Search with Learning

Babur De los Santos[1]
Ali Hortaçsu[2]
Matthijs Wildenbeest[3]

[1]Clemson University

[2]University of Chicago and NBER

[3]Kelley School of Business, Indiana University

EARIE 2015 Munich

Motivation

Typical search model

- Tradeoff between gains from search and cost of searching to determine whether to continue searching (or how many times to search).
- Gains from search are derived using the assumption that consumers "know" the distribution of prices or wages.

Search and learning

- Starting from Rothschild (1974), a number of papers diverge from this view and analyze optimal search behavior when consumers are uncertain about the distribution (in addition to actual draws).
- Important, since search behavior is sensitive to the assumed distribution (Gastwirth, 1976).

Objectives of the paper

Objective of the paper

Develop a method to estimate search costs for differentiated products when consumers learn about the utility distribution.

More specifically:

- We use observed search patterns to obtain bounds on the consumer's search cost.
- Bounds are conditional on utility parameters; estimate these and search cost parameters such that probability that consumers search cost is within these bounds is maximized.
- Model applies to settings where purchase decisions and search histories are observed.
- Application: MP3 players sold online.

Search and learning model

- Search is sequential.
- Consumers have imperfect information about the utility distribution.
- Consumers learn by Bayesian updating their prior on the unknown utility distribution.
- Cost of each search c_i is consumer specific.

Dirichlet setup

Learning model is easiest explained assuming N options available in the market, with utility $u = \{u_1, u_2, \dots, u_N\}$.

- Probability of sampling each utility is given by vector $\rho = (\rho_1, \rho_2, \dots, \rho_N)$, with $\sum_n \rho_n = 1$.
- Utility values are known to consumers, buy probabilities are not.
- Instead, probabilities are considered random variables distributed according to a Dirichlet distribution with concentration parameters $a = (a_1, a_2, \dots, a_N)$.

- Prior expected value of ρ_n is $E[\rho_n] = \frac{a_n}{W}$, where $W = \sum_n a_n$ is the weight put on the initial prior.
- Dirichlet distribution is the conjugate prior of the multinomial distribution, so posterior will be Dirichlet as well:

$$E[\rho_n] = \begin{cases} \dfrac{a_n}{W+1} & \text{if } n \text{ is not sampled;} \\ \\ \dfrac{a_n+1}{W+1} & \text{if } n \text{ is sampled.} \end{cases}$$

Bayesian updating

Model

Example

- Three options, uninformative prior, i.e., $a = (1, 1, 1)$.
- Prior expected value of sampling each option given by $E[\rho] = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.
- If consumer searches and samples option 2, we get $a = (1, 2, 1)$.
- Posterior expected values are $E[\rho] = \left(\frac{1}{4}, \frac{2}{4}, \frac{1}{4}\right)$.

What makes this learning environment attractive?

- When searching, consumers make a trade-off between cost of an additional search and expected gains from search.
- What matters for the expected gains from search is the (posterior) probability of sampling a higher utility than utility of best alternative observed so far.
- With Dirichlet distribution, this posterior probability only depends on:
  - The initial utility distribution.
  - The weight put on the initial prior, W.
  - How many alternatives have been sampled to date, t.

Bayesian updating

Model

- Since a continuous distribution of utilities is more applicable in our setting, we use the case of a Dirichlet *process*.
- Generalization of the Dirichlet distribution to a continuous distribution.
- Attractive features of Dirichlet distribution carry over:
  - gains from search are only a function of base distribution H with updated weight $W/(W+t)$.
- Gain from search at utility $\hat{u}_{it}$ are then

$$G(\hat{u}_{it}) = \frac{W}{W+t} \int_{\hat{u}_{it}}^{\infty} (u - \hat{u}_{it}) \cdot h(u)\, du,$$

  where $h(u)$ is the density of the base distribution.
- The term $W/(W+t)$ reflects consumers' updating process: less weight is put on offers that exceed $\hat{u}_{it}$ every time a utility is drawn that is lower than $\hat{u}_{it}$.

Product differentiation

Model

Consumer i's indirect utility for product j, sold by retailer k, is given by

$$u_{ijk} = \underbrace{\alpha p_j + X_j\beta + X_k\gamma}_{\text{mean utility } \delta_{jk}} + \varepsilon_{ijk},$$

where X_j are product characteristics, X_k are firm characteristics, and ε_{ijk} is a utility shock from a type I extreme value distribution.

We making the following assumptions:

- Consumers do not know ε_{ijk} before searching.
- Actual values of the attributes are not observed either and have to be learned.
- Parameters of the utility function are known to consumers but not to the econometrician.
- Simplification: consumers know the joint distribution of attributes.
  - Implies consumers know the available variety of mean utilities δ_{jk}, but do not know which firm is offering which mean utility level until they start searching.

Product differentiation

Model

- To a consumer alternatives are ex-ante identical and only ex-post differentiated (see also Hortaçsu and Syverson, 2004), and as a result consumers will search randomly.
- We assume that by visiting a retailer a consumer observes prices and characteristics of all products sold by retailer k.
- Let there be K retailers and J products.
- The indirect utility distribution in the market follows a mixture distribution of $J \times K$ type I extreme value distributions with density

$$f(u) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{J} \sum_{j=1}^{J} \exp(-(u - \delta_{jk} + \exp(-(u - \delta_{jk})))).$$

  - location parameter $\delta_{jk} = \alpha p_j + X_j \beta + X_k \gamma$
  - scale parameter 1

Priors
Model

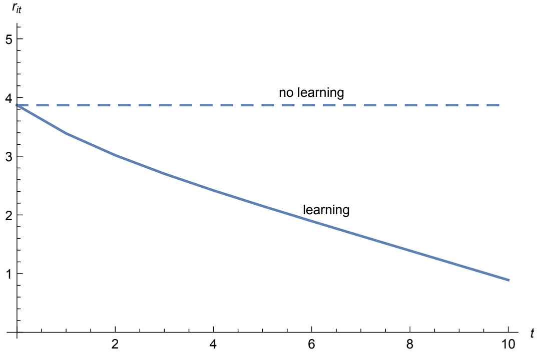
- Data in our application does not allow us to identify consumers' priors.
- However, our model is flexible enough to capture different priors.
- In our main specification we use an informative prior, which means consumers have correct initial beliefs.
- Allows us to write the gains from search equation as:

$$G(\hat{u}_{it}) = \frac{W}{W + t \cdot J} \sum_{k=1}^{K} \frac{1}{K} \left(\underbrace{\gamma + \log \left[\sum_{J} \exp\left(\delta_{jk}\right) \right]}_{\text{expected max utility of search}} - \hat{u}_{it} + \underbrace{\int_{\sum_J \exp[\delta_{jk} - \hat{u}_{it}]}^{\infty} e^{-x}/x \, dx}_{\text{option value of sticking to } \hat{u}_{it}} \right)$$

where γ is the Euler constant.

Non-increasing reservation utilities

Model



- Reservation utilities are non-increasing in the number of alternatives sampled.
- Consumers are more likely to accept a sampled alternative over time than in the no-learning model.
- Consumers may also recall, which is useful for explaining the data in our application.

Estimation strategy

- Use information on a consumer's sequence of searches to derive expressions for bounds on the search cost that rationalizes the consumer' observed search behavior.
- Expressions will depend on the parameters of the utility distribution.
- We pick the parameters of the utility and search cost distribution such that the probability that a consumer's search cost is within the observed search cost bounds is maximized.

Search cost bounds

Estimation

Search cost lower bound:

- Buying a product corresponds to a decision not to continue searching.
- Therefore, $c_i > G(\hat{u}_{it})$.
- Lower bound on search cost is $\underline{c}_i = G(\hat{u}_{it})$.

Search cost upper bound:

- If consumer samples more than once, gains from search in previous period were higher than search cost.
- Therefore, $c_i \leq G(\hat{u}_{it-1})$.
- Upper bound on search cost is $\overline{c}_i = G(\hat{u}_{it-1})$.

Search costs

Estimation

- Search cost are $\ln c_i = \beta X_i + \eta_i$, where X_i is a vector of consumer demographics and η_i is a standard normal distributed error term.
- The probability that consumer i's search cost is within the relevant search cost bounds is then given by

$$P\left(\underline{c}_i < c_i \leq \overline{c}_i\right) = \Phi\left(\ln \overline{c}_i - \beta X_i\right) - \Phi\left(\ln \underline{c}_i - \beta X_i\right),$$

  where $\Phi(\cdot)$ is the standard normal CDF.
- In case a consumer searches only once we set $\Phi\left(\ln \overline{c}_i - \beta X_i\right) = 1$.

Likelihood function

Estimation

- Estimate by (simulated) maximum likelihood
- Chosen product should also be the preferred product among the set of products sold by the seller.
- Log-likelihood function takes both the stopping decision and the product choice decision into account.

ComScore Web-Behavior Panel

Data

- Online browsing and transaction data from 91,689 randomly selected internet users in 2007.
- Date, time, and duration of each visit to a website or domain.
- Price, quantity, and description of product purchased during a session.
- Sample consists of purchases of MP3 players from online retailers: 731 transactions.
- Visit histories of purchasers: 9,742 visits to online retailers within 7 days prior to transactions.

Transactions and visits by retailer

Data

Firms	Market Share %	Search visits %
apple.com	59.10	13.56
amazon.com	16.69	19.91
circuitcity.com	9.71	9.80
target.com	3.15	9.10
bestbuy.com	2.33	9.96
walmart.com	2.33	12.55
overstock.com	1.50	4.54
Other Electronic Stores	3.69	9.77
Other Retailers	1.50	10.80
Observations	731	9,742

Terminology and facts

Data

Search is measured by number of visits to different retailers prior to purchase. To identify a user's visit as search related to a particular transaction, we link visits up to 7 days before a transaction.

- Average number of online stores visited is 2.82.
- Percentage of consumers visiting two or more stores ranges from 45% (same day window) to 69% (7 day window).

Retailers searched by product

Data

Product	Mean number of firms searched	Prices of searched firms Mean	Std. Dev.	CV
iPod Nano 4Gb	2.79	161.81	11.29	0.07
iPod Shuffle 1Gb	2.62	75.86	0.62	0.01
iPod Nano 8Gb	2.88	194.29	3.36	0.02
iPod 80Gb	2.91	262.30	28.21	0.11
iPod 30Gb	2.60	244.78	2.68	0.01
iPod Nano 2Gb	2.47	134.96	6.49	0.05
Zune 30Gb	3.46	168.25	13.84	0.08
iPod Touch 8Gb	3.23	294.82	3.39	0.01
Sandisk Sansa Shaker MP3	4.15	31.60	0.49	0.02
Sandisk Sansa E250 MP3	1.75	86.57	3.18	0.04
Total	2.82	168.52	68.13	0.40

Price Endogeneity

Results

- Unobserved product and retailer characteristics may be correlated with prices.
- Use retailer and brand fixed effects to capture most of the correlation.
- Use a control function approach (Petrin and Train, 2010).
  - First stage: regress prices on all exogenous variables and instruments.
  - Second stage: use residuals from first stage as additional control.
  - Use BLP-type instruments.

Baseline results

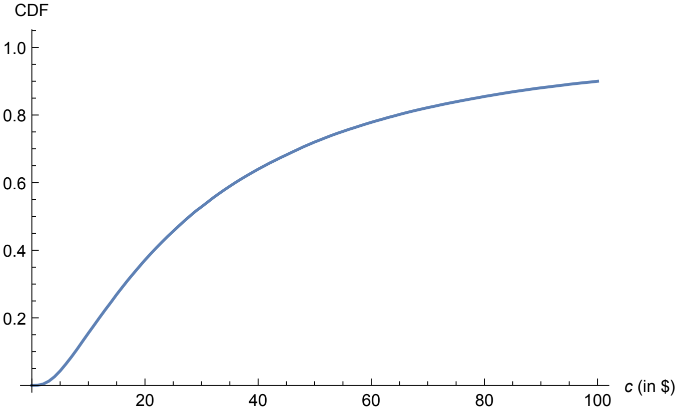
Results

	(1) No Control Function		(2) Control Function	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.
Search Cost				
Constant	-1.086	(0.186)***	-1.103	(0.169)***
Broadband	-0.171	(0.166)	-0.183	(0.156)
Age 60+	-0.086	(0.118)	-0.076	(0.108)
Income >75k	-0.134	(0.085)	-0.132	(0.080)*
Household size	-0.005	(0.027)	0.002	(0.026)
Product Attributes				
Zune	0.207	(0.194)	0.142	(0.198)
Sansa	-1.418	(0.313)***	-1.960	(0.285)***
Storage/weight	0.012	(0.017)	0.066	(0.015)***
Price	-0.400	(0.090)***	-0.942	(0.093)***
Control Function				
Unobserved attributes			0.915	(0.150)***
Log-likelihood	-2,655.05		-2,623.01	

Notes: *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Bootstrapped standard errors within parentheses. The number of observations is 731. The number of simulated consumers is 200 per observation. The value of the smoothing parameter of the logit-smoothed AR simulator is 0.10. Weight on the initial prior is $W = 62$. All specifications have retailer specific dummies.

Estimated search cost distribution

Results



- Relatively large search cost estimates: median search costs are $27.86 and for 25 percent of consumers search costs exceed $54.63.
- However, consumers search for *all* characteristics of the products for *all* products sold by a retailer.

No learning

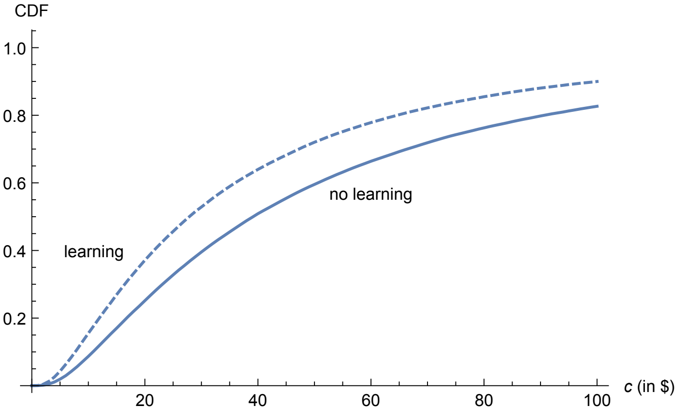
Results

Variable	Learning Coeff.	Learning Std. Err.	No Learning Coeff.	No Learning Std. Err.
Search Cost				
Constant	-1.103	(0.169)***	-0.777	(0.154)***
Broadband	-0.183	(0.156)	-0.133	(0.132)
Age 60+	-0.076	(0.108)	-0.054	(0.099)
Income >75k	-0.132	(0.080)*	-0.084	(0.072)
Household size	0.002	(0.026)	-0.003	(0.024)
Product Attributes				
Zune	0.142	(0.198)	-0.173	(0.190)
Sansa	-1.960	(0.285)***	-1.383	(0.349)***
Storage/weight	0.066	(0.015)***	0.061	(0.017)***
Price	-0.942	(0.093)***	-0.981	(0.093)***
Control Function				
Unobserved attributes	0.915	(0.150)***	0.891	(0.139)***
Log-likelihood	-2,623.01		-2,763.86	

Notes: *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Bootstrapped standard errors within parentheses. The number of observations is 731. The number of simulated consumers is 200 per observation. The value of the smoothing parameter of the logit-smoothed AR simulator is 0.10. Weight on the initial prior is $W = 62$. All specifications have retailer specific dummies.

Search costs with and without learning

Results



- With no updating median search costs are with \$39.15 more than 40 percent higher.
- Reservation utilities are decreasing in t in learning model, which makes consumers search less. So to explain the observed number of searches, no-learning model requires higher search costs.

Robustness

Results

Variable	(1) Three-Day Window Coeff.	Std. Err.	(2) Uninformative Prior Coeff.	Std. Err.
Search Cost				
Constant	-0.663	(0.199)***	-0.699	(0.310)**
Broadband	-0.230	(0.179)	-0.273	(0.252)
Age 60+	-0.168	(0.117)	-0.131	(0.263)
Income >75k	-0.120	(0.103)	-0.477	(0.228)**
Household size	-0.022	(0.030)	-0.001	(0.057)
Product Attributes				
Zune	0.258	(0.180)	0.083	(0.278)
Sansa	-2.138	(0.243)***	-1.717	(0.354)***
Storage/weight	0.065	(0.016)***	0.058	(0.024)**
Price	-0.971	(0.118)***	-0.938	(0.120)***
Control Function				
Unobserved attributes	0.949	(0.177)***	0.916	(0.137)***
Log-likelihood	-2,405.92		-2,601.56	

Notes: *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Bootstrapped standard errors within parentheses. The number of observations is 731. The number of simulated consumers is 200 per observation. The value of the smoothing parameter of the logit-smoothed AR simulator is 0.10. All specifications have retailer specific dummies.

Objectives of Monte Carlo experiments:

- Confirm that our estimation procedure is able to recover the unknown parameters of the search cost distribution and the utility function.
- Study robustness to some of our assumptions.

Results Monte Carlo

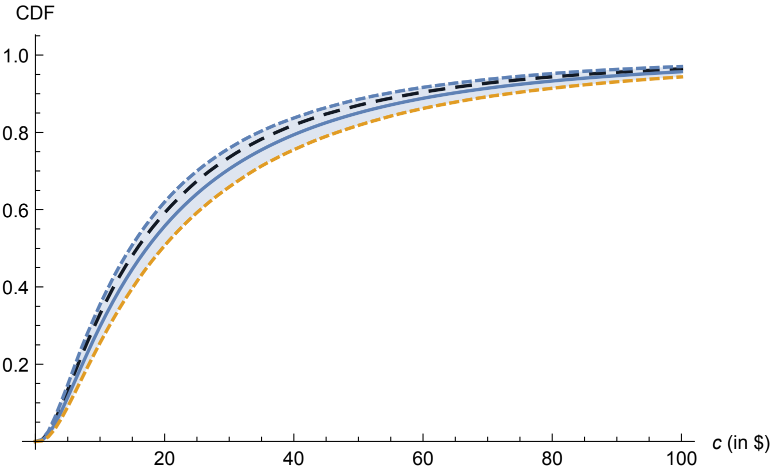
Monte Carlo Experiments

	(1) True	(2) Learning		(3) No Learning	
Variable	Coeff.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Search Cost					
Constant	-1.000	-0.931	(0.052)	-0.279	(0.044)
Broadband	-0.500	-0.439	(0.084)	-0.278	(0.078)
Utility					
Firm 1	-2.000	-2.049	(0.179)	-1.773	(0.160)
Firm 2	-1.500	-1.553	(0.162)	-1.404	(0.154)
Firm 3	-1.000	-1.048	(0.152)	-0.972	(0.143)
Firm 4	-0.500	-0.526	(0.133)	-0.498	(0.126)
Product 2	-1.000	-0.984	(0.105)	-0.947	(0.120)
Product 3	1.000	1.009	(0.108)	0.979	(0.124)
Price	-2.000	-2.003	(0.193)	-1.918	(0.222)

Notes: Number of observations is 1,000. The number of simulated consumers is 200 per observation. Weight on the initial prior $W = 15$. Simulated prices for product 1 are uniform $U(100, 175)$, prices for product 2 are uniform $U(75, 125)$, and prices for product 3 are uniform $U(125, 225)$.

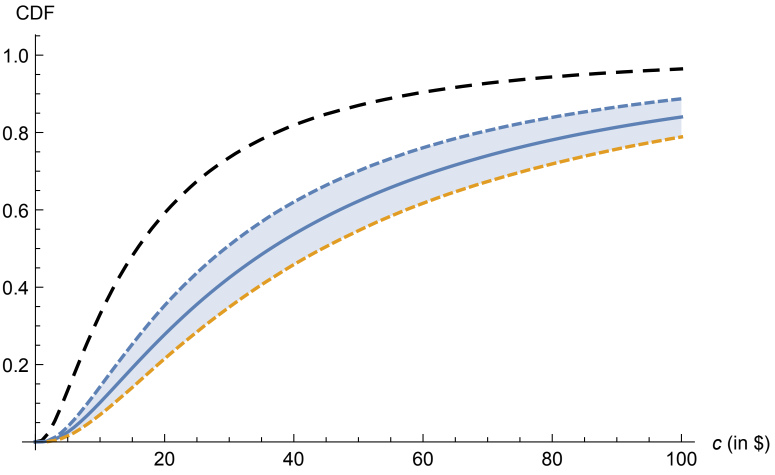
Estimated search costs (learning)

Monte Carlo Experiments



Estimated search costs (no learning)

Monte Carlo Experiments



Estimated elasticities

Monte Carlo Experiments

	Learning	No Learning
Firm 1	-1.735	-1.385
Firm 2	-1.631	-1.343
Firm 3	-1.474	-1.241
Firm 4	-1.271	-1.104
Firm 5	-1.022	-0.932

Notes: Firms are ordered by increasing market shares.

- Elasticity estimates are biased towards zero for all retailers.
- Biased is most severe for firms with the lowest market shares.

Weaknesses of the data:

- We only observe browsing behavior at the domain level.
  - Prior visits during a set search window are treated as related to the search.
  - We therefore may wrongfully attribute a visit as related to the wrong product.
  - On the other hand, consumers may get information from non-retailers.
- Prices are measured with error.
  - Prices are obtained from transactions, and we infer the prices at the other retailers from recent transaction from other consumers.
  - Prices may change over time, which means there is potential for measurement error in prices.

Robustness

Monte Carlo Experiments

	(1) True	(2) Noise Choice Set		(3) Noise Prices	
Variable	Coeff.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Search Cost					
Constant	-1.000	-1.101	(0.049)	-0.942	(0.055)
Broadband	-0.500	-0.368	(0.083)	-0.425	(0.093)
Utility					
Firm 1	-2.000	-1.781	(0.185)	-2.035	(0.173)
Firm 2	-1.500	-1.341	(0.153)	-1.520	(0.190)
Firm 3	-1.000	-0.903	(0.155)	-1.006	(0.144)
Firm 4	-0.500	-0.444	(0.126)	-0.493	(0.130)
Product 2	-1.000	-0.967	(0.108)	-0.795	(0.110)
Product 3	1.000	0.992	(0.107)	0.800	(0.098)
Price	-2.000	-1.960	(0.195)	-1.436	(0.152)

Notes: Number of observations is 1,000. The number of simulated consumers is 200 per observation. Weight on the initial prior $W = 15$. In column (3), the noise term has a mean of 1.005 and a variance of 0.010, which implies that 95 percent of draws is between 0.8 and 1.2.

Summing up

Conclusions

- We have presented a method to estimate search costs in a environment in which consumers are uncertain about the utility distribution.
- Model provides expressions for individual-specific search cost bounds that rationalize observed search and purchasing behavior.
- We have applied the methodology to MP3 players sold online, and find search cost to be sizable.