On the Identification of the Costs of Simultaneous Search *

José Luis Moraga-González †
Zsolt Sándor ‡
Matthijs R. Wildenbeest §

July 2010

Abstract

This paper studies the identification of the costs of simultaneous search in a class of (portfolio) problems studied by Chade and Smith (2006). We show that aggregate data from a single market, or disaggregate data from a single market segment, do not provide sufficient information to identify the costs of simultaneous search in any reasonable interval. We then show that by pooling aggregate data from multiple markets, or disaggregate data from multiple market segments, the econometrician can identify the costs of simultaneous search in a non-empty interval. Within the context of specific examples, we illustrate that identification of the search cost distribution in its full support may easily be obtained.

Keywords: search costs, portfolio choice, non-parametric identification

JEL Classification: C14, D83, J64

*Financial support from Marie Curie Excellence Grant MEXT-CT-2006-042471 is gratefully acknowledged.

†ICREA, IESE Business School and University of Groningen, E-mail: jose.l.moraga@gmail.com.

‡University of Groningen, E-mail: z.sandor@rug.nl.

§Kelley School of Business, Indiana University, E-mail: mwildenb@indiana.edu.

1 Introduction

This paper studies the nonparametric identification of the costs of simultaneous search in a class of (portfolio) problems examined by Chade and Smith (2005, 2006). This class embeds a number of important decision problems in economics. In these problems a decision maker must simultaneously choose among a set of ranked stochastic options; each choice is costly and only the best realized option is finally exercised. This problem arises for example when students apply for colleges (Gale and Shapley, 1962; Kelso and Crawford, 1982; Roth and Sotomayor, 1989), when consumers search for differentiated products (Stigler, 1961; Wolinsky, 1986; Anderson and Renault, 1999), or when workers search for employment (Burdett et al., 2001; Albrecht et al., 2006; Kircher, 2009).

In all these papers, an important issue is the study of the extent to which the costs of simultaneous search drive a wedge between the market outcome and the social optimum. Assessing public policy measures aimed at shortening the gap between the market and the social outcomes (for instance policies to increase social mobility, minimum wage policy or merger policy), or estimating the social value of new schools or new products requires the development of methods to identify and estimate the costs of simultaneous search. This paper is a contribution to this objective.

We consider markets where the costs of search vary across the decision makers. In these types of market, we study the conditions under which the econometrician can identify the search cost distribution nonparametrically. We first show that detailed data from a single market (such as the aggregate market share and the utility distribution corresponding to each of the available options) do not provide sufficient information to identify the costs of simultaneous search in any reasonable interval. The problem originates from the fact that the sequence of critical search costs that can be identified from the data is convergent to zero so the set of search cost values the econometrician can identify is not dense in the support of the search cost distribution. As a result, nonparametric identification of the search cost distribution at quantiles other than the lowest fails. The convergence-to-zero property of the search cost cutoffs is related to the fact that the marginal gains from searching $k + 1$ options rather than k options are typically decreasing in k (see our Proposition 1).[1]

The paper proceeds by providing conditions under which the use of aggregate data from multiple markets, or disaggregate data from multiple market segments, helps identify the search cost distri-

[1] Stigler (1961) made a related point within the specific context of consumer search for homogeneous products.

bution. We propose to gather (market shares and utility distributions) data from several markets (segments) with the same search technology. This helps in that every market (segment) generates a distinctive set of search cost values for which the econometrician can retrieve the density of search costs. By pooling market share data from many markets (segments) one forces the search cost distribution to be uniquely determined for a much larger set of points. Gathering the appropriate data is relatively easy for the econometrician. For example, in the college problem, one can take data from different cities, with typically distinct numbers of colleges, different success rates and distinct earning distributions. In the case of consumer search for differentiated products, one can pool data from markets where different product qualities are available, or, within a single market, data from consumer segments with different demographics.[2]

In the remainder of the paper, we first describe the class of problems we study, then we provide our identification results and finally we apply them to a couple of specific examples.

2 The model

We study the identification of the costs of simultaneous search in a class of (portfolio) problems inspired by the work of Chade and Smith (2006). In this class of problems, a decision maker must make a simultaneous choice among ranked stochastic options; each choice is costly and only the best realized alternative is finally consumed. Chade and Smith (2005, 2006) prove that a greedy algorithm finds the solution in two classes of problems: problems with downward recursive (DR) payoff functions, and non-DR problems with prize distributions ordered by a second-order stochastic dominance condition. For the purpose of our paper, we can treat these two classes of problems together within a single framework.

We assume that in a market (or in a market segment) there is a continuum of heterogeneous decision makers who can choose to consume prizes/options from a total of N options. Each option i gives a payoff u_i with probability $\alpha_i \in (0,1]$, where u_i is a random variable with probability distribution Ψ_i with support $[\underline{u}_i, \overline{u}_i], i = 1, 2, ..., N$. The scalar α_i measures the probability with which an option succeeds to yield a payoff. We assume that the random variables $u_1, u_2, ..., u_N$

[2] Consumer search for differentiated products is indeed an exciting area of active research. Several recent papers have estimated models of search for differentiated products by assuming parametric search cost distributions (Mehta, Rajiv, and Srinivasan, 2003; Honka, 2010; Moraga-González, Sándor, and Wildenbeest, 2010). The results of our paper contribute to the nonparametric identification of the search cost distribution in such settings.

are independent. Let the interval $[0, \overline{u}]$ contain the union of all the options' supports $[\underline{u}_i, \overline{u}_i]$, $i = 1, 2, ..., N$. Define $\nu_i = I_{\alpha_i} u_i$, where I_{α_i} is a Bernoulli random variable that takes value 1 with probability α_i. The distribution of ν_i is $G_i = (1 - \alpha_i) + \alpha_i \Psi_i$ and its support will be denoted by Φ_i. When $\alpha_i < 1$ then $\Phi_i \equiv \{0\} \cup [\underline{u}_i, \overline{u}_i]$, otherwise $\Phi_i \equiv [\underline{u}_i, \overline{u}_i]$. We assume that:

Assumption 1. *The distributions G_i can be ranked according to the quasi-second order stochastic dominance criterion, that is: $\int_x^{\overline{u}} G_i(u)du \leq \int_x^{\overline{u}} G_{i+1}(u)du$, $i = 1, 2, ..., N-1$, for all $x \in [0, \overline{u}]$ with strict inequality at $x = 0$.*

This assumption implies that $E[u_{i+1}] > E[u_i]$ so (*ex-ante*) the best single option is option 1, the second best single option is option 2, etc.

We shall assume that strictly dominated options, if any, have been already discarded. That is, suppose there exists an option ℓ that succeeds with probability 1 and that pays always more than some other option k; in this case ℓ strictly dominates k and k can be discarded. If there does not exist an option that succeeds surely, then no option can be *ex-ante* discarded.

Assumption 2. *Let N_F be the set of options that can fail with strictly positive probability. Let $\overline{N_F} = N \setminus N_F$. We assume $\cap_{i \in \overline{N_F}} \Phi_i$ is a non-empty interval.*

It is clear that if all options fail with strictly positive probability, then all options will have a positive market share provided the costs of search are not too high. The role of this assumption is to introduce some (*ex-post*) "competition" among the options that yield a payoff surely.

Each decision maker is characterized by her cost of searching an option. Let c be the cost of search of a decision maker. If the decision maker searches the subset of options S, her total cost is $c|S|$, where $|S|$ denotes the cardinality of the subset S. Assume c is drawn independently from a common atomless distribution $H(c)$ with support $\Omega = [0, \overline{c}]$; let $h(c)$ denote the corresponding density. It will be convenient to assume $\overline{c}$ is sufficiently large, which ensures that there always exist decision makers who do not search at all.[3]

Every decision maker ultimately consumes one single option (like in discrete choice models). For a given set of options $S \subseteq N$, let $f(S)$ denote the expected (gross-of-cost) payoff:

$$f(S) \equiv E[\max\{u_i : i \in S\}] = \int_0^{\overline{u}} \left(1 - \prod_{i \in S} G_i(u)\right) du. \tag{1}$$

[3] This assumption can easily be relaxed (see Section 3.1). We do not allow the highest search cost to be low here because this requires an amount of additional notation that makes the presentation of the ideas somewhat cumbersome.

We adopt the convention $f(\varnothing)=0$. The problem of a decision maker with cost of searching options equal to c is to choose a (sub-)set of options $S \subseteq N$ to maximize her expected payoff:

$$\max_{S\subseteq N}\left\{f(S)-c|S|\right\}. \tag{2}$$

Under Assumption 1, a solution to the problem in (1) for decision maker with search cost c is a set of options $\{1,2,...,i^*(c)\}$. Chade and Smith (2006) provide an algorithm, namely the *Marginal Improvement Algorithm* (MIA), that identifies the solution to this problem via an inductive procedure. In essence, the MIA algorithm works as follows:

Step 1. Add option 1 to the optimal choice set if $f(\{1\})\geq c$, otherwise choose the empty set and stop.

Step 2. Add option 2 to the optimal choice set if $f(\{1,2\})-f(\{1\})\geq c$, otherwise stop.

...

Step i. Add option i to the optimal choice set if $f(\{1,2,...,i\})-f(\{1,2,...,i-1\})\geq c$, otherwise stop (for all $i=3,4,..,N$).

In their papers Chade and Smith (2005, 2006) prove that the MIA algorithm encounters such an optimal solution $\{1,2,...,i^*(c)\}$.

In describing the market outcome this problem generates, it is useful to prove that the (gross-of-cost) payoff in (1) has decreasing returns in i.

Proposition 1 *Under Assumption 1,*[4] $f(\{1,2,...,i\})-f(\{1,2,...,i-1\})\geq f(\{1,2,...,i+1\})-f(\{1,2,...,i\})$, $i=2,...,N$.

Proof. Consider first the case where the CDFs Ψ_i are not degenerated. For this case, we need to prove that

$$\int_0^{\overline{u}}\left(1-\prod_{j=1}^{i}G_j(u)\right)du-\int_0^{\overline{u}}\left(1-\prod_{j=1}^{i-1}G_j(u)\right)du-\int_0^{\overline{u}}\left(1-\prod_{j=1}^{i+1}G_j(u)\right)du+\int_0^{\overline{u}}\left(1-\prod_{j=1}^{i}G_j(u)\right)du$$

$$=\int_0^{\overline{u}}\left(\prod_{j=1}^{i-1}G_j(u)-2\prod_{j=1}^{i}G_j(u)+\prod_{j=1}^{i+1}G_j(u)\right)du=\int_0^{\overline{u}}\left(\prod_{j=1}^{i-1}G_j(u)\left(1-2G_i(u)+G_i(u)G_{i+1}(u)\right)\right)du\geq 0.$$

We now argue that

$$\int_0^{\overline{u}}\left(\prod_{j=1}^{i-1}G_j(u)\left(1-2G_i(u)+G_i(u)G_{i+1}(u)\right)\right)du\geq\int_0^{\overline{u}}\left(\prod_{j=1}^{i-1}G_j(u)(1-G_i(u))^2\right)du\geq 0.$$

[4] In fact, we do not need the strict inequality condition at $x=0$.

For this, it suffices that

$$\int_0^{\overline{u}} \beta(u)\gamma(u)du \geq 0, \tag{3}$$

where $\beta(u) = \prod_{j=1}^{i} G_j(u)$ and $\gamma(u) = G_{i+1}(u) - G_i(u)$. Note that β is monotonically increasing in u, is absolutely continuous on $(0, \overline{u}]$, with $\beta(0) = 0$ and $\beta(\overline{u}) = 1$. Moreover, by Assumption 1, $\int_x^{\overline{u}} \gamma(u)du \geq 0$ for all $x \in [0, \overline{u}]$. Let

$$\Gamma(x) = -\int_x^{\overline{u}} \gamma(u)du.$$

Integrating (3) by parts gives

$$\begin{aligned}\int_0^{\overline{u}} \beta(u)\gamma(u)du &= \beta(\overline{u})\Gamma(\overline{u}) - \beta(0)\Gamma(0) - \int_0^{\overline{u}} \Gamma(u)\frac{d\beta(u)}{du}du \\ &= -\prod_{j=1}^{i}(1-\alpha_j)\Gamma(0) - \int_0^{\overline{u}} \Gamma(u)\frac{d\beta(u)}{du}du \geq 0.\end{aligned}$$

and the result follows.

Consider now the case where the CDFs Ψ_i are degenerated. In such a case, the payoff structure is DR (for the definition see section 2.1) and the result follows from Lemma 5 in Chade and Smith (2006). ■

We can use Proposition 1 to describe the market outcome this problem generates as follows. Let us denote by q_i the probability that a randomly selected decision maker chooses the set $\{1, 2, ..., i\}$, $i = 1, 2, .., N$. By construction $\sum_{i=0}^{N} q_i = 1$ where q_0 is the probability of choosing the empty set. Since c is distributed according to H, these probabilities can readily be computed:

$$\begin{aligned} q_0 &= 1 - H(c_0) \text{ and} \\ q_i &= H(c_{i-1}) - H(c_i),\ i = 1, 2, ..., N, \end{aligned}$$

where

$$\begin{aligned} c_0 &= f(\{1\}), \\ c_i &= f(\{1, 2, ..., i+1\}) - f(\{1, 2, ..., i\}),\ i = 1, 2, ..., N-1 \\ c_N &= 0. \end{aligned} \tag{4}$$

The number c_i is the cutoff value of the search cost distribution that makes a decision maker indifferent between choosing i and $i+1$ options. By Proposition 1, the sequence of cutoff values $\{c_i\}_{i=0}^{N}$ is decreasing and therefore the probabilities q_i are well-defined.

Let s_i denote the market share of option i. Computing the market share of an option i involves summing over all q's the probability option i turns out to be the most attractive. In general this market share can be written as follows:

$$\begin{aligned}
s_0 &= q_0 + q_1 \Pr(\nu_1 = 0) + q_2 \Pr(\nu_1 = \nu_2 = 0) + ... + q_N \Pr(\nu_1 = \nu_2 = ... = \nu_N = 0), \\
s_1 &= q_1 \Pr(\nu_1 > 0) + q_2 \Pr(\nu_1 > \max\{0, \nu_2\}) + ... + q_N \Pr(\nu_1 > \max\{0, \nu_2, \nu_3 ..., \nu_N\}), \\
s_2 &= q_2 \Pr(\nu_2 > \max\{0, \nu_1\}) + ... + q_N \Pr(\nu_2 > \max\{0, \nu_1, \nu_3 ..., \nu_N\}), \qquad (5) \\
&... \\
s_N &= q_N \Pr(\nu_N > \max\{0, \nu_1, \nu_2, ..., \nu_{N-1}\}).
\end{aligned}$$

2.1 Examples

Some leading examples generate payoff structures similar to that in (1).

Example 1. The college problem.

In the college problem the distributions Ψ_i are degenerated, for all i. A student must choose a set $S \subseteq N$ of colleges to apply for admission and this costs her $c|S|$. Given Assumption 1, we have $\alpha_1 u_1 > \alpha_2 u_2 > ... > \alpha_N u_N$ so *ex-ante* the best college is college 1, the second best is college 2, etc.[5] The expected (gross-of-cost) payoff from applying to the set of colleges S is

$$f(S) = \sum_{i \in S} \alpha_i u_i \prod_{j<i, j \in S} (1 - \alpha_j) \qquad (6)$$

Chade and Smith (2006) point out that this payoff function is *downward recursive* (DR). That is for any two sets U, L in N with $U \sqsupseteq L$ (i.e. the worst option in U beats the best in L) we have $f(U + L) = f(U) + \rho(U) f(L)$ where $\rho(U) \equiv \prod_{j \in U} (1 - \alpha_j)$ is the failure chance of all the options in the set U.

This problem is similar to the *directed* search problem in labor economics studied by Burdett, Shi and Wright (2001), Albrecht et al. (2006) and Kircher (2009).

Example 2. Search for differentiated products.

[5] This follows from the fact that $\int_0^{\overline{u}} G_i(u) du = \overline{u} - \alpha_i u_i$.

In the model of search for differentiated products, which generalizes Stigler (1961) and Burdett and Judd (1983) to the case of non-identical prizes, $\alpha_i = 0$ for all i. A consumer must visit a set S of shops to learn the utility she derives from the different products. Once the utility is learnt, the consumer picks the product giving her the highest utility. Given Assumption 1, *ex-ante* the best product is product 1, the second best product is product 2 etc. The expected (gross-of-cost) payoff from visiting the set of shops S is

$$f(S) \equiv \int_0^{\overline{u}} \left(1 - \prod_{i \in S} \Psi_i(u)\right) du \tag{7}$$

Chade and Smith (2005) note that the payoff structure in (7) is *not* DR. However, their MIA algorithm also works in this case.

In recent research on oligopolistic competition with differentiated products, it is typically assumed that consumers differ in the (gross-of-cost) utilities they derive from the various products. We note that the payoff structure in (7) should then be interpreted as the corresponding payoff to the consumers in a particular consumer segment. Later we will come back to this point when discussing identification using disaggregate data from various consumer segments.

Example 3. Capacity-constrained firms and consumer search / Random job search with multiple applications.

The generality of the payoff structure in (1) captures situations in which firms may be capacity constrained and the exact prizes ex-ante unknown by consumers. This situation arises for example when a consumer searches for differentiated products sold by capacity-constrained firms. A consumer needs to visit shops not only to learn the exact utility he/she derives from the products but also to check their availability. Another situation is when a worker randomly searches in the labor market for a better-paid job (Gautier and Moraga-González, 2004). Here workers apply for multiple jobs but firms have only a limited number of vacancies and wages are drawn from an atomless probability distribution.

3 Identification results

The econometrics problem consists of estimating the costs of simultaneous search, i.e., the CDF H of search costs. A crucial requirement for consistent estimation is that the search cost distribution is identified. In what follows, we study whether such identification is possible using aggregate data from a single market segment. In particular, we assume that:

Assumption 3. *The econometrician observes:*

1. A set of options with corresponding distributions $\{G_i\}_{i=1}^N$ *over the set of prizes* $\{\Phi_i\}_{i=1}^N$ *each option can deliver, where the* G_i*'s* satisfy Assumption 1.

2. The aggregate market shares of each of the options, denoted $\{s_i\}_{i=0}^N$*, where* s_0 *denotes the market share of the "outside" option.*

In the examples above, Assumption 3.2 requires observing the numbers of students accepted at the different colleges, observing the market shares of the various products in a single market segment, or the number of persons hired by the different firms in a market. Assumption 3.1 implies that the researcher knows the utility (earning) distributions each option can yield. Admittedly, this is a significant amount of information and therefore Assumption 3 represents a case quite favorable for identification of the search cost distribution. Even in this favorable situation, we shall point to some identification challenges (which of course would remain if the utility distributions were not known to the econometrician).

Our first result states that the data described in Assumption 3 allow for the identification of the search cost distribution at the cutoff values defined in (4).

Proposition 2 *Under Assumptions 1, 2 and 3 the search cost cutoff values* $\{c_i\}_{i=0}^N$ *and the corresponding values of the CDF of search cost* $\{H(c_i)\}_{i=0}^N$ *are identified.*

Proof. The cutoff values can be computed from the set of equalities

$$\begin{aligned}
c_0 &= f(\{1\}) \\
c_i &= f(\{1, 2, ..., i+1\}) - f(\{1, 2, ..., i\}),\ i = 1, 2, ..., N-1, \\
c_N &= 0.
\end{aligned}$$

which form a decreasing sequence by Proposition 1.

Market shares satisfy the system of equations in (5). Note that this system of equations is triangular and it has strictly positive diagonal elements. To see this, consider the k-th diagonal element: $\Pr(\nu_k > \max\{0, \nu_1, \nu_2, ..., \nu_{k-1}\})$. Denote by $L \subset \{1, 2, ..., k-1\}$ the subset of options that fail with strictly positive probability. Let $\rho(L)$ be the probability that all options in L fail; by convention

$\rho(\varnothing) = 1$. Then $\Pr(\nu_k > \max\{0, \nu_1, \nu_2, ..., \nu_{k-1}\}) \geq \rho(L)\Pr(\nu_k > \max\left\{0, \max\{\nu_i : i \in \overline{L}\}\right\})$. If option k succeeds with probability 1, then by Assumption 2, $\Pr(\nu_k > \max\{\nu_i : i \in \overline{L}\}) > 0$. If option k fails with strictly positive probability, by the free disposal of strictly dominated options assumption, it must be the case that the upper bound of the support of option k, $\overline{u}_k > \max\{\underline{u}_i : i \in \overline{L}\}$. As a result, $\Pr(\nu_k > \max\{\nu_i : i \in \overline{L}\}) > 0$.

Therefore, the system of equations (5) can be solved for $\{q_i\}_{i=0}^{N}$. Once we have obtained such solution, one can iteratively compute the corresponding values of the CDF of the costs of simultaneous search at the cutoffs:

$$H(c_i) = 1 - \sum\nolimits_{k=0}^{i} q_k \text{ where } i = 0, 1, ..., N.$$

■

Given that one can identify the sequence of points $\{c_i, H(c_i)\}_{i=0}^{N}$ of the search cost distribution, the question that arises is whether taking a market with many options ($N \to \infty$) can allow for the identification of the search cost distribution in any interval of interest.

Proposition 3 *Under Assumptions 1,2 and 3, the search cost distribution is not identified in any interval, even if* $N \to \infty$.

Proof. If $N \to \infty$ then, by Proposition 1 the sequence $\{c_i\}_{i=0}^{N}$ will be decreasing and convergent to zero because $\{f(\{1, 2, ..., i\})\}_{i=1}^{N}$ is convergent. Therefore, the set of points outside an arbitrarily small neighborhood around the limit point will be necessarily finite. The result follows. ■

Proposition 3 shows that using the type of aggregate data described in Assumption 3 only allows for identification of the costs of simultaneous search at low quantiles. This is a problem because any reasonable study of the effects of public policy measures aimed at improving the market outcome (for instance, introducing new college options, increasing school places, challenging a merger between options, etc.) in any of these settings would require the identification of search costs at higher quantiles. We next move to the study of identification using data from multiple market segments.

We propose to pool data from a (large) number M of distinct markets segments, all of them with the same underlying search cost distribution but with different characteristics. Let the features of market segment m be characterized by the vector θ_m, $m = 1, 2, ..., M$. Assume that $M \to \infty$ and define $\mathcal{N}^{\infty}$ as the set that includes all the numbers of options $\widehat{N}$ for which infinitely many market segments exist. For each $\widehat{N} \in \mathcal{N}^{\infty}$ let $\{c_i^{\widehat{N}}(\theta_m)\}_{i=0}^{\widehat{N}}$ denote the series of search cost cutoffs obtained from a market segment m (containing $\widehat{N}$ options).

Assumption 4. *The econometrician observes the following data:*

1. A number $M \to \infty$ of different market segments, indexed by m, each of them characterized by a vector θ_m randomly drawn from a distribution with bounded support Θ, and all of them having the same underlying search cost distribution H.

2. In every market segment m with number of options $\widehat{N} \in \mathcal{N}^{\infty}$ the data described in Assumption 3, i.e., $\{G_i^{\widehat{N}}(\cdot, \theta_m)\}_{i=1}^{\widehat{N}}$, where $G_i^{\widehat{N}}$ is continuous in θ_m, *and $\{s_i^{\widehat{N}}(\theta_m)\}_{i=0}^{\widehat{N}}$, $m = 1, 2, ..., \infty$.*

Then we can state that:

Proposition 4 *Under Assumptions 1,2, and 4, the search cost distribution is identified in the following interval:*

$$\cup_{\widehat{N} \in \mathcal{N}^{\infty}} \cup_{i=0}^{\widehat{N}-1} [\min_{\theta \in \overline{\Theta}} c_i^{\widehat{N}}(\theta), \max_{\theta \in \overline{\Theta}} c_i^{\widehat{N}}(\theta)], \tag{8}$$

where $\overline{\Theta}$ is the closure of Θ.

Proof. For any $\widehat{N} \in \mathcal{N}^{\infty}$, $i \in \left\{0, 1, ..., \widehat{N}\right\}$ and $\theta_m \in \Theta$, the cutoff value $c_i^{\widehat{N}}(\theta_m)$ corresponding to i in market segment m with $\widehat{N}$ options and $H\left(c_i^{\widehat{N}}(\theta_m)\right)$ are identified by Proposition 2, for all market segments m with $\widehat{N}$ options. We know that

$$c_i^{\widehat{N}}(\theta_m) = \int_0^{\overline{u}} \left(\prod_{k=1}^{i} G_k(\cdot, \theta_m) - \prod_{k=1}^{i+1} G_k(\cdot, \theta_m) \right) du,$$

so $c_i^{\widehat{N}}(\cdot)$ is a continuous function of θ_m. Since there are infinitely many market segments m with $\widehat{N}$ options and θ_m is drawn randomly from Θ, the realizations $c_i^{\widehat{N}}(\theta_m)$ will be dense in the set $\left\{c_i^{\widehat{N}}(\theta) : \theta \in \overline{\Theta}\right\}$. By the continuity of $c_i^{\widehat{N}}(\cdot)$ in θ and by the compactness of $\overline{\Theta}$, this set is the compact interval $\left[\min_{\theta \in \overline{\Theta}} c_i^{\widehat{N}}(\theta), \max_{\theta \in \overline{\Theta}} c_i^{\widehat{N}}(\theta)\right]$. Therefore, the search cost distribution is identified on this interval. The result follows. ■

3.1 Identification in the full support

Proposition 4 gives set (8), in which the costs of simultaneous search are identified. Such set is the union of a number of intervals so the question that arises is whether such union covers the entire support of the search cost distribution. We now address this issue within the specific context put forward in the examples above. In particular, we provide simple sufficient conditions under which a full identification result obtains.

Example 1. The college problem.

In the college problem, suppose every market/city m has just two colleges. For simplicity, suppose that utilities are constant across markets but success rates vary from market to market. Let u_1 and u_2 denote the common utilities colleges offer the applicants. Let α_1^m and α_2^m denote the success rates of the two colleges in market m. Therefore, in the context of Proposition 4 $\theta_m = (\alpha_1^m, \alpha_2^m)$. Let us assume that θ_m is randomly distributed across markets, with support $\Theta = (0, 1] \times (0, 1]$.

By Assumption 1, in every market $\alpha_1^m u_1 > \alpha_2^m u_2$ so, from Proposition 4, we can identify the costs of simultaneous search in the interval $[0, u_1]$. Whether the distribution of the costs of simultaneous search is identified in its full support then depends on the relationship between u_1 and $\overline{c}$. If $\overline{c}$ is less than or equal to u_1, a full identification result obtains.

Example 2. Search for differentiated products.

Consider the model of search for differentiated products above where the utility from a product i is given by $u_i = \delta_i + \varepsilon_i$. The parameter δ_i represents quality of product i and the term ε_i is a consumer-specific match-value, uniformly distributed across consumers with support $[0, \overline{\varepsilon}]$. Assume match values are independent across consumers. Suppose there are just two products in every market m and assume they only vary in their quality δ_i. If instead of markets we have market segments the parameter δ_i should be seen as a random coefficient. Let $\theta_m = (\delta_1^m, \delta_2^m)$ and assume θ_m is randomly distributed across market (segments) with support $\Theta = \{\delta_1, \delta_2 \in [0, 1] \times [0, 1]$ such that $\delta_1 > \delta_2\}$ which means that Assumption 1 holds so in every market segment m, option 1 is *ex-ante* the best option. Moreover, note that

$$c_0^m = \delta_1^m + \frac{\overline{\varepsilon}}{2} \tag{9}$$

so using this information, we can identify the search cost distribution in the set $\left[\frac{\overline{\varepsilon}}{2}, \frac{\overline{\varepsilon}}{2} + 1\right]$ (see (8)). Notice that when $\overline{\varepsilon}$ is sufficiently large (so that $\frac{\overline{\varepsilon}}{2} + 1 > \overline{c}$), we can identify the search cost distribution at high quantiles. However, note that using the information provided by the series of c_0^m cutoffs, we cannot identify the search cost distribution in the set $\left[0, \frac{\overline{\varepsilon}}{2}\right]$. To gain further information we can use the series of c_1^m cutoffs (see (8)), given by

$$c_1^m = \frac{(\overline{\varepsilon} - (\delta_1^m - \delta_2^m))^3}{6\overline{\varepsilon}^2}$$

Using this information, we can additionally identify the search cost distribution in the set $[0, \frac{\overline{\varepsilon}}{6}]$. As a result, if across markets the products vary only in their quality, we can only obtain identification of the search cost CDF in the set $[0, \frac{\overline{\varepsilon}}{6}] \cup \left[\frac{\overline{\varepsilon}}{2}, \frac{\overline{\varepsilon}}{2} + 1\right]$.

To obtain additional variation so as to identify the search cost distribution in its full support, let us suppose that $\overline{\varepsilon}$ also varies across markets and let $\overline{\varepsilon}$ take on just two values, $\overline{\varepsilon}_1$ and $\overline{\varepsilon}_2$, with $\overline{\varepsilon}_1 \geq 3\overline{\varepsilon}_2$. This suffices to identify the search cost distribution in its full support.

4 Concluding remarks

In this paper we have studied the identification of the costs of simultaneous search in the class of (portfolio) problems studied by Chade and Smith (2006). We have shown that aggregate data from a single market, or disaggregate data from a single market segment, in particular, market shares and utility distributions corresponding to each of the available options, do not provide sufficient information to identify the search cost distribution at any reasonable interval. The reason is that the sequence of critical search cost values that is identified is convergent to zero and therefore not dense in the support of the search cost distribution.

We have provided conditions under which pooling aggregate data from multiple markets, or pooling disaggregate data from multiple market segments, can help identifying the cost of simultaneous search. Within the context of two specific examples, we have shown that full identification results can easily be obtained.

References

[1] Albrecht J., P. A. Gautier, and S. Vroman (2006): "Equilibrium Directed Search with Multiple Applications," *Review of Economic Studies* 73, 869–91.

[2] Anderson, S., and R. Renault (1999): "Pricing, Product Diversity, and Search Costs: a Bertrand-Chamberlin-Diamond model," *RAND Journal of Economics*, 30, 719-735.

[3] Burdett, K. and K. Judd (1983), “Equilibrium Price Dispersion,” *Econometrica* 51, 955-969.

[4] Burdett, K., S. Shi, and R. Wright (2001): “Pricing and Matching with Frictions.” *Journal of Political Economy* 109, 1060–85.

[5] Chade, H. and L. Smith (2005): “Simultaneous Search,” Mimeo, Michigan and Arizona State.

[6] Chade, H. and L. Smith (2006): “Simultaneous Search,” *Econometrica* 74, 1293-1307.

[7] Gale, D. and L. Shapley (1962): “College Admissions and the Stability of Marriage,” *American Mathematical Monthly* 69, 9-15.

[8] Gautier, P. A. and J. L. Moraga-González (2004): “Strategic Wage Setting and Coordination Frictions with Multiple Job Applications,” Tinbergen Institute Discussion Paper # 04-063/1, The Netherlands, June 2004.

[9] Honka, E. (2010): “Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry,” mimeo.

[10] Kelso, A. S., Jr., and V. P. Crawford (1982): “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica* 50, 1483-1504.

[11] Kircher, P. (2009): “Efficiency of Simultaneous Search,” *Journal of Political Economy* 117, 861-913.

[12] Mehta, N., S. Rajiv, K. Srinivasan (2003): “Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation,” *Marketing Science* 22, 58-84.

[13] Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2010): “Consumer Search and Prices in the Automobile Market,” mimeo.

[14] Roth, A. E. and M. Sotomayor (1989): “The College Admissions Problem Revisited,” *Econometrica* 57, 559-570

[15] Stigler, G. (1961): “The Economics of Information,” *Journal of Political Economy* 69, 213-25.

[16] Wolinsky, A. (1986): “True Monopolistic Competition as a Result of Imperfect Information,” *Quarterly Journal of Economics*, 101, 493-511.