TI 2006-019/1
Tinbergen Institute Discussion Paper

Maximum Likelihood Estimation of Search Costs

José Luis Moraga-González[1]
Matthijs R. Wildenbeest[2]

[1] *University of Groningen, Erasmus University Rotterdam, and Tinbergen Institute;*
[2] *Erasmus University Rotterdam, and Tinbergen Institute.*

Maximum Likelihood Estimation of Search Costs[*]

José Luis Moraga-González[†]
Matthijs R. Wildenbeest[‡]

January 2006

Abstract

In a recent paper Hong and Shum (forthcoming) present a structural methodology to estimate search cost distributions. We extend their approach to the case of oligopoly and present a maximum likelihood estimate of the search cost distribution. We apply our method to a data set of online prices for different computer memory chips. The estimates of the search cost distribution suggest that consumers have either quite high or quite low search costs so they either search for all prices in the market or for at most three prices. According to Kolmogorov-Smirnov goodness-of-fit tests, we cannot reject the null hypothesis that the observed prices are generated by the model.

Keywords: consumer search, oligopoly, price dispersion, structural estimation, maximum likelihood

JEL Classification: C14, D43, D83, L13

1   Introduction

There is substantial evidence that prices of seemingly homogeneous consumer goods are quite dispersed across stores (see e.g. Stigler, 1961; Dahlby and West, 1986; Pratt, Wise and Zeckhauser, 1979; Sorensen 2000; Brown and Goolsbee, 2002; Lach, 2002; Baye, Morgan and Scholten, 2004). During the last twenty five years, economists have dedicated significant theoretical effort to explain this empirical regularity as an equilibrium phenomenon. Price dispersion can be sustained in equilibrium when some consumers observe several prices while other consumers observe only one price. Such unequal distribution of price information across consumers often arises in the market as a result of costly search (see e.g. Burdett and Judd, 1983; Rob, 1985; Stahl, 1989; Varian, 1980).

In spite of this substantial theoretical effort, somewhat surprisingly, very little empirical work has focused on identifying and measuring search costs in real-world markets. From an applied point of view, this is certainly an omission because the predictions of the various theoretical models are quite sensitive to the nature of search costs in the market. Since competition policy recommendations may depend on the nature of search costs, there is a need to develop methods to quantify the search cost distribution.

In a recent paper Hong and Shum (forthcoming) develop a structural methodology to retrieve information on search costs in a market for homogeneous goods. They show that firm and consumer equilibrium behavior imposes enough structure on the data to allow for the estimation of search cost parameters using only observed prices. 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 product differentiation.

In the nonsequential search model studied by Hong and Shum (forthcoming), infinitely many firms operate in the market by setting prices. Heterogeneous search cost consumers search to discover prices and buy from the cheapest firm they observe. Some consumers do not search whereas others do and this leads to price dispersion. Hong and Shum formulate the estimation of the unknown search cost distribution as a two-step procedure. They first estimate the parameters of the equilibrium price distribution by *maximum empirical likelihood* (MEL). To do this, they exploit equilibrium behavior, which provides the econometrician

with an infinite number of moment conditions. The estimates of the parameters give the height of the search cost distribution evaluated at a series of cut-off points. Then, these cut-off points are estimated directly from the empirical cdf of prices. In practice, however, only a few parameters of the price distribution are estimated, which implies that low search cost consumers are ignored.

In this paper we introduce two novelties. First, we extend Hong and Shum's approach to the oligopoly case. Vis-à-vis the competitive case, the oligopoly model is far more useful for competition policy issues. In addition, the oligopoly model has the advantage that if one observes N firms operating in a market, then we know consumers will search up to a maximum of N prices. As a result we can estimate the N relevant cut-off points of the search cost distribution so we also learn about the distribution of search costs at relatively low quantiles, that is, we find out how much search is conducted by consumers who search thoroughly and how low their search costs are.[1]

Second, we present a *maximum likelihood* (ML) estimate of the search cost distribution. To do this, we first estimate the parameters of the price distribution by maximum likelihood and then introduce a method to calculate the cut-off points of the search cost distribution as a function of the ML estimate of the price distribution. In this way, the estimates of the cut-off points are also ML so the asymptotic theory for computing standard errors of and for conducting tests of hypotheses on the estimates of the search cost distribution remains standard.

To implement our method we only need to observe the prices firms charge over some period of market interaction. We perform Monte Carlo simulations and show that ignoring low search cost consumers leads to a large bias in the estimates: search costs are substantially overestimated and price-cost margins largely exaggerated. These biases result in a poor fit of the model to the data and goodness-of-fit tests reject the null hypothesis that the empirical and the estimated distribution of prices are equal.

If the fraction of low search cost consumers were negligible in real-world markets, this

---

[1]Brown and Goolsbee (2002) argue that prices of life insurance policies did not fall with rising Internet usage (which probably meant an upward shift of the search cost distribution) but with the emergence of price comparison sites (which most likely meant a dramatic change of the shape of the distribution). Picking up such an effect requires information on how the Internet has affected search costs for all quantiles.

would not be a problem. However, it turns out that the fraction of consumers searching intensively in real-world markets is sizable. We apply our method to a data set of prices for four personal computer memory chips. For all the products, we observe significant price dispersion as measured by the coefficient of variation. 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. Our 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.[2]

Consumer search behavior confers substantial market power to the firms. In spite of the fact that in each of the markets studied there are more than 20 stores operating, we estimate that the average price-cost margin ranges between 23% and 28%. This suggests that demand side characteristics might be even more important than market structure to assess market competitiveness (Waterson, 2003).

The validity of the theoretical model is tested by conducting Kolmogorov-Smirnov tests of the goodness of fit. According to the test results, we cannot reject the null hypothesis that the observed prices are generated by the model.

The structure of the paper is as follows. In the next section, we review and modify the nonsequential consumer search model studied in the paper of Hong and Shum. In Section 3 we discuss our maximum likelihood estimation method. In Section 4, we present a Monte Carlo study whereas in Section 5 we estimate the search cost distribution underlying price data from some online markets for memory chips. Finally, Section 6 concludes.

---

[2]Using a different methodology, Sorensen (2001) finds that between 5% to 10% of the consumers conduct an exhaustive search for prices in the market for prescription drugs.

2 The consumer search model

We study an oligopolistic version of the model proposed in Hong and Shum (forthcoming); their model generalizes the nonsequential consumer search model of Burdett and Judd (1983) by adding consumer search cost heterogeneity.[3] Assume there are N retailers selling a homogeneous good. Let r be the common unit selling cost of each retailer. There is a unit mass of identical buyers. Each consumer inelastically demands one unit of the good. Let $\bar{p}$ be the consumer valuation. Beyond the first price, a consumer incurs a search cost c to obtain further price information. Consumers differ in their search costs. Assume that the cost of a consumer is randomly drawn from a distribution of search costs F_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, with density $f_p(p)$. Let $\underline{p}$ and $\bar{p}$ be the lower and upper bound of the support of F_p.[4] Given firm behavior, the number of prices $i(c)$ a consumer with search cost c observes must be optimal, i.e.,

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

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

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

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

---

[3]The oligopoly case is also studied in Janssen and Moraga-González (2004) but with a two-point search cost distribution.

[4]It will become clear later that the upper bound of the price distribution must be equal to the consumer valuation.

q_i sampling i prices are simply

$$q_1 \;=\; 1 - F_c(\Delta_1); \tag{4a}$$

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

$$q_N \;=\; F_c(\Delta_{N-1}). \tag{4c}$$

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, i.e.,

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

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

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

As shown in Hong and Shum (forthcoming), equations (2) to (6) impose enough structure on the data to allow for the estimation of the search cost distribution using only price data. Since quantity information is often hard to obtain, this will also be the focus of our next section; at the end of it, however, we discuss how to use quantity data when it is available to the researcher.

3   Maximum likelihood estimation

Assume the researcher observes the prices of the N firms operating in the market.[5] The objective is to estimate the collection of points $\{\Delta_i, q_i\}_{i=1}^{N}$ of the search cost distribution by maximum likelihood. Once we get these estimates we can construct an estimate of the search cost distribution by spline approximation. A difficulty here is that equation (5) cannot be

---

[5]In practice, sometimes not all the firms are observed by the researcher; our Monte Carlo study in Section 4 examines the implication of this lack of data.

solved for the equilibrium price distribution F_p and this makes it difficult to calculate

$$\Delta_i = \int_{\underline{p}}^{\overline{p}} p[(i+1)F_p(p) - 1](1 - F_p(p))^{i-1} f_p(p) dp, \ i = 1, 2, \ldots, N - 1.$$

Hong and Shum (forthcoming) propose to use the empirical price distribution to calculate Δ_i. Even though this approach is practical, it does not necessarily provide minimal variance estimates. We proceed differently and obtain ML estimates of the cut-off points. To do this, we rewrite Δ_i as a function of the ML estimates of the parameters of the price distribution. This has the advantage that the asymptotic theory for computing the standard errors of Δ_i and for conducting tests of hypotheses remains standard.[6] We first rewrite the cut-off points as (by integration by parts)

$$\Delta_i = \int_{\underline{p}}^{\overline{p}} F_p(p)(1 - F_p(p))^i dp, \ i = 1, 2, \ldots, N - 1. \tag{7}$$

Since the distribution function $F(p)$ is monotonically increasing in p, its inverse exists. Using equation (6) we can find the inverse function:

$$p(z) = \frac{q_1(\overline{p} - r)}{\sum_{i=1}^{N} i q_i (1 - z)^{i-1}} + r. \tag{8}$$

Using this inverse function, a change of variables in equation (7) yields:

$$\Delta_i = \int_0^1 p(z)[(i+1)z - 1](1 - z)^{i-1} dz, \ i = 1, 2, \ldots, N - 1. \tag{9}$$

If we obtain ML estimates of $r, \overline{p}, \underline{p}$ and $q_i, i = 1, 2, \ldots, N$, then we can use equations (8) and (9) to calculate ML estimates of the cut-off points of the search distribution so we obtain and 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 the researcher has only price data, which will often be the situation. Since the price density cannot be obtained in closed form, we apply the implicit function theorem

---

[6]Note that for this we need that prices are independently and identically distributed in different periods, that the number of firms is fixed and finite and that the number of periods goes to infinity.

to equation (5), which yields

$$f_p(p) = \frac{\sum_{i=1}^{N} i q_i (1 - F_p(p))^{i-1}}{(p - r) \sum_{i=1}^{N} i(i-1) q_i (1 - F_p(p))^{i-2}}. \tag{10}$$

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.[7] 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:

$$r = \frac{p_1 \sum_{i=1}^{N} i q_i - q_1 p_M}{\sum_{i=2}^{N} i q_i}. \tag{11}$$

We then solve the maximum likelihood estimation problem

$$\max_{\{q_i\}_{i=1}^{N}} \sum_{\ell=2}^{M-1} \log f_p(p_\ell; q_1, q_2, ..., q_N)$$
$$\text{s.t.} \quad r = \frac{p_1 \sum_{i=1}^{N} i q_i - q_1 p_M}{\sum_{i=2}^{N} i q_i} \text{ and } \sum_{i=1}^{N} q_i = 1,$$

numerically, which yields the desired ML estimates. We note that in this formulation the estimate of r is obtained from equation (11) as a function of the estimates of the other parameters. This procedure introduces some dependence between the price observations. Our Monte Carlo study next shows that this approach works reasonably well, as the upper and lower bounds of the price distribution converge to the true values at a super-consistent rate.

The standard errors of the estimates of q_i, $i = 1, 2, \ldots, N-1$ are calculated in the standard way, i.e., by taking the square root of the diagonal entries of the inverse of the negative Hessian matrix evaluated at the optimum. Since $q_N = 1 - \sum_{i=1}^{N-1} q_i$, we can calculate the standard error of the estimate of q_N using the Delta method. The same applies to the standard errors of the estimates of the marginal cost r and the Δ_i's since they are obtained as linear transformations of the estimated q_i's.

---

[7]See also Donald and Paarsch (1993) on using order statistics to estimate the lower and upper bound of bid distributions.

Assume now that the researcher is in the possession of quantity data as well. Then we can write the expected quantities sold in equilibrium $s(p)$ as a function of prices, i.e.,

$$s(p) = \sum_{i=1}^{N} \frac{iq_i}{N}(1 - F_p(p))^{i-1}. \tag{12}$$

The only unknowns in this equation are the fractions of consumers q_i, $i = 1, 2, \ldots, N$. Since the parameters q_i's enter equation (12) linearly, we can estimate them by Ordinary Least Squares (OLS). Once the q_i's are estimated we can proceed as before and, using the lowest and highest prices as estimates of the bounds of the price distribution, retrieve an estimate of r from equation (11). Finally, the cut-off points of the search cost distribution follow from equations (8) and (9).

4   A Monte Carlo study

The study in this section has various purposes. First, we investigate how the maximum likelihood estimates of the price distribution and of the search cost distribution behave when the number of price observations is limited. In particular we are interested in the type of bias introduced by the estimation of the upper and lower bound of the price distribution by the maximum and the minimum prices observed in the data. Secondly, we investigate the impact of some measurement error in the number of firms N operating in the market. Finally, we study how the estimates of the search cost distribution change when low search cost consumers are ignored. To implement this idea we study a market operated by a relatively large number of firms and assume that consumers conduct only a limited search for prices, say, for 1, 2, 3 or for 4 prices at most.

The general setup of the Monte Carlo experiment is as follows. We consider a market operated by $N = 25$ firms. Market interaction evolves over a finite number of T periods.[8] We assume that consumers' search costs are drawn independently from a log-normal distribution with mean 0.5 and standard deviation 5. Further, the value of the product $\bar{p}$ is assumed to be 100 and the unit cost r to be 50. For our first simulations we set $T = 4$ so we have a

---

[8]The equilibrium of the static game described in Section 2 is also equilibrium of the repeated game with finite horizon.

total of 100 price observations.

To solve for equilibrium, we compute numerically the fractions $\{q_1, q_2, \ldots, q_N\}$ for which equations (4a)-(4c) and (9) hold simultaneously. Next, we use these parameters to construct the equilibrium price distribution, implicitly defined by equation (5). After this, we draw a random sample from the cdf of prices, which serves as an input for the maximum likelihood estimation procedure described in the previous section. We replicate the experiment 1,000 times and report the mean and the 90% confidence interval of the estimates.

The first column of Table 1 gives the true parameter values. We see that equilibrium price dispersion is substantial. In particular, the lowest price of the equilibrium price distribution is 51.68, which is about half the maximum price, thus providing large gains from search. A firm charging the minimum price has a price-cost margin of only 3.3% while a firm charging the maximum price has a price-cost margin of 50%.

	TRUE	EXP1	EXP2	EXP3
T	-	4	10	20
N	25	25	25	25
M	-	100	250	500
r	50.00	47.50 (6.53)	49.16 (2.03)	49.47 (1.31)
$\underline{p}$	51.68	52.06 (0.42)	51.83 (0.16)	51.75 (0.07)
v	100.00	99.91 (0.09)	99.97 (0.03)	99.98 (0.02)
q_1	0.380	0.451 (0.154)	0.417 (0.098)	0.405 (0.072)
q_2	0.032	0.037 (0.017)	0.035 (0.012)	0.035 (0.009)
q_3	0.026	0.025 (0.032)	0.024 (0.026)	0.025 (0.022)
q_4	0.022	0.025 (0.041)	0.024 (0.033)	0.027 (0.032)
q_5	0.020	0.022 (0.042)	0.027 (0.042)	0.022 (0.034)
q_6	0.018	0.022 (0.046)	0.018 (0.036)	0.016 (0.030)
q_7	0.016	0.018 (0.046)	0.015 (0.035)	0.016 (0.032)
q_8	0.015	0.011 (0.038)	0.015 (0.036)	0.016 (0.034)
q_9	0.014	0.013 (0.046)	0.012 (0.033)	0.013 (0.032)
q_{10}	0.013	0.013 (0.049)	0.014 (0.041)	0.013 (0.033)
q_{11}	0.013	0.011 (0.051)	0.010 (0.035)	0.011 (0.031)
q_{12}	0.012	0.013 (0.054)	0.011 (0.037)	0.011 (0.033)
q_{13}	0.011	0.010 (0.044)	0.010 (0.032)	0.009 (0.027)
q_{14}	0.011	0.009 (0.043)	0.011 (0.036)	0.010 (0.032)
q_{15}	0.010	0.013 (0.052)	0.017 (0.055)	0.010 (0.033)
q_{16}	0.010	0.011 (0.045)	0.015 (0.046)	0.013 (0.042)
q_{17}	0.009	0.012 (0.057)	0.013 (0.043)	0.014 (0.041)
q_{18}	0.009	0.016 (0.072)	0.014 (0.044)	0.017 (0.045)
q_{19}	0.008	0.012 (0.053)	0.014 (0.044)	0.019 (0.050)
q_{20}	0.008	0.011 (0.052)	0.017 (0.053)	0.020 (0.052)
q_{21}	0.008	0.010 (0.049)	0.017 (0.051)	0.022 (0.056)
q_{22}	0.007	0.011 (0.058)	0.016 (0.050)	0.021 (0.054)
q_{23}	0.007	0.012 (0.069)	0.018 (0.056)	0.018 (0.050)
q_{24}	0.007	0.012 (0.060)	0.029 (0.066)	0.019 (0.047)
q_{25}	0.314	0.199 (0.216)	0.186 (0.175)	0.197 (0.182)

Notes:
Standard errors in parenthesis.

Table 1: True and estimated parameter values

In equilibrium most of the consumers search either for only one price (about 38% of the buyers) or for all the prices in the market (about 31% of the consumers). The fractions of consumers searching for an intermediate number of prices (from 2 to 24 firms) are pretty small, in all cases less than about 3% and often close to zero.[9]

The second column of Table 1 gives the results of our first set of experiments. The numbers reported are the mean of the estimates of the parameters with corresponding standard errors in parenthesis. We observe that the estimate of the fraction of consumers who search for one price only is about 45% and highly significant. This estimate is about 7% higher than the true value so the fraction of consumers who do not compare prices at all is overestimated. The estimate of the fraction of consumers searching for two prices is also significant and again overestimated. The estimates of the rest of the parameters are not significantly different from zero (at the 5% level). However since the true parameters are close to zero, this does not have an important impact on the estimates of the price and the search cost distribution. The estimate of the fraction of consumers searching for all prices in the market is about 20%, somewhat lower than the true parameter (31.4%). In sum we see that the fractions of consumers searching little are overestimated while the fractions of consumers searching a lot are underestimated. Arguably the implication of these biases is that the estimate of the search cost distribution will be biased towards high search costs.

The first column of Table 2 reports the true cut-off points of the search cost distribution. The second column gives the estimated ones. We see that all the estimates of the cut-off points are highly significant, and quite close to the true ones.

The estimates of the equilibrium price distribution and of the search cost distribution are plotted in Figure 1(a) and 1(b) respectively. In both graphs the black curves are the true distributions while the dashed-black curves show the mean of the 1,000 estimated distributions. The dashed red curves are respectively the 5% percentile and the 95% percentile of the estimates. We observe that the estimate of the price distribution is remarkably close to the true price distribution. However, the estimate of the search cost distribution lies below

---

[9]Of course, this partition of the set of consumers in equilibrium has to do with the choice of search cost distribution. For example, a mixture of a log-normal (e.g. with parameters 0.5 and 0.2) and a gamma distribution (with parameters 0.5 and 0.2) leads to a very different equilibrium, with most consumers searching intensively and very few consumers not searching at all.

	TRUE	EXP1	EXP2	EXP3
T	-	4	10	20
N	25	25	25	25
M	-	100	250	500
Δ_1	7.60	7.40 (0.63)	7.54 (0.40)	7.57 (0.28)
Δ_2	5.01	4.85 (0.34)	4.95 (0.21)	4.98 (0.15)
Δ_3	3.59	3.47 (0.21)	3.55 (0.13)	3.57 (0.09)
Δ_4	2.71	2.62 (0.16)	2.68 (0.09)	2.69 (0.07)
Δ_5	2.12	2.04 (0.13)	2.09 (0.08)	2.10 (0.06)
Δ_6	1.69	1.64 (0.12)	1.67 (0.07)	1.68 (0.05)
Δ_7	1.38	1.33 (0.11)	1.36 (0.07)	1.36 (0.05)
Δ_8	1.14	1.11 (0.11)	1.12 (0.07)	1.13 (0.05)
Δ_9	0.95	0.93 (0.10)	0.94 (0.07)	0.94 (0.05)
Δ_{10}	0.80	0.79 (0.10)	0.80 (0.06)	0.80 (0.05)
Δ_{11}	0.69	0.68 (0.09)	0.68 (0.06)	0.68 (0.04)
Δ_{12}	0.59	0.58 (0.09)	0.59 (0.06)	0.59 (0.04)
Δ_{13}	0.51	0.51 (0.08)	0.51 (0.05)	0.51 (0.04)
Δ_{14}	0.45	0.45 (0.08)	0.44 (0.05)	0.44 (0.04)
Δ_{15}	0.39	0.39 (0.07)	0.39 (0.05)	0.39 (0.04)
Δ_{16}	0.34	0.35 (0.07)	0.34 (0.04)	0.34 (0.03)
Δ_{17}	0.31	0.31 (0.06)	0.31 (0.04)	0.30 (0.03)
Δ_{18}	0.27	0.28 (0.06)	0.27 (0.04)	0.27 (0.03)
Δ_{19}	0.24	0.25 (0.06)	0.24 (0.04)	0.24 (0.03)
Δ_{20}	0.22	0.22 (0.05)	0.22 (0.04)	0.22 (0.03)
Δ_{21}	0.20	0.20 (0.05)	0.20 (0.03)	0.20 (0.02)
Δ_{22}	0.18	0.19 (0.05)	0.18 (0.03)	0.18 (0.02)
Δ_{23}	0.16	0.17 (0.05)	0.16 (0.03)	0.16 (0.02)
Δ_{24}	0.15	0.15 (0.04)	0.15 (0.03)	0.15 (0.02)

Notes:
Standard errors in parenthesis.

Table 2: True and estimated critical search cost values

the true distribution. In spite of this, the true distribution falls (for its most part) within the 90% confidence interval.[10]

The fact that the estimate of the search cost distribution lies below the true one might be a reflection of the fact that the fractions of consumers searching little are overestimated. Since the estimate of the upper bound of the price distribution is lower than the true one while the estimate of the lower bound of the price distribution is higher than the true one, the price distribution might very well be less dispersed than the true one. This implies that gains from search are lower than in equilibrium, which is consistent with estimates of the search cost distribution being biased towards higher search costs.

To see whether the downward bias of the estimated search cost distribution can be attributed to the biased estimation of the upper and lower bound of the price distribution we conducted the following experiment. We took the sample and set the upper and lower

---

[10]The last cut-off point of the search cost distribution we estimate is c_1 and therefore we do not have information about search costs beyond that point.

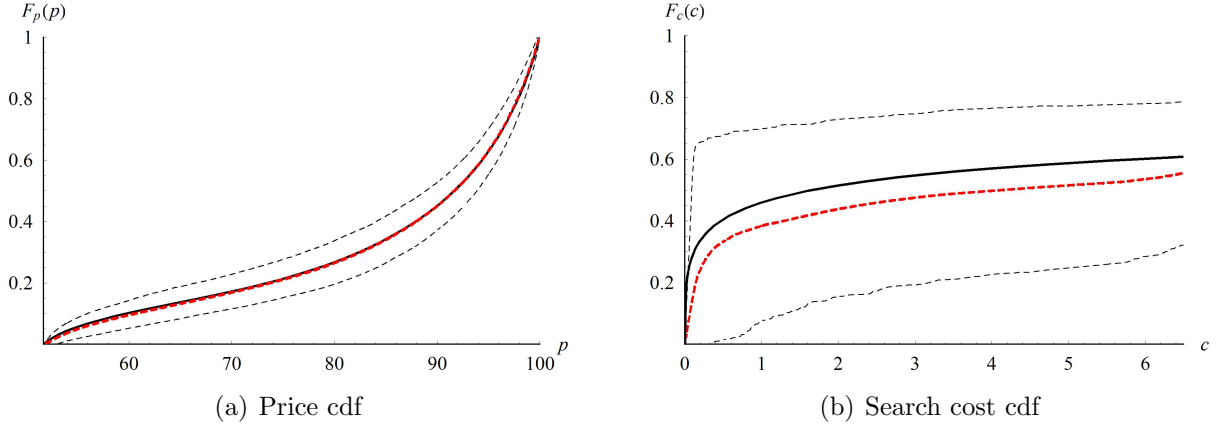(a) Price cdf   (b) Search cost cdf

Figure 1: Estimated equilibrium price and search cost cdf $M = 100$ ($T = 4$)

bound of the distribution equal to the true values and estimated the rest of parameters of the model. The dashed red curve in Figure 2(a) shows the estimated search cost distribution. To compare with the previous estimates, we also plot in gray the search cost distribution when the upper and the lower bounds are estimated by the minimum and the maximum price. The graph reveals that the new estimate is much closer to the true distribution than the previous one. The average estimate of q_1 goes down 0.04 points and this results in an upward shift of the estimated search cost distribution.

(a) $\underline{p}$ and $\bar{p}$ at true values   (b) N underestimated ($N = 20$)
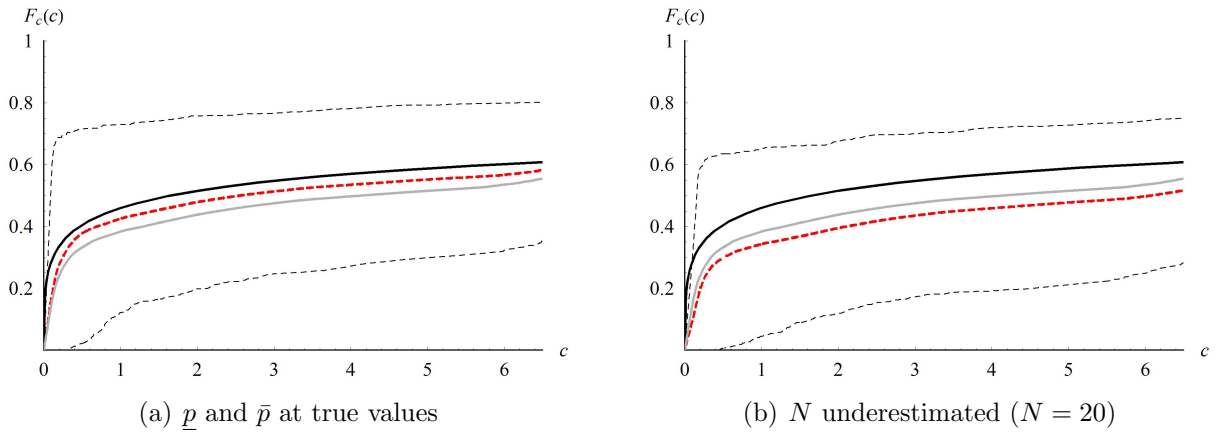
Figure 2: Estimated search cost cdf $M = 100$ ($T = 4$)

An alternative explanation is that the bias we observe is simply due to the fact that the maximum likelihood estimator is biased for finite samples. Indeed, this problem weakens as we increase the number of observations. This can be seen in columns 3 and 4 of Tables 1

and 2 where the number of periods over which the market develops is set equal to 10 and 20, respectively (the number of observations in each simulation goes up to 250 and 500). The Tables show that the estimates of the parameters of the price distribution (including the upper and lower bound) become more precise and this leads to more accurate estimates of the search cost distribution. The effect of an increase in the number of observations on the estimates can be seen in Figures 3(a) and 3(b).
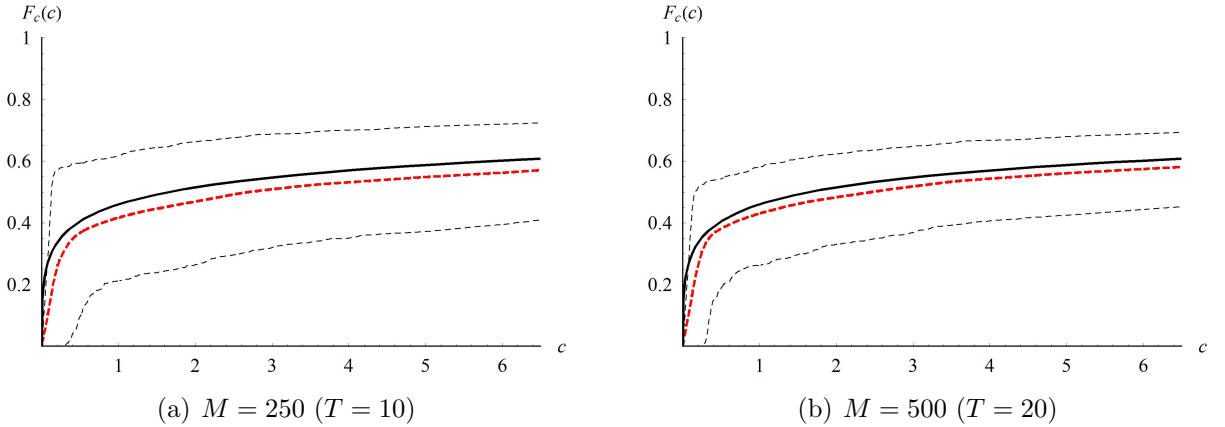

(a) $M = 250$ ($T = 10$)   (b) $M = 500$ ($T = 20$)

Figure 3: Estimated search cost cdf

To investigate the impact of measurement error in the number of firms N, we conducted an experiment in which we set N equal to 20 instead of equal to the true value 25. The results can be seen in Figure 2(b). As the graph shows, the underestimation of the number of firms did not change the shape of the search cost distribution. The average estimate of q_1 went up from 0.451 to 0.487 and this led to a greater downward bias in the estimate of the search cost distribution.

Finally, we study how the estimates of the search cost distribution change when low search cost consumers are ignored altogether. Implementing the idea that consumers conduct only a limited search for prices is equivalent to underestimating N. Figure 4 gives the estimation results when we set N equal to 4 instead of equal to 25. In Figure 4(a) we have plotted the estimated price cdf and the true empirical price distribution. Clearly, the fit is much worse than if we would have estimated N correctly (see Figure 1(a)). This has a large impact on the estimates of search cost and marginal cost parameters. As can be seen in Figure 4(b), the estimated search cost cdf is far from the true one. In particular, the estimates lead to the

$F_p(p)$

(a) price cdf
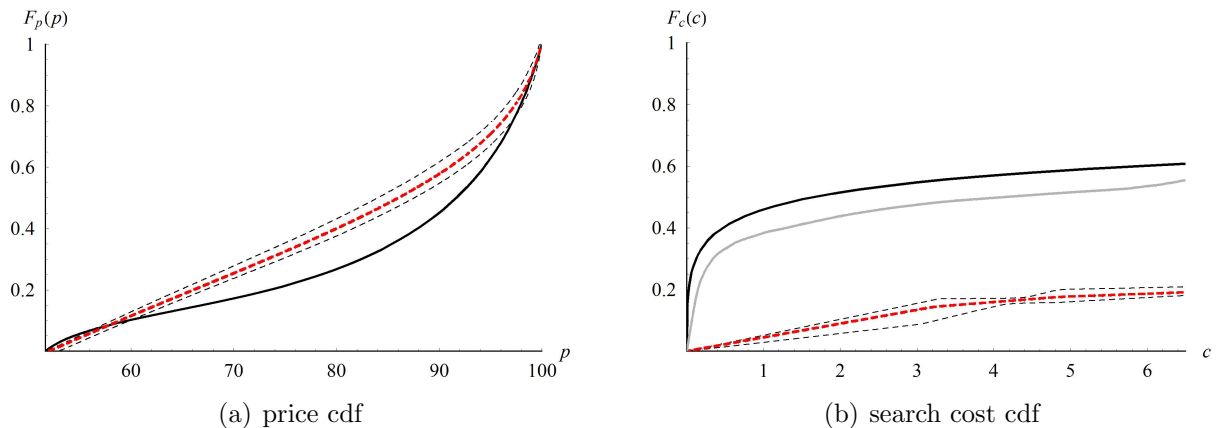
$F_c(c)$

(b) search cost cdf

Figure 4: Estimated equilibrium price and search cost cdf ($N = 4$) $M = 100$ ($T = 4$)

wrong conclusion that search costs are much higher than what they actually are. Likewise, this translates into an average price-cost margin being largely exaggerated; in particular around 100%, while the true price-cost margin of a typical firm is 42%. In their empirical section, Hong and Shum (forthcoming) estimate a similar model taking N much lower than observed in order to make their MEL estimation procedure work.[11] We conclude that this should be avoided.

5 Empirical application

5.1 Data and empirical issues

In this section we apply our estimation procedure to data obtained from real-world markets. Before presenting the results, we discuss the data set and, following Lach (2002), check one by one the assumptions of the theoretical model.

The focus is on the on-line markets for personal computer memory chips. Fitting the model in Section 2 to data from on-line markets assumes implicitly that consumers search for prices nonsequentially. This type of search is optimal in some circumstances, for example when several price quotations arrive at a time.[12] When consumers use websites, web-forums, and search engines to find price information, nonsequential search is a good approximation

---

[11]We note that our estimation procedure is quite robust. We observed neither problems of convergence nor sensitivity to starting values.

[12]For details on the optimality of nonsequential search see Morgan and Manning (1985).

of buyer behavior.

We selected four memory chips all manufactured by Kingston, which is by far the largest producer in the sector (the 2004 market share of Kingston was 27.0%, while the second biggest producer of memory chips Smart Modular Technologies had a 2004 market share of 8.1%). The details of these four products are given in Table 3.

Ellison and Ellison (2004) have pointed out that in this type of markets firms often engage in "bait and switch" strategies. We selected the memory chips to avoid this problem: chosen chips were at the moment of data collection somewhat at the top of the product line, exhibiting relatively large storage capacity (512 megabytes) and fast speed of operation (above 266 MHz). Two of the memory chips are of the SO-DIMM (Small Outline Dual In-line Memory Modules) type, which are intended for notebooks only.

It may be argued that different memory chips are in the same relevant market so a differentiated products market model is more appropriate than the model in Section 2. To avoid this problem to the extent possible, we included in the analysis only memory chips intended for particular PC's. In particular, we chose two memory chips for notebooks, one intended for Toshiba notebooks and the other for Dell Inspiron notebooks. Consumers who own for example a Toshiba Satellite 5105 notebook contemplating to extend its memory by 512 MB would most likely consider to buy only the Kingston KTT3614 memory chip (see www.toshiba.com).[13] The other two memory chips are intended for Dell desktop computers, in particular for the Dimension series.

The typical consumer in these markets is expected to buy a single chip to upgrade the memory capacity of his/her computer. Actually, often computers have just a single slot available for memory upgrades. As a result, the inelastic demand assumption of the model seems reasonable in these markets.

At the time of data collection, the four memory chips were sold in stores advertising prices on the price comparison site shopper.com. Shopper.com is one of the largest price comparison sites on the Internet. Internet shops get listed on shopper.com by subscribing to

---

[13]We are not experts in these markets but the information available at www.toshiba.com suggests that Kingston memory chips are the original parts used by Toshiba. For many consumers buying the same part as the original part is important (see Delgado and Waterson's (2003) study of the UK tyre market).

Product name	KTT3614	KTDINSP8200	KTD4400	KTD8300
Manufacturer	Kingston	Kingston	Kingston	Kingston
PC	Toshiba notebooks	Dell Inspiron 8200	Dell Dimension 4400	Dell Dimension 8300
MB	512MB DDR SDRAM	512MB DDR SDRAM	512MB DDR SDRAM	512MB DDR SDRAM
Memory Speed	PC2100 (266 MHz)	PC2100 (266 MHz)	PC2100 (266 MHz)	PC3200 (333 MHz)
Type	SO-DIMM	SO-DIMM	DIMM	DIMM

Notes:
SO-DIMM memory chips are for notebooks while DIMM memory chips are for desktop computers.

Table 3: List of products

CNET, the owner of shopper.com. Stores can choose between three types of listing schemes, general, preferred or premier. Preferred or premier listing allows a shop to add a store logo. Shops can provide once or twice a day price information by uploading a so-called price feed, but it is not necessary to do so if a shop does not desire to alter its price. The feed is collected four times a day and published on shopper.com. Shops are required to fill in eight fields in the feed: credit card price, manufacturer name, manufacturer Stock Keeping Unit (SKU), product URL, product name, availability, shipping and handling cost and category.

By using a so-called "spider" computer code, we automatically collected this information for the four memory chips directly from shopper.com, from the beginning of August 2004 till the end of September 2004. Unfortunately we could not collect more data because the IP address of the computer we were using to download the data was blocked by the system managers of shopper.com at the end of September 2004.

Product name	KTT3614	KTDINSP8200	KTD4400	KTD8300
Total No. of Stores	25	24	24	23
Mean No. of Stores (Min, Max)	22.4 (20, 24)	21.8 (19, 23)	21.8 (19, 23)	20.3 (17, 22)
Mean Weeks in Sample (Std)	7.16 (1.72)	7.25 (1.70)	7.25 (1.73)	7.04 (1.74)
No. of Observations	179	174	174	162
Mean Price (Std)	142.96 (24.34)	142.09 (21.33)	117.56 (18.34)	126.02 (20.58)
Max. and Min. Prices	208.90, 115.00	200.50, 109.20	170.50, 96.00	182.50, 102.00
Coefficient of Variation (as %)	17.02	15.01	15.60	16.33

Notes:
Prices are in US dollars. Pooled data is used for the estimates of the mean, max. and min. prices and the coefficient of variation.

Table 4: Summary statistics

The summary statistics of the data set can be found in Table 4. We found distinct numbers of stores operating in different markets but in all cases the number was quite high. For the KTT3614 memory chip, 25 firms were seen quoting prices over the period under

study; for the KTDINSP8200 and KTD4400 chips we collected prices from 24 different stores and for the KTD8300 chip we found 23 stores.

In our study, we estimated N by the total number of firms which were seen operating in each market on shopper.com. This number is probably lower than the true number of stores in the relevant market but, as shown in our Monte Carlo simulations, the results are relatively robust to small measurement errors in N.[14]

Not every firm was quoting a price every week. For example, for the KTT3614 memory chip we saw an average of 22.4 stores quoting a price in a typical week. The lowest number of stores for this product was 20 and the highest number of stores was 24. Similar numbers were found for the other products (see Table 4). There might be several reasons for this variation. For some stores there were missing values somewhere in the middle of the period of data collection. This might be due to technical problems when uploading the price feed to shopper.com. We also observed that some stores appeared in the sample only after some weeks had passed. In any case, on average, a typical firm was quoting prices more than 88% of the sample period (7 weeks out of 8).

The estimations were conducted under the assumption that firms play a repeated game of finite horizon so the data should reflect the equilibrium of the static game analyzed in Section 2. This implies that prices should be dispersed and that there should be variation in the position of a typical firm in the price ranking. We now examine how these two equilibrium features appear in the data.

Table 4 shows the mean price and corresponding standard deviation of prices, for each product. As expected, memory chips for notebooks are on average more expensive than those intended for desktop computers; moreover, the KTD8300 chip is more expensive than the KTD4400 chip due to its faster speed of operation. For all the products, we observe significant price dispersion as measured by the coefficient of variation. On average, relative to buying from one of the firms at random, the gains from being fully informed in this market are sizable, ranging from 21.56 to 32.89 US dollars.

A careful examination of the data reveals that most stores certainly change their price

---

[14]For robustness purposes, we also estimated the model taking 5 more firms than those seen in the data. The qualitative nature of our results did not change significantly.

from time to time, but we observe that they do not do it synchronously, that is, the length of time between price revisions changes from firm to firm. For example, in the market for the KTT3614 memory chip, 20 stores out of 25 changed their price at least once during the period under study. On average, a typical firm selling the KTT3614 chip changed the price once every 5 weeks but while some firms did change their prices several times over the sample period (up to 5 times), other firms did not. For the other memory chips, we found similar patterns.[15] The reason for this variation may be due to menu cost dispersion across firms.

We also observe some variation in the price ranking of a typical firm. For example, for the KTT3614 memory chip the standard deviation of the ranking of a firm ranges from 0 to 3.77. This is somewhat smaller than what we would expect on the basis of the theoretical model. Again one reason for this might be the short length of time of the sample period because some of the firms did not alter their prices.[16] To check this hypothesis, we gathered prices at the time of writing this paper and compared the current ranking of a typical firm with that at the time of data collection (one year ago). For example, for the KTT3614 memory chip, we found 21 stores quoting prices so some stores are no longer active in this market. This is not surprising since this market evolves very rapidly so after one year a product may be somehow outdated. Of these 21 stores, 16 stores were either higher or lower in the ranking compared to one year ago. The difference in ranking ranged from 0 to 7 and was on average of 2.48. Finally, 9 stores out of 21 are now in a different quartile of the ranking distribution. Similar figures apply to the other memory chips.[17] While not being fully conclusive, this evidence suggests that these markets resemble well the hypotheses of the model when data collection extends over a long period of time.

The prices used for our estimations include neither shipping costs nor sales taxes. One reason for not including shipping costs in the main analysis is that we do not have the data

---

[15]A typical firm selling the KTDINSP8200 chip changed its price once every 6 weeks, once every 7 weeks for the KTD4400 chip and once every 6 weeks for the KTD8300 memory chip.

[16]Lach (2002) examined the Israeli markets for chicken, coffee, flour and refrigerators during 48 months. The median duration of a store's ranking in a given quartile ranged from 1 month for coffee and chicken to 2 to 3 months for flour and refrigerators; in that period most of the firms were seen quoting prices in all quartiles of the price distribution.

[17]For the KTDINSP8200 memory chip 11 out of 22 stores were in a different quartile, with an average difference in ranking of 3.35. For the KTD4400 chip 14 out of 23 stores were in a different quartile, with an average difference in ranking of 3.39. Finally, for the KTD8300 chip 11 out of 22 stores were in a different quartile, while the average difference in ranking is 3.55.

for all the stores.[18] Another reason is that shipping costs and sales taxes depend on the state in which the consumer lives, which makes it difficult to compare total prices. However, for robustness purposes, we estimated the model neglecting sales taxes but using the shipping costs as if we were living in New York. Since a store not providing shipping cost information cannot be considered to ship for free (otherwise they would announce it as a promotional strategy), we attributed average shipping costs to the missing values. The qualitative nature of the results did not change.

Some of the variation in prices may be due to store differentiation. Consumers might view some stores more appealing than others and base this view on observable store characteristics like firm reputation, return policies, stock availability, order fulfillment, payment methods, etc. Unfortunately, we do not have information on all of these indicators. But we do have information on whether the item was in stock or not, on whether firms disclosed shipping cost on shopper.com or not and on the CNET certified ranking of a store, which is a store quality index computed by CNET on the basis of consumer feedback. To see the impact of these (observable) variables on the prices of each memory chip in our data set, we estimated the following model:

$$PRICE_{jt} = \beta_0 + \beta_1 \cdot RATING_{jt} + \beta_2 \cdot SHIP_{jt} + \beta_3 \cdot STOCK_{jt} + \varepsilon_{jt}, \qquad (13)$$

where, for each product, $PRICE_{jt}$ is the list price of store j in week t, $RATING_{jt}$ is the CNET certified ranking of store j in week t, $SHIP_{jt}$ is a dummy for whether shop j disclosed shipping cost in week t, and $STOCK_{jt}$ is a dummy for whether shop j had the item in stock in week t. We estimated equation (13) by OLS. The resulting R-squared values indicate that only between 6% and 17% of the total variation in prices can be attributed to observable differences in store characteristics.[19] This suggests that the rest of the price variation can

---

[18]Actually stores may choose to report blank in the shipping and handling cost field of the price feed form. As a result, shopper.com reports "See Site" in the shipping and handling column for that particular store.

[19]For all memory chips, the OLS estimates of the coefficient of $SHIP_{jt}$ are negative and highly significant. The estimates of the coefficient of $RATING_{jt}$ are positive and significant on a 1% level for the KTDINSP8200 chip, significant on a 10% level for the KTT3614 and KTD4400 chips and not significant for the KTD8300 chip. The coefficient of $STOCK_{jt}$ was not significant for any of the products, but this could be due to the lack of variation of this variable in our data (upon reporting on shopper.com, almost all stores had the product in stock).

be due to strategic price setting in the presence of search costs.

One could argue that differences in marginal costs between the stores, or unobserved heterogeneity across shops (e.g. branding) play an important role in explaining the prices. In principle, this can be tested by adding firm dummies to equation (13). However, doing this is not very meaningful here because, as argued above, for some stores we do not observe much variability in the prices they charge over time. Including store dummies in the regression equation would lead to significant dummies but we would not be able to distinguish whether this is because of the short period of data gathering or because of unobserved heterogeneity between stores. Moreover, the finding that quite a few stores do change their price often and also that store rankings change from week to week already gives an indication that store heterogeneity cannot be the most important factor in explaining price setting behavior.

5.2   Estimation results

The estimation results for the four different memory chips are presented in Table 5. An interesting observation is that even though the products differ in their characteristics, the estimates are quite similar across memory chips. This suggests that the consumers acquiring these products have similar search cost distributions.

	KTT3614	KTDINSP8200	KTD4400	KTD8300
$\underline{p}$	115.00	109.20	96.00	102.00
v	208.90	200.50	170.50	182.50
N	25	24	24	23
M	179	174	174	162
q_1	0.22 (0.05)	0.29 (0.04)	0.24 (0.05)	0.30 (0.04)
q_2	0.39 (0.15)	0.58 (0.02)	0.68 (0.01)	0.66 (0.03)
q_3	0.31 (0.14)	0.00	0.00	0.00
q_4	0.00	0.00	0.00	0.00
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
q_{N-1}	0.00	0.00	0.00	0.00
q_N	0.08 (0.07)	0.13 (0.05)	0.09 (0.05)	0.04 (0.01)
r	109.69 (1.43)	103.15 (0.84)	90.91 (1.16)	90.55 (1.91)
LL	715.42	677.81	644.64	616.39
KS	0	0	0	0

Notes:
Estimated standard errors in parenthesis. $KS = 0$ indicates that the null hypothesis that both distributions (empirical and estimated price cdf) are the same cannot be rejected.

Table 5: Estimation results

The estimates of the share of consumers who search once, q_1, range from 22% to 30%

and are all highly significant.[20] These consumers do not compare prices and thus confer monopoly power to the firms. Firms compete for the rest of the consumers, who happen to search for 2 or 3 prices or for all the prices in the market. In particular, the estimates of q_2 range from 39% for the KTT3614 memory chip to 68% for the KTD4400 chip and are highly significant as well. The KTT3614 has also a sizable share of consumers comparing three prices, about 31%. For all the products, the estimates of parameters q_4 till q_{N-1} are all approximately zero. Finally, the estimates of the fraction of consumers comparing all the prices in the market, q_N, range from 4% to 13% and are, except for the KTD3614 memory chip, significant at a 5% level.[21]

These results suggest a clear picture of consumer search costs. The entire consumer population can roughly be grouped into three subsets: buyers who do not search, buyers who compare at most three prices and buyers who compare all the prices in the market. This is consistent with the view that consumers have either quite high search costs or quite low search costs.

The estimated cut-off points of the search cost distribution, Δ_i, with corresponding standard errors are presented in Table 6. All the cut-off points are highly significant and notice again that there is very little variation in the estimates across products. The estimated critical search cost values in combination with the estimated shares of consumers searching i times allow us to construct estimates of the search cost distributions underlying firm and consumer behavior.

Figure 5 gives the estimated cumulative search cost distributions for the four memory chips. For example for the KTT3614 memory chip we see that around 22% of the consumers have search costs higher than 12.26 US dollars; these costs are so high that these consumers only search once in equilibrium. Around 70% of the consumers have search costs in between 2.21 and 12.26 US dollars and for these consumers it is worth to search 2 or 3 times. Finally, around 8% of the buyers have search costs that are at most 9 dollar cents; these costs are so low that these buyers check the prices of all vendors. In sum, these estimates imply that

---

[20]To be able to calculate the standard errors, we deleted the columns and rows of the Hessian for which the corresponding parameter estimates were zero.

[21]In a study of the consumer click-through behavior online, Johnson *et al.* (2004) also point out that many consumers search quite little.

	KTT3614	KTDINSP8200	KTD4400	KTD8300
Δ_1	12.26 (1.42)	10.94 (0.28)	8.34 (0.47)	9.76 (0.28)
Δ_2	4.41 (1.19)	4.25 (0.27)	2.91 (0.32)	3.49 (0.17)
Δ_3	2.21 (0.81)	2.37 (0.20)	1.50 (0.20)	1.78 (0.10)
Δ_4	1.34 (0.58)	1.59 (0.16)	0.96 (0.14)	1.10 (0.07)
Δ_5	0.92 (0.44)	1.19 (0.13)	0.69 (0.11)	0.77 (0.05)
Δ_6	0.68 (0.35)	0.94 (0.11)	0.54 (0.09)	0.58 (0.04)
Δ_7	0.53 (0.29)	0.78 (0.09)	0.44 (0.07)	0.46 (0.03)
Δ_8	0.43 (0.24)	0.67 (0.08)	0.37 (0.06)	0.38 (0.02)
Δ_9	0.36 (0.21)	0.58 (0.07)	0.32 (0.06)	0.32 (0.02)
Δ_{10}	0.31 (0.18)	0.51 (0.06)	0.28 (0.05)	0.28 (0.02)
Δ_{11}	0.27 (0.16)	0.45 (0.06)	0.24 (0.04)	0.24 (0.02)
Δ_{12}	0.24 (0.15)	0.41 (0.05)	0.22 (0.04)	0.22 (0.01)
Δ_{13}	0.21 (0.13)	0.37 (0.05)	0.20 (0.04)	0.19 (0.01)
Δ_{14}	0.19 (0.12)	0.33 (0.04)	0.18 (0.03)	0.17 (0.01)
Δ_{15}	0.17 (0.11)	0.30 (0.04)	0.16 (0.03)	0.16 (0.01)
Δ_{16}	0.16 (0.10)	0.28 (0.04)	0.15 (0.03)	0.15 (0.01)
Δ_{17}	0.14 (0.09)	0.26 (0.03)	0.14 (0.03)	0.13 (0.01)
Δ_{18}	0.13 (0.08)	0.24 (0.03)	0.13 (0.02)	0.12 (0.01)
Δ_{19}	0.12 (0.08)	0.22 (0.03)	0.12 (0.02)	0.11 (0.01)
Δ_{20}	0.11 (0.07)	0.20 (0.03)	0.11 (0.02)	0.11 (0.01)
Δ_{21}	0.11 (0.07)	0.19 (0.02)	0.10 (0.02)	0.10 (0.01)
Δ_{22}	0.10 (0.06)	0.18 (0.02)	0.10 (0.02)	0.09 (0.01)
Δ_{23}	0.09 (0.06)	0.17 (0.02)	0.09 (0.02)	-
Δ_{24}	0.09 (0.06)	-	-	-

Notes:
Estimated standard errors in parenthesis.

Table 6: Estimated critical search cost values

typical consumers have either very high search costs or very low search costs.

In spite of having more than 20 stores operating in each of the markets, we observe that market power is substantial. The estimates of r indicate that unit costs are between 50% and 53% of the value of the product so the average price-cost margins range between 23% and 28%.[22] This is of course the consequence of search costs, suggesting that demand side characteristics might be even more important than supply side ones to assess market competitiveness (Waterson, 2003).

We finally test the goodness of fit of the model. To see how well the estimated price density function fits the data, we use the Kolmogorov-Smirnov test (KS-test) to compare the actual distribution to the fitted distribution. The KS-test is based on the maximum difference between empirical cumulative distribution and the hypothesized estimated cumulative distribution. The null hypothesis for this test is that they have the same distribution, the alternative hypothesis is that they have different distributions. As Table 5 shows, for all four memory chips we cannot reject that the price observations are drawn from the es-
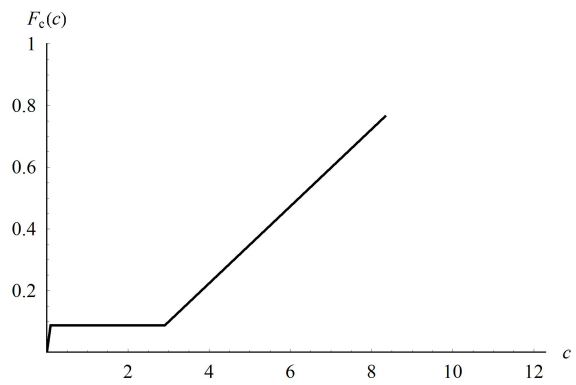

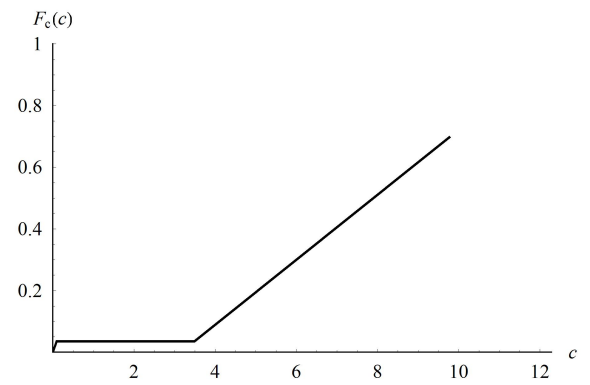[22]These margins are similar to those found in the book industry (Clay *et al.* 2001).
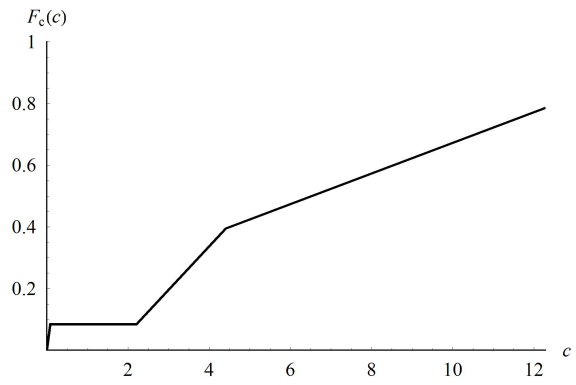
(a) KTT3614
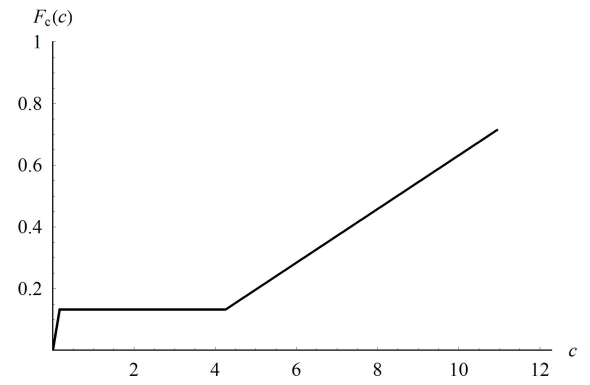
(b) KTDINSP8200

(c) KTD4400
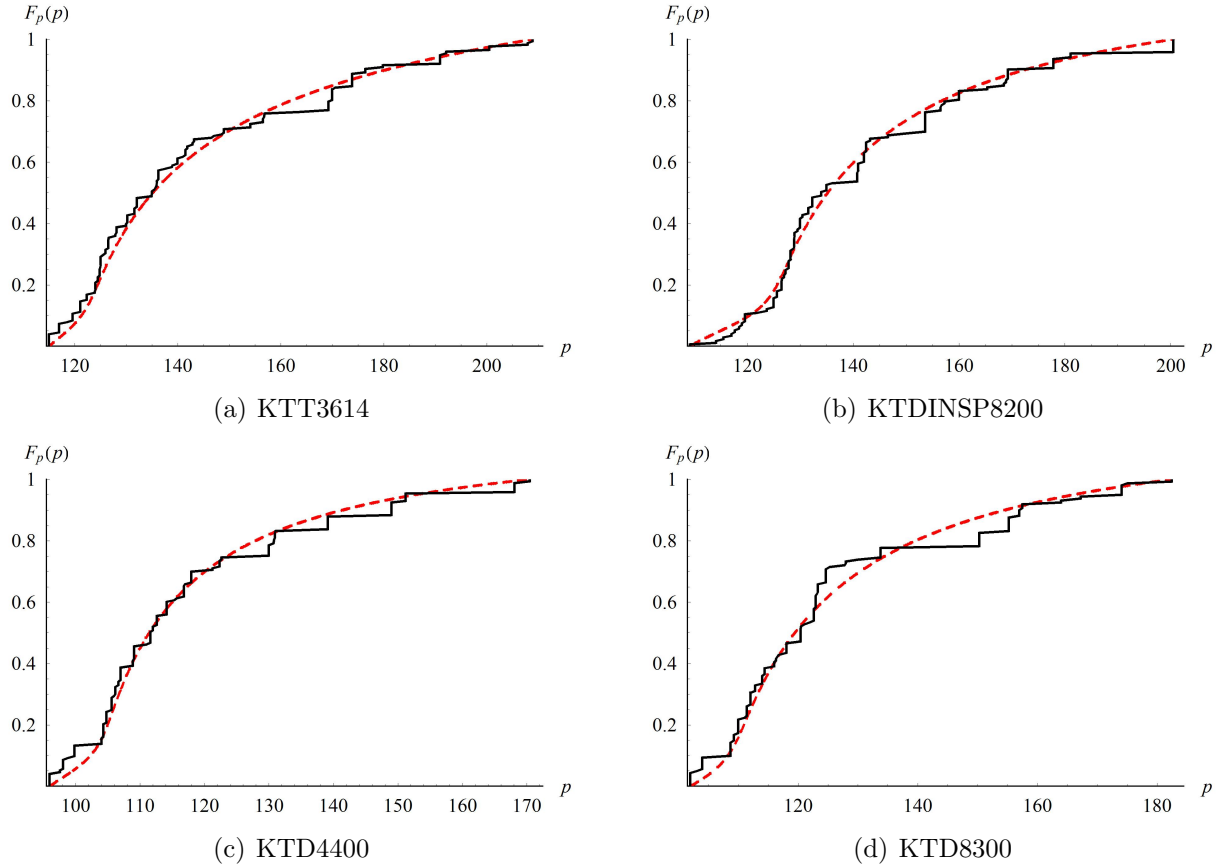
(d) KTD8300

Figure 5: Estimated search cost cdf

(a) KTT3614   (b) KTDINSP8200

(c) KTD4400   (d) KTD8300

Figure 6: Estimated and empirical price cdf

timated price cumulative distribution.[23] The goodness of fit of the model to the data can be visualized in Figure 6. A solid curve represents an empirical price distribution, while a dashed curve represents an estimated one.

6   Conclusions

Consumer search models have shown for example that, depending on the nature of search in the market, an increase in the number of firms can increase or decrease the price levels and price dispersion. Since competition policy recommendations may depend on the nature

---

[23]Because some of the parameters that enter the test are estimated we also calculated the Rao-Robson Statistic, which is a kind of chi-squared test corrected for the uncertainty involved in estimating some of the parameters of the distribution that has to be fitted (for more details see Moore, 1986). The Rao-Robson statistics for two of the four products are below their corresponding critical values (KTT3614 and KTD4400), which means that for these products we cannot reject the null hypothesis that the estimated and empirical price cdf are the same.

of the search cost distribution, there is a need to develop methods to identify and quantify search costs.

Hong and Shum (forthcoming) were the first to exploit the restrictions equilibrium search models place on the joint distribution of prices and search costs to structurally estimate unobserved search cost parameters. In this paper we have extended their approach to the oligopoly case and we have proposed a method to obtain maximum likelihood estimates of the search cost distribution.

Our method has two important features. One, if N firms operate in the market we can estimate N cut-off points of the search cost distribution. In this way, we also learn about the extent of thorough search in markets and about the height of the search costs faced by the consumers who search intensively. Indeed, using a data set of prices for four memory chips we find that between 4% and 13% of the consumers search for all prices in the market. These consumers have a search cost of at most 17 US dollar cents and obtain sizable gains relative to buying from one of the firms at random, namely, from 21 to 33 US dollars. The second feature is that our method yields maximum likelihood estimates of the search cost distribution, which allows for standard asymptotic theory and hypothesis testing.

Our estimates of the consumer search cost density underlying the price observations for the memory chips suggest that consumers have either quite low or quite high search costs. Search costs confer significant market power to the firms. The estimates reveal that average price-cost margins range from 23% to 28%. Finally, according to the Kolmogorov-Smirnov goodness-of-fit test, we cannot reject the null hypothesis that the price observations were drawn from the distribution functions specified by the theoretical search model.

References

[1] Michael Baye, John Morgan and Patrick Scholten: "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics* 52, 463-96, 2004.

[2] Jeffrey R. 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.

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

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

[5] Bev Dahlby and Douglas S. West: "Price Dispersion in an Automobile Insurance Market," *Journal of Political Economy* 94, 418-38, 1986.

[6] Juan Delgado and Michael Waterson: "Tyre price dispersion across retail outlets in the UK," *Journal of Industrial Economics* 51, 491-509, 2003.

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

[8] Glenn Ellison and Sara Fisher Ellison: "Search, Obfuscation, and Price Elasticities on the Internet," *Mimeo*, 2004.

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

[10] 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.

27

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

[12] Eric J. Johnson, Wendy W. Moe, Peter S. Fader, Steven Bellman and Gerald L. Lohse: "On the Depth and Dynamics of On-line Search Behavior," *Management Science* 50, 299-308, 2004.

[13] 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.

[14] Saul Lach: "Existence and Persistence of Price Dispersion: An Empirical Analysis," *Review of Economics and Statistics* 84, 433-44, 2002.

[15] David S. Moore: "Tests of Chi-Squared Type," in: Ralph B. D'Agostino and Michael A. Stephens (Editors), *Goodness-of-Fit Techniques*, New York: Marcel Dekker, 1986.

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

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

[18] K.C. Rao and D.S. Robson: "A Chi-Square Statistic for Goodness-of-Fit within the Exponential Family," *Communications in Statistics* 3, 1139-53, 1974.

[19] Rafael Rob: "Equilibrium Price Distributions," *Review of Economic Studies* 52, 487-504, 1985.

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

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

[22] Alan T. Sorensen: "An Empirical Model of Heterogeneous Consumer Search for Retail Prescription Drugs," NBER Working Paper No. 8548, 2001.

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

[24] Hal R. Varian: "A Model of Sales," *American Economic Review*, 70, 65159, 1980.

[25] Michael Waterson: "The role of consumers in competition and competition policy," *International Journal of Industrial Organization* 21, 129-50, 2003.