0.5	0.3891***
		(0.0056)
n_0	−0.5	−0.5652***
		(0.0142)
List page information		
Price	−4	−3.5527***
		(0.0017)
Rating	1	0.8993***
		(0.0026)
Restaurant page information		
Number of photos	0.5	0.4249***
		(0.0046)
Search cost		
Constant	−4	−3.0726***
		(0.0700)
Log-likelihood		-2,723
Observations		10,000

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

So far, we have described the model and its estimation and identification. In the next section, we present estimation results from this model.

6 Results

In this section, we present estimation results using the sequential search model and estimation procedure presented in previous sections. For the estimation, we restrict the sample as follows. First, we only consider queries for which we observe search duration information for all clicks and in which consumers did not revisit a previously clicked restaurant. Second, we perform the estimation only on queries without extreme prices and in which the number of search results on the page equals 10 (73% of the data). The resulting estimation samples has 12,364 observations

[8]The results are obtained with (inverse) scaling factor $1/\lambda$ equal to 10. However, upon request, we can share results for simulations with $1/\lambda$ ranging from 1 to 20.

and 1,124 queries.[9]

Table 6: Main Estimation Results

	Estimated values
Learning	
μ_0	-0.8983^{***}
	(0.0070)
μ	-0.0588
	(0.0175)
n_0 (exp)	0.1018^{***}
	(0.0030)
List page information	
Price (<20 RMB)	-0.3623^{***}
	(0.0320)
Price (20-50 RMB)	-0.1507^{***}
	(0.0151)
Price (51-79 RMB)	0.0558^{***}
	(0.0089)
Price (80-119 RMB)	0.0861^{***}
	(0.0113)
Price (120-200 RMB)	-0.1216^{***}
	(0.0236)
Price (>200 RMB)	-0.0646
	(0.0410)
Rating	-0.0095
	(0.0050)
Restaurant page information	
Number of photos	0.0396
	(0.0201)
Search cost	
Constant (exp)	3.5595^{***}
	(0.0454)
Log-likelihood	-9297.6
Observations	11,000

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 presents our main estimation results. Generally, we find utility and search cost estimates that are economically meaningful and significant. In particular, a higher price decreases utility and more information revealed on the restaurant pages increases utility. In addition, the learning parameters are significant and they increase utility. In contrast, we find a negative effect of the restaurant's quality on utility, possibly due to small differentiation in quality across restaurants in our sample. Most importantly, this model allows us to interpret search costs in terms of consumers' opportunity cost of time. We find an opportunity cost of time of \$1.30 per minute, as measured by the ratio of the search cost estimate to the price estimate, accounting for the exchange rate from RMB to US Dollar.[10] The estimated opportunity cost of time is in line with the result in Seiler & Pinna (2017) of \$2.10 per minute. Our result shows how studying search duration aids in

[9] To make the estimation feasible, we randomly select a subsample of queries for estimation.
[10] We use US Dollar= 0.15$*$RMB as conversion rate.

interpreting search model results.

7 Counterfactual

Table 7: Counterfactual results

	Current (simulated)		Full information in one search	
	Mean	Median	Mean	Median
Panel A: Estimated parameters				
Number of restaurants searched (per consumer)	1.71	1.00	2.01	2.00
Duration of a click	1.52	1.00	1.00	1.00
Price (purchased restaurant)	74.97	69.81	74.19	69.12
Rating (purchased restaurant)	2.67	2.67	2.66	2.67
Panel B: Low search costs				
Number of restaurants searched (per consumer)	3.52	3.06	4.14	4.00
Duration of a click	4.07	2.00	1.00	1.00
Price (purchased restaurant)	74.38	69.32	74.56	69.25
Rating (purchased restaurant)	2.66	2.67	2.66	2.67
Panel C: High search costs				
Number of restaurants searched (per consumer)	1.21	1.00	1.32	1.00
Duration of a click	1.14	1.00	1.00	1.00
Price (purchased restaurant)	75.23	70.48	75.42	70.88
Rating (purchased restaurant)	2.66	2.67	2.67	2.68

Notes: Low search costs are a third of the estimated value, while high search costs are three times as large.

8 Conclusion

In this article, we study consumers' decision to spend time searching, in addition to the decision of which products to search and whether to purchase. Using data on consumers searching for restaurants on a Chinese review website, we document that search duration is considerable and that restaurants that are searched longer are more likely to be purchased. In addition, we are the first to estimate a sequential search model that captures both the intensive and the extensive margins of search, using the optimal search rules for the full set of decisions made by consumers during search: which products to search, how much time to search each product, and whether to purchase. Our approach provides a general framework to study consumer engagement with a product through search.

References

[1] Banks, J., R. Sundaram. 1994. Switching Costs and the Gittins index. *Econometrica* **62** 687-694.

[2] Branco, F., M. Sun, J. M. Villas-Boas. 2016. Too Much Information? Information Provision and Search Costs. *Marketing Sci.* **35**(4) 605-618.

[3] Branco, F., M. Sun, J. M. Villas-Boas. 2012. Optimal Search for Product Information. *Management Sci.* **58**(11) 2037-2056.

[4] Brezzi, M., T. Lai. 2002. Optimal Learning and Experimentation in Bandit Problems. *J. Economic Dynam. Control* **27**(1) 87-108.

[5] Brezzi, M., T. Lai. 2006. Optimal learning and experimentation in bandit problems. *J. Economic Dynam. Control* **27**(1) 87-108.

[6] Chen, Y., S. Yao. 2016. Sequential Search with Refinement: Model and Application with Click-Stream Data. *Management Sci.* **34**(4) 606-623.

[7] Chernoff, H., S. N. Ray. 1965. A Bayes Sequential Sampling Inspection Plan. *Ann. Math. Statist.* **36**(5) 1387-1407.

[8] Chick, S., N. Gans. 2009. An Economic Analysis of Simulation Selection Problems. *Management Sci.* **55**(3) 421-437.

[9] Chick, S., P. Frazier. 2012. Sequential Sampling with Economics of Selection Procedures. *Management Sci* **58**(3), 1-16.

[10] De los Santos, B., A. Hortaçsu, M. R. Wildenbeest. 2015. Search With Learning for Differentiated Products: Evidence from E-Commerce. *Journal of Business & Economic Statistics*, **0**(0), 1-16.

[11] De los Santos, B., S. Koulayev. 2016. Optimizing Click-through in Online Rankings for Partially Anonymous Consumers. Working paper.

[12] Ghose, A., P. Ipeirotis, B. Li. 2012. Surviving Social Media Overload: Predicting Consumer Footprints on Product Search Engines. Working paper.

[13] Gittins, J., D. Jones. 1974. A Dynamic Allocation Index for the Sequential Design of Experiments. Gani J., Sarkadi K., Vince I., eds. *Progress in Statistics* (North-Holland, Amsterdam), 241-266.

[14] Gittins, J. 1979. Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Sta-*

tistical Society, Series B 41, 148-177.

[15] Gittins, J. 1989. Multi-Armed Bandit Allocation Indices. *John Wiley & Sons, New York.*

[16] Hong, H., M. Shum. 2006. Using Price Distributions to Estimate Search Cost. *RAND J. Econom.* **37**(2) 257-275.

[17] Honka, E. 2014. Quantifying search and switching costs in the U.S. auto insurance industry. *RAND J. Econom.* **45**(4) 847-884.

[18] Honka, E., P. Chintagunta. 2017. Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry. *Marketing Sci.* **36**(1) 21-42.

[19] Ke, T., J.M. Villas-Boas. 2017. Optimal Learning Before Choice. Working paper.

[20] Ke, T., Z. M. Shen, J.M. Villas-Boas. 2016. Search for Information on Multiple Products. *Management Sci.* **62**(12) 3576-3603.

[21] Kim, J. B., P. Albuquerque, B. J. Bronnenberg. 2010. Online Demand under Limited Consumer Search. *Marketing Sci.* **29**(6) 1001-1023.

[22] Kim, J. B., P. Albuquerque, B. J. Bronnenberg. 2016. The Probit Choice Model Under Sequential Search with an Application to Online Retailing. *Management Sci.* Forthcoming.

[23] Koulayev, S. 2013. Search With Dirichlet Priors: Estimation and Implications for Consumer Demand. *Journal of Business & Economic Statistics.* **31** 226-239.

[24] Koulayev, S. 2014. Search for Differentiated Products: Identification and Estimation. *RAND J. Econom.* **45**(3) 553-575.

[25] Moraga-Gonzalez, J., and M. Wildenbeest (2008): "Maximum Likelihood Estimation of Search Cost," European Economic Review, 52, 820-848.

[26] Moraga-Gonzalez, J.L., Z. Sandor, and M. Wildenbeest (2015): "Consumer Search and Prices in the Automobile Market," Working paper.

[27] Lai, T. 1987. Adaptive Treatment Allocation and the Multi-armed Bandit Problem. *Ann. Statist.* **15**(3) 1091-1114.

[28] Lin, S., J. Zhang, J. R. Hauser. 2014. Learning from Experience, Simply. *Marketing Sci.* **34**(1) 1-19.

[29] McFadden, D. 1989. A Method of Simulating Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica.* **57**(5) 995-1026.

[30] Seiler, S. 2013. The Impact of Search Costs on Consumer Behavior: A Dynamic Approach.

Quant. Marketing Econom. **11**(2) 155-203.

[31] Seiler, S., F. Pinna. 2017. Estimating Search Benefits from Path-Tracking Data: Measurement and Determinants. *Marketing Sci.* (forthcoming).

[32] Weitzman, M. 1979. Optimal search for the best alternative. *Econometrica.* **47**(3) 641-654.

[33] Wu, C., H. Che, T. Chan, X. Lu. 2015. The Economic Value of Online Reviews. *Marketing Science* **34**(5) 739-754.