Estimation of Search Costs[*]

José Luis Moraga-González[†]

March 2006

Abstract

This paper reviews the recent contributions on the structural estimation of search costs. We first discuss some of the theoretical and empirical literature on price dispersion and consumer search. We then argue that optimal design of competition policy needs the development of methods to identify and estimate search costs. We finally discuss the methods that have been proposed to date and the results obtained.

Keywords: consumer search, oligopoly, non-parametric estimation

JEL Classification: C14, D43, D83, L13

[*]This paper has been written for the 2006 ENCORE workshop on "Consumers and competition." Maarten Janssen invited me to talk at the workshop in Rotterdam. In my talk and in this paper I have freely drawn from joint work with Matthijs Wildenbeest and Zsolt Sandor (see references). Matthijs also provided comments on this draft. I am indebted to the three of them.

[†]University of Groningen, E-mail: `j.l.moraga.gonzalez@rug.nl`

1 Introduction

Since the seminal article of Stigler (1961) on the economics of information, a great deal of theoretical and empirical work has revolved around the existence of price dispersion in homogeneous product markets. The theoretical work, probably inspired by Rothschild's (1973) criticism on Stigler's partial-partial equilibrium approach, first focused on finding a rationale for price dispersed equilibria. In later work, researchers focused on the properties of markets with search costs, including how prices and price dispersion depend on market characteristics like the height of search costs or the number of firms. The early empirical work focused on documenting the scope of price dispersion, in particular in markets for goods that are seemingly standardized; in later work, applied economists tried to test some of the predictions obtained from consumer search theory. We will explain in some detail a number of important theoretical and empirical contributions to the consumer search literature in Section 2.

In spite of the relatively large body of theoretical and empirical work, somewhat surprisingly, very little attention has been given to the identification and measurement of search costs in real-world markets. We will argue that this is a serious omission in Section 3. The reason is that the predictions of various theoretical models are quite sensitive to the details of the search cost distribution. As a matter of fact, competition policy recommendations might depend on the nature of search costs. As a result, there is a need to develop methods to quantify the search cost distribution. Later in the section, we will review the emerging empirical literature trying to cover this gap by providing methods to identify and estimate search costs. In Section 4, we will offer some concluding remarks.

2 Related literature

Stigler's seminal article spawned a great deal of theoretical work that focused on finding conditions under which price dispersed equilibria can be sustained in markets for homogenous goods. Following his ideas, an explanation that gained interest quickly was based on consumer costly search.[1] Some authors considered sequential consumer search, i.e., a search strategy by which a consumer decides in a initial stage whether to observe a first price or not and then, after seeing the first price, whether to continue searching or not, and so on. Reinganum (1979) presents a simple model of sequential consumer search where a continuum of firms with heterogeneous costs quotes prices to consumers. Consumers engage in costly sequential search to discover prices. Reinganum proves that an equilibrium with price dispersion can be sustained whenever consumers hold elastic demands; in equilibrium each firm's profit-maximizing price is a constant markup over its cost and this induces a distribution of prices in the market. In Reinganum's model the source of price dispersion was thus marginal cost heterogeneity across firms.

Stahl (1989) also studied a model of sequential consumer search but his explanation for price dispersion relied on consumer search costs variation. In his model, some consumers search at no cost while others have positive search costs. In equilibrium, firms randomize their prices in an attempt to appeal to both types of consumers and positive search cost consumers do not search. One of the most interesting results in his paper is that entry of firms is anti-competitive. As expected, increases in search costs lead to higher prices. Stahl makes the assumption that consumers obtain the first price quotation for free. Janssen, Moraga-González and Wildenbeest (2005) show that when sequential search is truly costly, an equilibrium need not involve full consumer participation. If this is the case, somewhat counter-intuitively, then search cost increases can lead to price reductions.

Burdett and Judd (1983) were the first to show that price dispersion can arise in a model with no ex-ante heterogeneity whatsoever. The key assumption is that consumers search non-sequentially. Under non-sequential search a consumer decides once and for all how much to search. This search procedure differs from sequential search in that the consumer does not condition his/her decision to continue searching or not on the outcome of previous searches.[2]

[1]The literature has offered at least three other explanations for price dispersion in homogeneous product markets. One, price dispersion arises in models where firms engage in costly price advertising (see Butters, 1977; Stahl, 1994). Two, price dispersion is also an equilibrium phenomenon in "clearinghouse models", i.e., where some consumers gain access to a list of prices and others do not (e.g., Varian, 1980; Rosentahl, 1980; Salop and Stiglitz, 1977; Baye and Morgan, 2001). Three, price dispersion arises in Bertrand competition models with unknown costs of production (Spulber, 1995).

[2]Non-sequential search is also called fixed-sample-size search in the literature. Morgan and Manning (1984) show that optimal search strategies are typically a hybrid of sequential and non-sequential.

Burdett and Judd show that an equilibrium must necessarily involve consumer randomization so that, ex-post, some consumers observe one price quotation and some consumers observe two price quotations. As a response to such consumer behavior, firms find it optimal to mix in prices so price dispersion materializes in equilibrium. Janssen and Moraga-González (2004) also study non-sequential search but their focus is on how search intensity and the equilibrium price distribution depend on the number of firms in the market. They find that the effects of entry on search behavior, expected prices, price dispersion and welfare are sensitive to the status quo number of firms. For example, more firms around results in more intensive search and lower prices when the number of competitors in the market is low to begin with, but in less search and higher prices when the number of competitors is large.

The theoretical literature on consumer search is by now extensive and the reader is referred to McMillan and Rothschild (1994) for a survey. In more recent research, search costs have been proven to have a bearing on an array of other issues, including wage and technology dispersion (Burdett and Mortensen, 1997; Acemoglu and Shimer, 2000), excessive product diversity (Anderson and Renault, 2000; Wolinsky, 1984), inefficient investments (Wolinsky, 2005), asymmetric price-cost adjustments (Lewis, 2003), and the likelihood of different price institutions (Bester, 1994).

The early empirical work focused first on documenting the existence of price dispersion in homogeneous product markets, and second on explaining it on the basis of consumer search. Stigler (1961) provided himself a couple of examples, one on 27 prices for Chevrolets in the Chicago area and another on 14 prices for anthracite coal delivered in Washington. In his own words, "in both cases the range of prices was significant on almost any criterion." As a matter of fact, the coefficient of variation for the Chevrolets was 1.72 while that for anthracite was 6.8.

Marvel (1976) is one of the first studies of price dispersion in retail markets; using gasoline pump prices collected by US officials for use in preparation of the Consumer Price Index, he found that price ranges vary significantly across cities, more than what could reasonably be attributed to differences in market structures and taxes across cities. Pratt, Wise and Zeckhauser (1979) made perhaps the earliest systematic documentation of price dispersion in seemingly competitive markets. The authors compiled a sample of 39 randomly chosen products from the Yellow Pages in the Boston area. Prices for these goods were collected by making phone calls to the list of sellers that appeared in the phone directory. For all products, they found substantial price differences, being the maximum price over twice the lowest price for 18 of the 39 items.

Carlson and Pescatrice (1980) tested one of Stigler's propositions that goods that are frequently purchased should have smaller price differentials as compared to goods more seldomly bought. The statement was borne out by their data. Dahlby and West (1986) studied the prices of automobile insurance in Alberta (Canada) from 1974 to 1981. They concluded that observed price dispersion could be based on costly consumer search. In particular, they found that insurance premiums are less dispersed in driver classes where search intensity is likely to be high. Van Hoomissen (1988) studied prices for 13 goods sold in Israel from 1971 to 1984. She argued that actual or perceived differences in stores' quality, location, after-sale service, etc. could not explain price differentials across outlets in these markets so she concluded that price dispersion in her sample was strongly based on the existence of an unequal distribution of information in the market, which is consistent with search theory. More recently, Sorensen (2000), in a study of a market for prescription drugs, also presents evidence that mean prices and price dispersion are sensitive to the characteristics of the drug therapy in the way search theory suggests. In particular, mean prices and price dispersion of long (multi-month) prescriptions are lower than those of regular drug therapies because the incentives to search for prices in the former category of goods are larger than in the later.

Consumer search models have produced price dispersion based on pure-strategy equilibria or price dispersion based on mixed strategy equilibria. Some authors have argued that if price dispersion is permanent, pure-strategy equilibria cannot be the explanation for it. This is because if some stores permanently charge lower prices than other stores, as it would happen in a pure-strategy equilibrium, consumers would learn which stores offer discounts and therefore the persistence of price dispersion would not sustain. By contrast, if observed price dispersion is the result of firms playing mixed strategies, then consumers would not be able to learn too much from experience. This is simply because by the time a consumer returns to a shop the information he/she might have collected has probably expired. In this situation, price dispersion is expected to persist over time.

To the best of my knowledge, Lach (2002) was the first to study the persistence and the nature of price dispersion over time. He argued that price dispersion based on pure-strategy equilibria could not be what we observe in real-world markets. Using a data set of prices of four homogeneous goods (chicken, coffee, flour and a refrigerator) sold in Israel from January 1993 to June 1994, Lach found that price dispersion exists and prevails over time. Further, he noticed that the position of a typical store in the price ranking changed up and down so knowledge of a store's price position in a given month was not useful to predict the store's price position 6 months later. As a result, he concluded that consumers learning about which stores have consistently low prices is rather limited so the data seems to be consistent with stores' use of mixed strategies.

With the arrival of the Internet, the collection of price data in online markets has become quite easy. As a result, studies of price competition in electronic markets have proliferated. One of the most researched themes is the impact of the Internet on competitiveness. The Internet, supposedly, reduces search costs significantly, so several authors have tried to trace out its effects from price data. The results of these studies are somewhat mixed. Researchers do not seem to agree on whether competition has weakened or strengthened with the introduction of the Internet. In fact, concerning price levels in electronic markets, some studies, for instance Bailey's (1998) on books and CDs, find that they are higher than corresponding prices in conventional markets. Other analyses on book markets, namely, Friberg *et al.* (2000) and Clay *et al.* (2000a), report that prices in on-line and physical stores are similar. Finally, Brynjolfsson and Smith (1999) in a study on books and CDs find lower prices in electronic markets. In the market for life insurance policies, Brown and Goolsbee (2000) argue that there is no evidence that Internet usage reduced prices before comparison websites emerged and proliferated. On price dispersion, the effect of moving markets on-line also seems to be ambiguous empirically.

3 Estimation of search costs

In spite of the large body of theoretical and empirical work, somewhat surprisingly, very little attention has been given to the identification and measurement of search costs in real-world markets. From a practical point of view, this is a serious omission because the predictions of the various theoretical models are quite sensitive to the details of the search cost distribution. For example, in Janssen and Moraga-González (2004) mergers of firms may increase prices and price dispersion if the cost of search is high; otherwise they lead to lower prices. From this observation it follows that competition policy recommendations might depend on the nature of search costs; as a result, there is a need to develop methods to quantify the search cost distribution.

At the time of writing this survey, to the best of our knowledge, there are only three papers dealing with this issue.[3] Hong and Shum (forthcoming) were the first to develop a structural methodology to retrieve information on search costs. They focus on markets for homogeneous goods, with sequential as well as with non-sequential search. One of their main messages is that price data alone may suffice for the estimation of search cost parameters. Since data on market shares are often hard to obtain, this finding is of great importance. Moraga-González and Wildenbeest (2006) extend the approach of Hong and Shum to the oligopoly case, arguably a more relevant market structure when it comes to antitrust applications, and they provide a maximum likelihood estimate of the search cost distribution. Hortaçsu and Syverson (2004) show that when price and quantity data are available, this methodology can be extended to industries where search frictions coexist with (vertical) product differentiation.

In the remainder of this section, we will describe these contributions in some more detail. Hong and Shum (forthcoming) examine two distinct models of firm competition under consumer search. In their first model consumers search non-sequentially while in their second model consumers search sequentially. We concentrate here on the nonsequential search model of Hong and Shum because Hortacsu and Syverson (2004) also consider a model with

[3] To be sure, Sorensen (2001) also tries to estimate search costs but he uses a model of consumer choice where the supply side is exogenous.

sequential consumer search and we will discuss it later.

3.1 A model of non-sequential search and oligopolistic pricing

The model of Hong and Shum (forthcoming) generalizes the non-sequential consumer search model of Burdett and Judd (1983) by adding consumer search cost heterogeneity. The details of the model are as follows. Assume there are infinitely many retailers selling a homogeneous good with a common unit selling cost denoted by r. There is a very large number of consumers; assume that the number of consumers per firm is μ. Consumers hold inelastic demands, i.e., each buyer demands one unit of the good or nothing. All consumers place the same value on the good, denoted $\overline{p}$. Consumers obtain the first price quotation for free but further price information only comes at a search cost c per price quote. Buyers differ in their search costs. Assume that the cost of a consumer is randomly drawn from a distribution of search costs $F_c(c)$. A consumer with search cost c sampling ℓ firms incurs a total search cost ℓc.

Denote the symmetric mixed strategy equilibrium by the distribution of prices $F_p(p)$, with density $f_p(p)$. Let $\underline{p}$ and $\overline{p}$ be the lower and upper bound of the support of $F_p(p)$. Given firm behavior, the number of prices $i(c)$ a consumer with search cost c observes must be optimal, i.e., it must be the solution to the problem of minimizing the total pay for the item:

$$i(c) = \arg\min_{i>1} c(i-1) + \int_{\underline{p}}^{\overline{p}} ip(1-F_p(p))^{i-1} f_p(p)dp. \tag{1}$$

Since $i(c)$ must be an integer, the problem in equation (1) induces a partition of the set of consumers into subsets of size q_i, $i = 1, 2, ..., \infty$, with $\sum_{i=1}^{\infty} q_i = 1$; thus, the number q_i is the fraction of buyers sampling i firms.

This partition is calculated as follows. Let $Ep_{1:i}$ be the expected minimum price in a sample of i prices drawn from the price distribution $F_p(p)$. Then

$$\Delta_i = Ep_{1:i} - Ep_{1:i+1}, \; i = 1, 2, ..., \infty \tag{2}$$

denotes the search cost of the consumer indifferent between sampling i prices and sampling

$i+1$ prices. Note that Δ_i is a decreasing function of i. Using this, the fractions of consumers q_i sampling i prices are simply

$$q_1 = 1 - F_c(\Delta_1); \tag{3}$$

$$q_i = F_c(\Delta_{i-1}) - F_c(\Delta_i), \; i = 2, 3, ..., \infty. \tag{4}$$

Given consumer search behavior it is indeed optimal for firms to mix in prices. The upper bound of the price distribution must be $\overline{p}$ because a firm which charges the upper bound sells only to the consumers who do not compare prices (consumers in q_1), who would also accept $\overline{p}$. The equilibrium price distribution follows from the indifference condition that a firm should obtain the same level of profits from charging any price in the support of $F_p(p)$, i.e.,

$$(p - r)\left[\sum_{i=1}^{\infty} iq_i(1 - F_p(p))^{i-1}\right] = q_1(\overline{p} - r). \tag{5}$$

From equation (5) it follows that the minimum price charged in the market is

$$\underline{p} = \frac{q_1(\overline{p} - r)}{\sum_{i=1}^{\infty} iq_i} + r. \tag{6}$$

Hong and Shum (forthcoming) show how to use equations (2) to (6) to estimate the search cost distribution using only price data. This is certainly good news given the fact that quantity information is often very hard to obtain.

The first problem one encounters when taking the model to data is that, typically, there is a finite number of firms operating in a market. Suppose one observes N firms selling the product in a market, and the prices they charge. Let $\hat{F}_p(p)$ be the empirical distribution of prices. Ordering the prices in ascending order we have

$$\underline{p} = p_1 < p_2 < ... < p_N = \overline{p} \tag{7}$$

Let K be the maximum number of firms sampled by a consumer (in principle, one would expect that K should be equal to N). Then the indifference condition (5) becomes

$$(p_j - r)\left[\sum_{i=1}^{K} iq_i(1-\hat{F}_p(p_j))^{i-1}\right] = q_1(\overline{p}-r),\ j = 1, 2, ..., N-1. \tag{8}$$

Using the observed prices, these equations can be used to obtain estimates of the parameters $\{q_1, q_2, ..., q_K\}$ of the price distribution. For example, if one knows the marginal cost of the firms, then using the fact that $q_K = 1 - \sum_{i=1}^{K-1} q_i$, the expressions in (8) form a system of $N-1$ equations that can be used to estimate (by OLS) the rest of the unknowns of interest in the price distribution, namely, $\{q_1, q_2, ..., q_{K-1}\}$. Alternatively, if we don't know the marginal cost r, then we can isolate it from the equation in (8) obtained when $j = 1$, which gives $r = (\underline{p}\sum_{i=1}^{N} iq_i - q_1\overline{p})/\sum_{i=2}^{N} iq_i$, and plug it into the rest of the equations in (8). In this case we lose one equation so we need to assume that $K \leq N-1$ and use the $N-2$ equations left in (8) to estimate (by non-linear least squares) the parameters $\{q_1, q_2, ..., q_{K-1}\}$.

Hong and Shum proceed differently and formulate the estimation of the parameters of the price distribution as an empirical likelihood problem. They argue that once the estimates of the q_i's are obtained, one can use the empirical distribution of prices in (2) to get the cut-off points Δ_i. The collection of points $\{\Delta_i, q_i\}$ yields an estimate of the search cost distribution, which can be constructed by using some interpolation method. This way, their estimate of the search cost distribution is a hybrid of empirical likelihood and kernel nonparametric for the cutoff points.

To obtain minimum variance estimates, Moraga-González and Wildenbeest (2006) have proposed a maximum likelihood estimation method. Their procedure has the advantage that the asymptotic theory for computing the standard errors of Δ_i and for conducting tests of hypotheses remains standard. To estimate the search cost distribution by maximum likelihood, we rewrite Δ_i as a function of the ML estimates of the parameters of the price distribution. To do this, we first rewrite the cut-off points as (by integration by parts)

$$\Delta_i = \int_{\underline{p}}^{\overline{p}} F_p(p)(1-F_p(p))^i dp,\ i = 1, 2, \ldots, N-1. \tag{9}$$

Using the inverse of the price distribution

$$p(z) = \frac{q_1(\overline{p} - r)}{\sum_{i=1}^{N} i q_i (1-z)^{i-1}} + r. \tag{10}$$

a change of variables in equation (9) yields:

$$\Delta_i = \int_0^1 p(z)[(i+1)z - 1](1-z)^{i-1} dz, \ i = 1, 2, \ldots, N-1. \tag{11}$$

If we obtain ML estimates of $r, \overline{p}, \underline{p}$ and q_i, $i = 1, 2, \ldots, N$, then we can use equations (10) and (11) to calculate ML estimates of the cut-off points of the search distribution so we obtain a ML estimate of the search cost distribution $F_c(c)$.

We now discuss how to estimate r, $\overline{p}$ and q_i, $i = 1, 2, \ldots, N$ by maximum likelihood, assuming that only price data are available. Since the price density cannot be obtained in closed form, we apply the implicit function theorem to equation (5), which yields

$$f_p(p) = \frac{\sum_{i=1}^{N} i q_i (1 - F_p(p))^{i-1}}{(p - r) \sum_{i=1}^{N} i(i-1) q_i (1 - F_p(p))^{i-2}}. \tag{12}$$

Moraga-González and Wildenbeest (2006) assume the researcher observes the prices firms charge over some period of time. Let $\{p_1, p_2, \ldots, p_M\}$ be the vector of observed prices. Without loss of generality, let $p_1 < p_2 < \ldots < p_M$. Following Kiefer and Neumann (1993) we take the minimum price in the sample p_1 and the maximum one p_M to estimate the lower and upper bounds of the support of the price distribution $\underline{p}$ and $\overline{p}$, respectively. These estimates of the bounds of the price distribution converge super-consistently to the true bounds. Using the estimates of $\underline{p}$ and $\overline{p}$, equation (6) can be solved to obtain the marginal cost r as a function of the other parameters:

$$r = \frac{p_1 \sum_{i=1}^{N} i q_i - q_1 p_M}{\sum_{i=2}^{N} i q_i}. \tag{13}$$

We then can solve the maximum likelihood estimation problem

$$\max_{\{q_i\}_{i=1}^{N}} \sum_{\ell=2}^{M-1} \log f_p(p_\ell; q_1, q_2, ..., q_N) \tag{14}$$

$$\text{s.t.} \sum_{i=1}^{N} q_i = 1, \text{ (8) and (13)}, \tag{15}$$

numerically, which yields the desired ML estimates.[4][5]

Moraga-González and Wildenbeest (2006) apply their method to a data set of prices for four personal computer memory chips. Prices were obtained from *cnet.com*, an American web-based search engine.[6] For all the products, observed price dispersion, measured by the coefficient of variation, was significant. To give an idea of the extent of price dispersion, on average, relative to buying from one of the firms at random, the gains from being fully informed in these markets are sizable, ranging from 21.56 to 32.89 US dollars. Their estimates of the parameters of the price distribution yield an interesting finding: consumers either search for all prices in the market (between 4% and 13% of the consumers) or search very little, namely for at most three prices. Almost no consumer searches for an intermediate number of prices. The search cost distribution consistent with these estimates implies that consumers have either quite high or quite low search costs. The estimates suggest that the search cost of consumers who sample all prices in the market is at most 17 US dollar cents.

Given that quite a few consumers search so little, buyer behavior must confer a significant amount of market power to the firms. Their estimates of the average price-cost margins range between 23% and 28%. Note that these price-cost margins are quite high from the perspective that there were more than 20 firms quoting prices in each of the markets studied. Moraga-González and Wildenbeest (2006) conduct Kolmogorov-Smirnov tests of the goodness of

[4]Note that since the estimate of r is obtained from equation (13) as a function of the estimates of the other parameters, this procedure introduces some dependence between the price observations. This is not a problem because the upper and lower bounds of the price distribution converge to the true values at a super-consistent rate.

[5]The standard errors of the estimates of q_i, $i = 1, 2, \ldots, N-1$ can be calculated by taking the square root of the diagonal entries of the inverse of the negative Hessian matrix evaluated at the optimum. The standard error of the estimate of q_N can be calculated using the Delta method and the same applies to the standard errors of the estimates of the marginal cost r and the Δ_i's.

[6]Non-sequential search is not perhaps the most appropriate search protocol to describe search on the Internet. Nonetheless, the model fits the data pretty well.

fit to check the validity of the theoretical model. According to their test results, the null hypothesis that price data are generated by the model cannot be rejected. This implies that the non-sequential search model works quite well when it comes to predicting prices in markets for homogeneous goods.

3.2 A model of sequential search and oligopolistic pricing

Hortacsu and Syverson (2004) propose a model of sequential consumer search to study search costs and price dispersion in the mutual fund industry. In their model, consumers are all identical except in search costs, which are drawn from a common distribution $G(c)$. Consumers search sequentially, i.e., a consumer decides in a initial stage whether to observe a first price or not and then, after seeing the first price, whether to continue searching or not, and so on. Depending on the assumptions made, sequential search may make computations quite complicated; to make things relatively accessible, they assume that consumers search with replacement and, more importantly, that consumers know the distribution of realized utilities before any search is actually conducted.[7]

In their model, there are N firms offering vertically differentiated products. All firms produce with a constant returns to scale technology but firms' unit costs need not be the same. They assume that observed prices and quantities are the outcome of Nash equilibrium strategies: that is, an individual firm sets its price to maximize profits, taking other firms' prices as well as consumer search behavior as given. Likewise, consumers continue searching till the expected gains from search fall short of search costs.

Let the consumer's utility from buying good j at price p_j be

$$u_j = W_j\beta - p_j + \xi_j \tag{16}$$

where W_j denotes a vector of observable product attributes and ξ_j denotes an unobservable (by the econometrician) attribute.

A consumer's decision is a stopping rule. Consider a consumer i who has found so far

[7] This assumption is somewhat dubious because it assumes that consumers have a lot of information without having yet engaged in actual search.

a good providing utility u^*. Let $H(u)$ be the distribution of product utilities, with support $[\underline{u}, \overline{u}]$. Then buyer i's optimal search rule implies to continue searching as long as

$$c_i \leq \int_{u^*}^{\overline{u}} (u - u^*) dH(u) \tag{17}$$

Assuming that consumers know the *empirical* distribution of utilities offered in the market, we can, without loss of generality, order the utilities offered by the different goods as follows: $u_1 < u_2 < u_2 < ... < u_N$. Then the empirical distribution of product utilities is

$$H(u) = \frac{1}{N} \sum_{j=1}^{N} I\{u_j \leq u\} \tag{18}$$

Consider a consumer who has sampled product j. This consumer will continue searching if $c_j \leq \sum_{k=j}^{N} k(u_k - u_j)/N$. Therefore we can find the search cost of that consumer who is indifferent between stopping searching and searching further once he/she has found product j. This reasoning gives us N cut-off points of the search cost distribution

$$c_j = \sum_{k=j}^{N} \frac{1}{N}(u_k - u_j), \ j = 1, 2, ..., N \tag{19}$$

If the econometrician observed utilities, we could compute these cutoff points directly. However, in general, the researcher ignores utilities.[8]

In what follows we show how Hortacsu and Syverson use prices and market shares to estimate the search cost distribution. First, we show how quantities identify the height of the search cost distribution at the critical points $\{c_j\}_{j=1}^{N}$; then, we describe how to estimate these critical points.

Consider the firm offering the lowest utility product: product 1. The market share of this firm is given by those consumers who happen to find product 1 at first and who happen

[8]There is one case where we can use this equation to calculate the cutoff points directly: the homogeneous goods case. In that case utility from product j equals $v - p_j$ so we have $c_j = \sum_{k=j}^{N} \frac{1}{N}(p_k - p_j)$ and then we can calculate the c_j's from observed prices.

not to find worthwhile to continue searching. Therefore:

$$q_1 = \frac{1}{N}(1 - G(c_1)) \tag{20}$$

For product 2 demand is made up of consumers whose search cost is above c_1 and find product 2 in their first search, plus consumers whose search cost lies in between c_2 and c_1 and happen to visit firm 2 at first, or visit firm 2 after sampling first firm 1, or visit firm 2 after sampling firm 1 two times, or three times and so on. Therefore:

$$q_2 = \frac{1}{N}\left[1 - G(c_1) + \left(1 + \frac{1}{N} + \frac{1}{N}\frac{1}{N} + ...\right)(G(c_1) - G(c_2))\right] \tag{21}$$

$$= \frac{1}{N}\left[1 + \frac{\frac{1}{N}G(c_1)}{1-\frac{1}{N}} - \frac{G(c_2)}{1-\frac{1}{N}}\right] \tag{22}$$

And so on and so forth, for product j we get:

$$q_j = \frac{1}{N}\left[1 + \frac{\frac{1}{N}G(c_1)}{1-\frac{1}{N}} + \frac{\frac{1}{N}G(c_2)}{(1-\frac{1}{N})(1-\frac{2}{N})} + \sum_{k=3}^{j-1}\frac{\frac{1}{N}G(c_k)}{(1-\frac{k-1}{N})(1-\frac{k}{N})} - \frac{G(c_j)}{(1-\frac{j-1}{N})}\right] \tag{23}$$

$j = 3, 4, ..., N$.

If we observe market shares $q_1, q_2, ..., q_N$ we can use the above linear system of equations to estimate (by OLS) the heights of the search cost distribution evaluated at the cutoff points c_j's, i.e., $\{G(c_1), G(c_2), ..., G(c_{N-1})\}$.

Due to the fact that we have truly differentiated products, we need to use firms' optimal pricing decisions to identify the cutoff points c_j's. Taking the first order conditions, we get

$$q_j(\cdot) + (p_j - mc)\frac{\partial q_j}{\partial p_j} = 0 \tag{24}$$

where

$$\frac{\partial q_j}{\partial p_j} = -\frac{(\frac{1}{N})^3 g(c_1)}{1-\frac{1}{N}} - \frac{(\frac{1}{N})^3 g(c_2)}{(1-\frac{1}{N})(1-\frac{2}{N})} - \sum_{k=3}^{j-1}\frac{(\frac{1}{N})^3 g(c_k)}{(1-\frac{k-1}{N})(1-\frac{k}{N})} - \frac{(\frac{1}{N})^{N-j+1} g(c_j)}{1-\frac{j-1}{N}} \tag{25}$$

$j = 1, 2, ..., N$.

Assuming that we know the unit cost of the firms, and assuming a value for $g(c_N) = g(0)$ (the density at zero search costs), the first order conditions in (24) constitute a linear system of equations that can be used to estimate (by OLS) the density of the search cost distribution evaluated at the cutoff points c_j's, i.e., $\{g(c_1), g(c_2), ..., g(c_{N-1})\}$.

Once we have the cdf and pdf values $G(c_j)$'s and $g(c_j)$'s, we can calculate the cutoff points c_j's as follows. Note that

$$G(c_{j-1}) - G(c_j) = \int_{c_j}^{c_{j-1}} g(c)dc \tag{26}$$

which can be approximated by the trapezoid method:

$$G(c_{j-1}) - G(c_j) = \frac{1}{2}[g(c_{j-1}) + g(c_j)](c_{j-1} - c_j) \tag{27}$$

so we get

$$c_{j-1} - c_j = \frac{2G(c_{j-1}) - G(c_j)}{g(c_{j-1}) + g(c_j)} \tag{28}$$

Once we have got the cutoff points c_j's and the heights $G(c_j)$'s, we can construct an estimate of the search cost distribution by spline interpolation.

Interestingly, after we have obtained the cutoff points c_j's, using the expression

$$c_j = \sum_{k=j}^{N} \frac{1}{N}(u_k - u_j), \; j = 1, 2, ..., N \tag{29}$$

we can estimate non-parametrically the utilities u_j's. Finally, using the estimate of the utilities, the coefficients of the observable characteristics can be estimated from the regression:

$$u_j + p_j = W_j\beta + \xi_j \tag{30}$$

Hortacsu and Syverson (2004) apply the estimation methods described above to a data set of prices and quantities for S&P 500 index funds. Price dispersion in the mutual fund industry and in the S&P 500 index funds in particular is significant; for example, in 2000 the highest-price fund charged annualized fees nearly 30 times higher than the lowest-price fund.

As argued by Hortacsu and Syverson, all funds in this sector try to imitate the financial performance of the S&P 500 index. As a result, price dispersion should not be the result of portfolio differentiation but differentiation in other attributes.

Their results confirm that relatively low search costs and vertical product characteristics play an important role in explaining the large variation in observed prices. Further, their estimates of the search cost distributions suggest that consumers' search costs have changed over time, in particular, search costs have fallen in the lower three quartiles of the distribution and increased in the upper quartile during the period 1996-2000. Hortacsu and Syverson argue that this might have been due to the change in the relative composition of demand driven by entry of many novice mutual fund investors during the period under study.

4 Conclusions

Theoretical work on consumer search suggests that meaningful competition policy measures (such as the challenge of a merger, the introduction of price controls or, more generally, the liberalization of a market) in environments where search costs are important must be based on accurate and reliable information on search costs. This leads to the question how to identify and estimate consumer search costs.

This paper has surveyed the recent work on search cost estimation using demand and supply equilibrium models. An important feature of estimated demand and supply models is that one can easily simulate counterfactuals, which helps design optimal competition policy measures. It has been shown that in homogeneous product markets, price information suffices to estimate search costs. In markets with (vertical) product differentiation, by contrast, a combination of price and quantity data are required. In spite of the fact that the models discussed here are relatively simple, they seem to explain the data relatively well.

Firm interaction in real-world markets involves a number of interesting features not accounted for in the models considered here. For example, in real-world markets we observe marginal cost heterogeneity, heterogeneous consumer valuations, multi-product firms, dynamic interaction, etc. Further research should extend the models presented here in those directions.

References

[1] Daron Acemoglu and Robert Shimer: "Wage and Technology Dispersion," *Review of Economic Studies* 67, 585-608, 2000.

[2] Simon Anderson and Regis Renault: "Pricing, Product Diversity and Search Costs: A Bertrand-Chamberlin-Diamond Model," *Rand Journal of Economics* 30-4, 719-735, 1999.

[3] Joseph Bailey: "Electronic Commerce: Prices and Consumer Issues for three Products: Books, Compact Discs, and Software," Organization for Economic Co-operation and Development, OCDE/GD(98)4, 1998.

[4] Michael R. Baye and John Morgan: "Price Dispersion in the Lab and on the Internet: Theory and Evidence," *Rand Journal of Economics* 35, 448-66, 2004.

[5] Steve Berry, Jim Levinsohn and Ariel Pakes: "Automobile Prices in Market Equilibrium," *Econometrica* 63, 841-90, 1995.

[6] Helmut Bester: "Price Commitment in Search Markets," *Journal of Economic Behavior and Organization* 25, 109-120, 1994.

[7] Jeffrey Brown and Austan Goolsbee: "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy* 110, 481-507, 2002.

[8] Erik Brynjolfsson and Michael Smith: "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science* 46, 563-85, 2000.

[9] Kenneth Burdett and Kenneth L. Judd: "Equilibrium Price Dispersion," *Econometrica* 51, 955-69, 1983.

[10] Kenneth Burdett and Dale Mortensen: "Wage Differentials, Employer Size and Unemployment, *International Economic Review*, 39, 257-73, 1998.

[11] John A. Carlson and Donn R. Pescatrice: "Persistent Price Distributions," *Journal of Economics and Business* 33-1, 21-27, 1980.

[12] Karen Clay, Ramayya Krishnan and Eric Wolff: "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics* 49, 521-39, 2001.

[13] Bev Dahlby and Douglas S. West: "Price Dispersion in an Automobile Insurance Market," *Journal of Political Economy* 94-2, 418-438, 1986.

[14] Peter Diamond: "A Model of Price Adjustment," *Journal of Economic Theory* 3, 156-68, 1971.

[15] Stephen G. Donald and Harry J. Paarsch: "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review* 34, 121-48, 1993.

[16] Chaim Fershtman and Arthur Fishman: "Price Cycles and Booms: Dynamic Search Equilibrium," *American Economic Review* 82, 1221-33, 1992.

[17] Richard Friberg, Mattias Ganslandt and Mikael Sandstrom: "E-commerce and Prices: Theory and Evidence," SSE/EFI Working Paper Series in Economics and Finance No. 389, Stockholm School of Economics, 2000.

[18] Han Hong and Matthew Shum: "Using Price Distributions to Estimate Search Costs," *Rand Journal of Economics*, forthcoming.

[19] Theresa van Hoomissen: "Price Dispersion and Inflation: Evidence from Israel," *Journal of Political Economy* 96-6, 1303-14, 1988.

[20] Ali Hortaçsu and Chad Syverson: "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: a Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics* 119, 403-56, 2004.

[21] Maarten C. W. Janssen and José Luis Moraga-González: "Strategic Pricing, Consumer Search and the Number of Firms," *Review of Economic Studies* 71, 1089-118, 2004.

[22] Maarten C. W. Janssen, José Luis Moraga-González and Matthijs Wildenbeest: "Truly Costly Sequential Search and Oligopolistic Pricing," *International Journal of Industrial Organization* 23, 451-466, 2005.

[23] Nicholas M. Kiefer and George R. Neumann: "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model," in: Henning Bunzel, Peter Jensen and Niels Westergård-Nielsen (Editors), *Panel Data and Labour Market Dynamics*, Amsterdam: North-Holland, 1993.

[24] Saul Lach: "Existence and Persistence of Price Dispersion: An Empirical Analysis," *Review of Economics and Statistics* 84, 433-444, 2002.

[25] M. Lewis: "Asymmetric Price Adjustment and Consumer Search: An Examination of teh Retail Gasoline Market," Discussion Paper 0407010, Center for the Study of Energy Markets working paper series, 2003.

[26] John McMillan and Michael Rothschild: "Search," *Handbook of game theory with economic applications,* vol. 2., 905-27, 1994.

[27] Peter Morgan and Richard Manning: "Optimal Search," *Econometrica* 53, 923-44, 1985.

[28] John W. Pratt, David A. Wise and Richard J. Zeckhauser: "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics* 93, 189-211, 1979.

[29] Jeniffer F. Reinganum: "A Simple Model of Equilibrium Price Dispersion," *Journal of Political Economy* 87, 851-58, 1979.

[30] Robert W. Rosenthal: "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica* 48, 1575-79, 1980.

[31] Michael Rothschild: "Models of Market Organization with Imperfect Information: A Survey," *Journal of Political Economy* 81-6, 1283-1308, 1973.

[32] Steven. C. Salop and Joseph. E. Stiglitz, "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents", *American Economic Review* 79, 1121-1130, 1982.

[33] Alan T. Sorensen: "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy* 108, 833-50, 2000.

[34] Alan T. Sorensen: "Price Dispersion and Heterogeneous Consumer Search for Retail Prescription Drugs," NBER Working Paper w8548, 2001.

[35] Daniel F. Spulber: "Bertrand Competition When Rivals' Costs Are Unknown," *Journal of Industrial Economics* 43-1, 1-11, 1995.

[36] Dale O. Stahl: "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review* 79, 700-12, 1989.

[37] George Stigler: "The Economics of Information," *Journal of Political Economy*, 69, 213-25, 1961.

[38] Hal R. Varian: "A Model of Sales," *American Economic Review* 70, 651-59, 1980.

[39] Asher Wolinsky: "Product Differentiation with Imperfect Information," *Review of Economic Studies*, 51-1, 53-61, 1984.

[40] Asher Wolinsky: "Procurement via Sequential Search," *Journal of Political Economy* 113-4, 785-810, 2005.