Estimating Search Costs using Equilibrium Models*

Han Hong
Dept. of Economics
Princeton University

Matthew Shum
Dept. of Economics
Johns Hopkins University

Preliminary and incomplete: April 19, 2002
Please do not quote.

Abstract

In this paper, we propose methodologies for estimating consumer search costs, using just observed price data. We differ from the existing literature by exploiting the *equilibrium* restrictions of several theoretical price search models to identify and estimate the population search cost distribution: in this sense, our estimation methodology is structural. We consider both equilibrium models of *sequential* and *nonsequential* search, and discuss identification and estimation for each model. We present a short illustration using price data for a number of online electronics and book markets.

1 Introduction

Ever since the seminal paper 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, two of the most prominent of which have been equilibrium unemployment (Mortensen (1986), Albrecht and Axell (1984)) and equilibrium price dispersion in otherwise homogeneous product markets (Stiglitz (1989), Burdett and Judd (1983), Rob (1985)).

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. Recent studies by Brynjolfsson and Smith (2000), Clay, Krishnan, and Wolff (2001), and Goolsbee (2001) have attempted to understand the nature of search costs from price dispersion in homogeneous good markets by comparing the degree of price dispersion between online and traditional retail markets. Sorensen (2000a) and Sorensen (2000b) also gauge the ability of search models to explain observed price dispersion in retail

*This draft is preliminary and may contain errors. Please do not quote. We thank Matthew Langley and Wei Tan for research assistance. We thank J. Harrington, A. Goolsbee, S. Moorthy and seminar participants at Toronto, SITE, and the 2002 Winter Econometric Society Meetings in Atlanta for helpful comments. We gratefully acknowledge support from the NSF (SES-0079495, SES-0003352).

markets for prescription drugs. Finally, Scott-Morton, Zettelmeyer, and Silva-Risso (2001) consider directly the role of online middlemen in reducing transactions costs for automobiles.[1]

In this paper, we contribute to this literature by proposing new methods to estimate consumer search costs, using just observed price data (which is often, in practice, the only types of data available to researchers). We differ from the existing literature by exploiting the *equilibrium* restrictions of several theoretical price search models to identify and estimate the population search cost distribution: in this sense, our estimation methodology is structural. We consider equilibrium models of *sequential* and *nonsequential* search, which have been the two most important search technologies considered in the theoretical literature, and also suggests ways to test between them for the various online retail markets that we consider. As far as we are aware, this work constitutes the first attempt to estimate search costs in this structural manner, and test between competing search theories.

For the nonsequential search model, it turns out that equilibrium restrictions alone allow us to nonparametrically identify and estimate the population search cost distribution, without additional functional form restrictions. This is not true for the sequential search model, where we need to make additional functional form assumptions in order to identify the search cost distribution. In order to test between the two models, therefore, we suggest (non-nested) model selection tests for parametric versus moment condition models which we have developed in Chen, Hong, and Shum (2001).

Our results to date indicate that the nonsequential search models perform better than the sequential search models, in most of the markets that we consider. For this preferred specification, our search cost estimates indicate that median search costs (ie. the cost that a consumer pays to obtain one additional price sample) to be generally less than 10% of the median retail price: for the Palm Pilot *IIIxe*, for instance, we estimate median search costs to be about 5.9% of the median retail price, or around $11.00. The rather large absolute values for these search costs indicate that factors not accommodated in the models we consider, such as retailer reputation or other forms of differentiation among retailers, may also be important in explaining equilibrium price dispersion.

In the next section, we describe our proposed nonparametric estimation procedure for nonsequential search models. Section 3 contains details on our parametric estimation procedure for sequential search models. In Section 4, we present estimation results using online price in several consumer electronics and book categories. In Section 5, we describe and implement the model selection tests to decide between the sequential and nonsequential search models, for the product markets we consider. Section 6 concludes.

At this point we wish to point out that the goal of this paper is to present a methodology for estimating search costs by exploiting the equilibrium restrictions from search models often employed

[1] In related work, Ellison and Ellison (2001) focuses on strategic "obfuscation" policies which internet retailers may engage in to avoid destructive price competition, when consumers obtain prices from search engines. In this paper we abstract away from issues such as product availability, which (as pointed out by Ellison and Ellison) may be an avenue of obfuscation by online retailers. We focus instead on the nonparametric identification of equilibrium search models, from data on prices alone.

in theoretical work. The markets which fit the assumptions of these models are those where a large number (i.e., a continuum) of small firms offer a homogeneous product to consumers who demand only a single unit of the good. While we have attempted to choose product categories which we believe to conform to these ideal markets, clearly there are features of real world markets which we have not accommodated. This caveat should be kept in mind when interpreting the results.

Existing literature To our knowledge there has been no previous attempt to estimate search costs in product markets using only observed price data by exploiting the equilibrium restrictions of theoretical search models. As such, our work bears similarities to three existing applied research areas where equilibrium modeling also plays an important role. First is the equilibrium job search literature in applied labor economics, such as the seminal papers by Eckstein and Wolpin (1990), Ridder and van den Berg (1998), and Bontemps, Robin, and van den Berg (1999).

Second, a long literature on differentiated-product oligopolies in empirical industrial organization (cf. Bresnahan (1981), Berry (1994), and Berry, Levinsohn, and Pakes (1995)) has demonstrated that one can obtain estimates of firms' unobserved marginal costs using data just on prices and market shares, by exploiting oligopoly equilibrium restrictions. Finally, the structural empirical auction literature (cf. Paarsch (1992), Laffont, Ossard, and Vuong (1995), Guerre, Perrigne, and Vuong (2000)) has demonstrated how one can estimate the population distributions of bidders' private valuations from observed bid data, using only the first-order conditions defining bidders' equilibrium strategies as the estimating equations. These two literatures have repeatedly emphasized the potential of equilibrium restrictions in identifying and estimating equilibrium models, and we take a similar approach throughout this paper.

In this paper, we take consumers' search technologies as given, and do not consider the optimality of these search strategies. For a discussion of these issues, see the treatment in Morgan and Manning (1985). Generally speaking, nonsequential (of "fixed sample size") search strategies can be optimal when there is a fixed-cost component to engaging in 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.

2 Nonsequential Search

In this section, we consider a nonsequential search model, which is an extension of the model in Burdett and Judd (1983) (sect. 3.2) to heterogeneous consumer search costs. Consumers who search nonsequentially are those who commit to buying from the lowest-priced store after obtaining a random sample of n (≥ 1) prices. Following Burdett and Judd (1983), we assume that there are a continuum of firms and consumers, and interpret the equilibrium price distribution F_p as the symmetric equilibrium mixed strategy employed by all firms. We assume that consumers have inelastic demand for a single unit of the good. Consumers incur a search cost c to receive a single price quote, and we assume that search costs are *i.i.d.* draws across consumers from a distribution

F_c.

Finally, let $\underline{p}$ and p^* denote, respectively, the lower and upper bound of the support of F_p, and r denote the common unit production cost of each firm. Since all firms produce homogeneous products, only search frictions (arising from consumers' imperfect information about stores' prices) generate price dispersion in this market.

A consumer with search cost c chooses the number of stores n to canvass to minimize her total expected cost, which is the sum of her search costs as well as the price she expects to pay for the product:

$$n^*(c) \equiv \text{argmin}_{n>1}\mathcal{C}(n;c) \equiv c*(n-1) + \int_{\underline{p}}^{p^*} np\left(1-F(p)\right)^{n-1} f(p)dp. \tag{1}$$

We need the following proposition.

Proposition 1 *$n^*(c)$ is monotonically decreasing in c.*

Proof: We show that the objective $\mathcal{C}(n;c)$ has increasing differences in $(n;c)$. $\forall c' > c$, the difference at n is $\Delta\mathcal{C}(n;c) \equiv (c'-c)(n-1) + \int_{\underline{p}}^{p^*} np\left(1-F(p)\right)^{n-1} f(p)dp$. Now the difference $Ep_{(1:n)} - Ep_{(1:n+1)} = \int pf(p)\left(1-F(p)\right)^{n-1}(-F(p))\,dp$ is nonincreasing in n, which implies that the sequence $\Delta\mathcal{C}(n+1;c) - \Delta\mathcal{C}(n;c) = (c'-c) - (Ep_{1:n} - Ep_{1:n+1})$ is nondecreasing in n. Hence, $\mathcal{C}(n;c)$ has increasing differences. Then the desired monotonicity obtains via a standard result (cf. Vives (1999), Theorem 2.3). ■

2.1 Nonparametric identification

Given Proposition 1, we can recover a nonparametric estimate of the population search cost distribution F_c just from the empirical distribution of prices, $\hat{F}_p$. Next we discuss the intuition behind the nonparametric identification strategy for this model. Since consumers are assumed to draws *i.i.d.* samples from the equilibrium price distribution F_p, the marginal expected savings from searching $i+1$ versus i stores is just

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

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

Since the sequence of marginal expected savings $c_i,\ i = 1, 2, \ldots$ is nonincreasing in i for any price distribution F_p,[2] while the cost per search is constant, a consumer with search cost c will determine $n^*(c)$ as $\max_i c_i$ such that $c_i > c$: search as long as the marginal expected savings c_i exceeds the marginal search cost c.

[2] Note that the sequence $c_1, c_2, \ldots$ is nonincreasing for every price distribution $F(p)$. This is because the difference $c_k \equiv Ep_{(1:k)} - Ep_{(1:k+1)} = \int pf(p)\left(1-F(p)\right)^{k-1}(-F(p))\,dp$ is nonincreasing in k.

Therefore, the sequence of marginal expected savings $c_1, c_2, \ldots$ can also be interpreted as the search costs of the "indifferent" consumers: c_i is the search cost faced by the consumer who is indifferent between searching $i+1$ and i stores. See figure 1. The area of the regions A,B,C, and D are the measure of consumers who search at (respectively) one, two, three, or four stores. c_1, c_2, c_3, and c_4 are the search costs of the indifferent consumers, where a consumer with search costs c_k is indifferent between searching k and $k+1$ stores.

Furthermore, we note that we can obtain estimates of these indifference points from the empirical price distribution $\hat{F}_p$ as follows, via the relation (2). Now define:

$$\begin{aligned} \tilde{q}_1 &\equiv 1 - F_c(c_1) \text{ :the proportion of consumers who search only one store} \\ \tilde{q}_2 &\equiv F_c(c_1) - F_c(c_2) \text{ :the proportion of consumers who search two stores} \\ \tilde{q}_3 &\equiv F_c(c_2) - F_c(c_3) \text{ :the proportion of consumers who search three stores, etc.} \end{aligned} \tag{3}$$

Note that if we could estimate the sequence $\tilde{q}_1, \tilde{q}_2, \ldots$, then we would immediately be able to recover an estimate of the search cost distribution F_c, by solving the equations (3) above for the values of the cumulative distribution F_c corresponding to the indifferent consumers $c_1, c_2, \ldots$.

It turns out that 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 $F_p(\cdot)$ is

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

for all $p \in [\underline{p}, p^*]$. Following Burdett and Judd (1983), the characterization of the equilibrium price dispersion starts with the mixed strategy condition that firms be indifferent between charging the monopoly price p^* (and selling only to people who canvass one store) and any other price p in the interior of the equilibrium support $[\underline{p}, p^*]$:

$$(p^* - r)\,\tilde{q}_1 = (p - r)\left[\sum_{k=1}^{\infty} \tilde{q}_k k \left(1 - F(p)\right)^{k-1}\right]. \tag{4}$$

Now we show that, just as in the empirical auction literature cited above, the optimality equation (4) allows us to recover a nonparametric estimate of the search cost distribution F_c from the empirical price distribution $\hat{F}_p$ alone. Let $\underline{\hat{p}}$ and $\hat{p}^*$ denote the lowest and highest observed prices, respectively. Next, pick K prices, $p_1, \ldots, p_K \in [\underline{\hat{p}}, \hat{p}^*)$, where $p_1 = \underline{\hat{p}}$. The indifference condition (equation 4) for each of these prices is:

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

This constitutes K equations from which we can solve for the K unknowns $\{r, \tilde{q}_1, \ldots, \tilde{q}_{K-1}\}$.

After deriving estimates of $\tilde{q}_i,\ i = 1, \dots, K$ and r, we can obtain an estimate of $F(p)$ as follows.[3] Let $\phi(s) \equiv F^{-1}(s)$ denote the inverse CDF for the equilibrium prices. Then, for each $s \in [0, 1]$, we rewrite the firm indifference condition (4) as

$$\begin{aligned} (p^* - r)\,\mu\tilde{q}_1 &= (\phi(s) - r)\,\mu\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s)^{k-1}\right] \Longrightarrow \\ \phi(s) &= r + \frac{(p^* - r)\,\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s)^{k-1}\right]}. \end{aligned} \tag{6}$$

Clearly, given estimates of $\tilde{q}_i,\ i = 1, \dots, K$ and r, we can compute (6) at all $s \in [0, 1]$ to derive an estimate of the (inverse) equilibrium price distribution.

Subsequently, through equation (3), we use the values of $\tilde{q}_1, \dots, \tilde{q}_{K-1}$ to solve for $F_c(c_1), \dots, F_c(c_{K-1})$, the cumulative distribution function of the search costs evaluated at the indifference points $c_1, \dots, c_{K-1}$. Asymptotically, as the number of price observations grows large, we can recover an arbitrarily accurate estimate of F_c by taking larger and larger values for K.

In order to obtain standard errors for the estimates, we formulate the estimation problem as an empirical likelihood problem with estimating equations given by (4). As noted by Imbens, Spady, and Johnson (1998) and Kitamura and Stutzer (1997), empirical likelihood 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 identical to that for the efficient GMM estimates (Owen (2001) provides more details). Complete details on the maximum empirical likelihood procedure is given in the Appendix.

A first example: Palm Pilots In Figure 2 we plot the histogram and empirical CDF of 31 prices for the Palm Pilot *IIIxe* at different online retailers, as retrieved by the MySimon (`www.mysimon.com`) search engine on May 11, 2001. The absence of a prominent tail at the lower end indicates that the searching strategies may not lead to enough "lucky" consumers (i.e., those who are lucky enough to find one of the relative small number of stores charging relatively low prices) to make the strategy of charging very low prices profitable. This suggests that, in our nonsequential search model with heterogeneous search costs, we should not find very many consumers with "low" search costs.

We illustrate the estimation procedure for nonsequential search models using the Palm Pilot data. The indifferent search costs values were computed directly from equations (2), using the observed empirical distribution of prices. They were:

---

[3] We are grateful to Geert Ridder for this suggestion.

c_1	\$10.06
c_2	\$3.34
c_3	\$1.83
c_4	\$1.22
c_5	\$0.96
c_6	\$0.74
c_7	\$0.59
c_8	\$0.45
c_9	\$0.42
c_{10}	\$0.36

In Figure 3, we graph the estimated population search cost distribution, for different values of K. Note that all four graphs are similar in shape, and all indicate that the search cost distribution has a thick upper tail, with about 30% of consumers having search costs exceeding ten dollars. This implies that there will be a large enough group of consumers who search only one store, ensuring a nondegenerate equilibrium price distribution.

The estimated production costs corresponding to these models are:

K	Prod cost estimate r (\$)
3	151.71
5	157.48
8	159.99
10	160.18
15	162.56

Note that the production cost estimates are relatively stable, for different values of K. However, they are quite high, implying very small margins for most internet retailers. On the other hand, these magnitudes could be reasonable, since a vertical market with relatively intense retailer competition would generate small retail margins (and thereby allow the upstream monopolist to obtain the full monopoly profits).[4]

Given these of the search cost distribution and the production costs, we were able to recover the equilibrium price distribution $F(p)$ by inverting the firm indifference condition, as in equation (6) above. The estimated equilibrium price distribution, as well as the observed empirical price distribution, are both graphed in Figure 4.

We note that the estimated equilibrium price distribution remains very stable, for different values of K. Furthermore, they are similar in shape to the empirical price distribution, which is reported in Table 2.

In summary, then, all the results appear reasonable, except for the thin margins. To what extent

[4] However, rather than implying high production costs, these cost estimates may simply reflect the effect of an implicit "price floor". Although explicit resale price maintenance is illegal, it is well-known that Palm (3Com) chooses its resellers with quite stringent conditions.

are the results due to the assumption of non-sequential search by consumers? Later in this paper, we turn to sequential search models.

3 Sequential Search

The sequential search technology differs from the non-sequential 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.

This case is more complicated than the non-sequential search framework, but is the standard model in job search models in labor economics (cf. the survey in Mortensen (1986)). A standard result in the literature is that the consumers' problem is analogous to an "optimal stopping" problem, and that the stationary optimal search strategy takes the form of a reservation price ($\bar{p}$ in the Burdett and Judd (1983) paper) below which consumers will purchase with certainty.

The starting point of all sequential search models is the negative result in Diamond (1971) (pg. 161, lemma 2) that the equilibrium price distribution in a sequential search model with positive search costs is degenerate at the monopoly price (i.e., the reservation price of consumers). Papers by Burdett and Judd (1983), Albrecht and Axell (1984), and Stahl (1989) have tried to formulate sequential search models in which you get a nondegenerate equilibrium price distribution.

Burdett and Judd (1983) get around the Diamond result by positing a "noisy" search technology, in which the number of prices that you get during each step of the sequential search is a random variable, not necessarily 1 with probability one (as in the Diamond model). Specifically, it is a random variable with support on the positive integers: q_k denotes the probability that the search yields k prices, and $\sum_{k=1}^{\infty} q_k = 1$.

Note that Theorem 4 of their paper, part (c) states that in order to get a nondegenerate equilibrium price distribution, you need q_1 strictly between 0 and 1: so "noisy" sequential search is a necessary condition for a nondegenerate price equilibrium.[5]

However, the noisy search assumption does not seem appropriate for empirical applications; therefore, in our empirical sequential search model, we avoid both the Diamond result as well as the assumption of noisy search by assuming heterogeneity in consumer search costs, which is very reasonable in most real-world markets. Intuitively, in sequential search models, the existence of a non-degenerate equilibrium price distribution requires heterogeneity in consumers' reservation prices, so that low price firms "cater" to consumers with low search costs (and therefore low reservation prices), and high price firms cater to high search cost consumers. Without consumer heterogeneity

[5] Stahl (1989) gets around the Diamond result by assuming two types of consumers: the *shoppers* with zero search costs (who essentially know the prices at all the stores) and the *searchers* with strictly positive search costs, who engage in (non-noisy) sequential search. Furthermore, he considers an oligopolistic situation with N firms, and an aggregate demand function $D(p)$.

in reservation prices, it cannot be a optimal response for firms to charge any price other than the reservation price, so that the profit-maximizing price will be the monopoly price.

Caveat: Our analysis and choice of parametric search cost distributions do not yet accommodate the conditions in Rob (1985) for the existence of non-degenerate price equilibria.

Hence we adapt the sequential search model in Burdett and Judd (1983) (sect. 3.3) to accommodate heterogeneity in consumer search costs. We will focus on the case where $q_1 = 1$ (i.e., search is no longer "noisy", but rather one search yields one price with probability 1). Let F_c denote the population distribution of search costs, and let F_p denote the equilibrium distribution of prices, with support $[\underline{p}, p^*]$.

We start with the characterization of consumers' reservation prices: note that the heterogeneity in search costs leads to heterogeneity in reservation prices. For consumer i, who has search costs c_i, let $z^*(c_i)$ denote the price z which satisfies the indifference condition

$$c_i = \int_0^z (z - p) F_p(dp) = \int_0^z F_p(p) dp \tag{7}$$

where the second equality follows from integration by parts. Note that $z^*(c)$ is increasing in c.

Now, for each cost c_i, the reservation price

$$\bar{p}_i \equiv \bar{p}(c_i) \equiv \min\left(z^*(c_i), p^*\right). \tag{8}$$

Let $G_{\bar{p}}$ denote the distribution of reservation prices in the population, given F_c and the mapping (8). Note that there is a mass $1 - G_{\bar{p}}(p^*)$ of consumers for whom the reservation price is p^*.

Next we characterize the firms' decision problems. As before, we use an indifference condition to define the equilibrium price distribution. We need to first derive an expression for an individual firm's equilibrium demand, at a price $\hat{p} \in [\underline{p}, p^*]$. In doing this, we follow the derivation in Eckstein and Wolpin (1990) (top, pg. 789) for a discrete distribution of consumer types. Consider consumer i, who has reservation price $\bar{p}_i$. The probability that this consumer encounters a firm charging a price $\hat{p} < \bar{p}_i$ (which she accepts) is

$$\begin{aligned} & F(\bar{p}_i) + (1 - F(\bar{p}_i))F(\bar{p}_i) + (1 - F(\bar{p}_i))^2 F(\bar{p}_i) + \cdots \\ = & \sum_{n=1}^{\infty} (1 - F(\bar{p}_i))^{n-1} F(\bar{p}_i) \\ = & \frac{1}{F(\bar{p})} F(\bar{p}) = 1. \end{aligned} \tag{9}$$

Now the firm charging $\hat{p}$ will only sell to consumers i for whom $\hat{p} < \bar{p}_i$, and (by assumption) each firm services a proportion μ of the "leftover" consumers each period. Therefore, the demand at price

$\hat{p}$ is:

$$\begin{aligned} D(\hat{p}) &= \int_{\hat{p}}^{\infty} \mu dG_{\bar{p}}(\omega) \\ &= \mu * (G_{\bar{p}}(p^*) - G_{\bar{p}}(\hat{p})) + \mu\,(1 - G_{\bar{p}}(p^*)) \\ &= \mu * (1 - G_{\bar{p}}(\hat{p}))\,. \end{aligned} \tag{10}$$

Note that $D(p^*) = \mu\,(1 - G_{\bar{p}}(p^*))$. Therefore the firms' indifference condition is:

$$\begin{aligned} (p^* - r)D(p^*) = (p - r)D(p) \Longleftrightarrow \\ (p^* - r) * (1 - G_{\bar{p}}(p^*)) = (p - r) * (1 - G_{\bar{p}}(p)) \end{aligned} \tag{11}$$

for each $p \in [\underline{p}, p^*)$.

Identification The equilibrium sequential model with heterogeneous search cost is fundamentally unidentified.

This is because the equations (11) define $n-1$ equations, for each of the $n-1$ observed prices (with p^* excluded).[6] However, there are $n+1$ unknowns: r, and $G_{\bar{p}}(p_1 = \underline{p}), G_{\bar{p}}(p_2), \ldots, G_{\bar{p}}(p_{n-1}), G_{\bar{p}}(p_n = p^*)$. Therefore the model is underidentified without some additional assumptions. In the next section, we describe a parametric (MLE) approach to estimating this model, which we employ to obtain our estimates.

3.0.1 Parametric Estimation Method

As we discussed above, $G_{\bar{p}}$ and r are not separately identified from observed price data without more assumptions. Here we consider parametric estimation of this model, assuming that the search cost distribution is parameterized $F_c(\cdot;\theta)$. Furthermore, we let $\alpha \equiv (1 - G_{\bar{p}}(p^*))$ be another parameter. Next we derive the likelihood for this model.

Given an estimate of α and r, we can estimate the τ-th quantile of the reservation price distribution, denoted $G_{\bar{p}}^{-1}(\tau;\alpha,r)$, using the firm indifference condition (11):

$$\begin{aligned} (p^* - r)\alpha = \left(G_{\bar{p}}^{-1}(\tau;\alpha,r) - r\right)(1-\tau) \Leftrightarrow \\ G_{\bar{p}}^{-1}(\tau;\alpha,r) = \alpha\frac{(p^* - r)}{(1-\tau)} + r. \end{aligned} \tag{12}$$

Let $F_c^{-1}(\tau;\theta)$ denote the τ-th quantile of the parameterized cost distribution. By the consumers' reservation price condition, we know that

$$F_c^{-1}(\tau;\theta) = \int_{\underline{p}}^{G_{\bar{p}}^{-1}(\tau;\alpha,r)} F_p(p)dp \tag{13}$$

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

and therefore

$$(F_c^{-1})'(\tau;\theta) = F_p\left(\alpha\frac{(p^*-r)}{(1-\tau)}+r\right)\frac{(p^*-r)\alpha}{(1-\tau)^2} \Leftrightarrow$$
$$\frac{1}{f_c\left(c(\tau;\theta);\theta\right)} = F_p\left(\alpha\frac{(p^*-r)}{(1-\tau)}+r\right)\frac{(p^*-r)\alpha}{(1-\tau)^2}$$

where $c(\tau;\theta)$ denotes the τ-th quantile of $F_c(\cdot;\theta)$.

Next, changing variables from τ to $t \equiv \alpha\frac{(p^*-r)}{(1-\tau)} + r$, we can derive the price CDF corresponding to θ, α, and r:

$$F_p(t;\theta,\alpha,r) = \frac{(t-r)}{\alpha\,(p^*-r) * f_c\left(c\left(1-\alpha\frac{p^*-r}{t-r};\theta\right);\theta\right)} \tag{14}$$

with a corresponding density function $f_p(t;\theta,\alpha,r)$ which can be derived by differentiating the above with respect to t:

$$f_p\left(t;\theta,\alpha,r\right) = \frac{\alpha*(p^*-r)*f_c\left(c\left(1-\frac{p^*-r}{t-r};\theta\right);\theta\right) - \alpha^2\frac{(p^*-r)^2}{t-r}*f_c'\left(c\left(1-\alpha\frac{p^*-r}{t-r};\theta\right);\theta\right)*c'\left(1-\alpha\frac{p^*-r}{t-r};\theta\right)}{\left\{\alpha^2\,(p^*-r)*f_c\left(c\left(1-\alpha\frac{p^*-r}{t-r};\theta\right);\theta\right)\right\}^2}$$
$$= \frac{1}{\alpha\,(p^*-r)*f_c\left(c\left(1-\frac{p^*-r}{t-r};\theta\right);\theta\right)} - \frac{f_c'\left(c\left(1-\alpha\frac{p^*-r}{t-r};\theta\right)\right)}{(t-r)*\left[f_c\left(c\left(1-\alpha\frac{p^*-r}{t-r};\theta\right);\theta\right)\right]^3} \tag{15}$$

Now given the observed p^*, the α corresponding to θ, r, denoted $\alpha(\theta,r;p^*)$, is defined by the implicit equation

$$\alpha * f_c\left(c(1-\alpha);\theta\right) = 1 \tag{16}$$

which is just equation (14) evaluated at p^*.

The likelihood function, then, is just

$$\log L(\theta,r) = \sum_{i=1}^{n}\log f_p(p_i;\theta,r,\alpha(\theta,r;p^*)).$$

Issues Complicating matters is the possibility that we need to impose some very nonlinear restrictions of the parameters θ and r to ensure that all the observed prices have positive likelihood. These are:

- $\alpha(\theta,r;p^*)$ solving equation (16) lies within $[0,1]$.
- Furthermore, for every $p \in [\underline{p},p^*]$, the quantile $F_c\left(1-\alpha(\theta,r;p^*)\frac{p^*-r}{p-r};\theta\right) \in [0,1]$.

- Similarly, we need to ensure that the price density in equation (15) is positive.

For a fixed (θ, r), these conditions also imply that the support of the price distribution on the particular parameter values, which raises difficulties for maximum likelihood estimation.

4 Estimation results

Summary statistics for the prices used in the estimation are given in Table 6. For these preliminary results, we use prices which *do not include shipping and handling fees.*

There are noticeable qualitative differences between the books and electronics prices. First, we have fewer price observations for books, and there is also more clumping in book prices (i.e., fewer distinct prices). Furthermore, the median price is very close to the list price and, except for the Stokey–Lucas text, the upper bound p^* is very close to the list price (so that there is no upper tail). Second, the low price dispersion observed for books implies that there may be little benefit to searching. We expect therefore to estimate high value of $\tilde{q}_1$ for books. Generally, we observe more price dispersion for the electronic goods. This implies that there are benefits to searching, so that we might expect $\tilde{q}_1$ to be small.

Tables 7 and 8 contains estimates for Model A, the nonsequential search model with heterogeneous search costs. Finally, Table 9 presents the maximum likelihood estimates for Model B, using three parametric assumptions for the search cost distribution: exponential, gamma, and Weibull. The exponential distribution is nested by both the gamma and Weibull distributions, but the gamma and Weibull distributions are not nested. Table 11 summarizes the differences in the magnitudes of the estimated search costs, across the two models.

Magnitudes of search costs Generally, the magnitudes of the estimated search costs from sequential search models are higher than those obtained from the nonsequential search models. How reasonable are the magnitudes of search costs? To facilitate the interpretation of the search cost as an opportunity cost, we present some data on the average hourly wages in different professions (taken from the *1998 National Compensation Survey* collected by the Bureau of Labor Statistics).

Profession	Average hourly wage
All	15.72
White Collar	19.39
Engineers, architects, surveyors	28.97
Physicians	38.55
Economics teachers	46.28
Lawyers	36.30
Blue Collar	12.90

See Tables 1, 2, and 3 for imputed minutes per search, using the interpretation of search costs as opportunity costs of labor.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

4.1 Online vs. traditional retail markets: A bit of evidence

We also conducted a small phone survey of electronics retailers in the Baltimore metropolitan area, to canvass their Palm Pilot prices (for the models considered in this paper). Somewhat to our surprise, we found practically *no* price dispersion across the prices for each Palm Pilot model.

In the equilibrium search models (both nonsequential as well as sequential), zero price dispersion can arise in equilibrium for two reasons: (1) search costs are zero, and prices represent the Bertrand (zero profit) equilibrium; (2) search costs are prohibitively high, and the observed prices represent the equilibrium in which all firms charge the high monopoly price. Furthermore, a third possibility not considered by these models is collusion. Since all of these scenarios lead to observationally equivalent pricing outcomes, there is no way to test between them without additional data on costs.

5 Additional Results and Implications

In this section, we consider additional estimates of Models A and B for a series of prices collected on February 5, 2002 from the `pricescan.com` website, for a subset of the same products as before. (There were very fewer price quotes available for the Palm III*xe* and the Sony MP3 player, so these two products were omitted.) For these products, both the prices with and without shipping and handling charges was collected, and the models estimated separately for each price series, in order to gauge the importance of shipping and handling charges.

A summary of the median search cost estimates for these data is given in Table 4. The magnitudes of the estimated median search costs is roughly similar to (but slightly higher than) the estimates considered in the previous section, for price data from about six months earlier. Furthermore, we see that including the shipping charges does not affect the magnitude of the median search costs appreciably, in most markets. This may not be surprising for the electronics markets, where the shipping charges account for a very small portion of the total price, but may not have been expected for the books, where the shipping charges constitute a larger chunk of the final price (especially for an inexpensive paperback such as the *English Patient*).

Most importantly, we see that, just as before, the median search cost estimates in the sequential model are uniformly higher than the estimates for the nonsequential models. The earlier findings, then, appear rather robust in this sense.

[Table 4 about here.]

While this paper is primarily *not* a paper about search costs in online markets, the estimated search costs from Model A (especially for the book markets) are not implausible (and, indeed, may be very plausible in traditional retail markets). Stronger evidence against the search cost explanation for the observed price dispersion is given by the implications of the results for each firm's market share. Recalling that, in the nonsequential search model, we use q_i to denote the population proportion of consumers who choose to obtain an i-sized sample before buying, we can write the demand for a firm charging a price $\tilde{p}$ in equilibrium as:

$$\begin{aligned} D(\tilde{p}) \equiv & \mu \left[q_1 + q_2 * \text{Prob}\left(\tilde{p} \leq p_{1:1}\right) + q_3 * \text{Prob}\left(\tilde{p} \leq p_{1:2}\right) \right. \\ & \left. + \cdots + q_i * \text{Prob}\left(\tilde{p} \leq p_{1:i-1}\right) + \cdots \right] \end{aligned}$$

where $p_{1:i}$ denotes the random variable which is the minimum of i draws from the equilibrium price distribution F_p, and μ is a constant describing the measure of consumers at each firm (cf. Burdett and Judd (1983), pg. 957). For each firm, we can estimate the implied market share as its demand divided by the total implied market demand.[7]

[Table 5 about here.]

6 Testing between the Sequential and Non-sequential Search Models [INCOMPLETE]

In this final section, we consider the possibility of employing model selection tests in order to gauge whether the nonsequential search model (specification B) outperforms the sequential search model (specification D) in generating the observed price data. While both models were estimated using likelihood based methods, the maximized likelihood functions between the two models are not directly comparable, since the EL procedure for Model B assumes a discrete (multinomial) density for the prices, but the parametric ML procedure used for Model D assume a contionuous price density function.

In ongoing work (Chen, Hong, and Shum (2001)), we are extending the testing principle in Vuong (1989), which is based on the likelihood ratio for two (not necessarily nested) competing parametric models, to the present case, in which one of the alternative models is parametric, and the other is a semiparametric model where the data generating process is only partially specified via population moment restrictions. In this paper, however, we do not perform a formal model selection exercise. Rather, we undertake a more informal model specification exercise by deriving a continuous price density function for Model B, which we use to generate a likelihood function for the observed prices which we compare to the log-likelihood function values generated from Model D (and reported in the right-most column of Table 9).

We start by considering the general estimation problem for Model B, which is to derive an estimate

[7] By construction of the mixed-strategy equilibrium, the share of industry profits for each firm is identically $\frac{1}{n}$.

of the price density $f_p(\cdot)$ which satisfies the population moment conditions

$$\int m(p;\gamma_0)f_p(p)dp = 0.$$

One can derive an estimate $\hat{f}_{p,n}(p)$ of $f_p(p)$ using the general maximum likelihood principle, extended to infinite-dimensional (functional) parameter spaces (cf. Grenander (1981), chap. 8). For a fixed vector θ_n, one can choose the function $f_{p,n}(x;\theta_n)$ (from some class of functions) which maximizes the sample log-likelihood subject to the moment conditions (evaluated at θ_n); in other words:

$$f_{p,n}(p;\theta_n) \equiv \text{argmax}_{\bar{f}_p(\cdot)} \sum_{i=1}^{n} \log \bar{f}_p(p_i) \text{ s.t. } \int m(p;\theta_n)\bar{f}_p(p)dp = 0.$$

Now the maximum likelihood estimate $\hat{f}_{p,n}(x)$ is derived via the maximization

$$\hat{f}_{p,n}(p) \equiv \text{argmax}_{\theta_n} \sum_{i=1}^{n} f_{p,n}(p_i;\theta_n).$$

Clearly, the maximum likelihood estimate for $f_p(p)$ is difficult to derive in practice for arbitrary function spaces for the density function. However, EL provides a simple solution to this maximization problem for the case when $f_{p,n}(p)$ is restricted to be the family of n-point multinomial distributions.[8] In fact (as derived above),

$$f_{p,n}(p) = \begin{cases} \frac{1}{n}\frac{1}{1+\hat{\tau}_n' * m(p;\hat{\theta}_n)} & p \in \{p_1,\ldots,p_n\} \\ 0 & \text{otherwise} \end{cases}$$

where $\left(\hat{\tau},\hat{\theta}_n\right)$ are the solutions of the following saddle-point problem:

$$\max_{\gamma}\min_{\tau}\sum_{i=1}^{n}\log\left(1+\hat{\tau}_n' * m(p_i;\hat{\theta}_n)\right).$$

Therefore, we estimate the continuous price density $\tilde{f}_{p,n}(p)$ for Model B by taking the derivative of a smoothed version of the CDF from the EL estimates:

$$\tilde{f}_{p,n}(p) \equiv \frac{d\tilde{F}^{EL}_{p,n}(p)}{dp}$$

where

$$\tilde{F}^{EL}_{p,n}(p) = \sum_{i=1}^{n} f_{p,n}(p_i)\tilde{\mathbf{1}}\left(p_i \leq p\right)$$

[8] Note that the EL solution also applies when considering the family of multinomial distributions with countable (but possibly infinite) points of supports; for a dataset of $p_1,\ldots,p_n$, EL would never place positive probability on any x's besides the observed ones.

and $\tilde{\mathbf{1}}(\cdots)$ denotes some smoothed (and differentiable in p) indicator function. In practice, we use the standard normal CDF as a smoother:

$$\tilde{\mathbf{1}}(p_i \leq p) = \Phi\left(\frac{p - p_i}{h}\right)$$

where h is a small smoothing parameter (in the calculations below, we used $h = 0.01$).

For more than half of the products considered, the smoothed log-EL for the nonsequential search Model B exceeds the likelihoods for any of the parametric specifications of Model D. For example, the smoothed maximum log-EL for the Palm Pilot Vx is 72.7, whereas the maximized parametric log-likelihood for the Weibull specification is 58.5.

The exceptions are the Palm Pilot Vx, the Sony Mavica and NW-MS9 MP3 player, and the Lazear and Billingsley books. Perhaps it is not surprising that the non-sequential search model performs marginally better, across the 12 product categories we consider, because we would expect that a nonsequential search strategy is more appropriate in markets where search costs are relatively low, and *a priori* one would consider these markets to be low cost markets, especially in relation to (say) job markets, in which high search costs probably render a sequential search strategy more likely.

7 Conclusions

In this paper, we proposed new methods to estimate consumer search costs, using just observed price data. In identifying and estimating the population search cost distribution, we exploit the *equilibrium* restrictions of several theoretical price search models: in this sense, our estimation methodology is structural, and is related to much of the work on estimating differentiated product oligopoly and auction models. We considered equilibrium models of *sequential* and *nonsequential* search, which have been the two most important search technologies considered in the theoretical literature, and also test between them for a number of online retail markets. As far as we are aware, this work constitutes the first attempt to estimate search costs, and test between competing search paradigms.

We have focused on the case where equilibrium restrictions help us to identify search costs in the absence of any data besides observations of equilibrium prices. A challenging extension would be to combine the equilibrium search models considered in this paper with rich individual-level datasets (such as supermarket scanner panel datasets or the drug purchase dataset used in Sorensen (2000b)). It would be interesting to investigate both the usefulness as well as the ways to exploit the equilibrium restrictions of the theoretical search models in identifying consumer search costs in a much more data-rich environment.

References

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

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Bontemps, C., J.-M. Robin, and G. van den Berg (1999): "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms," *International Economic Review*, 40, 1039–1074.

Bresnahan, T. (1981): "Departures from Marginal Cost Pricing in the American Automobile Industry," *Journal of Econometrics*, 17, 201–227.

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

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

Chen, X., H. Hong, and M. Shum (2001): "Testing between Parametric and Moment Condition Models," in progress.

Clay, K., R. Krishnan, and E. Wolff (2001): "Pricing Strategies on the Web: Evidence from the Online Book Industry," mimeo., Carnegie-Mellon University.

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

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

Ellison, G., and S. Ellison (2001): "Search, Obfuscation, and Price Elasticities on the Internet," mimeo, MIT.

Goolsbee, A. (2001): "Competition in the Computer Industry: Online vs. Retail," *Journal of Industrial Economics*, 49, 463–486.

Grenander, U. (1981): *Abstract Inference*. John Wiley and Sons.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–74.

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

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

Laffont, J. J., H. Ossard, and Q. Vuong (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63, 953–980.

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

MORTENSEN, D. (1986): "Job Search and Labor Market Analysis," in *Handbook of Labor Economics, Vol. II.*, ed. by O. Ashenfelter, and R. Layard, pp. 849–919. North-Holland.

OWEN, A. (2001): *Empirical Likelihood.* Chapman and Hall/CRC.

PAARSCH, H. (1992): "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191–215.

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

RIDDER, G., AND G. VAN DEN BERG (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66, 1183–1221.

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

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

SORENSEN, A. (2000a): "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, 108, 833–850.

——— (2000b): "Price Dispersion and Heterogeneous Consumer Search for Retail Prescription Drugs," mimeo., University of California, San Diego.

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

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

STIGLITZ, J. (1989): "Imperfect Information in the Product Market," in *Handbook of Industrial Organization, Vol. 1*, ed. by R. Schmalensee, and R. Willig. North Holland.

VIVES, X. (1999): *Oligopoly Pricing.* MIT Press.

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

A Maximum Empirical Likelihood (MEL) estimation procedure for non-sequential search model

Direct maximum likelihood estimation of the model with heterogeneous search costs is no longer feasible, since the equilibrium price distribution is implicitly defined by the equation (4), and cannot be expressed in closed form. In this section we sketch an 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.

Note that, potentially, the firms' equilibrium indifference condition (4) imposes an infinite number of moment conditions, as $K \to \infty$. However, in estimation, we will only use a finite number $I < \infty$ of these conditions. This naturally suggests that there are always restrictions "left over" which can be used as the basis for specification tests.

Our data consist of n prices, p_i, $i = 1, \dots, n$. Consider a discrete price distribution with n points of support, at p_i, $i = 1, \dots, n$, with probability weight π_i at point p_i. Thus $F(p) = \frac{1}{n}\sum_{i=1}^{n} \pi_i \mathbf{1}(p_i \leq p)$. Let $\theta \equiv \{r, \tilde{q}_1, \dots, \tilde{q}_K\}$ denote the unknown parameters we want to estimate. As before, $\underline{p}$ and p^* can be estimated at a superconsistent rate, so in what follows we treat these parameters as known and non-stochastic.

Using the discrete distribution for $F(p)$, we obtain, for all $i = 1, \dots, K$:

$$(p^* - r)\,\tilde{q}_1 = (p_i - r)\left[\sum_{k=1}^{K} \tilde{q}_k k \left(1 - \left[\frac{1}{n}\sum_{j=1}^{n} \pi_j \mathbf{1}(p_j \leq p_i)\right]\right)^{k-1}\right]. \tag{17}$$

From equation (5), evaluated at $\underline{p}$, we rewrite r as a function of p^*, $\underline{p}$, and $\tilde{\mathbf{q}} \equiv \tilde{q}_1, \dots, \tilde{q}_K$:

$$\tilde{\rho}(\tilde{\mathbf{q}}) \equiv \frac{\underline{p} * \left[\sum_{k=1}^{K} \tilde{q}_k k\right] - p^* * \tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\right] - \tilde{q}_1} \tag{18}$$

This can be plugged into each of the equations in (17) above in order to eliminate r from the estimating equations.

So, given some value of θ, the empirical log-likelihood function is

$$\sum_{i=1}^{n} \log \pi_i. \tag{19}$$

and the empirical likelihood problem is to maximize (19) with respect to π_i, $i = 1, \dots, n$ and θ, subject to the K restrictions (17) and the summing up condition $\sum_{i=1}^{n} \pi_i = 1$.

We can transform the restrictions (17) into estimating equations (or moment conditions) of the form

$Ef(x;\theta) = 0$, as follows. For $s_1, \ldots, s_I \in [0,1]$ and $I \geq K$:

$$(17) \Leftrightarrow (p^* - r)\,\tilde{q}_1 = \left(F_p^{-1}(s_i) - \tilde{\rho}(\tilde{\boldsymbol{q}})\right)\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s_i)^{k-1}\right]$$

$$\Rightarrow F_p^{-1}(s_i) = \tilde{\rho}(\tilde{\boldsymbol{q}}) + \frac{(p^* - r)\,\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s_i)^{k-1}\right]} \equiv g_{s_i}(\tilde{\boldsymbol{q}}).$$

This population quantile restriction (that the s_i-th quantile of $F(p)$ equals $g_{s_i}(\tilde{\boldsymbol{q}})$) can be written as a population mean restriction (cf. Owen (2001), pg. 43)

$$E\left\{1\left(p_i \leq \rho(\tilde{\boldsymbol{q}}) + \frac{(p^* - \rho(\tilde{\boldsymbol{q}})\,\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s_i)^{k-1}\right]}\right) - s_i\right\} = 0,\; i = 1, \ldots, I,\; I \geq K. \tag{20}$$

The sample analogs are

$$\frac{1}{n}\sum_{j=1}^{n} \pi_j \left[\mathbf{1}\left(p_j \leq \tilde{\rho}(\tilde{\boldsymbol{q}}) + \frac{(p^* - r)\,\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s_i)^{k-1}\right]}\right) - s_i\right] = 0. \tag{21}$$

So the empirical likelihood estimate $\hat{\theta}$ of $\theta = (q_1, \ldots, q_{k-1})$ can be solved from the following saddle-point problem:

$$\max_{\theta} \min_{t} \sum_{i=1}^{n} \log\left(1 + t'\left[1\left(p_i \leq \rho(\tilde{\boldsymbol{q}}) + \frac{(p^* - \rho(\tilde{\boldsymbol{q}})\,\tilde{q}_1}{\left[\sum_{k=1}^{K} \tilde{q}_k k\,(1-s_i)^{k-1}\right]}\right) - s_i\right]\right)$$

$$\equiv \max_{\theta} \min_{t} \sum_{i=1}^{n} \log\left(1 + t' m\,(p_i; \theta)\right)$$

After deriving the estimates, the EL estimate of the equilibrium price distribution is simply

$$F^{EL}(p) = \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_i \cdot \mathbf{1}\,(p_i \leq p) \tag{22}$$

where $\hat{\pi}_i$ is the estimated weight on p_i. Given this, we can derive the quantiles of the search cost distribution using the consumer indifference conditions (2). Specifically, define $\hat{w}_i \equiv \sum_{j=1}^{n} \hat{\pi}_j \mathbf{1}\,(p_j \leq p_i)$. Note that $w_i^{1:s} \equiv s\hat{\pi}_i\,(1-\hat{w}_i)^{s-1}$, for any integer $s > 1$, is (approximately, as n grows large and the possibility of ties disappears) the weight on p_i in the discrete distribution of $p_{1:s}$, the smallest price out of s independent draws from the discrete distribution $F^{EL}(p)$. Given this notation, note that $Ep_{1:s} = s\sum_i p_i \hat{\pi}_i\,(1-\hat{w}_i)^{s-1}$. Therefore, the indifference conditions (2) can be rewritten as:

$$c_k = \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_i \left(1 - k\,(1-\hat{w}_i)^{k-1}\right) p_i,\; k = 1, \ldots, K. \tag{23}$$

Asymptotic theory for MEL estimates Using results from Qin and Lawless (1994) (pg. 318), we obtain the limiting distribution of the above empirical likelihood estimate:

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

where the matrixes in the limiting distribution is given by

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

$$B = \begin{bmatrix} s_1\left(1-s_1\right) & \cdots & \cdots & \cdots \\ \min\left(s_1,s_2\right)-s_1s_2 & s_2\left(1-s_2\right) & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \min\left(s_1,s_I\right)-s_1s_I & \cdots & \cdots & s_I\left(1-s_I\right) \end{bmatrix}$$

As noted above, the variance-covariance matrix for the MEL estimates corresponds to the variance-covariance matrix for the GMM estimates of θ using the I moment restrictions (20) and the optimal weighting matrix.

Note also that A can be consistently estimated by taking numerical derivatives of the following expression, where G denotes a smoothing function and h is a bandwidth:

$$\frac{\partial}{\partial\theta}\left[\frac{1}{n}\sum_{i=1}^{n}G\left(\frac{\rho\left(\tilde{q}\right)+\frac{\left(p^*-\hat{r}\right)\hat{\tilde{q}}_1}{\left[\sum_{k=1}^{K}\tilde{q}_k k\left(1-s_i\right)^{k-1}\right]}}{h}\right)\right].$$

In our estimation, we take G to be the standard normal CDF, and h to be 0.0001.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

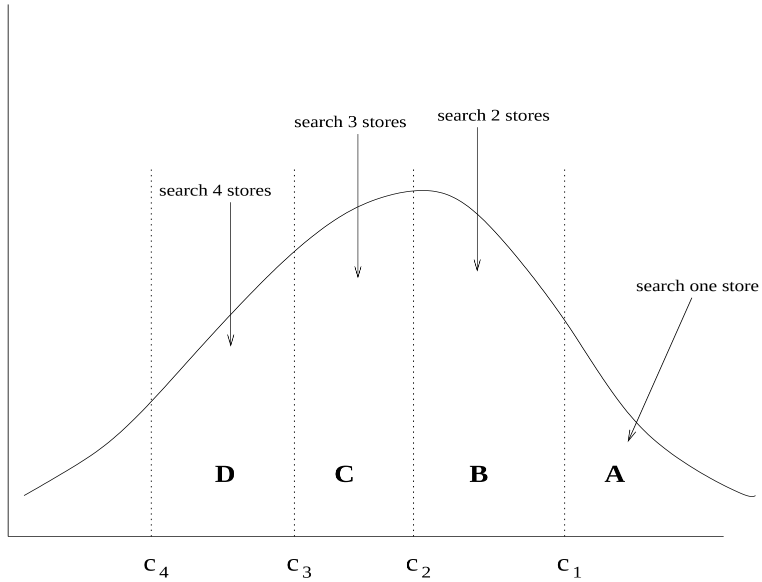
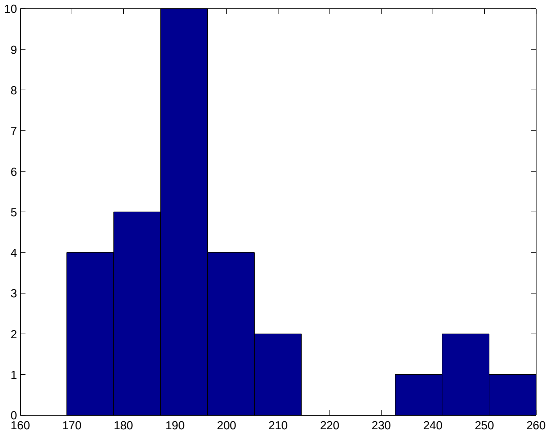
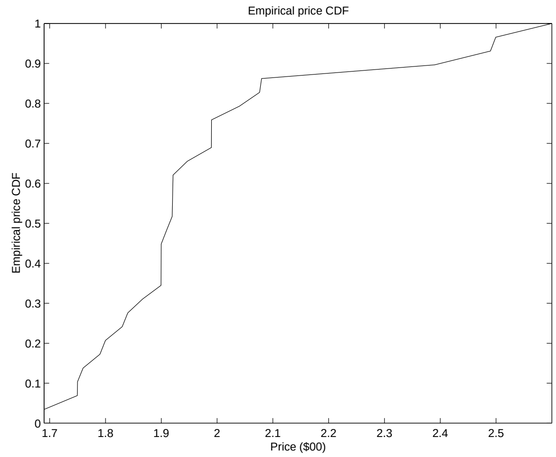


Figure 1: Identification scheme: Population distribution of consumer search costs F_c



Histogram of observed prices



Empirical CDF of Prices

Figure 2: Histogram and Empirical CDF of Palm Pilot IIIxe Prices

Data consists of 29 prices from online retailers; retrieved on May 11, 2001 by MySimon search engine.

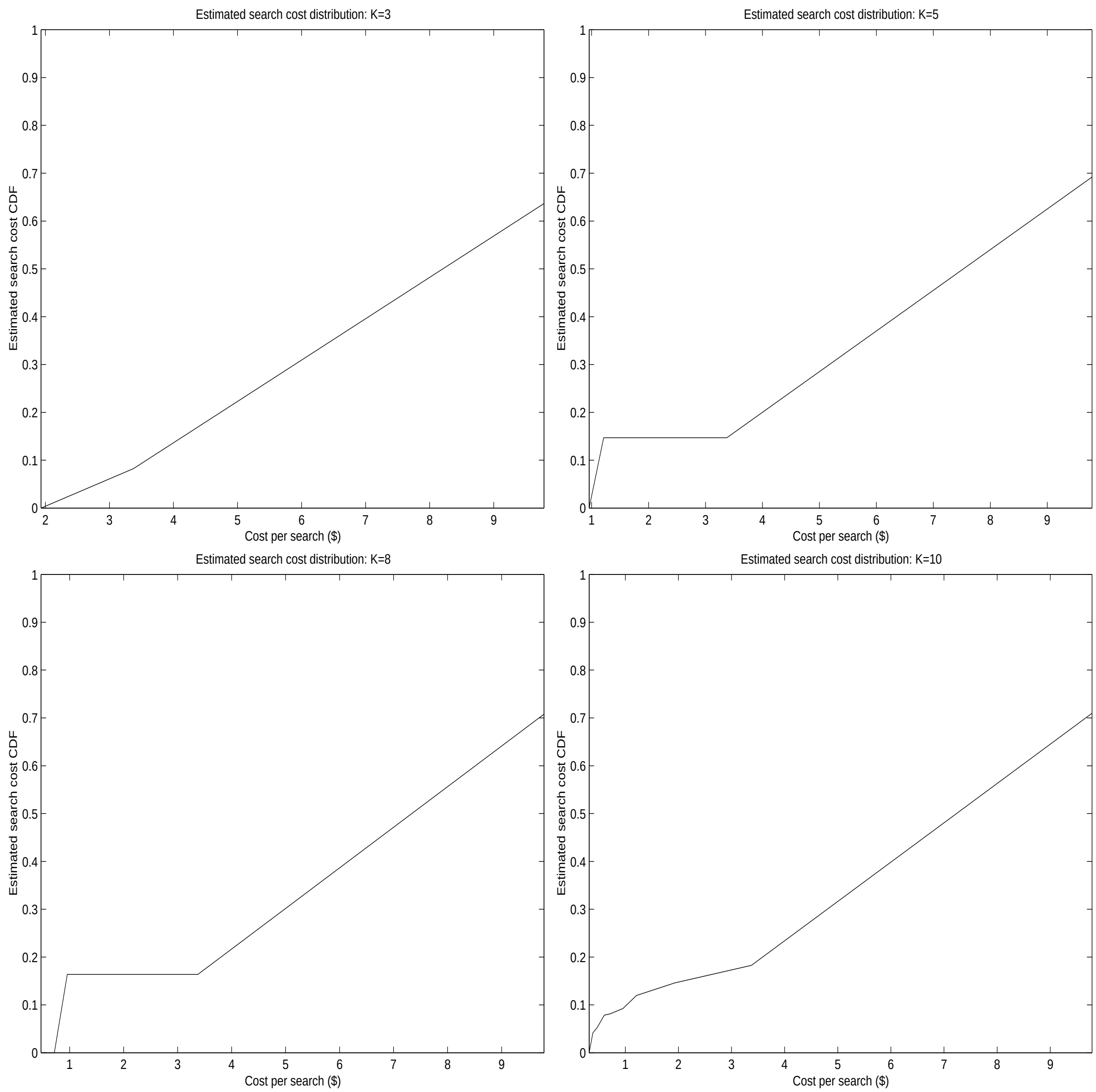


Figure 3: Estimated search cost distribution: Nonsequential search model, with heterogeneous search costs

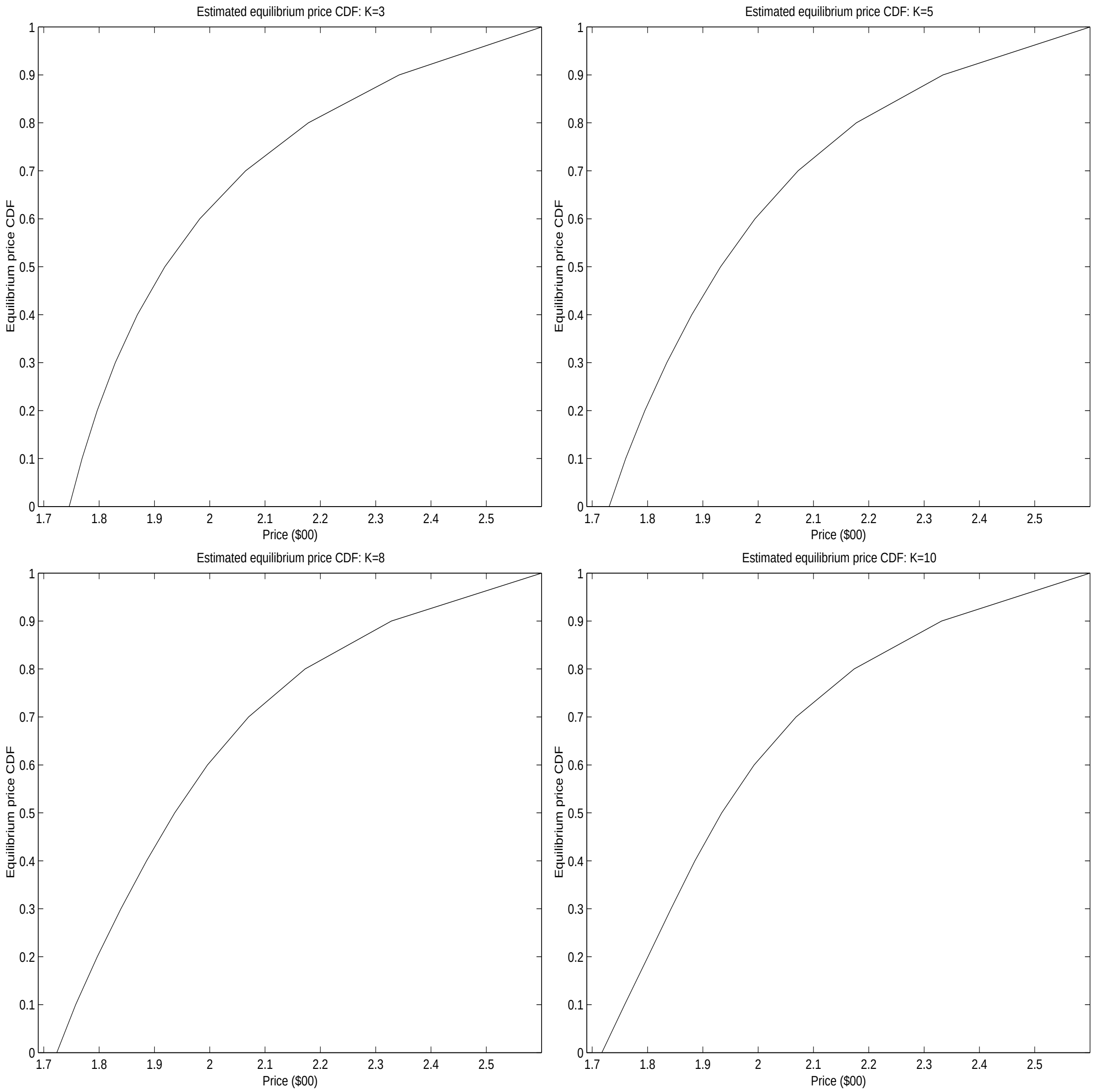


Figure 4: Estimated Equilibrium Price Distribution: Heterogeneous Search Costs Estimates
Data consists of 29 prices from online retailers; retrieved on May 11, 2001 by MySimon search engine.

60*(Ratio of estimated median search cost to average wage for All Workers)
Average wage: $15.72 (1998 *National Compensation Survey*).

Product	Model A	Model B[a]
Electronics:		
Palm IIIxe	43.1	295.4
Palm IIIc	60.9	211.6
Palm Vx	95.5	313.0
Palm VIIx	153.0	178.7
HP Jornada	64.3	65.8
Sony Mavica	$\in [60.3, 120.4]$	460.2
Sony MP3 NW-MS9	41.2	537.9
Books:		
Stokey-Lucas	$\in [6.7, 11.6]$	16.8
Ondaatje	> 2.2	12.0
Lazear	6.0	22.8
Billingsley	10.1	44.9
Duffie	14.8	24.1

Table 1: Imputed Minutes per Search: All Workers

[a]Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.

60*(Ratio of estimated median search cost to average wage for Lawyers)
Average wage: $36.30 (1998 *National Compensation Survey*).

Product	Model A	Model B[a]
Electronics:		
Palm IIIxe	18.7	127.9
Palm IIIc	26.4	91.6
Palm Vx	41.4	135.6
Palm VIIx	66.2	77.4
HP Jornada	46.4	28.5
Sony Mavica	$\in [26.1, 52.1]$	199.3
Sony MP3 NW-MS9	17.8	232.9
Books:		
Stokey-Lucas	$\in [2.9, 5.0]$	7.3
Ondaatje	> 1.0	5.2
Lazear	2.6	9.9
Billingsley	4.4	19.5
Duffie	6.4	10.4

Table 2: Imputed Minutes per Search: Lawyers

[a]Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.

60*(Ratio of estimated median search cost to average wage for Economics Teachers)
Average wage: $46.28 (1998 *National Compensation Survey*).

Product	Model A	Model B[a]
Electronics:		
Palm IIIxe	14.6	100.3
Palm IIIc	20.7	71.9
Palm Vx	32.5	106.3
Palm VIIx	52.0	60.7
HP Jornada	21.9	22.3
Sony Mavica	$\in [20.5, 40.9]$	156.3
Sony MP3 NW-MS9	14.0	182.7
Books:		
Stokey-Lucas	$\in [2.3, 3.9]$	5.7
Ondaatje	> 1.0	4.1
Lazear	2.0	7.8
Billingsley	3.4	15.3
Duffie	5.0	8.2

Table 3: Imputed Minutes per Search: Economics Teachers

[a]Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.

Median Search Cost Estimates, for Models A and B

Product	Model A		Model B [a]	
	with SH	w/o SH	with SH	w/o SH
Electronics:				
Palm IIIc	> 21.29	> 24.74	70.66	102.78
Palm Vx	> 20.80	> 21.97	105.26	75.67
Palm VIIx	16.78	15.66	136.50	42.24
HP Jornada	35.37		42.16	49.20
Sony Mavica	$\in [20.99, 45.77]$	$\in [23.64, 56.43]$	82.89	56.20
Books:				
Stokey-Lucas	1.16	1.20	21.35	19.36
Ondaatje		0.86	8.18	8.35
Lazear	1.07		11.37	11.04
Billingsley	2.04	1.99	7.74	12.20
Duffie	2.07		10.71	11.21

Table 4: Comparisons of Search Cost Estimates, for Models A and B

[a] Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.

Retailer	with shipping $\tilde{p}$	with shipping Est. MktSh	without shipping $\tilde{p}$	without shipping Est. MktSh
		Stokey Lucas		
alldirect	63.35	0.10	59.90	0.13
a1books	66.00	0.06	60.50	0.07
alphacraze	64.45	0.08	60.50	0.07
amazon	64.98	0.07	60.50	0.07
bn.com	64.45	0.08	60.50	0.07
booksamillion	64.45	0.08	60.50	0.07
booksfree	63.99	0.10	60.50	0.07
bookvariety	64.70	0.07	60.50	0.07
fatbrain	64.45	0.08	60.50	0.07
reiters	65.50	0.06	60.50	0.07
wordsworth	65.45	0.07	60.50	0.07
varsity	65.90	0.06	60.95	0.06
doublediscount	65.90	0.05	66.55	0.06
1bookstreet	71.05	0.05	72.60	0.06
		Palm IIIc		
americawestdigital	311.39	0.08	294.99	0.07
buyandmore	342.69	0.07	334.90	0.06
buydig.com	254.95	0.10	235.00	0.09
club pda			279.00	0.08
compuamerica	255.60	0.10	239.00	0.08
First-In-Technology	318.28	0.08	286.50	0.07
moreaudiovideo.com	229.99	0.11	229.99	0.09
My Mobile Devices	310.62	0.08	299.95	0.07
officemax	299.98	0.09	299.98	0.06
pcrush	266.98	0.10	254.35	0.08
regencycamera	244.39	0.11	229.99	0.09
statestreetdirect	336.86	0.08	329.00	0.06
universal computers			175.00	0.01

Table 5: Implied market shares: Nonsequential search models

Product	# obs	List	Mean	Median	$\underline{p}$	p^*
Electronics:						
Palm IIIxe	29		197.79	191.95	169.00	259.99
Palm IIIc	35		294.50	299.00	235.00	399.99
Palm Vx	38		285.25	299.00	218.49	407.00
Palm VIIx	37		241.45	199.00	164.00	419.95
Sony Mavica MVC-CD300	32		920.49	918.75	809.00	1033.00
HP Jornada 547	27		399.38	393.95	325.00	449.99
Sony MP3 Player NW-MS9	21		332.93	348.00	298.00	350.00
Books:						
Stokey/Lucas	12	57.50	55.37	57.50	46.00	69.00
Ondaatje: *The English Patient*	13	13.00	11.24	11.49	8.55	13.00
Lazear: *Personnel Economics*	13	31.95	30.15	31.95	25.56	35.15
Billingsley: *Probability and Measure*	14	94.95	91.05	94.70	80.95	94.99
Duffie: *Dynamic Asset Pricing Theory*	14	59.50	53.32	59.49	40.50	59.50

Table 6: Summary statistics on prices for different products
Price data for all products downloaded from `pricescan.com`

Product	K	I	$\tilde{q}_1$	$\tilde{q}_2$	$\tilde{q}_3$	Prod'n Cost	MEL value	Smooth MEL[a]
Electronics:								
Palm IIIxe	3	7	0.490 (0.160)	0.053 (0.816)	0.015 (2.440)	145.74 (22.46)	178.14	72.72
Palm IIIc	3	7	0.570 (0.261)	0.273 (0.095)	0.000 (0.034)	154.95 (11.68)	88.28	84.74
Palm Vx	3	7	0.466 (0.158)	0.256 (0.124)	0.133 (0.189)	159.68 (53.36)	123.12	51.67
Palm VIIx	3	7	0.488 (0.092)	0.237 (0.159)	0.032 (0.433)	83.09 (39.96)	340.79	50.81
HP Jornada	2	5	0.458 (0.061)	0.268 (0.034)		282.81 (8.98)	158.01	81.37
Sony Mavica	2	5	0.246 (0.787)	0.654 (1.811)		774.81 (193.73)	5.03	40.15
Sony NW-MS9	2	5	0.434 (0.222)	0.295 (0.175)		281.91 (17.05)	36.47	46.33
Books:								
Stokey-Lucas	2	5	0.311 (0.545)	0.456 (0.794)		41.57 (13.76)	95.53	29.28
Ondaatje	2	5	0.761 (0.423)	0.239 (0.350)		1.46 (17.57)	72.20	53.75
Lazear	2	5	0.488 (0.141)	0.261 (0.059)		21.89 (2.68)	51.68	26.13
Billingsley	2	5	0.445 (0.232)	0.282 (0.092)		76.43 (6.23)	24.50	26.37
Duffie	2	5	0.491 (0.223)	0.258 (0.143)		33.13 (8.40)	24.02	25.65

Table 7: Estimates for Model A: Empirical likelihood approach
(Standard errors in parentheses.)

[a] See Section 6 for details.

Product	c_1	$F_c(c_1)$	c_2	$F_c(c_2)$	c_3	$F_c(c_3)$
Electronics:						
Palm IIIxe	11.30	0.510	3.96	0.457	2.15	0.442
Palm IIIc	15.96	0.430	8.20	0.157	5.63	0.157
Palm Vx	25.03	0.534	11.65	0.278	7.43	0.145
Palm VIIx	40.08	0.512	12.52	0.275	4.92	0.243
HP Jornada	16.85	0.542	8.13	0.274		
Sony Mavica	31.54	0.754	15.79	0.100		
Sony MP3 NW-MS9	10.78	0.566	7.05	0.271		
Books:						
Stokey-Lucas	3.03	0.689	1.75	0.233		
Ondaatje	0.58	0.239	0.34	0.000		
Lazear	1.56	0.512	0.89	0.251		
Billingsley	2.65	0.555	1.77	0.273		
Duffie	3.87	0.509	2.64	0.251		

Table 8: Search Cost Distribution Estimates for Model A: Empirical likelihood approach
Search cost estimates corresponding to parameter estimates from Table 7.
Indifferent points c_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.

Product	Search Cost Distn	Mean	Median	Prodn Cost	LL value
Electronics:					
Palm IIIxe	Exponential	89.70 (23.19)	62.18 (16.08)	122.80 (11.55)	56.39
	Gamma	133.08 (3.16)	77.39 (6.03)	111.29 (15.74)	63.07
	Weibull	183.67 (1.85)	100.29 (2.32)	97.56 (7.32)	62.61
Palm IIIc	Exponential	79.42 (7.77)	55.05 (5.39)	169.04 (6.82)	50.91
	Gamma	90.46 (2.30)	55.44 (3.40)	157.71 (2.45)	57.57
	Weibull	141.78 (2.11)	67.28 (21.61)	127.53 (3.89)	56.93
Palm Vx	Exponential	88.62 (10.20)	61.43 (7.06)	133.72 (11.86)	49.54
	Gamma	45.55 (0.94)	28.24 (1.17)	162.77 (1.54)	56.72
	Weibull	130.42 (14.37)	82.01 (31.33)	106.93 (2.12)	58.47
Palm VIIx	Exponential	18.57 (8.47)	12.87 (5.87)	12.08 (1.87)	33.81
	Gamma	78.15 (2.49)	46.83 (3.51)	54.30 (2.53)	42.04
	Weibull	70.63 (0.65)	38.05 (1.13)	59.79 (2.07)	38.60
HP Jornada 547	Exponential	37.21 (17.12)	25.79 (11.87)	295.09 (37.43)	54.50
	Gamma	55.08 (1.63)	28.90 (2.46)	279.78 (36.13)	56.65
	Weibull	27.31 (0.66)	17.23 (0.83)	298.38 (4.75)	58.11
Sony Mavica	Exponential	136.46 (15.64)	94.59 (10.84)	677.57 (71.19)	40.44
	Gamma	227.68 (1.69)	120.56 (2.16)	588.75 (24.47)	46.27
	Weibull	349.75 (101.87)	145.81 (34.19)	537.25 (86.37)	46.15
Sony MP3 NW-MS9	Exponential	36.81 (3.47)	25.52 (2.40)	258.64 (13.23)	59.16
	Gamma	248.83 (1.73)	140.93 (3.81)	243.05 (4.24)	62.70
	Weibull	446.04 (233.43)	172.17 (37.92)	224.63 (15.46)	62.56
Books:					
Stokey-Lucas:	Exponential	9.14 (0.64)	6.34 (0.44)	35.58 (1.31)	12.69
Recursive Methods ...	Gamma	6.77 (0.29)	4.39 (0.65)	37.45 (34.06)	23.09
	Weibull	21.41 (0.32)	11.78 (2.43)	25.59 (11.95)	22.35
Ondaatje:	Exponential	1.86 (0.24)	1.29 (0.16)	7.81 (0.71)	37.35
The English Patient	Gamma	4.89 (0.52)	3.15 (0.47)	7.21 (3.59)	40.70
	Weibull	18.93 (0.14)	10.49 (0.58)	4.97 (0.68)	39.54
Lazear:	Exponential	10.45 (1.54)	7.24 (1.06)	20.97 (1.74)	25.26
Personnel Economics	Gamma	9.26 (0.09)	5.98 (0.22)	21.13 (2.11)	35.85
	Weibull	19.49 (0.19)	12.16 (0.25)	17.74 (2.22)	35.01
Billingsley:	Exponential	14.63 (0.58)	10.14 (0.40)	72.24 (2.64)	24.95
Probability and Measure	Gamma	19.83 (0.33)	11.77 (0.20)	68.83 (1.11)	28.53
	Weibull	25.82 (3.56)	14.04 (6.91)	66.47 (20.23)	28.27
Duffie:	Exponential	3.40 (0.62)	2.35 (0.43)	36.17 (0.19)	21.58
Dynamic Asset ...	Gamma	10.16 (0.10)	6.30 (0.07)	30.90 (3.91)	25.07
	Weibull	19.60 (0.33)	10.50 (0.32)	24.82 (5.02)	24.66

Table 9: Estimates for Model B: Parametric approach
(Standard errors in parentheses.)

Median Search Cost Estimates, for Models A and B

Product	Model A	Model B[a]
Electronics:		
Palm IIIxe	11.30	77.39
Palm IIIc	15.96	55.44
Palm Vx	25.03	82.01
Palm VIIx	40.08	46.83
HP Jornada	16.85	17.23
Sony Mavica	$\in [15.79, 31.54]$	120.56
Sony MP3 NW-MS9	10.78	140.93
Books:		
Stokey-Lucas	$\in [1.75, 3.03]$	4.39
Ondaatje	> 0.58	3.15
Lazear	1.56	5.98
Billingsley	2.65	11.77
Duffie	3.87	6.30

Table 10: Comparisons of Search Cost Estimates, for Models A and B

[a]Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.

Ratio (%) of estimated median search cost to median observed price (from Table 6).

Product	Model A	Model B[a]
Electronics:		
Palm IIIxe	5.88	40.31
Palm IIIc	5.34	18.54
Palm Vx	8.37	27.43
Palm VIIx	20.14	23.53
HP Jornada	4.28	4.37
Sony Mavica	$\in [1.72, 3.43]$	13.12
Sony MP3 NW-MS9	3.10	40.50
Books:		
Stokey-Lucas	$\in [3.04, 5.27]$	7.63
Ondaatje	> 5.05	27.42
Lazear	4.88	18.72
Billingsley	2.83	12.43
Duffie	6.50	10.59

Table 11: Comparisons of Search Cost Estimates, for Models A and B

[a]Reported for parametric specification with the highest log-likelihood value, as reported in the rightmost column of Table 9.