RAND Journal of Economics
Vol. 37, No. 2, Summer 2006
pp. 257–275

Using price distributions to estimate search costs

Han Hong*

and

Matthew Shum**

We show how the equilibrium restrictions implied by standard search models can be used to estimate search-cost distributions using price data alone. We consider both sequential and nonsequential search strategies, and develop estimation methodologies that exploit equilibrium restrictions to recover estimates of search-cost heterogeneity that are theoretically consistent with the search models. We illustrate the method using online prices for several economics and statistics textbooks.

1. Introduction

■ Ever since the seminal article by Stigler (1964), search models have played an important role in economics. Search frictions resulting from agents' imperfect information about sellers' prices have been used to explain many economic phenomena, including equilibrium price dispersion in otherwise homogeneous product markets. While the search paradigm has been and continues to be very important in the theoretical literature, explicit measures of search costs are few and far between.[1]

In this article we develop a methodology for recovering search-cost estimates that requires only observed price data, and that is theoretically consistent with the equilibrium search models. By doing this, we can evaluate the ability of these theoretical models to explain observed patterns of price dispersion. We consider equilibrium models of sequential and nonsequential search, two important search strategies considered in the theoretical literature.

We illustrate our methods using prices obtained from a number of online booksellers. Our emphasis on estimating search costs that are theoretically consistent with equilibrium search

* Duke University; hanhong@econ.duke.edu.

** Johns Hopkins University; mshum@jhu.edu.

We thank the Editor, Rob Porter, and two anonymous referees for their extensive and valuable suggestions. We also thank Austan Goolsbee, Joe Harrington, Sridhar Moorthy, Geert Ridder, Seth Sanders, Randal Watson, and seminar participants at Johns Hopkins, Toronto, the 2002 SITE meetings at Stanford University, the 2002 Winter Econometric Society Meetings in Atlanta, and the 2002 Society of Economic Dynamics Meetings in New York for helpful comments. Matthew Langley and Wei Tan provided excellent research assistance. We gratefully acknowledge support from the NSF (grant nos. SES-0335331, SES-0003352).

[1] See Stigler's (1964) article as well as Pratt, Wise, and Zeckhauser (1979) for early documentations of the extent of price dispersion in many presumably homogeneous product markets.

models distinguishes our approach from existing empirical studies of search in online markets. For example, Brynjolfsson and Smith (2000), Clay, Krishnan, and Wolff (2001), and Goolsbee (2001) have attempted to understand the nature of search costs by comparing the degree of price dispersion between online and traditional retail markets.[2]

However, one difficulty in measuring search costs is that often the only data the researcher has at her disposal are prices. This is true for the articles cited above, and is especially true for online markets, in which price data are easily obtained but quantities are difficult to measure: indeed, recent work by Chevalier and Goolsbee (2003) illustrates how difficult it is to obtain quantity information about online commerce, and they use several methods—some quite costly, as they involve actual book purchases—to infer online book sales from sales rankings.

The main message of this article, however, is that sometimes price data alone suffice for estimating search costs. This is because the equilibrium supply-demand restrictions of the theoretical models place many restrictions on the observed price data, so that one can exploit these restrictions to recover estimates of search costs using only observed prices.[3] We illustrate our methodology by obtaining search-cost estimates implied by online price data for several economics and statistics textbooks.

Our emphasis on the identification of search costs using only price data requires some strong modelling assumptions. For instance, we assume that each observed price quote is "real" in the sense that it generates positive sales for the retailer, so that our empirical framework does not accommodate "bait and switch" strategies whereby consumers are lured by a low price to a website where they are then steered to a higher-priced product.[4]

The remainder of the article is organized as follows. In Section 2 we describe the two search paradigms we considered and describe our methodologies for recovering estimates of search costs from observed price data alone. In Section 3 we present illustrative estimation results using online prices for several economics and statistics textbooks. Section 4 concludes.

2. Equilibrium search models with heterogeneous consumer search

■ We begin by reviewing two main equilibrium search paradigms considered in the existing theoretical literature, and derive the equilibrium restrictions on prices. In each case, we discuss how we recover estimates of search costs that are consistent with those models. These two paradigms—the nonsequential and sequential search models—are distinguished by the differing assumptions made about consumers' search strategies.

In our analysis, we take consumers' search strategies as given and do not consider the optimality of these search strategies. For a discussion of these issues, see the treatment in Manning and Morgan (1985). Generally speaking, nonsequential (of "fixed sample size") search strategies can be optimal when there is a fixed-cost component to search (so that even if the per-sample search cost is constant, average search costs are decreasing in the number of samples taken). Sequential search can be optimal when such fixed costs are absent. We have not been able to obtain identification and estimation results for a more general model in which consumers choose their search strategy.

Throughout, we assume that the price dispersion observed in the data arises as an equilibrium outcome due to heterogeneity in consumers' search costs (i.e., their costs associated with discovering a given retailer's price). Furthermore, we maintain the assumption that all firms offer

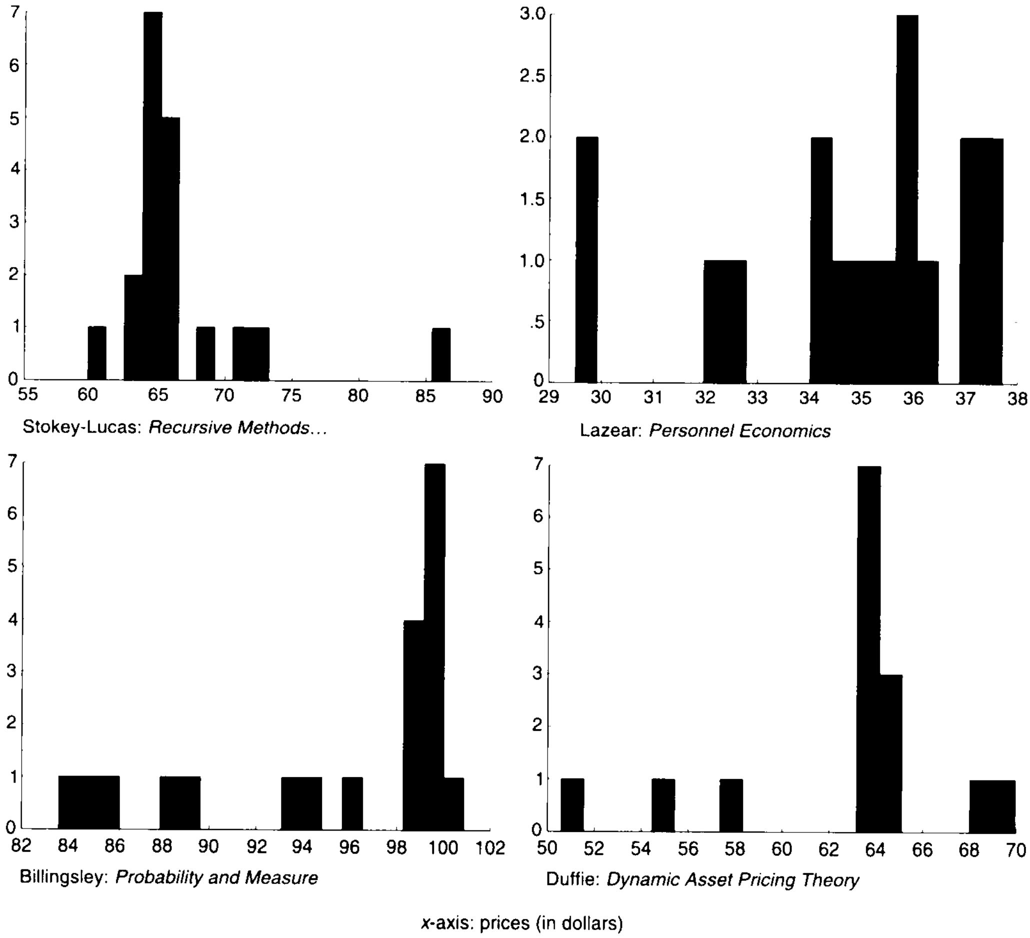
[2] Also see Brown and Goolsbee (2002) and Scott-Morton, Zettelmeyer, and Silva-Risso (2001) for studies of how the Internet has reduced transactions costs (part of which are price search costs) for insurance and automobiles.

[3] See Sorensen (2001) and Hortacsu and Syverson (2005) for structural analyses of search in traditional non-online markets, in which they employ unique datasets in which both price and quantity are observed. It is difficult to get this type of data for typical online markets. Similarly, the large empirical search literature in labor economics (e.g., Eckstein and Wolpin, 1990) often uses not only wage data, but also auxiliary data (such as unemployment and employment durations) to pin down "arrival rate" parameters in job-search models that are missing from price-dispersion models.

[4] See Ellison and Ellison (2001) for a study of such phenomena in online computer parts retailers.

FIGURE 1

RAW HISTOGRAMS OF ONLINE PRICES



homogeneous products, so that only search frictions (arising from consumers' imperfect information about stores' prices) and heterogeneity in search costs in the consumer population generate price dispersion in this market.

As a motivating example, we consider 20 online prices for the classic mathematical statistics textbook *Probability and Measure* (Billingsley, 1992), retrieved by the MySimon (www. mysimon.com) and Pricescan (www.pricescan.com) search engines, on February 5, 2002. A histogram of these prices is given in the third panel of Figure 1.[5]

The long left tail of the histogram suggests that consumers may have an incentive to search, because the potential cost savings can be over \$15 (the lowest price, \$85.58, is over \$15 less than the highest price of \$100.87). On the other hand, the large spike in the histogram around \$100 suggests that despite the low prices, consumers may not be searching very much, because otherwise firms would not find it optimal to "pile up" at a relatively high price. These arguments illustrate that the search models may imply conflicting interpretations of observed prices, if only the consumer or firm side is considered in isolation. For this reason, in this article we impose the optimality conditions for both consumer search and firm pricing in obtaining our estimates of search costs, thus rationalizing the observed prices as equilibrium outcomes for a given theoretical model.

[5] These prices include shipping costs.

Following Burdett and Judd (1983), we assume that there is a continuum of firms and consumers and interpret the equilibrium price distribution F_p as the symmetric equilibrium mixed strategy employed by all firms.[6] Let $\underline{p}$ and $\overline{p}$ denote, respectively, the lower and upper bound of the support of F_p. r denotes the common unit selling cost of each retailer.[7] A maintained assumption of our empirical models is that r is common across retailers. This seems reasonable because heterogeneity in r cannot, per se, generate price dispersion without heterogeneity in search costs (because high r generally leads to higher prices, so that high-search-cost consumers are required to generate demand for the high-r stores). Furthermore, the focus of this article is on estimating equilibrium search models using price data alone, and price data alone will not suffice to identify both the consumers' search cost and the retailers' selling cost distributions.

We assume that consumers have inelastic demand for a single unit of the good and incur a search cost c to retrieve each price quote after the first quote obtained, which, as is standard in the literature, we assume is obtained at no cost. Across the consumer population, search costs are heterogeneous and assigned via i.i.d. draws from a distribution F_c. The focus of this article is to develop methodologies for estimating F_c using only a sample of random prices drawn from F_p.

Nonsequential search. The first of the two search paradigms we shall consider is nonsequential search. Consumers who search nonsequentially are those who commit to buying from the lowest-priced store after obtaining a random sample of ℓ (≥ 1) prices. A consumer with per-price search cost c chooses the number of stores ℓ to canvass to minimize her total expected cost, which is the sum of her total search costs and the price she expects to pay for the product:

$$\ell^*(c) \equiv \underset{\ell \geq 1}{\operatorname{argmin}}\ c \cdot (\ell - 1) + \int_{\underline{p}}^{\overline{p}} \ell \cdot p(1 - F_p(p))^{\ell - 1} f(p)dp \equiv \mathcal{C}(\ell; c). \tag{1}$$

Since the cost c only enters the first term of the above expression, it is obvious that $\ell^*(c)$ is monotonically decreasing in c.

Nonparametric estimation of nonsequential-search model. The optimality condition (1) allows us to recover a nonparametric estimate of the population search-cost distribution F_c just from the observed prices. Since consumers are assumed to draw i.i.d. samples from the equilibrium price distribution F_p, the marginal expected savings from searching $i + 1$ versus i stores is simply

$$\Delta_i \equiv Ep_{1:i} - Ep_{1:i+1}, \qquad i = 1, 2, \ldots, \tag{2}$$

where $p_{1:i}$ denotes the lowest price out of i draws from the equilibrium price distribution F_p. That is, the expected savings is just the expected difference in the lowest out of $i + 1$ price quotes, and the lowest out of i price quotes.

Since the sequence of marginal expected savings Δ_i, $i = 1, 2, \ldots$ is nonincreasing in i for any price distribution F_p,[8] while the cost per search is constant, a consumer with search cost c will search as long as the marginal expected savings Δ_i exceeds his marginal search cost c, so that $\ell^*(c) = \operatorname{argmax}_i \Delta_i$ such that $\Delta_i > c$. Therefore, the sequence of marginal expected savings $\Delta_1, \Delta_2, \ldots$ can also be interpreted as the search costs of the "indifferent" consumers: Δ_i is the search cost faced by the consumer who is indifferent between searching $i + 1$ and i stores.

This is illustrated in Figure 2, where the areas of the regions A, B, C, and D are the measure of consumers who obtain (respectively) one, two, three, or four price quotes. Δ_1, Δ_2, Δ_3, and

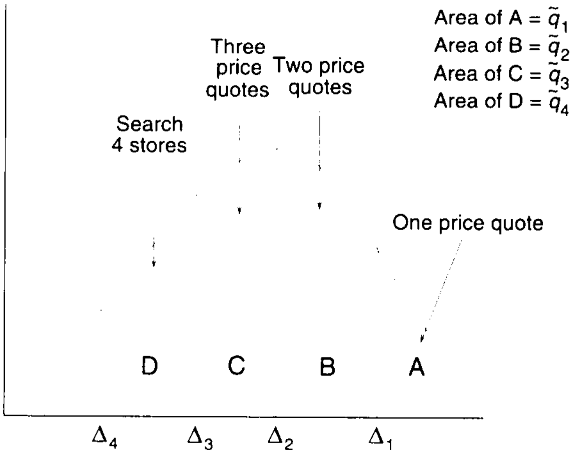
[6] This contrasts with Stahl (1989), in which a finite number of firms is considered.

[7] From the online retailer's viewpoint, this selling cost includes not only its wholesale cost, but also its selling costs—the labor involved in taking and fulfilling orders, the degradation of the firm's computer or warehouse capital, etc.

[8] This can be seen by noting that $Ep_{1:i} = \underline{p} + \int_{\underline{p}}^{\overline{p}} (1 - F(p))^i dp$, which is a nonincreasing and convex function of i (as can be seen by differentiation).

FIGURE 2

IDENTIFICATION SCHEME FOR SEARCH-COST DISTRIBUTION IN NONSEQUENTIAL-SEARCH MODEL



Δ_4 are the search costs of the indifferent consumers, where a consumer with search costs Δ_k is indifferent between obtaining k and $k+1$ price quotes.

Let $\hat{F}_p$ denote the empirical distribution of the observed prices. First, we note that we can obtain estimates of these indifference points from the empirical price distribution $\hat{F}_p$, via the relation (2). Second, define

$$\begin{aligned} \tilde{q}_1 &\equiv 1 - F_c(\Delta_1): &&\text{the proportion of consumers with one price quote;}\\ \tilde{q}_2 &\equiv F_c(\Delta_1) - F_c(\Delta_2): &&\text{the proportion of consumers with two price quotes;}\\ \tilde{q}_3 &\equiv F_c(\Delta_2) - F_c(\Delta_3): &&\text{the proportion of consumers with three price quotes.} \end{aligned} \quad (3)$$

We can estimate $\tilde{q}_1, \tilde{q}_2, \ldots$ by exploiting the firms' equilibrium pricing conditions. To see this, note that a firm's profits from following the mixed pricing strategy $F_p(\cdot)$ are (see Burdett and Judd, 1983)

$$\Pi(p) = (p - r)\left[\sum_{k=1}^{\infty} \tilde{q}_k k(1 - F_p(p))^{k-1}\right]$$

for all $p \in [\underline{p}, \overline{p}]$. The characterization of the equilibrium price distribution starts with the mixed-strategy condition that firms be indifferent between charging the monopoly price $\overline{p}$ (and selling only to people who never search but receive an initial free draw equal to $\overline{p}$) and any other price p in the equilibrium support $[\underline{p}, \overline{p}]$:

$$(\overline{p} - r)\tilde{q}_1 = (p - r)\left[\sum_{k=1}^{\infty} \tilde{q}_k k(1 - F_p(p))^{k-1}\right]. \quad (4)$$

The optimality equation (4) allows us to recover a nonparametric estimate of the search-cost distribution F_c from $\hat{F}_p$ alone, as we now show. Let $\hat{\underline{p}}$ and $\hat{\overline{p}}$ denote the lowest and highest observed prices, respectively. For convenience, we index the n observed prices in ascending order, so that

$$\hat{\underline{p}} = p_1 \leq p_2 \leq \cdots \leq p_{n-1} \leq p_n = \hat{\overline{p}}.$$

Let K ($\leq n - 1$) denote the maximum number of firms from which a consumer obtains price quotes in this market. Given this condition, the indifference condition (equation (4)) for each of the observed prices is

$$(\overline{p} - r)\tilde{q}_1 = (p_i - r)\left[\sum_{k=1}^{K} \tilde{q}_k k(1 - \hat{F}_p(p_i))^{k-1}\right], \qquad i = 1, \ldots, n-1. \tag{5}$$

Since $\tilde{q}_K = 1 - \sum_{k=1}^{K-1} \tilde{q}_k$, the above constitutes $n - 1$ equations from which we can solve for the K unknowns $\{r, \tilde{q}_1, \ldots, \tilde{q}_{K-1}\}$.[9]

Subsequently, through equation (3), we use the values of $\tilde{q}_1, \ldots, \tilde{q}_{K-1}$ to solve for

$$F_c(\Delta_1), \ldots, F_c(\Delta_{K-1}),$$

the cumulative distribution function of the search costs evaluated at the indifference points $\Delta_1, \ldots, \Delta_{K-1}$.

To obtain standard errors for the estimates, we formulate the estimation problem as an empirical likelihood problem. The firms' equilibrium indifference condition (4) implies a potentially infinite number of moment conditions, because it holds for every p (excepting $\overline{p}$) in the support of F_p. However, in estimation, we will use only a finite number $M < \infty$ of these conditions (with $M \geq K$). In practice, the values of K and M will be dictated by the number of observed prices, as will be illustrated in our empirical work. In deriving the asymptotic standard errors, we assume that K and M are fixed and finite, as the number of prices $n \to \infty$. Complete details are given in the Appendix.[10]

As noted by Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998), empirical likelihood has several advantages. First, it provides an alternative method of obtaining the efficient GMM estimates using an optimal weighting matrix. The variance-covariance matrix of the empirical likelihood estimates is asymptotically equivalent to that for the efficient GMM estimates (see Owen, 2001). Additionally, even if the underlying model is misspecified, the empirical likelihood estimates still have an interpretation as the estimates that minimize an information-theoretic criterion between the true (but unknown) model and the family of (potentially misspecified) models defined by the estimating moment conditions.

A first example: Billingsley (1992). We illustrate the estimation procedure for nonsequential-search models using Billingsley. The indifferent search-cost values were computed directly from equation (2), using the observed empirical distribution of prices. The first five cutoff points are Δ_1 = \$2.90; Δ_2 = \$2.00; Δ_3 = \$1.49; Δ_4 = \$1.04; Δ_5 = \$0.81.

In the top panel of Figure 3, we graph the population search-cost distribution, estimated in the manner detailed above, for $K = 3$ and $M = 5$. The corresponding estimates of $\tilde{q}_1, \tilde{q}_2$, and $\tilde{q}_3$ are .633, .309, and .058. Note that we cannot identify the shape of the search-cost distribution above the 35th quantile: the estimates indicate that 63.3% of the consumers only search once, so that all we can say about these consumers is that their search costs lie above Δ_1 = \$2.90. These costs rationalize a sample of 20 prices for Billingsley's *Probability and Measure*, from online retailers and were retrieved on February 5, 2002 by the MySimon and Pricescan search engines.

□ **Sequential search.** The second paradigm we consider is the sequential-search model. The sequential-search strategy differs from the nonsequential-search process in that, after each search, consumers can choose to purchase at the lowest price observed so far, or make an additional search. At any price, there is an option value associated with searching again, and the optimal search problem is analogous to an "optimal stopping" problem. Sequential search is the standard assumption made in job-search models in labor economics (see Mortensen's (1986) survey).

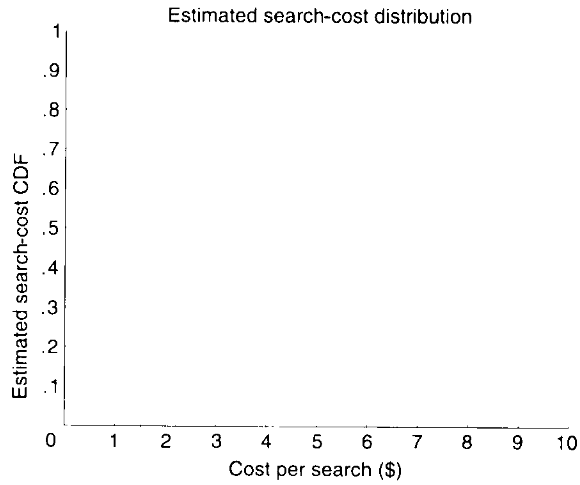
---

[9] For this reason, we required $K \leq n - 1$ above.

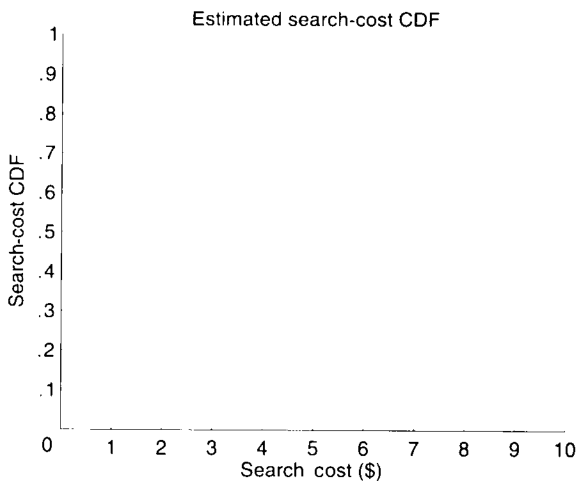
[10] Asymptotically, as n grows large, we can recover an arbitrarily accurate estimate of F_c by taking larger and larger values for K and M. This fully nonparametric case (where both K and M also approach infinity at some rate relative to n) is technically challenging (because the limit problem involves an infinite ($M \to \infty$) number of estimating moment conditions) and beyond the scope of this article.

FIGURE 3

ESTIMATED SEARCH-COST CDFs



Nonsequential search model ($K = 3$, $M = 5$)



Sequential-search model (gamma search-cost distribution)

An important early result in the sequential-search literature is the "Diamond (1971) paradox" that the equilibrium price distribution in a sequential-search economy with homogeneous consumers facing identical and positive search costs is degenerate at the monopoly price (i.e., the highest acceptable price of consumers). In this article, we follow the work by Albrecht and Axell (1984), Stahl (1989), and Rob (1985) in positing a model with heterogeneous search costs in order to generate a nondegenerate equilibrium price distribution. Intuitively, with heterogeneity in consumer search costs, low-price firms can "cater" to consumers with low search costs, and high-price firms cater to consumers with high search costs. However, consumer search-cost heterogeneity is not sufficient to ensure the existence of a continuous equilibrium price distribution. Rob (1985) provides a discussion of restrictions on the population search-cost distribution that are necessary for nondegenerate equilibrium price dispersion in the sequential-search model.[11]

[11] Specifically, Rob's Theorem 3 states that a sufficient condition for nonexistence of a continuous equilibrium price distribution is that the search-cost density vanishes in some positive interval $[0, A)$ with $A > 0$. We rule this out in

Specifically, in his Theorem 4, Rob shows that a continuous equilibrium price distribution, if it exists, must satisfy an equation analogous to (10) below. Hence, one can check for the existence of a continuous equilibrium price distribution by verifying that the function in (10) is a valid CDF; equivalently, we can check that the likelihood function (9) is a proper density function, which is positive along its support. We shall return to this issue below.

As before, let F_c denote the population distribution of search costs and F_p the equilibrium price distribution, with support $[\underline{p}, \overline{p}]$. A standard result in the sequential-search literature is that the consumers' optimal stationary-search strategy is a reservation price policy, where they search until they obtain a price that is no larger than some reservation price p^*, which is independent of the number of searches that have been made. The empirical question is whether observation of the price distribution F_p is enough to identify the search-cost distribution F_c.

For consumer i, who has per-price search costs c_i, let $z^*(c_i)$ denote the price z that satisfies the indifference condition

$$c_i = \int_{\underline{p}}^{z} (z - p) F_p(dp) = \int_{\underline{p}}^{z} F_p(p) dp, \tag{6}$$

where the second equality follows from integration by parts. This equation has a straightforward economic interpretation: the left-hand side is the cost of search, and the right-hand side is the expected benefit if the best currently held price is z. Note that $z^*(c)$ is increasing in c. Now, for each cost c_i, we define the reservation price as

$$p_i^* \equiv \overline{p}(c_i) \equiv \min(z^*(c_i), \overline{p}). \tag{7}$$

Let G denote the distribution of reservation prices in the population, given F_c and the mapping (7). Note that there is a mass $\alpha \equiv 1 - G(\overline{p})$ of consumers for whom the reservation price is $\overline{p}$, and that $G(\underline{p}) = 0$.

Turning to the firms' decision problems, we use (as before) an indifference condition to define the equilibrium price distribution. Suppose consumer i has reservation price p_i^*. The firm charging $\hat{p}$ will sell only to consumers i for whom $\hat{p} < p_i^*$. Since we assume that all firms are symmetric, a firm's demand at price $\hat{p}$ is proportional to $(1 - G(\hat{p}))$.[12] Note that at the highest price $\overline{p}$, the firm obtains demand proportional to α, the measure of consumers with reservation price equal to $\overline{p}$. Therefore the firms' indifference condition is, for each $p \in [\underline{p}, \overline{p})$,

$$(\overline{p} - r)D(\overline{p}) = (p - r)D(p) \Longleftrightarrow (\overline{p} - r) * \alpha = (p - r) * (1 - G(p)). \tag{8}$$

Before proceeding, we note that the indifference conditions above capture only the stationary search dynamics in this market and do not allow the consumer population to be distributed differently over time, due perhaps to the entry and exit of consumers. This simplification is a matter of convenience, given that what we exploit in estimation is a single cross-section of prices. If the population of consumers changed over time, firms would wish to change their pricing strategy over time. This would feed back to consumers, who would wish to change their reservation price over time. The resulting dynamic equilibrium involves modelling issues that we believe to be beyond the scope of this article.[13]

our empirical work below by assuming that search costs are distributed according to the gamma distribution, which has support $[0, \infty)$.

[12] This derivation relies on the assumption that there is an infinite number of firms. With only a finite ($N < \infty$) number of firms, the probability that a consumer with reservation price p_i^* encounters a firm charging a price she accepts is no longer 1 (as in the infinite-firm case) but rather $1 - (1 - F_p(p_i^*))^N$. Hence, the expected demand for a firm charging $\hat{p}$ would be proportional to $\int_{\hat{p}}^{\overline{p}} (1 - (1 - F_p(p^*))^N) dG(p^*)$, which is substantially more involved to compute than the expression for the continuum-firm case (which is $1 - G(\hat{p})$).

[13] We thank Randal Watson for this insight.

Estimation of the sequential-search model. Unlike the nonsequential-search model, it is difficult to estimate the equilibrium sequential model nonparametrically. This is because equation (8) defines $n-1$ equations, for each of the $n-1$ observed prices (with $\overline{p}$ excluded).[14] However, there are n unknowns: $G(p_1 = \underline{p}), G(p_2), \ldots, G(p_{n-1}), G(p_n = \overline{p})$. Therefore, the model cannot be estimated without some additional assumptions.

Hence, we find it natural to consider parametric maximum-likelihood estimation (MLE) of this model, assuming that the search-cost distribution $F_c(\cdot;\boldsymbol{\theta})$ is parameterized by a (finite-dimensional) vector $\boldsymbol{\theta}$.[15] After some algebraic manipulation, the likelihood function for each price can be derived as

$$f_p(p;\boldsymbol{\theta}) = -\frac{2\alpha(\overline{p}-r)}{(p-r)^3 * f_c\left(c\left(1-\frac{\overline{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)} - \frac{\alpha^2(\overline{p}-r)^2 f_c'\left(c\left(1-\alpha\frac{\overline{p}-r}{p-r};\boldsymbol{\theta}\right)\right)}{(p-r)^4 * \left[f_c\left(c\left(1-\alpha\frac{\overline{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)\right]^3}. \quad (9)$$

(See the Appendix for details.) The corresponding equilibrium price CDF is

$$F_p(p;\boldsymbol{\theta}) = \frac{\alpha(\overline{p}-r)}{(p-r)^2 * f_c\left(c\left(1-\alpha\frac{\overline{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)}. \quad (10)$$

In the above equations, we denote $c(\tau;\boldsymbol{\theta}) \equiv F_c^{-1}(\tau;\boldsymbol{\theta})$, the inverse CDF for the search-cost distribution. Given $\boldsymbol{\theta}$, the auxiliary parameters α and r can be solved for as follows: first, α, the proportion of consumers with reservation price equal to $\overline{p}$, is defined via the initial condition $G(\underline{p}) = 0$, which implies, via (8), that

$$(\overline{p}-r) * \alpha = (\underline{p}-r) * 1 \Longleftrightarrow \alpha = \frac{(\underline{p}-r)}{(\overline{p}-r)}. \quad (11)$$

Second, the selling cost parameter r can be determined by the restriction that $F_p(\overline{p}) = 1$ so that, given $\boldsymbol{\theta}$ and (10) and (11), r must satisfy

$$1 = F_p(\overline{p}) = \frac{(\underline{p}-r)}{(\overline{p}-r)^2 * f_c\left(c\left(1-\frac{\underline{p}-r}{\overline{p}-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)}. \quad (12)$$

The likelihood function for the whole sample of prices, then, is just $L(\boldsymbol{\theta}, r) = \prod_i f_p(p_i;\boldsymbol{\theta})$.

The discussion in Rob (1985) suggests that a necessary condition for the existence of a non-degenerate continuous equilibrium price distribution for this model within an interval $[\underline{p}, \overline{p}]$ is that the price CDF in (10) is nondecreasing in this range or, equivalently, that the likelihood function (9) be positive for all $p \in [\underline{p}, \overline{p}]$. This is a strong condition; examination of (9) suggests that it requires the search-cost density f_c to be strictly decreasing in the range $[0, c((\overline{p} - \underline{p})/(\overline{p} - r);\boldsymbol{\theta})]$.

In our empirical work, we use a gamma distribution for the search-cost distribution. The gamma distribution was chosen for its flexible shape and for the fact that it is one of the few two-parameter distributions that allows for the density to be strictly decreasing along its full support. For the gamma density,

$$f_c(c;\delta_1,\delta_2) = \frac{1}{\delta_2^{\delta_1}\Gamma(\delta_1)} c^{\delta_1 - 1} \exp(-c/\delta_2), \qquad \delta_1, \delta_2 > 0, \quad (13)$$

[14] If there are multiple observations of the same prices, then there are even fewer equations.

[15] As before, $\underline{p}$ and $\overline{p}$ can be (superconsistently) estimated from the data. See Donald and Paarsch (1993) for a discussion of maximum-likelihood estimation when a subset of the parameters can be superconsistently estimated.

the decreasing density requirement restricts $\delta_1 \leq 1$. We confirm that this restriction is satisfied at all the estimates we report below. Moreover, this restriction has important implications for the search-cost estimates that we obtain from this model, as we discuss below.

Example: Billingsley (1992). Returning to the example of prices for Billingsley considered earlier, we estimated the parameters of the gamma search-cost distribution. The estimated search-cost CDF is plotted in the bottom panel of Figure 3. Comparing the search-cost estimates in the top and bottom panels of Figure 3, one notices that the estimated search-cost distributions are similar in magnitude in the range of the quantiles that are identified, but the parametric specification of the sequential search model allows one to extrapolate the shape of the entire search-cost distribution. At the estimated parameter values, the median search cost implied by these estimates is $9.22 (and the mean search cost is even higher, at over $51). We return to the effects of parametric extrapolation below.

3. Empirical illustration: a snapshot of February 5, 2002

■ In this section we illustrate the use of our methodology to recover the search costs that are consistent with these two models, estimated for a series of online book prices collected on February 5, 2002 from the Pricescan.com and MySimon.com websites.[16] We attempted to select product markets that most resemble the prototypical markets considered in the search models, namely homogeneous-product markets in which consumers would demand only a single unit of the product. For this reason, we focus on several economics and statistics textbooks. An attempt was made to verify each price so that, in most cases, visits were made to each retailer's website to confirm not only availability but also shipping and handling information.[17]

Since our analysis is on a product-by-product basis, and given the computational issues related to the estimation technique (especially for the nonsequential-search models), we restrict our analysis to four books. We supplemented our empirical exercise with an informal survey of a dozen economists regarding their book-buying behavior. All the respondents stated that if they had to purchase these four books, they would do so online. Given these findings, we believe that the search-cost distributions we are recovering may pertain to the population of academic economists.

Summary statistics for the prices used in the estimation are given in Table 1. Throughout, we report results for prices inclusive of shipping and handling charges.[18] Raw histograms of the prices are given in Figure 1.

Note that the four histograms are quite varied in shape: there is some evidence of a long upper tail for Stokey-Lucas (1989) and long lower tails for Billingsley (1992) and Duffie (1992). There are prominent spikes around the list price for the Stokey-Lucas, Billingsley, and Duffie texts. Prices for the Lazear (1995) text are relatively uniformly distributed within their range.

In recent work, Ellison and Ellison (2001) suggested that online retailers employed "bait-and-switch" strategies with price search engines, by advertising low prices but overcharging for shipping and handling. We expect that these strategies might be reflected in the data by lower dispersion in the prices with shipping costs than without shipping costs, as well as a low correlation between prices with and without shipping costs. However, for each of the four books,

[16] We also estimated the models using similar data from June 2001 and June 2002, but because prices did not change substantially over this period, the results were not noticeably different. The search models we use are stationary in nature and cannot capture the price persistence across time that we observe. Accommodating these dynamics in a search context may require a model incorporating dynamic learning by consumers and dynamic pricing by firms, similar to the models explored in Bergemann and Välimäki (1997, 2004).

[17] Often the shipping and handling information was not directly available on the pricescan.com search engine's page. Also, our price data were collected before the two major online booksellers, amazon.com and barnesandnoble.com, instituted free shipping for orders over $25. Furthermore, we checked a number of booksellers' websites and confirmed that these books were available, although the estimated shipping times varied.

[18] We also estimated the models without the shipping costs but do not report them because the results were largely unchanged.

TABLE 1 **Summary Statistics on Prices for Different Products**

Product	n	List	Mean	Standard Deviation	Median	$\underline{p}$	$\overline{p}$
Stokey-Lucas	19	60.50	66.60	5.64	64.98	59.75	86.80
Lazear	17	31.95	34.73	2.48	35.27	29.51	37.70
Billingsley	20	99.95	95.48	5.87	98.90	83.58	100.87
Duffie	15	65.00	62.71	4.91	63.48	50.58	69.95

Note: Including shipping and handling costs. Price data for all products downloaded from Pricescan.com and MySimon.com: February 5, 2002. Summary price including S&H costs may not exceed the corresponding summary price without S&H costs, since we could not determine the shipping and handling charges from some of the websites.

the standard deviation of the prices with shipping costs was roughly the same magnitude as the prices without shipping costs. Moreover, the Spearman rank correlation statistics between the two sets of prices was around .90, indicating strong correlation. This evidence might cast some doubt on the possibility that online retailers are engaging in bait-and-switch strategies with regard to shipping costs.

Table 2 contains estimates for the nonsequential-search model. The estimates of the $\tilde{q}$'s indicate that in most markets, about half of the consumers never search (more precisely, they shop at the store where they received their initial "free" price). For the Stokey-Lucas text, 52%(= $100 * (1 - .480)$) of purchasers have search costs exceeding $\Delta_1 = \$2.32$, while over 60% of the Lazear book purchasers have search costs exceeding $\Delta_1 = \$1.31$. As was the case with the example considered above, the substantial proportion of people who don't search implies that we cannot identify the shape of the distribution for these people. For example, we know nothing about the shape of the search-cost distribution above the 52nd quantile for the Stokey-Lucas book, and above the 65th quantile for the Lazear book.

Table 3 presents the maximum likelihood estimates for the sequential-search model. Comparing these results to those obtained from the nonsequential-search models, we see that the sequential-search model predicts higher magnitudes for search costs: as an example, the median

TABLE 2 **Search-Cost Distribution Estimates for Nonsequential-Search Model**

Product	K[a]	M[b]	$\tilde{q}_1$[c]	$\tilde{q}_2$	$\tilde{q}_3$	Selling Cost r	MEL Value
Parameter estimates and standard errors: nonsequential-search model							
Stokey-Lucas	3	5	.480 (.170)	.288 (.433)		49.52 (12.45)	102.62
Lazear	4	5	.364 (.926)	.351 (.660)	.135 (.692)	27.76 (8.50)	84.70
Billingsley	3	5	.633 (.944)	.309 (.310)		69.73 (68.12)	199.70
Duffie	3	5	.627 (1.248)	.314 (.195)		35.48 (96.30)	109.13
Search-cost distribution estimates							
	Δ_1		$F_c(\Delta_1)$	Δ_2	$F_c(\Delta_2)$	Δ_3	$F_c(\Delta_3)$
Stokey-Lucas	2.32		.520	.68	.232		
Lazear	1.31		.636	.83	.285	.57	.150
Billingsley	2.90		.367	2.00	.058		
Duffie	2.41		.373	1.42	.059		

[a] Number of quantiles of search cost F_c that are estimated (see equation (5)). In practice, we set K and M to the largest possible values for which the parameter estimates converge. All combinations of larger K and/or larger M resulted in estimates that either did not converge or did not move from their starting values (suggesting that the parameters were badly identified).

[b] Number of moment conditions used in the empirical likelihood estimation procedure (see equation (17)).

[c] For each product, only estimates for $\tilde{q}_1, \ldots, \tilde{q}_{K-1}$ are reported; $\tilde{q}_K = 1 - \sum_{k=1}^{K-1} \tilde{q}_k$.

[d] Indifferent points Δ_k computed as $Ep_{(1,k)} - Ep_{(1,k+1)}$ (the expected price difference from having k versus $k+1$ price quotes), using the empirical price distribution. Including shipping and handling charges.

TABLE 3 Estimates of Sequential-Search Model

Product	δ_1	δ_2	Median[a] Search Cost	Selling Cost r	α[b]	$F_c^{-1}(1-\alpha;\theta)$	Log-L Value
Stokey-Lucas	.46 (.02)	1.55 (.03)	29.40 (1.45)	22.90 (1.31)	.58	19.19	31.13
Lazear	.40 (.01)	1.15 (.01)	16.37 (1.00)	11.31 (.79)	.69	4.56	34.35
Billingsley	.25 (.01)	2.01 (.04)	9.22 (.94)	65.37 (.83)	.51	8.43	23.73
Duffie	.21 (.02)	4.57 (.29)	10.57 (2.01)	28.24 (1.63)	.54	7.00	18.93

Note: Including shipping and handling charges. Standard errors in parentheses. δ_1 and δ_2 are parameters of the gamma distribution; see equation (13).

[a] As implied by estimates of the parameters of the gamma search-cost distribution.

[b] Proportion of consumers with reservation price equal to $\overline{p}$, implied by estimate of r (see equation (11)).

search cost for the Stokey-Lucas text consistent with the nonsequential-search model (as reported in the bottom panel of Table 2) is roughly \$2.32, whereas the corresponding number for the sequential-search model (as reported in Table 3) is \$29.40, over 10 times higher. Moreover, the estimates of r, the selling costs, are uniformly lower across the four books than the corresponding estimates in the nonsequential-search model.[19]

At first glance, the larger search-cost magnitudes for the sequential-search model might lead one to support the nonsequential-search model as a better descriptor of search behavior in the markets we consider. As pointed out by Morgan and Manning (1985), nonsequential-search strategies may be optimal when consumers face nonzero fixed costs of initiating a search, regardless of how many prices one obtains. Such a situation may describe the online market, since there may be nonzero search costs associated with, say, seeking out a computer, logging on, and so forth. Moreover, due to search engines, price quotes might tend to be obtained in groups, which may fit better with the nonsequential-search assumption.

However, as we remarked earlier, the large median search-cost estimates for the sequential model are due in part to parametric extrapolation. In the fourth column of Table 3, we report the implied estimates of α, the proportion of consumers with reservation price equal to $\overline{p}$ and, hence, the proportion of consumers who never search.[20] α exceeds .50 for all four products, suggesting that the high median search costs estimated in this specification are due in part to extrapolation based on the gamma functional form for the search-cost distribution. To see if this is true, we also computed, in the fifth column, implied estimates of $F_c^{-1}(1-\alpha)$, which are the search costs for the consumer who has a reservation price just equal to $\overline{p}$. We see that for three out of the four cases (excepting the Lucas-Stokey book), the value of $F_c^{-1}(1-\alpha)$ is comparable in magnitude to the search costs from the nonsequential model (and reported in the bottom panel of Table 2).[21]

□ **Specification checks for the sequential-search model.** On the other hand, we might worry that the sequential-search model may be biased toward large search-cost estimates, due to the necessary condition (remarked above) that the slope of the search-cost density $f_c(\cdots)$ be negative in order for the price density (9) to constitute a valid equilibrium price density. For the gamma density function in (13), we found that when $\delta_1 \leq 1$ (which was required for the density to be strictly decreasing), the hazard rate parameter δ_2 must be increased so that the tail of the density does not die off too quickly, which we need in order to fit the price data (which do not have thin

[19] As remarked above, we confirmed that at the reported estimates, the implied search-cost distribution had a strictly decreasing density. This was true for all four books. Furthermore, to verify the robustness of these estimates, we also reestimated the sequential-search models using a variety of starting values, but the reported results were very robust.

[20] α in the sequential model corresponds to $\tilde{q}_1$ in the nonsequential model.

[21] For the Lucas-Stokey text, the high search costs implied by the sequential-search model appear to be driven by the outlier price of \$86.80 (more than \$20 above the list price) charged by opengroup.com. We have confirmed that this price was not a temporary oversight: on February 7, 2004, the price on this site was still \$83.70 (without shipping costs). However, we note again that our modelling approach assumes that each price is "real," in the sense that it generates positive sales in equilibrium. This assumption may not apply to this website.

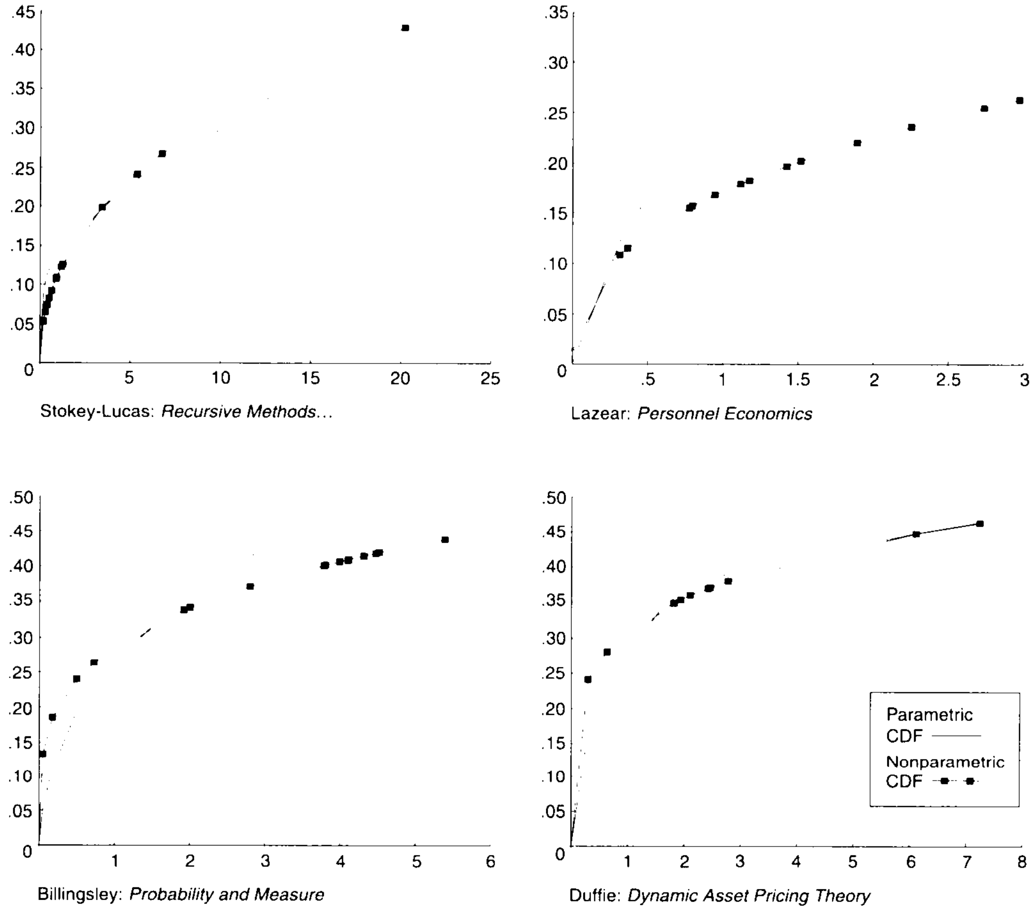
upper tails, as Figure 1 shows) well. This behavior of the gamma density function may lead to larger mean and median search-cost estimates.[22]

Hence, to evaluate the sensitivity of the results to the gamma distribution assumption, we also considered a variety of alternative parametric specifications for the search-cost distribution, including a lognormal, Weibull, a mixture of two gamma distributions, and truncated gamma and normal distributions. In all these cases, either we were not able to obtain estimates that satisfied the nonnegative density restriction for all of the observed prices, or the converged likelihood function was lower than that obtained under the gamma assumption.[23]

As an additional check of the influence that the parametric assumptions may be having on the search-cost estimates for the sequential model, we considered alternative search-cost estimates that rely less on parametric assumptions. Note that for a fixed value of r, we can solve for the

FIGURE 4

SEARCH-COST CDF FOR SEQUENTIAL-SEARCH MODEL: PARAMETRIC VERSUS NONPARAMETRIC ESTIMATES



[22] To confirm that the larger search costs may be related to the decreasing density restriction, we fit the nonparametric estimates of the search-cost distribution quantiles obtained for the nonsequential-search model (reported in the bottom of Table 2) to the gamma distribution. At the best-fitting parameter values, indeed, the mean and median search costs were smaller than the sequential model estimates, but the gamma densities evaluated at these parameters were characterized by an increasing density along part of its support, which would violate the equilibrium restrictions of the sequential model.

[23] According to Vuong (1989), when testing between two nonnested parametric models, a model with a higher converged log-likelihood function cannot be rejected in favor of a model with the lower converged log-likelihood function.

TABLE 4 **Nonparametric Estimates of Sequential-Search Model, Holding r Fixed**

Product	r	Search-Cost Estimates
***r* fixed at sequential-model estimates**[b]		
Stokey-Lucas	22.90	$F_c^{-1}(0.42) = 20.20$
Lazear	11.31	$F_c^{-1}(0.31) = 2.97$
Billingsley	65.37	$F_c^{-1}(0.49) = 5.39$
Duffie	28.24	$F_c^{-1}(0.46) = 7.24$
***r* fixed at nonsequential-model estimates**		
Stokey-Lucas	49.52	$F_c^{-1}(0.5) = 4.52$[d]
Lazear	27.76	$F_c^{-1}(0.5) = 0.26$
Billingsley	69.73	$F_c^{-1}(0.5) = 3.21$
Duffie	35.48	$F_c^{-1}(0.5) = 3.51$

[a] For fixed r, quantiles of search-cost distribution are obtained nonparametrically using equation (14).

[b] As reported in Table 2.

[c] As reported in Table 3.

[d] Median obtained by linear interpolation.

quantiles of the search-cost distribution nonparametrically via the relation

$$1 - G(p) = \frac{\underline{p} - r}{p - r} = 1 - F_c\left(\int_{\underline{p}}^{p} F_p(x)dx\right) \Rightarrow F_c\left(\int_{p}^{p} F_p(x)dx\right) = \frac{p - \underline{p}}{p - r}, \qquad (14)$$

which must hold for all $p = p_1, \ldots, p_n$. This allows us to solve for the values of the search-cost distribution F_c at the points $\int_{p}^{p_k} F_p(x)dx$ for $k = 2, \ldots, n$, which, from (6), denote the cutoff search costs for the consumers who have reservation prices exactly equal to the observed prices $p_2, \ldots, p_n$.

First, we perform this exercise by fixing r at the estimate obtained for the sequential-search model (and reported in Table 3). The results are reported in the top panel of Table 4. As expected, we were only able to estimate the shape of the search-cost distribution up to the $(1-\alpha)$th quantile, so that the nonparametric estimates of $F_c^{-1}(1-\alpha)$ are reported in Table 4. For all four books, these estimates are comparable to the parametric estimates of $F_c^{-1}(1-\alpha; \boldsymbol{\theta})$ reported earlier in Table 3. Furthermore, in Figure 4, we graph the search-cost distributions estimated using both the parametric and nonparametric methods (with r fixed at the sequential-model estimates). The nonparametric and parametric estimates coincide quite closely, suggesting that the gamma functional-form assumption is not unduly driving the estimates of the search-cost distribution, in the range below the $(1-\alpha)$th quantile.[24]

Second, to see whether the sequential-search model is capable of generating the lower search costs estimated for the nonsequential model, we performed the same exercise again but fixed r at the higher values estimated for the nonsequential-search specification (given in Table 2). The results, presented in the bottom panel of Table 4, show that once r is raised and fixed at the estimated levels of the nonsequential-search model, the implied median search costs for the sequential-search model are roughly of the same magnitude as the nonsequential-search estimates. For example, the median search cost for the Stokey-Lucas text is now \$4.52, not much higher than the \$2.32 obtained from the nonsequential-search model; furthermore, for the Lazear book, the median search cost of \$.26 is lower than the corresponding nonsequential estimate.[25] Hence, it appears

---

[24] In Chen, Hong, and Shum (2004), we derive formal likelihood-ratio tests for distinguishing between nonparametric and parametric competing models.

[25] For these results, we also verified informally that the implied search-cost density function is consistent with an equilibrium price density of the form in (9) by confirming visually that the search-cost CDF was concave, at least in the range defined by the relation (14).

that the sequential-search model can, in principle, generate estimates of search costs comparable to, or even lower than, those of the nonsequential-search model. (However, the parameters that maximize the likelihood function are those that imply larger search costs.[26])

4. Concluding remarks

■ In this article, we proposed new methods to estimate consumer search costs. An important benefit of our approach is that we require data on prices alone; this may make our methods useful for analyzing online markets, for which prices are often the only available data, and quantity data are difficult to obtain. In estimating the population search-cost distribution, we exploit the equilibrium restrictions of several theoretical price-search models.

Although we have focused in this article on search explanations for observed price dispersion, some recent work on online markets has emphasized nonsearch explanations for price dispersion in ostensibly homogeneous-good markets. For example, Ellison and Ellison (2001) focus on sellers' incentives to engage in "obfuscation" or "bait-and-switch" strategies. An assumption of our empirical approach is that all observed prices result in positive sales, which may not hold if retailers engage in bait and switch. Jin and Kato (2002) investigate the increased incidence of adverse selection in online markets due to the buyers' lack of opportunities to inspect products. These features, along with the potential unreliability of retailers to deliver products when promised, can enhance the importance of seller reputation in online markets, thereby leading to retailer differentiation and price dispersion even in the absence of consumer search costs and explicit product heterogeneity.

A challenging extension would be to combine the equilibrium search models considered in this article with rich individual-level datasets (such as supermarket scanner panel datasets or the individual-level drug purchase dataset used in Sorensen (2001)). It would be interesting to investigate how to exploit the equilibrium restrictions of the theoretical search models in identifying consumer search costs with more detailed data. Such additional data would allow us to fit richer models of supply and demand, including models of retailer differentiation (such as Wolinsky (1986)). Hortaçsu and Syverson (2005) use both price and market share data for mutual funds to assess the relative importance of search versus product differentiation in driving price dispersion.

We have not considered the possibility that consumers may substitute between online and traditional retailers. While such an extension involves data collection efforts that are beyond the scope of this article, we mention in conclusion that we also conducted a small telephone survey of electronics retailers in the Baltimore metropolitan area in February 2002 to canvass prices for Palm Pilots, a homogeneous product easily comparable across retailers. Somewhat to our surprise, we found practically no price dispersion for each Palm Pilot model.[27] In the equilibrium search framework, no price dispersion can arise, for two reasons: (i) search costs are zero, and prices represent the Bertrand (zero-profit) equilibrium; (ii) search costs are prohibitively high, and the observed prices represent the equilibrium in which all firms charge the monopoly price.[28] Furthermore, collusion is an additional possibility not considered by these models. Since all of these scenarios lead to observationally equivalent pricing outcomes, it is difficult to test between them without additional data (such as costs or quantities).

---

[26] Indeed, when we reestimated the sequential-search model using the higher nonsequential-search estimates for r as starting values, we converged again to the parameter values reported in Table 3.

[27] We could not do a similar exercise for the economics and statistics books used for the empirical illustration in this article, because none of the large brick-and-mortar bookstores in the Baltimore metropolitan area had them in stock.

[28] The uniformity of prices in brick-and-mortar stores could also be due to resale price maintenance. While explicit RPM is prohibited by law, it might be implemented in practice via "manufacturer's suggested retail prices," which, in fact, exist for Palm Pilots. (However, RPM cases are few and far between, perhaps due to the reluctance of the U.S. Department of Justice to prosecute them.)

Appendix

■ **Maximum empirical likelihood (MEL) estimation procedure for nonsequential search model.** Here, we describe our empirical likelihood approach for estimating the model parameters. As noted by Imbens, Spady, and Johnson (1998) and Kitamura and Stutzer (1997), empirical likelihood also provides an alternative (and potentially more convenient) method for obtaining the efficient GMM estimates of model parameters, based on moment restrictions on the data-generating process.

Our data consist of n prices, $p_i, i = 1, \ldots, n$. Consider a discrete price distribution with n points of support, at $p_i, i = 1, \ldots, n$, with probability weight π_i at point p_i. Thus $F_p(p) = \sum_{i=1}^{n} \pi_i \mathbf{1}(p_i \leq p)$. As before, $\underline{p}$ and $\overline{p}$ can be estimated at a superconsistent rate, so in what follows we treat these parameters as known and nonstochastic.

Using the discrete distribution for $F_p(p)$, we obtain, for all $i = 1, \ldots, n-1$ and fixed K:

$$(\overline{p} - r)\tilde{q}_1 = (p_i - r)\left[\sum_{k=1}^{K} \tilde{q}_k k \left(1 - \left[\sum_{j=1}^{n} \pi_j \mathbf{1}(p_j \leq p_i)\right]\right)^{k-1}\right]. \tag{A1}$$

From (5), evaluated at $\underline{p}$ (and using $F_p(\underline{p}) = 0$), we rewrite r as a function of $\overline{p}$, $\underline{p}$, and $\tilde{\mathbf{q}} \equiv \tilde{q}_1, \ldots, \tilde{q}_K$:

$$r(\tilde{\mathbf{q}}) = \frac{\underline{p} * \left[\sum_{k=1}^{K} \tilde{q}_k k\right] - \overline{p} * \tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\right] - \tilde{q}_1}. \tag{A2}$$

This can be plugged into each of the equations in (A1) in order to eliminate r.

Let

$$\boldsymbol{\theta} \equiv \{r, \tilde{q}_1, \ldots, \tilde{q}_K\}$$

denote the unknown parameters to be estimated. We can transform the restrictions (A1) into estimating equations (or moment conditions) of the form $Ef(x; \boldsymbol{\theta}) = 0$, as follows. For $s_m \in [0, 1]$, $m = 1, \ldots, M$ and $M > K$,

$$\text{(A1)} \Leftrightarrow (\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1 = \left(F_p^{-1}(s_m) - r(\tilde{\mathbf{q}})\right)\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]$$

$$\Rightarrow F_p^{-1}(s_m) = r(\tilde{\mathbf{q}}) + \frac{(\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]} \equiv g_{s_m}(\tilde{\mathbf{q}}).$$

This population quantile restriction (that the s_mth quantile of $F_p(p)$ equals $g_{s_m}(\tilde{\mathbf{q}})$) can be written as a population mean restriction (see Owen, 2001):

$$E\left\{\mathbf{1}\left(p_i \leq r(\tilde{\mathbf{q}}) + \frac{(\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]}\right) - s_m\right\} = 0, \qquad m = 1, \ldots, M, \; M \geq K. \tag{A3}$$

The sample analogs are

$$\sum_{i=1}^{n} \pi_i \left[\mathbf{1}\left(p_i \leq r(\tilde{\mathbf{q}}) + \frac{(\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]}\right) - s_m\right] = 0. \tag{A4}$$

Hence, the empirical likelihood program is to maximize

$$\sum_{i=1}^{n} \log \pi_i \tag{A5}$$

with respect to the weights $\pi_i, i = 1, \ldots, n$ and the parameters $\boldsymbol{\theta}$, subject to the sample moment restrictions (A4) and the summing-up condition $\sum_{i=1}^{n} \pi_i = 1$.[29]

[29] Because we use only M moment conditions for estimation, but the indifference condition in (5) holds for all prices in the support of F_p, there are "leftover" moment conditions not used in estimation that can be employed as the basis for specification tests. We do not consider this possibility in this article.

It is well known (see Qin and Lawless, 1994) that the empirical likelihood estimate $\hat{\boldsymbol{\theta}}$ can be solved from the following saddle-point problem:

$$\max_{\boldsymbol{\theta}} \min_{t} \sum_{i=1}^{n} \log \left(1 + t' \left[1 \left(p_i \leq r(\tilde{\mathbf{q}}) + \frac{(\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]} \right) - s_m \right] \right) \equiv \max_{\boldsymbol{\theta}} \min_{t} \sum_{i=1}^{n} \log(1 + t' m(p_i; \boldsymbol{\theta})),$$

where t denotes an M-vector of Lagrange multipliers associated with the sample moment conditions (A4).

Optimizing this objective function yields the empirical likelihood estimates of the parameters. Subsequently, via equation (3), we use the estimated values of $\tilde{q}_1, \ldots, \tilde{q}_{K-1}$ to solve for $F_c(\Delta_1), \ldots, F_c(\Delta_{K-1})$, the cumulative distribution function of the search costs evaluated at the indifference points $\Delta_1, \ldots, \Delta_{K-1}$:

$$\begin{aligned} F_c(\Delta_1) &= 1 - \tilde{q}_1 \\ F_c(\Delta_2) &= 1 - \tilde{q}_1 - \tilde{q}_2 \\ F_c(\Delta_3) &= 1 - \tilde{q}_1 - \tilde{q}_2 - \tilde{q}_3 \\ &\vdots \\ F_c(\Delta_{K-1}) &= 1 - \tilde{q}_1 - \tilde{q}_2 - \ldots - \tilde{q}_{K-1}. \end{aligned} \tag{A6}$$

Asymptotic theory for MEL estimates. Using results from Qin and Lawless (1994), we obtain the limiting distribution of the above empirical likelihood estimate:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, \left(AB^{-1}A\right)^{-1}\right),$$

where the matrices in the limiting distribution are given by

$$A = \frac{\partial}{\partial \boldsymbol{\theta}} F_p \left(r(\tilde{\mathbf{q}}) + \frac{(\overline{p} - r(\tilde{\mathbf{q}}))\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k (1 - s_m)^{k-1}\right]} \right)$$

$$B = \begin{bmatrix} s_1(1 - s_1) & \ldots & \ldots & \ldots \\ \min(s_1, s_2) - s_1 s_2 & s_2(1 - s_2) & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots \\ \min(s_1, s_M) - s_1 s_M & \ldots & \ldots & s_M(1 - s_M) \end{bmatrix}.$$

As noted above, the variance-covariance matrix for the MEL estimates corresponds to the variance-covariance matrix for the GMM estimates of $\boldsymbol{\theta}$ using the M moment restrictions (A3) and the optimal weighting matrix. The A matrix can be consistently estimated using numerical derivatives. Subsequently, the estimates of $F_c(\Delta_1), \ldots, F_c(\Delta_{K-1})$, which are linear transformations of the estimated q's, can be obtained using the delta method.

☐ **Derivation of likelihood function for the sequential-search model.** Given α and r, we can estimate the τth quantile of the reservation price distribution, denoted $G_{\overline{p}}^{-1}(\tau; \alpha, r)$, using the firm indifference condition (8):

$$(\overline{p} - r)\alpha = (G^{-1}(\tau; \alpha, r) - r)(1 - \tau) \Leftrightarrow G^{-1}(\tau; \alpha, r) = \alpha \frac{(\overline{p} - r)}{(1 - \tau)} + r. \tag{A7}$$

Let $F_c^{-1}(\tau; \boldsymbol{\theta})$ denote the τth quantile of the parameterized cost distribution, where $\boldsymbol{\theta}$ denotes the parameters of this distribution that we wish to estimate. By the consumers' reservation price condition, we know that

$$F_c^{-1}(\tau; \boldsymbol{\theta}) = \int_{\underline{p}}^{G^{-1}(\tau; \alpha, r)} F_p(p) dp \tag{A8}$$

and therefore

$$(F_c^{-1})'(\tau; \boldsymbol{\theta}) = F_p\left(\alpha \frac{(\overline{p} - r)}{(1 - \tau)} + r\right) \frac{(\overline{p} - r)\alpha}{(1 - \tau)^2}.$$

In what follows, let $c(\tau; \boldsymbol{\theta})$ denote the τth quantile of $F_c(\cdot; \boldsymbol{\theta})$ (i.e., $F_c(c(\tau; \boldsymbol{\theta}); \boldsymbol{\theta}) = \tau$). Changing variables from τ to $p \equiv \alpha[(\overline{p} - r)/(1 - \tau)] + r$, we can derive the price CDF corresponding to $\boldsymbol{\theta}$, α, and r,

$$F_p(p; \boldsymbol{\theta}) = \frac{\alpha(\overline{p} - r)}{(p - r)^2 * f_c(c(1 - \alpha \frac{\overline{p} - r}{p - r}; \boldsymbol{\theta}); \boldsymbol{\theta})}.$$

with a corresponding density function $f_p(p;\boldsymbol{\theta})$ that can be derived by differentiating the above with respect to p:

$$
\begin{aligned}
f_p(p;\boldsymbol{\theta}) = \Bigg\{ & -\alpha(\bar{p}-r) + \Bigg[2(p-r)\cdot f_c\left(c\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right) \\
& + (p-r)^2 \cdot f_c'\left(c\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)\cdot c'\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right)\cdot\alpha\frac{(\bar{p}-r)}{(p-r)^2}\Bigg]\Bigg\} \\
& \div \left\{(p-r)^2\cdot f_c\left(c\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)\right\}^2 \\
= {} & \frac{2\alpha(\bar{p}-r)}{(p-r)^3\cdot f_c\left(c\left(1-\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)} - \frac{\alpha^2(\bar{p}-r)^2 f_c'\left(c\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right)\right)}{(p-r)^4\cdot\left[f_c\left(c\left(1-\alpha\frac{\bar{p}-r}{p-r};\boldsymbol{\theta}\right);\boldsymbol{\theta}\right)\right]^3}.
\end{aligned}
$$

The maximum likelihood estimates for the $\boldsymbol{\theta}$ parameters are estimated by maximizing the sample log-likelihood function $\sum_i \log f(p_i;\boldsymbol{\theta})$. The variance-covariance matrix of the estimates is approximated by the inverse of the sample analog of the outer product of the gradient vector:

$$
V = \left[\sum_i \frac{\partial \log f(p_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \log f(p_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]^{-1},
$$

where the gradient vector for each observation i is, in turn, approximated by numerical derivatives.

References

Albrecht, J. and Axell, B. "An Equilibrium Model of Search Unemployment." *Journal of Political Economy*, Vol. 92 (1984), pp. 824–840.

Bergemann, D. and Välimäki, J. "Market Diffusion with Two-Sided Learning." *RAND Journal of Economics*, Vol. 28 (1997), pp. 773–795.

——— and ———. "Monopoly Pricing of Experience Goods." Working Paper no. 1463R, Yale University, 1992.

Billingsley, P. *Probability and Measure*. New York, NY: John Wiley & Sons, 1992.

Brown, J. and Goolsbee, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy*, Vol. 110 (2002), pp. 481–507.

Brynjolfsson, E. and Smith, M. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Marketing Science*, Vol. 46 (2000), pp. 563–585.

Burdett, K. and Judd, K. "Equilibrium Price Dispersion." *Econometrica*, Vol. 51 (1983), pp. 955–969.

Chen, X., Hong, K., and Shum, M. "Likelihood Ratio Model Selection Tests between Parametric and Moment Condition Models." Mimeo, NYU, 1983.

Chevalier, J. and Goolsbee, A. "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble." *Quantitative Marketing and Economics*, Vol. 1 (2003), pp. 203–222.

Clay, K., Krishnan, R., and Wolff, E. "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry." *Journal of Industrial Economics*, Vol. 49 (2001), pp. 521–540.

Diamond, P. "A Model of Price Adjustment." *Journal of Economic Theory*, Vol. 2 (1971), pp. 156–168.

Donald, S. and Paarsch, H. "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions." *International Economic Review*, Vol. 34 (1993), pp. 121–148.

Duffie, D. *Dynamic Asset Pricing Theory*. Princeton: Princeton University Press, 1992.

Eckstein, Z. and Wolpin, K. "Estimating a Market Equilibrium Search Model from Panel Data on Individuals." *Econometrica*, Vol. 58 (1990), pp. 783–808.

Ellison, G. and Ellison, S. "Search, Obfuscation, and Price Elasticities on the Internet." Manuscript, MIT, 2001.

Goolsbee, A. "Competition in the Computer Industry: Online vs. Retail." *Journal of Industrial Economics*, Vol. 49 (2001), pp. 487–499.

Hortaçsu, A. and Syverson, C. "Search Costs, Product Differentiation, and the Welfare Effects of Entry: A Case Study of S&P 500 Index Funds." *Quarterly Journal of Economics*, Vol. 119 (2004), pp. 403–456.

Imbens, G., Spady, R., and Johnson, P. "Information Theoretic Approaches to Inference in Moment Condition Models." *Econometrica*, Vol. 66 (1998), pp. 333–357.

Jin, G. and Kato, A. "Dividing Online and Offline: A Case Study." Manuscript, Department of Economics, University of Maryland, 2002.

Kitamura, Y. and Stutzer, M. "An Information-Theoretic Alternative to Generalized Method of Moments Estimation." *Econometrica*, Vol. 65 (1997), pp. 861–874.

Lazear, E.P. *Personnel Economics*. Cambridge, Mass.: MIT Press, 1995.

Morgan, P. and Manning, R. "Optimal Search." *Econometrica*, Vol. 53 (1985), pp. 923–944.

MORTENSEN, D. "Job Search and Labor Market Analysis." In O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics, Vol. II*. Amsterdam: North-Holland, 1986.

OWEN, A. *Empirical Likelihood*. Boca Raton, Fla.: Chapman and Hall/CRC, 2001.

PRATT, J., WISE, D., AND ZECKHAUSER, R. "Price Differences in Almost Competitive Markets." *Quarterly Journal of Economics*, Vol. 93 (1979), pp. 189–211.

QIN, J. AND LAWLESS, J. "Empirical Likelihood and General Estimating Equations." *Annals of Statistics*, Vol. 22 (1994), pp. 300–325.

ROB, R. "Equilibrium Price Distributions." *Review of Economic Studies*, Vol. 52 (1985), pp. 487–504.

SCOTT-MORTON, F., ZETTELMEYER, F., AND SILVA-RISSO, J. "Internet Car Retailing." *Journal of Industrial Economics*, Vol. 49 (2001), pp. 487–500.

SORENSEN, A. "Price Dispersion and Heterogeneous Consumer Search for Retail Prescription Drugs." NBER Working Paper no. 8548, 2001.

STAHL, D. "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, Vol. 79 (1989), pp. 700–712.

STIGLER, G. "The Economics of Information." *Journal of Political Economy*, Vol. 61 (1964), pp. 213–225.

STOKEY, N.L. AND LUCAS, R.E. JR. *Recursive Methods in Economic Dynamics*. Cambridge, Mass.: Harvard University Press, 1989.

VUONG, Q. "Likelihood-Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, Vol. 57 (1989), pp. 307–333.

WOLINSKY, A. "True Monopolistic Competition as a Result of Imperfect Information." *Quarterly Journal of Economics*, Vol. 101 (1986), pp. 493–511.