Department of Economics

Minimum Distance Estimation of Search Costs using Price Distribution

Fabio Sanches[1]
University of São Paulo

Daniel Silva Junior[1]
City, University of London

Sorawoot Srisuma[1]
University of Surrey

Department of Economics
Discussion Paper Series
No. 16/16

[1] Corresponding authors: Email: fmiessi@gmail.com; Daniel.Silva-Junior@city.ac.uk; s.srisuma@surrey.ac.uk

Minimum Distance Estimation of Search Costs using Price Distribution[*][†]

Fabio Sanches
University of São Paulo

Daniel Silva Junior
City University London

Sorawoot Srisuma
University of Surrey

September 9, 2016

Abstract

Hong and Shum (2006) show equilibrium restrictions in a search model can be used to identify quantiles of the search cost distribution from observed prices alone. These quantiles can be difficult to estimate in practice. This paper uses a minimum distance approach to estimate them that is easy to compute. A version of our estimator is a solution to a nonlinear least squares problem that can be straightforwardly programmed on softwares such as STATA. We show our estimator is consistent and has an asymptotic normal distribution. Its distribution can be consistently estimated by a bootstrap. Our estimator can be used to estimate the cost distribution nonparametrically on a larger support when prices from heterogeneous markets are available. We propose a two-step sieve estimator for that case. The first step estimates quantiles from each market. They are used in the second step as generated variables to perform nonparametric sieve estimation. We derive the uniform rate of convergence of the sieve estimator that can be used to quantify the errors incurred from interpolating data across markets. To illustrate we use online bookmaking odds for English football leagues' matches (as prices) and find evidence that suggests search costs for consumers have fallen following a change in the British law that allows gambling operators to advertise more widely.

JEL Classification Numbers: C13, C15, D43, D83, L13

Keywords: Bootstrap, Generated Variables, M-Estimation, Search Cost, Sieve Estimation

[*]We are very grateful to the authors of Moraga-González, Sándor and Wildenbeest (2013) for sharing their MATLAB code with us. We thank an anonymous referee for helpful comments and suggestions. We would also like to thank Valentina Corradi, Abhimanyu Gupta, Tatiana Komarova, Dennis Kristensen, Arthur Lewbel, and Paul Schrimpf for some helpful discussions.

[†]*E-mail addresses*: fmiessi@gmail.com; danielsjunior@gmail.com; s.srisuma@surrey.ac.uk

1 Introduction

Heterogenous search cost is one of the classic factors that can be used to rationalize price dispersion of homogenous products. E.g., see the seminal work of Stigler (1964). Various empirical models of search have been proposed and applied to numerous problems in economics depending on data availability. Hong and Shum (2006, hereafter HS) show that search cost distributions can be identified from the price data alone. The innovation of HS is very useful since price data are often readily available, for instance in contrast to quantities of products supplied or demanded.

We consider an empirical search model with non-sequential search strategies. HS show the quantiles of the search cost in such model can be estimated without specifying any parametric structure. Although there has been more recent empirical works that extend the original idea of HS to estimate more complicated models of search[1], there are still interests in the identification and estimation of the simpler search model nonparametrically. For examples, Moraga-González, Sándor and Wildenbeest (2013) show how data from different markets can be used to identify the search cost distribution over a larger support and Blevins and Senney (2014) consider a dynamic version of the search model we consider here.

The main insight from HS is that the equilibrium condition can be summarized by an implicit equation relating the price and its distribution, parameterized by the proportions of consumers searching different number of sellers. The latter can be used to recover various quantiles of the search cost distribution. Two main features of the equilibrium condition that lead to an interesting econometric problem are: (i) it imposes a continuum of restrictions since the mixed strategy concept leads to a continuous distribution of price in equilibrium; and, (ii) the observed price distribution is only defined implicitly and cannot be solved out in terms of terms of price and the parameters of interest.

In this paper we make two main methodological contributions that complement existing estimation procedures and make the empirical search model more accessible to empirical researchers.

First, when there are data from a single market, we provide an estimator for the quantiles on the cumulative distribution (cdf) of the search cost that is simple to construct and easy to perform inference on. Our estimator uses all information imposed by the equilibrium condition. We show under very weak conditions that our estimator is consistent and asymptotically normal at a parametric rate. We also show the distribution of our estimator can be approximated consistently by a standard nonparametric bootstrap. The ease of practical use is the distinguishing feature of our estimator compared to the existing ones in nontrivial ways. Its simplest version can be obtained by

[1] E.g. see Hortaçsu and Syverson (2004), De los Santos, Hortaçsu and Wildenbeest (2012), and Moraga-González, Sándor and Wildenbeest (2012)).

defining the distance function using the empirical measure that leads to a nonlinear least squares problem that can be implemented on STATA.

Second, when there is access to data from multiple markets, we propose a two-step sieve estimator that pools data across markets and estimate the cdf of the search cost as a function over a larger support. Single market data can only be used to identify a limited number of quantiles. Our sieve estimator provides a systematic way to combine quantiles from different individual markets. Any estimator in the literature can be used in the first stage, not necessarily the one we propose. The second stage estimation resembles a nonparametric series estimation problem with generated regressor *and* regressand. We provide the uniform rate of convergence for the sieve estimator. Since we know the rate of convergence of quantiles from each individual market, the uniform rate using pooled data can be used to quantify the cost of interpolation across markets.

For estimation HS takes a finite number of quantiles, each one to form a moment condition using the equilibrium restriction written in terms of quantiles, and develop an empirical likelihood estimator that has desirable theoretical properties such as efficiency and small finite sample bias (e.g. see Owens (2001) and Newey and Smith (2004)). However, a finite selection from infinitely many moment conditions may have implications in terms of consistent estimation and not just efficiency (Dominguez and Lobato (2004)). Some preliminary algebra suggests such issue may be relevant in the model of search under consideration. But at the same time, with finite data, it is also not advisable to use arbitrary many moment conditions for empirical likelihood estimation or any other optimal GMM methods due to the numerical ill-posedness associated with efficient moment estimation; see the discussion in Carrasco and Florens (2002). Particularly, a well-known problem with the empirical likelihood objective functions is they typically have many local minima, and the method is generally challenging to program and implement; see Section 8 in Kitamura (2007). Indeed HS also report some numerical difficulties in their numerical work; in their illustration they choose the largest number of quantiles that allow their optimization routine to converge.[2]

Partly motivated by the numerical issues associated with HS's approach, Moraga-González and Wildenbeest (2007, hereafter MGW) propose an inventive way to construct the maximum likelihood estimator by manipulating the equilibrium restriction. They suggest the likelihood procedure is easier to compute and, importantly, is also efficient. However, the numerical aspect in terms of the implementation of their estimator remains non trivial. The difficulty is due to the fact that the probability density function (pdf) of the price is defined implicitly in terms of its cdf, the latter in turns is only known as a solution of a nonlinear equation imposed by the equilibrium. This leads to a constrained likelihood estimation problem with many nonlinear constraints. A naïve programming approach to

[2]Hong and Shum (2006) illustrate their procedure using online price data of some well-known economics and statistics textbooks.

this optimization problem is to directly specify a nested procedure requiring an optimization routine on both the inner and outer loop, where the inner step searches over the parameter space and the outer step solves the nonlinear constraints. A more numerically efficient alternative may be possible by using constrained optimization solvers with algorithms that deal with the nonlinear constraints endogenously. See Su and Judd (2012) for a related discussion and further references.[3]

We take a different approach that is closely related to the asymptotic least squares estimation described in Chapter 9 of Gourieroux and Monfort (1995). Asymptotic least squares method, which can be viewed as an alternative representation to the familiar method of moment estimator, is particularly suited to estimate structural models as the objective functions can often be written to represent the equilibrium condition directly. For examples see the least squares estimators of Pesendorfer and Schmidt-Dengler (2008) and Sanches, Silva Junior and Srisuma (2016) in the context of dynamic discrete games.[4] However, the statistical theory required to derive the asymptotic properties of our estimator in this paper is more complicated than those used in the dynamic discrete games cited above since here we have to deal with a continuum of restrictions instead of a finite number of restrictions. We derive our large sample results using a similar strategy employed in Brown and Wegkamp (2002), who utilize tools from empirical process theory to derive analogous large sample results for a different minimum distance estimator.[5]

The estimator we propose focuses on the ease of practical use but not efficiency. There are at least two obvious ways to improve on the asymptotic variance of our estimator. As alluded above, the equilibrium restriction can also be stated as a continuum of moment conditions. Therefore an efficient estimation in the GMM sense can be pursued by solving an ill-posed inverse problem along the line of Carrasco and Florens (2000).[6] It is arguably even simpler to aim for the fully efficient estimator. For instance we can perform a Newton Raphson iteration once, starting from our easy compute estimate, using the Hessian from the likelihood based objective function proposed by MGW. Then such estimator will have the same first order asymptotic distribution as the maximum likelihood estimator (Robinson (1988)). But, of course, there is no guarantee the asymptotically efficient estimator will perform better than the less efficient one in finite sample.

[3]An important feature for the search model under consideration is that the number of constraints is large and grows with sample size, while many other well-known structural models, such as those associated with dynamic discrete decision problems and games, have a fixed and relatively small number of constraints.

[4]Pesendorfer and Schmidt-Dengler (2008) also illustrate how a moment estimator can be cast as an asymptotic least squares estimator.

[5]They consider a minimum distance estimator defined from a criterion based on a conditional independence condition due to Manski (1983).

[6]Also see a recent working paper of Chaussé (2011), who is extending the estimator of Carrasco and Florens (2000) to a generalized empirical likelihood setting.

When the data come from a single market, an inevitable limitation of the identifying strategy in HS is that only countable points of the distribution of the search cost can be identified. Particularly there is only one accumulation point at the lower support of the cost distribution. In order to identify higher quantiles of the cost distribution, and possibly its full support, Moraga-González, Sándor and Wildenbeest (2013) suggest combining data from different markets where consumers have the same underlying search distribution. In particular they provide conditions under which pooling data sets can be used for identification. In terms of estimation they suggest that interpolating data between markets can be difficult. In order to overcome this, they propose a semi-nonparametric maximum likelihood estimator for the pdf of the search cost. The cdf, which is often a more convenient object to make stochastic comparisons, can then be obtained by integration. However, their semi-nonparametric maximum likelihood procedure is complicated as it solves a highly nonlinear optimization problem with many parameters. They show their estimator can consistently estimate the distribution of the search cost where the support is identified but do not provide the convergence rate.[7]

Building on the semi-nonparametric idea, we propose a two-step sieve least squares estimator for the cdf of the search cost. The estimation problem involved can also be seen as an asymptotic least squares problem where the parameter of interest is an infinite dimensional object instead of a finite dimensional one. We show that sieve estimation is a convenient way to systematically combine data from different markets. It can be used in conjunction with any aforementioned estimation method, not necessarily with the minimum distance estimator we propose in this paper. In the first stage an estimation procedure is performed for each individual market. In the second stage we use the first-step estimators as generated variables and perform sieve least squares estimation. Our sieve estimator is easy to compute as it only involves ordinary least squares estimation. We provide the uniform rate of convergence for our estimator. The ability to derive uniform rate of convergence is important as it gives us a guidance on the cost of estimation the entire function compared to at just some finite points, which we know to converge at a parametric rate within each market.

The large sample properties of our sieve estimator are not immediately trivial to verify. In practice our second stage least squares procedure resembles that of a nonparametric regression problem with generated regressors and generated regressands. There has been much recent interest in the econometrics and statistics literature on the general theory of estimation problems involving generated regressors in the nonparametric regression context (e.g., see Escanciano, Jacho-Chávez and Lewbel (2012, 2014) and Mammen, Rothe and Schienle (2012, 2014)). Problems with generated variables on both sides of the equation seem less common. Furthermore, the asymptotic least squares

[7] Details can be found in the supplementary materials to Moraga-González, Sándor and Wildenbeest (2013), available at, http://qed.econ.queensu.ca/jae/2013-v28.7/moraga-gonzalez-sander-wildenbeest/.

framework generally differs from a regression model. We are not aware of any general results for an asymptotic least squares estimation of an infinite dimensional object. However, our problem is somewhat simpler to handle relative to the cited works above since our generated variables converge at the parametric rate rather than nonparametric. We derive the properties for our sieve estimator under the framework that the data have a pooled cross section structure. Our approach to derive the uniform rate of convergence is general and can be used in other asymptotic least squares problems.

We conduct a small scale Monte Carlo experiment to compare our proposed estimators with other estimators in the literature. Then we illustrate our procedures using real world data. We estimate the search costs using online odds, to construct prices, for English football leagues matches in the 2006/7 and 2007/8 seasons. There is an interesting distinction between the two seasons that follows from the United Kingdom (UK) passing of a well-known legistration that allows bookmakers to advertise more freely after the 2006/7 season has ended. We consider the top two English football leagues: the Premier League (top division) and the League Championship (2nd division). We treat the odds for matches from each league as coming from different markets. We find that the search costs generally have fallen following the change in the law as expected.

We present the model in Section 2, and then we define our estimator and briefly discuss its relation to existing estimators in the literature. Section 3 gives the large sample theorems for our estimator that uses data from a single market. Section 4 assumes we have data from different markets; we define our sieve estimator for the cdf of the search cost and give its uniform rate of convergence. Section 5 is the numerical section containing a simulation study and empirical illustrations. Section 6 concludes. All proofs can be found in the Appendix.

2 Model, Equilibrium Restrictions and Estimation

The empirical model in HS relies on theoretical result of Burdett and Judd (1983). The model assumes there are continuums of consumers and sellers. Consumers are heterogenous, differing by search costs drawn from some continuous distribution with a cdf, G. Sellers have identical marginal cost, r, and sell the same product; they only differ by the price they set. Each consumer has an inelastic demand for one unit of the product with the valuation of $\overline{p}$. Since search is costly, her optimal strategy is to visit the smallest number of sellers given her beliefs on the price distribution sellers use. In a symmetric mixed strategy equilibrium each seller sets a price that maximizes its expected profit given the consumers' search strategies, and the distribution of prices set by the sellers is consistent with the consumers' beliefs. Since the number of sellers observed in the data is often small we assume there are $K < \infty$ sellers. An equilibrium continuous price distribution, as the symmetric equilibrium strategy employed by all firms, is known to exist for a given set of primitives $(G, \overline{p}, r, K)$; see Moraga-González, Sándor and Wildenbeest (2010). We denote the cdf of the equilibrium price distribution by F. The constancy of the seller's equilibrium profit is our starting point:

$$\Pi(p,r) = (p-r)\sum_{k=1}^{K} kq_k (1-F(p))^{k-1} \quad \text{s.t. } \Pi(p,r) = \Pi(p',r) \text{ for all } p,p' \in \mathcal{S}_P, \tag{1}$$

where $\mathcal{S}_P = [\underline{p}, \overline{p}]$ is the support of P_i for some $0 < \underline{p} < \overline{p} < \infty$, and q_k is the equilibrium proportion of consumer searching k times for $1 \leq k \leq K$. Once $\{q_k\}_{k=1}^{K}$ are known, they can be used to recover the quantiles of the search cost distribution from the identity:

$$q_k = \begin{cases} 1 - G(\Delta_1), & \text{for } k = 1 \\ G(\Delta_{k-1}) - G(\Delta_k), & \text{for } k = 2, \ldots, K-1 \\ G(\Delta_{K-1}), & \text{for } k = K \end{cases}, \tag{2}$$

where $\Delta_k = E[P_{1:k}] - E[P_{1:k+1}]$ and $E[P_{1:k}]$ denotes the expected minimum price from drawing prices from $k-$ i.i.d. sellers, which is identified from the data. For further details and discussions regarding the model we refer the reader to HS, MGW and also Moraga-González, Sándor and Wildenbeest (2010).

The econometric problem of interest in this and the next sections is to first estimate $\{q_k\}_{k=1}^{K}$ from observing a random sample of equilibrium prices $\{P_i\}_{i=1}^{N}$, and then use them to recover identified points of the search cost distribution: $\{(\Delta_k, G(\Delta_k))\}_{k=1}^{K-1}$. First note that we can concentrate out the marginal cost by equating $\Pi(\overline{p}, r)$ and $\Pi(\underline{p}, r)$,

$$r(\mathbf{q}) = \frac{\overline{p} q_1 - \underline{p} \sum_{k=1}^{K} kq_k}{q_1 - \sum_{k=1}^{K} kq_k}. \tag{3}$$

Following HS, the equilibrium condition for the purpose of estimation can be obtained from equating from $\Pi(\overline{p}, r) = \Pi(p, r)$ for all p. In particular, this relation can be written as

$$p = r(\mathbf{q}) + \frac{(\overline{p} - r(\mathbf{q}))\, q_1}{\sum_{k=1}^{K} kq_k (1-F(p))^{k-1}} \text{ for all } p \in \mathcal{S}_P. \tag{4}$$

Before we introduce our estimator we now briefly explain how the equations above have been used for estimation in the literature.

Empirical Likelihood (Hong and Shum (2006))

Since P_i has a continuous distribution the inverse of F, denoted by F^{-1}, exists so that $p = F^{-1}(F(p))$ for all p. Note that equation (4) is equivalent to

$$F(p) = F\left(r(\mathbf{q}) + \frac{(\overline{p} - r(\mathbf{q}))\, q_1}{\sum_{k=1}^{K} kq_k (1-F(p))^{k-1}} \right).$$

Then choose finite V quantile points, $\{s_l\}_{l=1}^{V}$, so that $s_l \in [0,1]$ is the s_l-th quantile. HS develop an empirical likelihood estimator of $\mathbf{q}$ based on the following V moment conditions:

$$h(\mathbf{q}; s_l) = E\left[s_l - \mathbf{1}\left[P_i \leq r(\mathbf{q}) + \frac{(\overline{p} - r(\mathbf{q}))\, q_1}{\sum_{k=1}^{K} k q_k (1 - s_l)^{k-1}}\right]\right] \text{ for } l = 1, \ldots, V,$$

where $\mathbf{1}[\cdot]$ denotes an indicator function. Clearly one needs to choose $V \geq K - 1$, where the minus one comes from the restriction $\sum_{k=1}^{K} q_k = 1$.

In theory we would like to choose as many moment conditions as possible. However, there are practical costs and implementation issues in finite sample as explained in the Introduction. At the same time, in principle, choosing too few moment conditions can lead to an identification problem. We illustrate the latter point in the spirit of the illustrating examples in Dominguez and Lobato (2004).

Suppose $K = 2$, and we use $q_2 = 1 - q_1$, so $r(\mathbf{q}) = \frac{\overline{p} q_1 - \underline{p}(q_1 + 2(1-q_1))}{-2(1-q_1)}$. For any $s_0 \in [0,1]$, the moment condition becomes:

$$E\left[s_0 - \mathbf{1}\left[P_i \leq \frac{(\overline{p} - r)\, q_1}{q_1 + 2(1-q_1)(1-s_0)} - \frac{(\overline{p} - \underline{p})\, q_1 - 2\underline{p}(1-q_1)}{2(1-q_1)}\right]\right],$$

then for some p_0 that satisfies $s_0 = F(p_0)$, q_1 must satisfy

$$p_0 = \frac{\left(\overline{p} + \frac{\overline{p} q_1 - \underline{p}(q_1 + 2(1-q_1))}{2(1-q_1)}\right) q_1}{q_1 + 2(1-q_1)(1-F(p_0))} - \frac{(\overline{p} - \underline{p})\, q_1 - 2\underline{p}(1-q_1)}{2(1-q_1)}.$$

By multiplying the denominators across and re-arranging the equation above, by inspection, it is easy to see we have an implicit function $T(q_1, p_0, F(p_0)) = 0$ such that, for every pair $(p_0, F(p_0))$, $T(q_1, p_0, F(p_0))$ is a quadratic function of q_1. However this suggests there are potentially two distinct values for q_1 that satisfy the same moment condition for a given s_0, in which case it may not be possible give have a consistent estimator for q_1 based on one particular quantile.

More generally, when $K > 2$, each s_l leads to an equation for a general ellipse in $\mathbb{R}^{K-1}$ whose level set at zero can represent the values of proportions of consumers that satisfy the moment condition associated with each quantile level s_l.[8] Therefore, with any estimator based on (4), one may be inclined to incorporate all conditions for the purpose of consistent estimation.

Maximum Likelihood (Moraga-González and Wildenbeest (2007))

Let the derivative of F, i.e. the pdf, by f. By differentiating equation (4) and solve out for f, the implicit function theorem yields:

$$f(p) = \frac{\sum_{k=1}^{K} k q_k (1 - F(p))^{k-1}}{(p - r(\mathbf{q})) \sum_{k=2}^{K} k(k-1) q_k (1 - F(p))^{k-2}} \text{ for all } p \in \mathcal{S}_P. \tag{5}$$

[8] A natural restriction for a proportion can be used to rule out any complex value as well as other reals outside $[0,1]$.

MGW suggest a maximum likelihood procedure based on maximizing, with respect to $\mathbf{q}$, the following likelihood function:

$$\widetilde{f}\left(\mathbf{q};p\right)=\frac{\sum_{k=1}^{K}kq_{k}\left(1-\widetilde{F}\left(\mathbf{q};p\right)\right)^{k-1}}{\left(p-r\left(\mathbf{q}\right)\right)\sum_{k=2}^{K}k\left(k-1\right)q_{k}\left(1-\widetilde{F}\left(\mathbf{q};p\right)\right)^{k-2}}\text{ for }p\in\mathcal{S}_{P},\tag{6}$$

where $\widetilde{F}\left(\mathbf{q};p\right)$ is restricted to satisfy equation (4). In practice supposed the observed prices are $\{P_i\}_{i=1}^{N}$. Then for each candidate $\mathbf{q}'$, and each i, $\widetilde{F}\left(\mathbf{q}';P_i\right)$ can be chosen to satisfy the equilibrium restriction by imposing that it solves: $0=P_i-r\left(\mathbf{q}'\right)-\frac{\left(\overline{p}-r\left(\mathbf{q}'\right)\right)q_1'}{\sum_{k=1}^{K}kq_k'\left(1-\widetilde{F}\left(\mathbf{q}';P_i\right)\right)^{k-1}}$. However, it may not be a trivial numerical task to fully respect equation (4). Particularly $\widetilde{F}\left(\mathbf{q}';P_i\right)$ generally does not have a closed-form expression and is only known to be a root of some $(K-1)-$th order polynomial. Such polynomial always have multiple roots. The multiplicity issue can be migitated by imposing constraints that $\widetilde{F}\left(\mathbf{q}_0;P_i\right)$ must be real and take values between 0 and 1, and it must be non-decreasing in P_i.

Minimum Distance

We propose to use the equilibrium condition directly to define objective functions, rather than posing it as (a continuum of) moment conditions. In particular, in contrast to HS, we use equation (4) without passing the equilibrium restriction through the function F. This approach can be seen as a generalization of the asymptotic least squares estimator described in Gourieroux and Monfort (1995) when there is a continuum of restrictions. It will be also convenient to eliminate the denominators to rule out any possibilities of division near zero that may occur if q_1 is close to 0 and p approaches 1. We first substituting in for $r\left(\mathbf{q}\right)$, from (3), then equation (4) can be simplified to:

$$\left(\sum_{k=1}^{K}kq_k\left(1-F\left(p\right)\right)^{k-1}\right)\left(\left(p-\overline{p}\right)q_1-\left(p-\underline{p}\right)\sum_{k=1}^{K}kq_k\right)=q_1\left(\left(\underline{p}-\overline{p}\right)\sum_{k=1}^{K}kq_k\right).$$

Next we concentrate out q_K in the above equation by replacing it with $1-\sum_{k=1}^{K-1}q_k$, which leads to the following restriction:

$$\begin{aligned}0 &= q_1\left(\left(\underline{p}-\overline{p}\right)\left(K-\sum_{k=1}^{K-1}\left(k-K\right)q_k\right)\right) &(7)\\ &-\left(\begin{array}{c}K\left(1-\sum_{k=1}^{K-1}q_k\right)\left(1-F\left(p\right)\right)^{K-1}\\+\sum_{k=1}^{K-1}kq_k\left(1-F\left(p\right)\right)^{k-1}\end{array}\right)\times\left(\left(p-\overline{p}\right)q_1-\left(p-\underline{p}\right)\left(K-\sum_{k=1}^{K-1}\left(k-K\right)q_k\right)\right).\end{aligned}$$

this must hold for all $p\in\mathcal{S}_P$.

Note that the equation above can be re-written as a polynomial in $\mathbf{q}_{-K}\equiv\left(q_1,\ldots,q_{K-1}\right)$, which is always smooth independently of F. In contrast to the moment condition considered in HS that has $\mathbf{q}$ as an argument of the unknown function F; using the empirical cdf to construct the objective function in the latter case introduces non-smoothness in the estimation problem.

3 Estimation with Data from a Single Market

We use the equilibrium restiction (7) to define an econometric model $\{m(\cdot,\theta)\}_{\theta\in\Theta}$ so that $m(\cdot,\theta):\mathcal{S}_P\to\mathbb{R}$, where for all $p\in\mathcal{S}_P$:

$$\begin{aligned} m(p,\theta) &= \theta_1\left(\left(\underline{p}-\overline{p}\right)\left(K-\sum_{k=1}^{K-1}(k-K)\,\theta_k\right)\right) \qquad (8)\\ &- \left(\begin{array}{c} K\left(1-\sum_{k=1}^{K-1}\theta_k\right)(1-F(p))^{K-1} \\ +\sum_{k=1}^{K-1}\theta_k\,(1-F(p))^{k-1} \end{array}\right)\times\left(\left(p-\overline{p}\right)\theta_1-\left(p-\underline{p}\right)\left(K-\sum_{k=1}^{K-1}(k-K)\,\theta_k\right)\right), \end{aligned}$$

and $\theta=(\theta_1,\ldots,\theta_{K-1})$ is an element in the parameter space $\Theta=[0,1]^{K-1}$. We assume the model is well-specified so that $m(p,\theta)=0$ for all $p\in\mathcal{S}_P$ when θ equals $\mathbf{q}_{-K}$. Given a sample $\{P_i\}_{i=1}^N$, we define the empirical cdf as:

$$F_N(p)=\frac{1}{N}\sum_{i=1}^N\mathbf{1}\left[P_i\leq p\right]\quad\text{for all } p\in\mathcal{S}_P.$$

We then define $m_N(p,\theta)$ as the sample counterpart of $m(p,\theta)$ where F is replaced by F_N. And we propose a minimum distance estimator based on

$$\min_{\theta\in\Theta}\int\left(m_N(p,\theta)\right)^2\mu_N(dp),$$

where μ_N is a sequence of positive and finite, possibly random, measures.

We now define respectively the limiting and sample objective functions, and our estimator:

$$\begin{aligned} M_N(\theta) &= \int\left(m_N(p,\theta)\right)^2\mu_N(dp),\\ M(\theta) &= \int\left(m(p,\theta)\right)^2\mu(dp),\\ \widehat{\theta} &= \arg\min_{\theta\in\Theta}M_N(\theta). \end{aligned}$$

We shall denote the probability measure for the observed data by $\mathbb{P}$, and the probability measure for the bootstrap sample conditional on $\{P_i\}_{i=1}^N$ by $\mathbb{P}^*$. In what follows we use $\xrightarrow{a.s.}$, $\xrightarrow{p}$ and $\xrightarrow{d}$ to denote convergence almost surely, in probability, and in distribution respectively, with respect to $\mathbb{P}$ as $N\to\infty$. We let $\xrightarrow{a.s.^*}$, $\xrightarrow{p^*}$ denote convergence almost surely and in probability respectively, with respect to $\mathbb{P}^*$.

We adopt the same data generating environment as in HS and MGW.

Assumption A1. $\{P_i\}_{i=1}^{N}$ *is an i.i.d. sequence of continuous random variables with bounded pdf whose cdf satisfies the equilibrium condition in (4).*

Let θ_0 denote $\mathbf{q}_{-K}$. We now provide conditions for our estimator to be consistent and asymptotically normal.

Assumption A2. *(i)* $m(P_i,\theta) = 0$ *almost surely if and only if* $\theta = \theta_0$*; (ii)* μ_N *almost surely converges weakly to a non-random finite measure* μ *that dominates the distribution of* P_i*;*[9] *(iii)* $\int \frac{\partial}{\partial\theta} m(p,\theta_0) \frac{\partial}{\partial\theta^\top} m(p,\theta_0)\, \mu(dp)$ *is invertible.*

A2(i) is the point-identification assumption on the equilibrium condition. It is generally difficult to provide a more primitive condition for identification in a general nonlinear system of equations, e.g. see the results in Komunjer (2012) for a parametric model with finite unconditional moments. Our equilibrium condition presents a continuum of identifying restrictions. A2(ii) allows us to construct objective functions using random measures or otherwise, the domination condition ensures identification of θ_0 is preserved. Examples for measures that satisfy A2(ii) include the uniform measure on $\mathcal{S}_P$, in this case $\mu_N = \mu$ for all N, and a natural candidate for a random measure is the empirical measure from the observed data. For the latter, weak convergence is ensured if the class of functions under consideration is $\mu-$Glivenko-Cantelli, which can be verified using the methods discussed in Andrews (1994) and Kosorok (2008). A2(iii) assumes the usual local positive definiteness condition that follows from the Taylor expansion of the derivative of M around θ_0.

Theorem 1 *(Consistency). Under Assumptions A1, A2(i) and A2(ii),* $\widehat{\theta} \overset{a.s.}{\to} \theta_0$.

Theorem 2 *(Asymptotic Normality). Under Assumption A1 and A2,*

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \overset{d}{\to} \mathcal{N}\left(0, H^{-1}\Sigma H^{-1}\right), \text{ such that}$$

$$\Sigma = \mathrm{Var}\left(2\int \eta(p)\, \mathbb{B}(F(p))\, \mu(dp)\right), \tag{9}$$

$$H = 2\int \frac{\partial}{\partial\theta} m(p,\theta_0) \frac{\partial}{\partial\theta^\top} m(p,\theta_0)\, \mu(dp), \tag{10}$$

where $\mathbb{B}(F(p))$ *and* $\eta(p)$ *are defined in the Appendix (see equations (16) and (19) respectively).*

Given the regular nature of our criterion function, we obtain a root$-N$ consistency result as expected. The asymptotic variance of $\widehat{\theta}$ can be consistently estimated using its sample counterparts.

[9] Let $\mathcal{D}$ be a space of bounded functions defined on $\mathcal{S}_P$. We say μ_N almost surely converges weakly to a μ if $\int \varphi(p)\, \mu_N(dp) - \int \varphi(p)\, \mu(dp) \overset{a.s.}{\to} 0$ for every $\varphi \in \mathcal{D}$.

But this can be a cumbersome task. Specifically, although H is relatively easy to estimate, Σ requires estimating moments involving (a functional of) a Brownian bridge. Perhaps a more convenient method for inference can be performed base on resampling. Our next result shows the ordinary nonparametric bootstrap can be used to imitate the distribution of $\widehat{\theta}$ stated in Theorem 2.

Let $\{P_i^*\}_{i=1}^N$ denote a random sample drawn from the empirical distribution from $\{P_i\}_{i=1}^N$. For some positive and finite measure μ_N^*, let $M_N^*(\theta) = \int (m_N^*(p,\theta))^2 \mu_N^*(dp)$, where $m_N^*(p,\theta)$ is defined in the same way as $m_N(p,\theta)$ but based on the bootstrap sample instead of the original data set. We can then construct a minimum distance estimator using the bootstrap sample:

$$\widehat{\theta}^* = \arg\min_{\theta\in\Theta} M_N^*(\theta). \tag{11}$$

In addition we require is that μ_N^* has to be chosen in a similar manner to μ_N. The following statements are made conditioning on $\{P_i\}_{i=1}^N$.

Assumption A3. *μ_N^* almost surely converges weakly to μ_N.*

A3, as with A2(ii), is not necessary for nonrandom measures (cf. Brown and Wegkamp (2002)). However, it ensures the validity of a fully automated resampling process for instance when the empirical measure is use for μ_N and μ_N^*, with respect to the observed and resampled data.

Theorem 3 *(Bootstrap Consistency). Under Assumptions A1 to A3, $\sqrt{N}\left(\widehat{\theta}^* - \widehat{\theta}\right)$ converges in distribution to $\mathcal{N}(0, H^{-1}\Sigma H^{-1})$ in probability.*

Theorem 3 ensures that nonparametric bootstrap can be used to consistently estimate the distribution of $\sqrt{N}\left(\widehat{\theta} - \theta_0\right)$. Subsequently we can perform inference on θ_0 via bootstrapping.

By construction $\widehat{\theta}$ is the estimator for $\mathbf{q}_{-K}$. Then a natural estimator for q_K is $\widehat{\theta}_K \equiv 1 - \sum_{k=1}^{K-1}\widehat{\theta}_k$. The large sample properties of $\widehat{\theta}_K$ and the ability to bootstrap its distribution follow immediately from applications of various versions of the continuous mapping theorem.

Corollary 1 *(Large Sample Properties of $\widehat{\theta}_K$). Under Assumptions A1 and A2: (i) $\widehat{\theta}_K \xrightarrow{a.s.} q_K$; (ii) for some real $\sigma_K > 0$: $\sqrt{N}\left(\widehat{\theta}_K - q_K\right) \xrightarrow{d} \mathcal{N}(0, \sigma_K^2)$; and with the addition of Assumption A3, let $\widehat{\theta}_K^* = 1 - \sum_{k=1}^{K-1}\widehat{\theta}_k^*$ where $\widehat{\theta}^*$ is defined as in (11), then $\sqrt{N}\left(\widehat{\theta}_K^* - \widehat{\theta}_K\right)$ converges in distribution to $\mathcal{N}(0, \sigma_K^2)$ in probability.*

We state our Corollary 1, and subsequent Corollaries, without proof. The consistency result follows from Slutzky's theorem. The distribution theory can be obtained by the delta-method. Finally, the consistency of the bootstrap follows from the continuous mapping theorem. The validity of these smooth transformation results are standard, e.g. see Kosorok (2008). Although the asymptotic variances in all three Corollaries can be consistently bootstrapped, for completeness, we also provide their explicit forms in the Appendix.

We next turn to the distribution theory for the estimators of $\{\Delta_k\}_{k=1}^{K-1}$ and $\{G(\Delta_k)\}_{k=1}^{K-1}$. Recall that $\Delta_k = E[P_{1:k}] - E[P_{1:k+1}]$, so one candidate estimator for Δ_k is simply its empirical counterpart. Alternatively, in order to apply the same type of argument used for Corollary 1, we will employ an alternate identity for Δ_k that can be obtained from an integration by parts as shown in MGW (see equations (7) and (8) in their paper):

$$\begin{aligned} \Delta_k &= \int_{z=0}^{1} w(z;\mathbf{q})\left[(k+1)z-1\right](1-z)^{k-1}\,dz, \quad \text{for } k=1,\ldots,K-1 \text{ where} \\ w(z;\mathbf{q}) &= \frac{q_1(\overline{p}-r(\mathbf{q}))}{\sum_{k'=1}^{K} k' q_{k'}(1-z)^{k'-1}} + r(\mathbf{q}), \text{ for } z\in[0,1]. \end{aligned} \tag{12}$$

In what follows we define $\widehat{\Delta}_k$ as the feasible version of Δ_k in the above display where $w(\cdot;\mathbf{q})$ is replaced by $w(\cdot;\widehat{\theta})$, and with $\widehat{\theta}_K$ in place of q_K. Therefore $\widehat{\Delta}_k$ is necessarily a smooth function of $\widehat{\theta}$. For $\{G(\Delta_k)\}_{k=1}^{K-1}$, simple manipulation of equation (2) leads to:

$$\begin{aligned} G(\Delta_1) &= 1-q_1, \\ G(\Delta_2) &= 1-q_1-q_2, \\ \vdots &= \vdots \\ G(\Delta_{K-1}) &= 1-q_1-\ldots-q_{K-1}. \end{aligned} \tag{13}$$

The above system of equations can also be found in HS (equation A6, pp. 273). We define $\widehat{G}(\Delta_k)$ by replacing $\mathbf{q}$ by $\widehat{\theta}$, which is also a smooth function of $\widehat{\theta}$. Therefore the consistency and asymptotic distribution, as well as validity of the bootstrap, for $\{\widehat{\Delta}_k\}_{k=1}^{K-1}$ and $\{\widehat{G}(\Delta_k)\}_{k=1}^{K-1}$ immediately follow.

Corollary 2 *(Large Sample Properties of $\{\widehat{\Delta}_k\}_{k=1}^{K-1}$). Under Assumptions A1 and A2: (i) $\widehat{\Delta}_k \overset{a.s.}{\to} \Delta_k$ for all $k=1,\ldots,K-1$; (ii) for some positive definite matrix Ω_Δ:*

$$\sqrt{N}\left(\widehat{\mathbf{\Delta}}-\mathbf{\Delta}\right) \equiv \sqrt{N}\begin{pmatrix} \widehat{\Delta}_1-\Delta_1 \\ \vdots \\ \widehat{\Delta}_{K-1}-\Delta_{K-1} \end{pmatrix} \overset{d}{\to} \mathcal{N}(0,\Omega_\Delta);$$

and with the addition of Assumption A3, let $\{\widehat{\Delta}_k^*\}_{k=1}^{K-1}$ be defined using equation (12) where $w(z;\mathbf{q})$ is replaced by $w(z;\widehat{\theta}^*)$ and $\widehat{\theta}^*$ is defined in (11), then $\sqrt{N}\left(\widehat{\mathbf{\Delta}}^*-\widehat{\mathbf{\Delta}}\right)$ converges in distribution to $\mathcal{N}(0,\Omega_\Delta)$ in probability.

Corollary 3 *(Large Sample Properties of $\{\widehat{G}(\Delta_k)\}_{k=1}^{K-1}$). Under Assumptions A1 and A2: (i) $\widehat{G}(\Delta_k) \overset{a.s.}{\to} G(\Delta_k)$ for all $k = 1, \ldots, K-1$; (ii) for some positive definite matrix Ω_G:*

$$\sqrt{N}\left(\widehat{\mathbf{G}} - \mathbf{G}\right) \equiv \sqrt{N}\begin{pmatrix} \widehat{G}(\Delta_1) - G(\Delta_1) \\ \vdots \\ \widehat{G}(\Delta_{K-1}) - G(\Delta_{K-1}) \end{pmatrix} \overset{d}{\to} \mathcal{N}(0, \Omega_G);$$

and with the addition of Assumption A3, let $\{\widehat{G}^*(\Delta_k)\}_{k=1}^{K-1}$ be defined using equation (13) where θ is replaced by $\widehat{\theta}^*$, which is defined in (11), then $\sqrt{N}\left(\widehat{\mathbf{G}}^* - \widehat{\mathbf{G}}\right)$ converges in distribution to $\mathcal{N}(0, \Omega_G)$ in probability.

4 Pooling Data from Multiple Markets

In this section we show how data from different markets can be combined to estimate G. When the data come from a single market, we can only identify and estimate the cost distribution at finite cut-off points, $\{(\Delta_k, G(\Delta_k))\}_{k=1}^{K-1}$, since there is only a finite number of sellers that consumers can search from (see Proposition 1 in Moraga-González, Sándor and Wildenbeest (2013, hereafter MGSW)). Even if we allow the number of firms to be infinite, since Δ_k is decreasing in k and accumulates at zero, we would still not be able to identify any part of the cost distribution above Δ_1 (see the discussion in HS). One solution is to look across heterogenous markets. Proposition 2 in MGSW provides a sufficient set of conditions for the identification of G over a larger part, or possibly all, of its support based on using the data from different markets that are generated by consumers who endow the same search cost distribution of consumers but may differ in their valuations of the product, and the number of sellers and pricing strategy may also differ across markets.

MGSW suggest a semi-nonparametric method based on maximum likelihood estimation to estimate the cost distribution in one piece instead of combining different estimates of G across markets in some ad hoc fashion.[10] However, it is also quite simple to use estimates from individual markets to estimate G nonparametrically in a systematic manner. Here we describe one method based on using a sieve in conjunction with a simple least squares criterion.

Suppose there are T independent markets, where for each t we observe a random sample of prices $\{P_i^t\}_{i=1}^{N^t}$ with a common distribution described by a cdf F^t that is generated from the primitive $(G, \overline{p}^t, r^t, K^t)$. For each market t we can first estimate $\{q_k^t\}_{k=1}^{K^t}$ and use equation (2) to estimate $\{G(\Delta_k^t)\}_{k=1}^{K^t}$, where $\{q_k^t\}_{k=1}^{K^t}$ and $\{\Delta_k^t\}_{k=1}^{K^t}$ are the equilibrium proportions of search and the corresponding cut-off points in the cost distribution. Proposition 2 in MGSW provides conditions where G can be identified on $\mathcal{S}_C = \left[\underline{C}, \overline{C}\right]$, where $\underline{C} = \lim_{T\to\infty} \inf_{1\leq t\leq T} \Delta_1^t \leq \infty$ and

[10] They actually estimate the pdf of the search cost. It is then integrated to get the cdf.

$\overline{C} = \lim_{T\to\infty} \sup_{1\le t\le T} \Delta_1^t \le \infty$. In particular the degree of heterogeneity across different markets determines how close $\mathcal{S}_C$ is to the full support of the cost distribution.

Recall from (13) that each $G(\Delta_k^t)$ is expressed only in terms of $\{q_k^t\}_{k=1}^{K^t}$, particularly for each t we have:

$$\begin{aligned} G(\Delta_1^t) &= 1 - q_1^t, \\ G(\Delta_2^t) &= 1 - q_1^t - q_2^t, \\ \vdots &= \vdots \\ G(\Delta_{K^t-1}^t) &= 1 - q_1^t - \ldots - q_{K^t-1}^t. \end{aligned} \tag{14}$$

We define the squared Euclidean norm of the discrepancies for this vector of equations when G is replaced by any generic function g that belongs to some space of functions $\mathcal{G}$ by:

$$\psi^t(\mathbf{W}^t, g) = \sum_{k=1}^{K^t-1} \left(\left(1 - \sum_{k'=1}^{k} q_{k'}^t\right) - g(\Delta_k^t) \right)^2,$$

where $\mathbf{W}^t = (q_1^t, \ldots, q_{K^t-1}^t, \Delta_1^t, \ldots, \Delta_{K^t-1}^t)$. By construction $\psi^t(\mathbf{W}^t, G) = 0$. We can then combine these functions across all markets and define:

$$\Psi_T(g) = \frac{1}{T} \sum_{t=1}^{T} \psi^t(\mathbf{W}^t, g). \tag{15}$$

The key identifying condition for us is that $\Psi_T(G) = 0$. There are other distances that one can choose to define ψ^t, and also different ways to combine them across markets. We choose this particular functional form of the loss function for its simplicity. Particularly Ψ_T is just a sum of squares criterion that is similar to those studied in the series nonparametric regression literature (e.g. see Andrews (1991) and Newey (1997)) when $\left\{1 - \sum_{k'=1}^{k} q_{k'}^t\right\}_{k=1,t=1}^{K^t-1,T}$ and $\{\Delta_k^t\}_{k=1,t=1}^{K^t-1,T}$ are treated as regressands and regressors respectively. By using series approximation to estimate G our estimator is an example of a general sieve least squares estimator. An extensive survey on sieve estimation can be found in Chen (2007).

Before proceeding further we introduce some additional notations. For any positive semi-definite real matrix A we let $\underline{\lambda}(A)$ and $\overline{\lambda}(A)$ denote respectively the minimal and maximal eigenvalues of A. For any matrix A, we denote the spectral norm by $\|A\| = \overline{\lambda}\left(A^\top A\right)^{1/2}$, and its Moore-Penrose pseudo-inverse by A^-. We let $\mathcal{G}$ to denote some space of real-valued function defined on $\mathcal{S}_C$. We denote sieves by $\{\mathcal{G}_T\}_{T\ge1}$, where $\mathcal{G}_T \subseteq \mathcal{G}_{T+1} \subseteq \mathcal{G}$ for any integer T. For any function g in $\mathcal{G}_T$, or in $\mathcal{G}$, we let $|g|_\infty = \sup_{c\in\mathcal{S}_C} |g(c)|$. For random real matrices V_N and positive numbers b_N, with $N \ge 1$, we define $V_N = O_p(b_N)$ as $\lim_{\varsigma\to\infty} \limsup_{n\to\infty} \Pr[\|V_N\| > \varsigma b_N] = 0$, and define $V_N = o_p(b_N)$ as $\lim_{N\to\infty} \Pr[\|V_N\| > \varsigma b_N] = 0$ for any $\varsigma > 0$. For any two sequences of positive numbers b_{1N} and

b_{2N}, the notation $b_{1N} \asymp b_{2N}$ means that the ratio b_{1N}/b_{2N} is bounded below and above by positive constants that are independent of n.

Sieve Least Squares Estimation

We start with the infeasible problem where we assume to know $\{\mathbf{W}^t\}_{t=1}^T$. We estimate G on $\mathcal{S}_C$ using a sequence of basis functions $\{g_{lL}\}_{l=1}^L$ that span $\mathcal{G}_T$, where $g_{lL} : \mathcal{S}_C \to \mathbb{R}$ for all $l = 1, \ldots, L$ with L being an increasing integer in T, and L is short for $L(T)$. We use $g^L(c)$ to denote $(g_{1L}(c), \ldots, g_{LL}(c))^\top$ for any $c \in \mathcal{S}_C$, and $\mathbf{g} = \left(g^L(\Delta_1^1), \ldots, g^L\left(\Delta_{K^1-1}^1\right), \ldots, g^L\left(\Delta_1^T\right), \ldots, g^L\left(\Delta_{K^t-1}^T\right)\right)^\top$. Let $\boldsymbol{\iota}$ denote a $\sum_{t=1}^T (K^t - 1)$ $-$vector of ones, and $\mathbf{y} = \boldsymbol{\iota} - \left(q_1^1, \ldots, \sum_{k=1}^{K^1-1} q_k^1, \ldots, q_1^T, \ldots, \sum_{k=1}^{K^T-1} q_k^T\right)^\top$. Then the least squares coefficient from minimizing the sieve objective function is:

$$\widetilde{\beta} = \left(\mathbf{g}^\top \mathbf{g}\right)^- \mathbf{g}^\top \mathbf{y}.$$

We denote our infeasible sieve estimator for G by $\widetilde{G}$, where

$$\widetilde{G}(c) = g^L(c)^\top \widetilde{\beta}.$$

However, we do not observe $\{\mathbf{W}^t\}_{t=1}^T$. Our feasible sieve estimator can be constructed in two steps.

- *First step*: use the estimator proposed in the previous section we obtain $\widehat{\mathbf{W}}^t = (\widehat{q}_1^t, \ldots, \widehat{q}_{K^t-1}^t, \widehat{\Delta}_1^t, \ldots, \widehat{\Delta}_{K^t-1}^t)$ for every t.
- *Second step*: replace $(\mathbf{g}, \mathbf{y})$ by $(\widehat{\mathbf{g}}, \widehat{\mathbf{y}})$ where the latter quantities are constructed using $\{\widehat{\mathbf{W}}^t\}_{t=1}^T$ instead of $\{\mathbf{W}^t\}_{t=1}^T$. We define our sieve least squares estimator by:

$$\begin{aligned} \widehat{G}(c) &= g^L(c)^\top \widehat{\beta}, \text{ where} \\ \widehat{\beta} &= \left(\widehat{\mathbf{g}}^\top \widehat{\mathbf{g}}\right)^- \widehat{\mathbf{g}}^\top \widehat{\mathbf{y}}. \end{aligned}$$

Numerically our feasible estimation problem is identical to the estimator from a nonparametric series estimation of a regression function when the regressors and regressands used are based on $\{\widehat{\mathbf{W}}^t\}_{t=1}^T$. Notice, however, our sieve estimator is fundamentally different to a series estimator of a regression function since we have no regression error and the only source of sampling error (variance) comes from the generated variables we obtain from individual market in the first step.

We now state some assumptions that are sufficient for us to derive the uniform rate of convergence of $\widehat{G}$.

ASSUMPTION B1. *(i) For all $t = 1, \ldots, T$, $\mathbf{P}^t = \{P_i^t\}_{i=1}^{N^t}$ is an i.i.d. sequence of N^t random variables whose distribution satisfies the equilibrium condition in (4), where $(F^t, \overline{p}^t, r^t, K^t)$ is market* *specific; (ii)* $\mathbf{P}^t$ *and* $\mathbf{P}^{t'}$ *are independent for any* $t \neq t'$*; (iii) The analog of Assumption A2 holds for all markets.*

Assumptions B1(i) and B1(iii) ensure that $\left\| \widehat{\mathbf{W}}^t - \mathbf{W}^t \right\| = O_p\left(1/\sqrt{N^t}\right)$ for all t as $N^t \to \infty$. We impose independence in B1(ii) between markets for simplicity. In principles the recent conditions employed in Lee and Robinson (2014) to derive the uniform rates of series estimator under a weak form of cross sectional dependence can also be applied to our estimator.

Assumption B2. *(i)* $\{K^t\}_{t=1}^T$ *is an i.i.d. sequence with some discrete distribution with support* $\mathbf{K} = \{2, \ldots, \overline{K}\}$ *for some* $\overline{K} < \infty$*; (ii)* $\{\mathbf{\Delta}^t\}_{t=1}^T$ *is an independent sequence of random vectors such that* $\mathbf{\Delta}^t = \left(\Delta_1^t, \ldots, \Delta_{K^t-1}^t\right)$ *is a decreasing sequence of reals, where each variable has a continuous marginal distribution defined on* $\mathcal{S}_C$ *for all* t*. Furthermore, for any* $t \neq t'$ *such that* $K^t = K^{t'}$*,* Δ_k^t *and* $\Delta_k^{t'}$ *have identical distribution.*

Assumption B2 consists of conditions on the data generating process that ensure any open interval in $\mathcal{S}_C$ is visited infinitely often by $\{\mathbf{\Delta}^t\}_{t=1}^T$ as $T \to \infty$. This allows the repeated observations of data across markets to nonparametrically identify G on $\mathcal{S}_C$. Note that the size of $\mathbf{\Delta}^t$ is random since K^t is a random variable, so that $\{\mathbf{\Delta}^t\}_{t=1}^T$ (and thus $\{\mathbf{W}^t\}_{t=1}^T$) is an i.i.d. sequence. On the other hand, conditional on $\{K^t\}_{t=1}^T$, $\{\mathbf{\Delta}^t\}_{t=1}^T$ is an independent sequence but it does not have an identical distribution across t. In addition Δ_k^t and $\Delta_{k'}^t$ are neither independent nor have identical distribution for a given t. Therefore $\{\Delta_k^t\}_{k=1,t=1}^{K^t-1,T}$ is a $\overline{K}-$dependent process due to the independence across t.

Assumption B3. *(i) For* $K = 2, \ldots, \overline{K}$*,*

$$\begin{aligned} \min_{1 \leq k \leq K} \underline{\lambda}\left(E\left[g^L\left(\Delta_k^t\right) g^L\left(\Delta_k^t\right)^\top |K^t = K\right]\right) &> 0, \text{ and} \\ \max_{1 \leq k \leq K} \overline{\lambda}\left(E\left[g^L\left(\Delta_k^t\right) g^L\left(\Delta_k^t\right)^\top |K^t = K\right]\right) &< \infty; \end{aligned}$$

(ii) There exists a deterministic function $\zeta(L)$ *satisfying* $\sup_{c \in \mathcal{S}_C} \left\|g^L(c)\right\| \leq \zeta(L)$ *for all* L *such that* $\zeta(L)^4 L^2/T \to 0$ *as* $T \to \infty$*; (iii) For all* L *there exists a sequence* $\beta^L = (\beta_1, \ldots, \beta_L) \in \mathbb{R}^L$ *and some* $\alpha > 0$ *such that*

$$\left|G - g^{L\top}\beta^L\right|_\infty = O\left(L^{-\alpha}\right).$$

Assumption B3 consists of familiar conditions from the literature on nonparametric series estimation of regression functions, e.g. see Andrews (1991) and Newey (1997). B3(i) implies that redundant bases are ruled out, and that the second moment matrices are uniformly bounded away from zero and infinity for any distribution of Δ_k^t under consideration. The bounding of the moments from above and below is also imposed in Andrews (1991), who consider independent but not identically distributed sequence of random variables. Assumption B3(ii) controls the magnitude of the series terms. Since G is bounded, the bases can be chosen to be bounded and non-vanishing in which case it is easy to see from the definition of the norm that $\zeta(L) = O(\sqrt{L})$. Some examples for other rates of $\zeta(L)$, such as those of orthogonal polynomials or B-splines can be found in Newey (1997, Sections 5 and 6 respectively). Assumption B3(iii) quantifies the uniform error bounds for the approximation functions. For example if G is $s-$times continuously differentiable and the chosen sieves are splines or polynomials, then it can be shown $\alpha = s$.

Assumption B4. *For the same $\zeta(L)$ as in B3(ii): (i) for all L and $l = 1, \ldots, L$, $g_{lL} \in \mathcal{G}_T$ is continuously differentiable and $\sup_{c \in S_{\mathcal{C}}} \left\| \partial g^L(c) \right\| \leq \zeta(L)$ for all L, and $\partial g^L(c)$ denotes $\left(\frac{d}{dc} g_{1L}(c), \ldots, \frac{d}{dc} g_{LL}(c)\right)^\top$; (ii) $\zeta(L) = o\left(\sqrt{N_T}\right)$ as $T \to \infty$, where N_T denotes $\min_{1 \leq t \leq T} N^t$.*

Assumption B4 imposes some smoothness conditions that allow us to quantify the effect of using generated variables obtained from the first-step estimation. B4(i) assumes the bases of the sieves to have at least one continuous derivative that are bounded above by $\zeta(L)$. This is a mild condition since most sieves used in econometrics are smooth functions with at least one continuous derivative, even for piece-wise smooth functions where differentiability can be imposed at the knots; see Section 2.3 in Chen (2007) for examples. B4(ii) ensures the upper bound of the basis functions and their derivatives does not grow too quickly over any $1/\sqrt{N_T}$ neighborhood on $S_{\mathcal{C}}$. Note that $1/\sqrt{N_T}$ is the rate that $\max_{1 \leq t \leq T} \left\| \widehat{\mathbf{W}}^t - \mathbf{W}^t \right\|$ converges to zero.

Theorem 4 *(Uniform rates of convergence of $\widehat{G}$). Under Assumptions B1 - B4, as N_T and T tend to infinity*:

$$\left| \widehat{G} - G \right|_\infty = O_p\left(\zeta(L) \left[\zeta(L) N_T^{-1/2} + L^{-\alpha} \right] \right),$$

The additive components of the convergence rate of $\left| \widehat{G} - G \right|_\infty$ come from two distinct sources. The first is the variance that comes from the first stage estimation and the latter is the approximation bias from using sieves. The order of the bias term is just the numerical approximation error and is the same as that found in the nonparametric series regression literature. The convergence rate for our variance term is inherited from the rates of the generated variables, which is parametric with respect to N_T. Our estimator has no other sampling error. This is due to the fact that, unlike in a regression context, $\mathbf{W}^t$ is completely known when $\{q_k^t\}_{k=1}^{K^t - 1}$ is known hence there is no variance component associated with regression error. The intuition for the expression $\zeta(L) N_T^{-1/2}$ is simple. This term effectively captures the rate of convergence of the difference between the feasible and infeasible least squares coefficients of the sieve bases. In particular the difference can be linearized and well approximated by the product between the derivatives of the basis functions and the sampling errors of the generated variable, which are respectively bounded by $\zeta(L)$ and $N_T^{-1/2}$.

The leading term for the uniform convergence rate of $\widehat{G}$ depends on whether the sampling error from generated variables across different markets is larger or smaller than the numerical approximation bias. If $\zeta(L)\, N_T^{-1/2} = o(L^{-\alpha})$, then the effect from first stage estimation is negligible for the rate of convergence of the sieve estimator. If, on the other hand the reciprocal relation holds, then the dominant term on the rate of convergence comes from the generated variables.

The uniform rate of convergence in Theorem 4 quantifies the magnitude for the errors we incur from fitting a curve since the sampling error from point estimation from each market is at most $N_T^{-1/2}$. However, the asymptotic distribution theory for a sieve estimator of an unknown function is often difficult to obtain and general results are only known to exist in some special cases. We refer the reader to Section 3.4 of Chen (2007) for some details. The development of the distribution theory for our estimator of G is beyond the scope of this paper.

5 Numerical Section

The first part of this section reports a small scale simulation study to compare our estimator with other estimators in the literature in a controlled environment. The second part illustrates our proposed estimator using online betting odds data.

5.1 Monte Carlo

We first consider the case when the data come from a single market. Here we adopt an identical design to the one used in Section 4.3 of MGW, where they study the small sample properties of their estimator and that of HS. In particular the consumers' search costs are drawn independently from a log-normal distribution with location and scale parameters set at 0.5 and 5 respectively. The other primitives of the model are: $(\overline{p}, r, K) = (100, 50, 10)$. We solve for a mixed strategy equilibrium and take 100 random draws from the corresponding price distribution, which can be interpreted as observing a repeated game played by the 10 sellers 10 times. We refer the reader to MGW for the details of the data generation procedure that is consistent with an equilibrium outcome as well as other discussions on the Monte Carlo design.

We simulate the data according to the description above and estimate the model 1000 times. We report the same statistics as those in MGW. We estimate the parameters using our minimum distance estimator and are able to replicate the maximum likelihood results in MGW. We focus our discussion on our estimator and MGW's since the latter has been shown to generally perform favorably relative to the empirical likelihood estimator. The comments provided by MGW in this regard are also applicable for our estimator and can be found in their paper. In particular our Tables 1 and 2 can be compared directly with their Tables 3(a) and 3(b) respectively. We also provide analogous statistics associated with the estimator for the cdf of the search cost evaluated at the cutoff points in Table 3.

Parameter		MLE			MDE		
	True	*Mean*	*St Dev*	*MSE*	*Mean*	*St Dev*	*MSE*
$r(\mathbf{q})$	50	48.384	4.276	20.896	49.535	3.112	9.900
q_1	0.37	0.413	0.111	0.014	0.378	0.114	0.013
q_2	0.04	0.043	0.019	0.000	0.039	0.014	0.000
q_3	0.03	0.033	0.038	0.001	0.024	0.021	0.001
q_4	0.03	0.025	0.046	0.002	0.021	0.028	0.001
q_5	0.03	0.025	0.066	0.004	0.027	0.031	0.001
q_6	0.02	0.038	0.096	0.009	0.031	0.031	0.001
q_7	0.02	0.041	0.110	0.012	0.029	0.027	0.001
q_8	0.02	0.050	0.131	0.018	0.025	0.024	0.001
q_9	0.02	0.059	0.141	0.022	0.020	0.019	0.000
q_{10}	0.42	0.274	0.239	0.079	0.404	0.158	0.025

Table 1: Properties of maximum likelihood (MLE) and minimum distance (MDE) estimators for $r(\mathbf{q})$ and $q_1, \ldots, q_{10}$.

Parameter		MLE			MDE		
	True	*Mean*	*St Dev*	*MSE*	*Mean*	*St Dev*	*MSE*
Δ_1	8.640	8.481	0.472	0.248	8.539	0.539	0.300
Δ_2	5.264	5.139	0.204	0.057	5.215	0.221	0.051
Δ_3	3.484	3.394	0.155	0.032	3.455	0.183	0.034
Δ_4	2.428	2.365	0.151	0.027	2.408	0.184	0.034
Δ_5	1.756	1.714	0.145	0.023	1.742	0.177	0.031
Δ_6	1.309	1.281	0.134	0.019	1.299	0.163	0.027
Δ_7	0.999	0.982	0.122	0.015	0.992	0.148	0.022
Δ_8	0.779	0.770	0.110	0.012	0.775	0.132	0.018
Δ_9	0.619	0.614	0.098	0.010	0.616	0.118	0.014

Table 2: Properties of maximum likelihood (MLE) and minimum distance (MDE) estimators for $\Delta_1, \ldots, \Delta_9$.

Parameter		MLE			MDE		
	True	*Mean*	*St Dev*	*MSE*	*Mean*	*St Dev*	*MSE*
$G(\Delta_1)$	0.630	0.631	0.117	0.014	0.587	0.111	0.014
$G(\Delta_2)$	0.592	0.589	0.125	0.016	0.544	0.120	0.017
$G(\Delta_3)$	0.559	0.561	0.128	0.016	0.511	0.123	0.018
$G(\Delta_4)$	0.531	0.544	0.138	0.019	0.486	0.131	0.019
$G(\Delta_5)$	0.505	0.523	0.160	0.026	0.462	0.145	0.023
$G(\Delta_6)$	0.482	0.493	0.195	0.038	0.423	0.175	0.034
$G(\Delta_7)$	0.460	0.456	0.226	0.051	0.382	0.207	0.049
$G(\Delta_8)$	0.440	0.413	0.252	0.064	0.332	0.226	0.063
$G(\Delta_9)$	0.422	0.378	0.265	0.072	0.274	0.239	0.079

Table 3: Properties of maximum likelihood (MLE) and minimum distance (MDE) estimators for $G(\Delta_1), \ldots, G(\Delta_9)$.

Tables 1 and 2 contain the true mean and standard deviation of various parameters for each estimator as reported in MGW,[11] in addition we include the mean square errors for the ease of comparison between our results and theirs (that include the empirical likelihood estimator). We provide the same statistics for the estimator of the cdf evaluated at the cutoff points in Table 3. Our estimator performs comparably well with respect to the maximum likelihood estimator. Particularly our estimator generally has smaller bias, but also higher variance. However, there is no dominant estimator with respect to the mean square errors, at least for this design and sample size. Our estimator appears to generally perform better for the parameters in Table 1. The maximum likelihood estimation is better for those in Table 2. The results are more mixed for Table 3.

Next we consider the estimation of data that come from several markets. Here we adopt the same design to the one used to generate results in the Supplementary Appendix that accompanies MGSW. The data are drawn from 10 heterogeneous markets. The consumers have the same search cost distribution in every market while sellers' marginal costs can vary and thus imply different equilibrium price distribution. For each simulation we draw 35 prices from the equilibrium price distribution from each market. So the total sample size is 350. Other details can be found in MGSW.

We simulate the data and estimate the model 1000 times. We estimate our estimator using Bernstein polynomials as the basis functions. Specifically, suppose $\mathcal{S}_C = [0,1]$. The basis functions

[11] The supporting numerical results for the consistency of the bootstrap for our estimator are available upon request.

that define Bernstein polynomials of order L consists of the following $L+1$ functions:

$$g_{lL}(c) = \frac{L!}{l!\,(L-l)!} c^l (1-c)^{L-l}, l = 0, \ldots, L.$$

We choose Bernstein polynomials due to its well-behaved uniform property as well as the simplicity to impose shape restrictions one expects from a cdf[12]. See Lorentz (1986) for further details. For a generic support, $\mathcal{S}_C = \left[\underline{C}, \overline{C}\right]$, we can scale the support of functions in $\mathcal{G}_T$ accordingly. We impose monotonicity in estimating our sieve estimator in the simulation study and the application. For the estimator of MGSW we use Hermite polynomials as the basis as done in their paper. We report in Table 4 the integrated mean square error (imse), defined as $\int E\left[\widehat{G}(c) - G(c)\right]^2 dG(c)$, for our estimator and theirs for the first corresponding 10 basis terms.

L	SLSE	SNMLE
	imse	$(\times 10^{-2})$
1	0.345	6.984
2	0.126	1.596
3	0.120	0.389
4	0.117	0.513
5	0.122	0.430
6	0.126	0.258
7	0.127	0.302
8	0.125	0.300
9	0.123	0.193
10	0.121	0.259

Table 4: Imse for sieve least squares (SLSE) and semi-nonparametric maxmimum likelihood (SNMLE) estimators using L basis functions.

We note that it would not be appropriate to compare our reported statistics and Table 1 of MGSW. In particular the imse we use is different to their integrated squared error. There are two

[12]For any continuous function g:

$$\lim_{L\to\infty} \sum_{l=0}^{L} g\left(\frac{l}{L}\right) \frac{L!}{l!\,(L-l)!} c^l (1-c)^{L-l} = g(c),$$

holds uniformly on $[0,1]$. Furthermore for $\mathcal{G}_T = \left\{g : g = g^{L\top} b \text{ for some } b = (b_0, \ldots, b_L)\right\}$, elements in $\mathcal{G}_T$ will be non-decreasing under the restrictions that $b_l \leq b_{l+1}$ for $l = 0, \ldots, L$, and the range of functions in $\mathcal{G}_T$ can be set by choosing b_0 and b_L to be the minimum and the maximum values respectively.

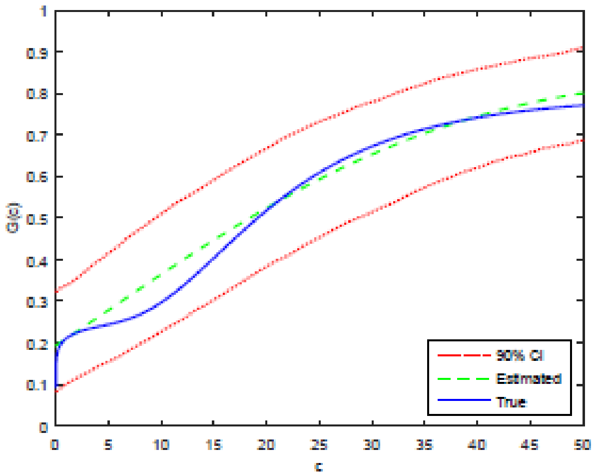

Figure 1: Sieve estimator of the cost cdf with $L = 4$.

differences. First, their integrated error is calculated by integrating the squared error between the Monte Carlo average of $\widehat{G}$ and the true. Second, their integrator is the identity function and ours is G; i.e. we use $\int [\cdot]\, dG(c)$ rather than $\int [\cdot]\, dc$.

We find that our estimator seems to perform slightly better with respect to the imse criterion for the number of polynomial terms considered. We certainly do not claim our estimator is necessarily better based on Table 4. It is generally difficult to compare any two estimators in finite sample, and in particular for nonparametric estimators using different basis functions. Note that Table 4 suggests the imse is minimized for our estimator when $L = 4$ and theirs when $L = 9$. As a visual illustration, we also plot the mean and the 5th and 95th percentiles of our estimator and theirs. Figures 1 and 2 represent our estimator with $L = 4$ and $L = 9$ respectively. Figures 3 and 4 correspond to MGSW counterparts.

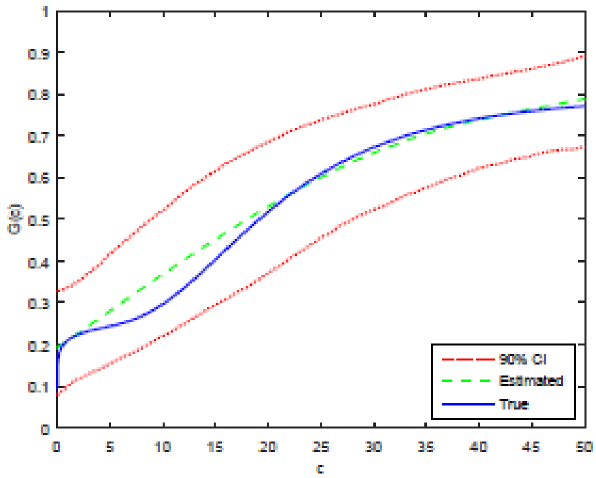

Figure 2: Sieve estimator of the cost cdf with $L = 9$.

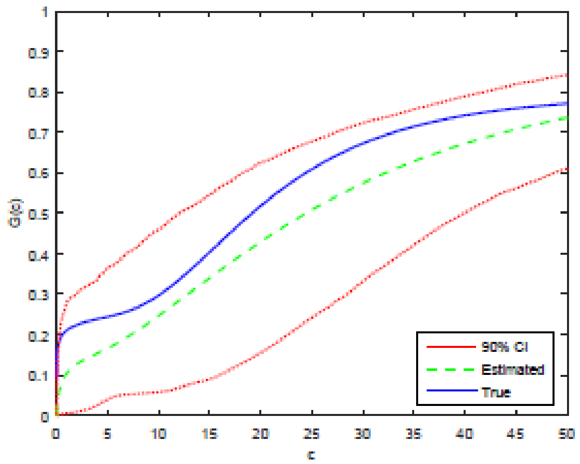

Figure 3: MGSW's estimator of the cost cdf with $L = 4$.

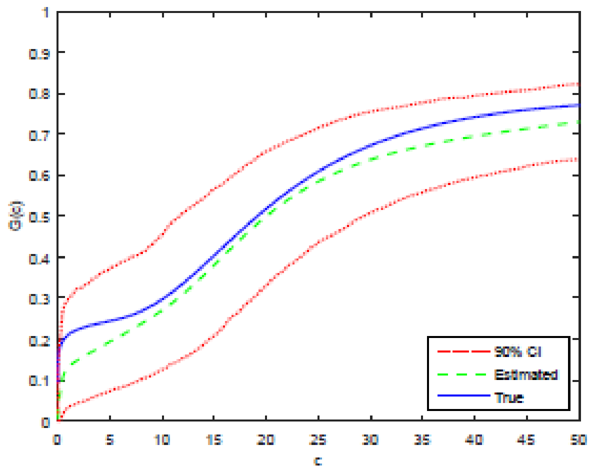

Figure 4: MGSW estimator of the cost cdf with $L = 9$.

5.2 Empirical Illustration

Background and Data

Gambling in the UK is regulated by the Gambling Commission on behalf of the government's Department for Culture, Media and Sport under the Gambling Act 2005. In addition to the moral duty to prevent the participation of children and the general policing against criminal activities related to gambling in the UK, another main goal of the Act is to ensure that gambling is conducted in a fair and open way. One crucial component of the Act that has received much attention in the media takes place in September 2007, which permits gambling operators to advertise more widely.[13] Its intention is to raise the awareness for the general public about potential bookmakers in the market in order to increase the competition between them.

In this section we illustrate the use of our estimators proposed in earlier parts of the paper. We assume the search model described in Section 2 serves as a (very) crude approximation of the true mechanism that generates the prices that we see in the data.[14] We focus on the booking odds set at different bookmakers for the top two professional football leagues in the UK, namely the Premier League and the League Championship, for the 2006/7 and 2007/8 seasons. We consider the odds for what is known as a "2x1 bet", where there are three possible outcomes for a given match: home (team) wins, away wins or they draw. We construct the price for each bookmaker from the odds we observe. Since the odd for each event is the inverse of its perceived probability, we define our price from each bookmaker as: 1/(home-win odd) + 1/(draw odd) + 1/(away-win odd). The sum of theses probabilities always exceeds 1 since consumers never get to play a fair game. This excess probability represents what is called the bookmaker's *overround*. The higher the overround, the more unfair and expensive is the bookmaker's price.

We obtain the data from http://www.oddsportal.com/, which is an open website that collects

[13] Gambling operators have been able to advertise on TV and radio from 1st of September 2007. Previously the rules for advertising for all types of gambling companies, including casinos and betting shops have been highly regulated. Traditional outlets for advertising are through magazines and newspapers, or other means to get public attention such as sponsoring major sporting events. Further information on the background and impact of the Gambling Act 2005 can be found in the review produced by the Committees of Advertising Practice at the request by the Department for Culture, Media and Sport, http://www.cap.org.uk/News-reports/~/media/Files/CAP/Reports%20and%20surveys/CAP%20and%20BCAP%20Gambling%20Review.ashx

[14] We highlight three underlying assumptions of the theoretical model. First, products are homogeneous. Second, consumers perform a non-sequential search. Third, each consumer purchases only one unit of the product. In the context of betting it is not unreasonable to assume products are homogeneous as consumers are only interested in making monetary profit. Our prices are based on online odds therefore non-sequential search strategy may also provide a reasonable approximation of consumers' behavior who conduct search online. However, assuming each consumer only purchases one unit, in this case translating to everyone having the same wager, is not realistic; also experienced and organized bettors often bet on multiple matches at the same time.

data from the main online bookmakers from a number of different events. In the tables and figures below, we use PL and LC to respectively denote Premier League and League Championship, and 06/07 and 07/08 respectively for the 2006/7 and 2007/8 seasons. We begin with Table 5 that gives some summary statistics on the data.

Group	Matches	Bookmakers			Overrounds		
		Mean	*Median*	*St Dev*	*Mean*	*Median*	*St Dev*
PL 06/07	380	21.58	22	2.93	9.55	9.90	2.09
PL 07/08	380	35.24	36	2.25	8.45	8.48	2.36
LC 06/07	557	20.62	21	2.19	11.13	11.14	1.09
LC 07/08	557	28.10	29	3.59	10.36	10.71	2.02

Table 5: Summary statistics on the data from different leagues and seasons.

We partition the data into four product groups. One for each league and season. The numbers of bookmakers we observe vary between matches as occasionally odds for some bookmakers have not been collected. The average overrounds between the two seasons indicate that prices have fallen after the change of law. Relatedly, we also see an increase in the average number of bookmakers as well.[15]

For each group we take the number of sellers to be the average number of bookmakers (rounded to the nearest integer). We treat the observed price for every match as a random draw from an equilibrium price distribution. We assume the distribution of the consumers' search cost to be the same for both leagues within each season. Our main interest is to see if there is any evidence the distribution of the search costs differ between the two seasons.

Single Market

We provide four sets of point estimates. One for each group using the estimator described in Section 3. For the following tables, the bootstrap standard errors are reported in parentheses.

[15] The total number of bookmakers for the 2006/7 season is 32, and for the 2007/8 season is 40.

Group	K	$\widehat{q}_1$	$\widehat{q}_2$	$\widehat{q}_K$	$r(\widehat{\mathbf{q}})$	$\underline{p}$	$\overline{p}$
PL 06/07	23	0.77	0.21	0.01	80.36	100.09	118.99
		(0.03)	(0.02)	(0.00)	(2.20)		
PL 07/08	36	0.40	0.55	0.05	96.40	100.04	125.62
		(0.10)	(0.08)	(0.00)	(1.13)		
LC 06/07	22	0.67	0.30	0.03	97.51	105.03	118.04
		(0.10)	(0.08)	(0.00)	(4.78)		
LC 07/08	29	0.20	0.73	0.07	97.87	101.23	159.30
		(0.11)	(0.10)	(0.01)	(0.98)		

Table 6: Estimates of search proportions, selling costs and range of prices

Over 90% of consumers search at most twice for every product group. Other proportions of consumers' search that are not reported are very close to zero. It is very noticeable that the proportions of consumers searching just once drop, following the law change, transferring mostly to searching twice. We now relate these to the search cost distribution.

Group	K	$\widehat{\Delta}_1$	$\widehat{G}(\Delta_1)$	$\widehat{\Delta}_2$	$\widehat{G}(\Delta_2)$	$\widehat{\Delta}_{K-1}$	$\widehat{G}(\Delta_{K-1})$
PL 06/07	23	2.49	0.23	1.24	0.01	0.07	0.01
		(0.05)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)
PL 07/08	36	3.17	0.60	1.26	0.05	0.03	0.05
		(0.40)	(0.09)	(0.10)	(0.01)	(0.00)	(0.01)
LC 06/07	22	1.72	0.33	0.83	0.03	0.05	0.03
		(0.08)	(0.08)	(0.08)	(0.01)	(0.01)	(0.01)
LC 07/08	29	6.07	0.80	1.98	0.07	0.04	0.07
		(1.18)	(0.12)	(0.24)	(0.01)	(0.00)	(0.01)

Table 7: Estimates of search cost distribution

We do not report the estimated cdf values for other cut-off points since they are almost identical to $\widehat{G}(\Delta_2)$ and $\widehat{G}(\Delta_{K-1})$. Since the cut-off values for each group differ, it is more convenient to make this comparison graphically. We next estimate the cdf as a function.

Pooling Data Across Markets

We combine the data between the two leagues for each time period using the sieve estimator proposed in Section 4. We use Bernstein polynomials as the base functions for our sieve estimator;

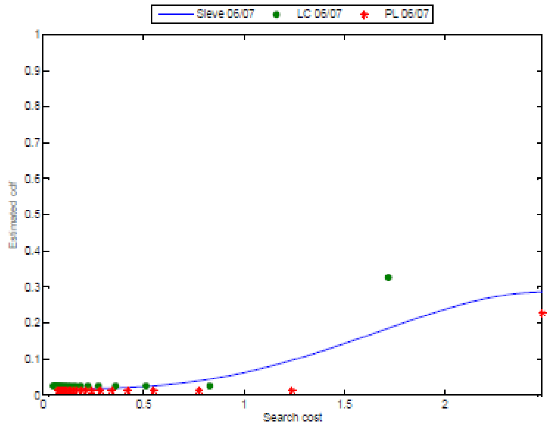

Figure 5: Scatter plots of the point estimates of cost quantiles for the two football leagues, and the corresponding sieve estimates of the cost cdf using data from the 2006/7 season.

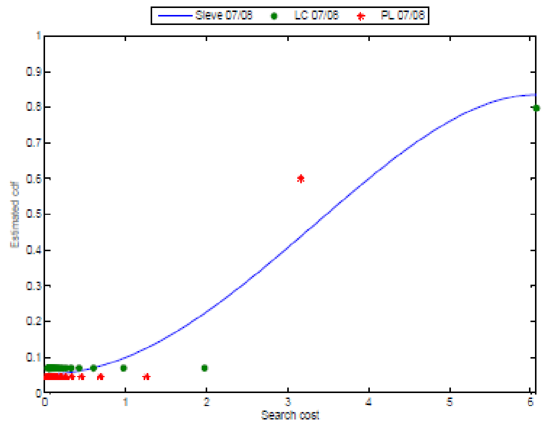

Figure 6: Scatter plots of the point estimates of cost quantiles for the two football leagues, and the corresponding sieve estimates of the cost cdf using data from the 2007/8 season.

see the description above. To construct our estimates for the cdfs we only impose monotonicity on the coefficients to ensure the estimates are non-decreasing. We fit the data using $L = 4$.

Figures 5 and 6 illustrate how sieve estimation interpolates data across markets. They provide scatter plots of the point estimates of quantiles for the two leagues and the corresponding sieve estimates for the 2006/7 and 2007/8 seasons respectively. Figure 7 plots the two curves together. We see that the estimate from the 2007/8 season takes higher value than the cdf from the 2006/7 season almost uniformly where their supports overlap. This display of a first order stochastic dominance behavior indicates the cost of search has fallen since the implementation of the new advertising law.

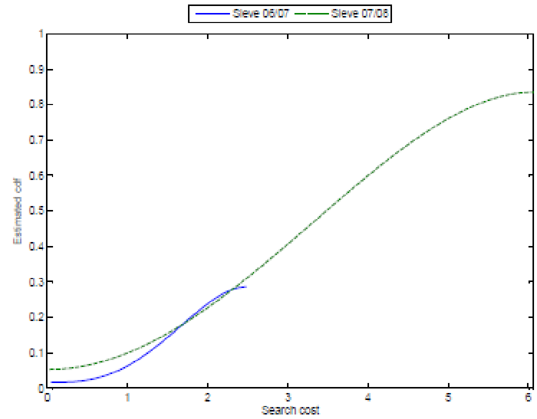

Figure 7: Sieve estimates of the cost cdf using data from the 2006/7 and 2007/8 seasons.

6 Conclusion

We propose a minimum distance estimator to estimate quantiles of search cost distribution when only the price distribution is available. We derive the distribution theory of our estimator and show it can be consistently bootstrapped. It is easier to estimate and perform inference with our estimator than previous methods. Our point estimator can be readily used to estimate the cdf of the search cost by the method of sieve. We provide the uniform convergence rate for our sieve estimator. The rate can be used to quantify the errors from interpolating quantiles across markets when such data are available. Both our estimators perform reasonably well relatively to other existing estimators in a simulation study with small sample. We also illustrate the ease of use for our estimators with real world data. We use online odds to construct bookmakers' prices for online betting for professional football matches in the UK for the two seasons either side of the change in the advertising law that allows gambling operators to advertise more freely. This particular change in the law marks a well-known event that has since been reported to increase competition amongst bookmakers by several measures, as intended by the Gambling Act 2005. One aspect of this outcome is supported by our simple model of search that suggests that consumers search more often, which can be attributed at least partly to the reduction of search costs. We expect the minimum distance approach in this paper can be adapted to offer a computationally appealing way to estimate more complicated search models.

Appendix

Preliminary Notations

The proofs of our Theorems make use of some results from empirical processes theory. We do not define basic terms and definitions from empirical processes theory here for brevity. We refer the reader to the book by Kosorok (2008) for such details.

Firstly, with an abuse of notation, it will be convenient to introduce a function $m(\cdot,\theta,\gamma):\mathcal{S}_P\to\mathbb{R}$ that depends respectively on finite and infinite dimensional parameters $\theta\in\Theta$ and $\gamma\in\Gamma$. Recall that $\Theta=[0,1]^{K-1}$. Here we use Γ to denote a set of all cdfs with bounded densities defined on $\mathcal{S}_P$. So that for $p\in\mathcal{S}_P$, $\theta\in\Theta$ and $\gamma\in\Gamma$, we define:

$$
\begin{aligned}
m(p,\theta,\gamma) &= \theta_1\left(\left(\underline{p}-\overline{p}\right)\left(K-\sum_{k=1}^{K-1}(k-K)\theta_k\right)\right) \\
&-\left(\begin{array}{c} K\left(1-\sum_{k=1}^{K-1}\theta_k\right)(1-\gamma(p))^{K-1} \\ +\sum_{k=1}^{K-1}\theta_k(1-\gamma(p))^{k-1}\end{array}\right)\times\left((p-\overline{p})\theta_1-\left(p-\underline{p}\right)\left(K-\sum_{k=1}^{K-1}(k-K)\theta_k\right)\right).
\end{aligned}
$$

Comparing the above to the function $m(\cdot,\theta)$ used in the main text (see e.g. (8)), we have that $m(\cdot,\theta)$ and $m(\cdot,\theta,F)$ are precisely the same objects.

We denote a space of bounded functions defined on $\mathcal{S}_P$ equipped with the sup-norm by $\mathcal{D}$. We view $m(\cdot,\theta,\gamma)$ as an element in $\mathcal{D}$, which is parameterized by $(\theta,\gamma)\in\Theta\times\Gamma$. Also since γ is defined in $m(p,\theta,\gamma)$ pointwise for each p, it will be useful in the proofs below for us to occasionally write $m(p,\theta,\gamma(p))\equiv m(p,\theta,\gamma)$ in defining some derivatives for clarity. In particular, pointwise for each p, using an ordinary derivative, for any γ let: $D_\gamma m(p,\theta,\gamma(p))\equiv\lim_{t\to0}\left|\frac{m(p,\theta,\gamma(p)+t)-m(p,\theta,\gamma(p))}{t}\right|$ and $D_\gamma\frac{\partial}{\partial\theta_k}m(p,\theta,\gamma(p))\equiv\lim_{t\to0}\left|\frac{\frac{\partial}{\partial\theta_k}m(p,\theta,\gamma(p)+t)-\frac{\partial}{\partial\theta_k}m(p,\theta,\gamma(p))}{t}\right|$ for all k. It is easy to see that $m(\cdot,\theta,\gamma)$, $\frac{\partial}{\partial\theta_k}m(\cdot,\theta,\gamma)$, $D_\gamma m(\cdot,\theta,\gamma)$ and $D_\gamma\frac{\partial}{\partial\theta_k}m(\cdot,\theta,\gamma)$ are elements in $\mathcal{D}$ for any (θ,γ) in $\Theta\times\Gamma$. In the main text we have denoted the sup-norm for any real value function defined on $\mathcal{S}_C$ by $|\cdot|_\infty$. In this Appendix we will also use $|\cdot|_\infty$ to denote the sup-norm for any real value function defined on $\mathcal{S}_P$ as well. We do not index the norm further to avoid additional notation. There should be no ambiguity whether the domain for the function under consideration is $\mathcal{S}_P$ or $\mathcal{S}_C$. We define the following constants that will be helpful in guiding the reader through our proofs:

$$
\begin{aligned}
\underline{\kappa}_m &= \sup_{(\theta,\gamma)\in\Theta\times\Gamma}|m(\cdot,\theta,\gamma(\cdot))|_\infty, \underline{\kappa}_{\frac{\partial}{\partial\theta}m}=\max_{1\le k\le K}\sup_{(\theta,\gamma)\in\Theta\times\Gamma}\left|\frac{\partial}{\partial\theta_k}m(\cdot,\theta,\gamma(\cdot))\right|_\infty, \\
\underline{\kappa}_{D_Fm} &= \sup_{(\theta,\gamma)\in\Theta\times\Gamma}|D_Fm(\cdot,\theta,\gamma(\cdot))|_\infty, \underline{\kappa}_{D_F\frac{\partial}{\partial\theta}m}=\max_{1\le k\le K}\sup_{(\theta,\gamma)\in\Theta\times\Gamma}\left|D_F\frac{\partial}{\partial\theta_k}m(\cdot,\theta,\gamma(\cdot))\right|_\infty.
\end{aligned}
$$

Other generic positive and finite constants that do not depend on the sample size are denoted by $\underline{\kappa}_0$, which can take different values in different places.

Lemmas

Lemmas 1 - 8 are used to prove Theorems 1 - 3 from Section 3. Lemmas 9 - 17 are used to prove Theorem 4 from Section 4.

LEMMA 1. Under Assumptions A2(i) and A2(ii), $M(\theta)$ has a well-separated minimum at θ_0.

PROOF OF LEMMA 1. Under A2(i) and the domination condition in A2(ii), M has a unique minimum at θ_0. Since M is continuous on Θ, the minimum is well-separated.■

LEMMA 2. Under Assumptions A2(i) and A2(ii), $\sup_{\theta\in\Theta}|M_N(\theta) - M(\theta)| \stackrel{a.s.}{\to} 0$.

PROOF OF LEMMA 2.

$$\begin{aligned}
& M_N(\theta) - M(\theta) \\
= & \int m(p,\theta,F_N)^2 (\mu_N(dp) - \mu(dp)) + \int m(p,\theta,F_N)^2 - m(p,\theta,F)^2 \, d\mu \\
= & I_1(\theta) + I_2(\theta).
\end{aligned}$$

For $I_1(\theta)$, using the bound for m, $|I_1(\theta)| \leq \underline{\kappa}_m^2 \int (\mu_N(dp) - \mu(dp))$. The convergence of measure follows from A2(ii) so that $\sup_{\theta\in\Theta}|I_1(\theta)| \stackrel{a.s.}{\to} 0$. For $I_2(\theta)$, we have

$$\begin{aligned}
|I_2(\theta)| & \leq 2\underline{\kappa}_m \int |m(p,\theta,F_N) - m(p,\theta,F)| \, \mu(dp) \\
& \leq 2\underline{\kappa}_m \underline{\kappa}_{D_F m} \int \mu(dp) \, |F_N - F|_\infty .
\end{aligned}$$

The second inequality follows from taking pointwise mean value expansion about F. Then $\sup_{\theta\in\Theta}|I_2(\theta)| \stackrel{a.s.}{\to} 0$ by Glivenko-Cantelli theorem. The proof then follows from the triangle inequality.■

Let

$$\begin{aligned}
H(\theta) &= \int 2\frac{\partial}{\partial\theta} m(p,\theta,F) \frac{\partial}{\partial\theta^\top} m(p,\theta,F) \, \mu(dp), \\
H_N(\theta) &= \int 2\frac{\partial}{\partial\theta} m(p,\theta,F_N) \frac{\partial}{\partial\theta^\top} m(p,\theta,F_N) \, \mu_N(dp), \\
H_N^*(\theta) &= \int 2\frac{\partial}{\partial\theta} m(p,\theta,F_N^*) \frac{\partial}{\partial\theta^\top} m(p,\theta,F_N^*) \, \mu_N^*(dp),
\end{aligned}$$

where F_N^* is the empirical cdf with respect to the bootstrap sample.

LEMMA 3. Under Assumption A2(ii), for any θ_N such that $\|\theta_N - \theta_0\| \stackrel{a.s.}{\to} 0$ then $\|H_N(\theta_N) - H(\theta_0)\| \stackrel{a.s.}{\to} 0$.

Proof of Lemma 3. First show $\sup_{\theta\in\Theta}\|H_N(\theta) - H(\theta)\| \xrightarrow{a.s.} 0$. Using the same strategy in the proof of Lemma 2, let $h(p,\theta,\gamma) = 2\frac{\partial}{\partial\theta}m(p,\theta,\gamma)\frac{\partial}{\partial\theta^\top}m(p,\theta,\gamma)$, we have:

$$\begin{aligned} & H_N(\theta) - H(\theta) \\ = & \int h(p,\theta,F_N)\,\mu_N(dp) - \int h(p,\theta,F)\,\mu(dp) \\ = & \int h(p,\theta,F_N)\left(\mu_N(dp) - \mu(dp)\right) + \int h(p,\theta,F_N) - h(p,\theta,F)\,\mu(dp) \\ = & J_1(\theta) + J_2(\theta). \end{aligned}$$

Then $\sup_{\theta\in\Theta}\|J_1(\theta)\| \le \underline{\kappa}_0\underline{\kappa}^2_{\frac{\partial}{\partial\theta}m}\int\left(\mu_N(dp) - \mu(dp)\right) \xrightarrow{a.s.} 0$, and $\sup_{\theta\in\Theta}\|J_2(\theta)\| \le \underline{\kappa}_0\underline{\kappa}^2_{D_F\frac{\partial}{\partial\theta}m}\int\mu(dp)\,|F_N - F|_\infty$ 0. Uniform almost sure convergence then follows from the triangle inequality.

By the continuity of $H(\theta)$ and Slutzky's theorem, $|H(\theta_N) - H(\theta_0)| \xrightarrow{a.s.} 0$.

The desired result holds by using the triangle inequality to bound $H_N(\widehat{\theta}) - H(\theta_0) = H_N(\widehat{\theta}) - H(\widehat{\theta}) + H(\widehat{\theta}) - H(\theta)$.■

Lemma 4. Under Assumption A2(ii), $\frac{\partial}{\partial\theta}M_N(\theta_0) \xrightarrow{d} \mathcal{N}(0,\Sigma)$.

Proof of Lemma 4. From its definition, $\frac{\partial}{\partial\theta}M_N(\theta_0) = 2\int\frac{\partial}{\partial\theta}m(p,\theta_0,F_N)\,m(p,\theta_0,F_N)\,\mu_N(dp)$, by adding nulls we have

$$\begin{aligned} & \sqrt{N}\frac{\partial}{\partial\theta}M_N(\theta_0) \\ = & 2\int\frac{\partial}{\partial\theta}m(p,\theta_0,F)\sqrt{N}m(p,\theta_0,F_N)\,\mu(dp) \\ & +2\int\left(\frac{\partial}{\partial\theta}m(p,\theta_0,F_N) - \frac{\partial}{\partial\theta}m(p,\theta_0,F)\right)\sqrt{N}m(p,\theta_0,F_N)\,\mu(dp) \\ & +2\int\frac{\partial}{\partial\theta}m(p,\theta_0,F)\sqrt{N}m(p,\theta_0,F_N)\left(\mu_N(dp) - \mu(dp)\right) \\ & +2\int\left(\frac{\partial}{\partial\theta}m(p,\theta_0,F_N) - \frac{\partial}{\partial\theta}m(p,\theta_0,F)\right)\sqrt{N}m(p,\theta_0,F_N)\left(\mu_N(dp) - \mu(dp)\right) \\ = & J_1 + J_2 + J_3 + J_4. \end{aligned}$$

We first show the desired distribution theory is delivered by J_1.

By Donskers' theorem the empirical cdf converges weakly to a standard Brownian bridge of F, denoted by $\left(\mathbb{B}(F(p))\right)_{p\in\mathcal{S}_P}$. So that for $p,p'\in\mathcal{S}_P$,

$$\begin{aligned} \mathbb{B}(F(p)) &\sim \mathcal{N}\left(0, F(p)\left(1 - F(p)\right)\right), \text{ and} \\ \mathrm{Cov}\left(\mathbb{B}(F(p)), \mathbb{B}(F(p'))\right) &= F(\min\{p,p'\}) - F(p)F(p'). \end{aligned} \tag{16}$$

In this proof, it will be convenient to define $m^\dagger(\cdot,\gamma) = m(\cdot,\theta_0,\gamma)$ as an element in $\mathcal{D}$ indexed by just

γ. Next we calculate the directional derivative of $m^\dagger$ at F in the direction ξ, which gives for all p:

$$\begin{aligned}
&\lim_{t\to 0}\left|\frac{m^\dagger\left(p,F\left(p\right)+t\xi\left(p\right)\right)-m^\dagger\left(p,F\left(p\right)\right)}{t}-\delta\left(p\right)\xi\left(p\right)\right|=0,\ \text{ where} \qquad (17)\\
\delta\left(p\right) &= \left(\left(K-1\right)K\left(1-\sum_{k=1}^{K-1}\theta_k\right)\left(1-F\left(p\right)\right)^{K-2}+\sum_{k=2}^{K-1}\left(k-1\right)\theta_k\left(1-F\left(p\right)\right)^{k-2}\right)\\
&\times\left(\left(p-\overline{p}\right)\theta_1-\left(p-\underline{p}\right)\left(K-\sum_{k=1}^{K-1}\left(k-K\right)\theta_k\right)\right).
\end{aligned}$$

It is clear that δ is an element in $\mathcal{D}$, and $m^\dagger$ is Hadamard differentiable at F. Consequently the linear functional $\gamma\mapsto 2\int\frac{\partial}{\partial\theta}m\left(p,\theta_0,F\right)m^\dagger\left(p,\gamma\right)\mu\left(dp\right)$ is also Hadamard differentiable at F. In particular its derivative is represented by a linear operator, which we denote by T_F:

$$\begin{aligned}
T_F &: \mathcal{D}\to\mathbb{R}\ \text{ such that for any } \xi, \qquad (18)\\
T_F\xi &= 2\int\eta\left(p\right)\xi\left(p\right)\mu\left(dp\right),\ \text{where}\\
\eta\left(p\right) &= \frac{\partial}{\partial\theta}m\left(p,\theta_0,F\right)\delta\left(p\right)\ \text{ for all } p\in\mathcal{S}_P.
\end{aligned}$$

Hence we can apply the functional delta method and the continuous mapping theorem by letting $\xi=\sqrt{N}\left(F_N-F\right)$:

$$\begin{aligned}
&2\int\frac{\partial}{\partial\theta}m\left(p,\theta_0,F\right)\sqrt{N}m^\dagger\left(p,F_N\right)\mu\left(dp\right) \qquad (19)\\
=\ &2\int\eta\left(p\right)\sqrt{N}\left(F_N\left(p\right)-F\left(p\right)\right)\mu\left(dp\right)+o_p\left(1\right)\\
\xrightarrow{d}\ &2\int\eta\left(p\right)\mathbb{B}\left(F\left(p\right)\right)\mu\left(dp\right).
\end{aligned}$$

It remains to show that $\|J_j\|\xrightarrow{p}0$ for $j=2,3,4$. We will repeatedly use the fact that any linear functional of $\sqrt{N}m^\dagger\left(\cdot,F_N\right)$ is asymptotically tight and is therefore also bounded in probability.

Consider the $k-$th component of J_2, $\left(J_2\right)_k$:

$$\begin{aligned}
\left|\left(J_2\right)_k\right| &\leq 2\int\left(\frac{\partial}{\partial\theta_k}m\left(p,\theta_0,F_N\right)-\frac{\partial}{\partial\theta_k}m\left(p,\theta_0,F\right)\right)^2\mu\left(dp\right)\int\left(\sqrt{N}m\left(p,\theta_0,F_N\right)\right)^2\mu\left(dp\right)\\
&\leq \underline{\kappa}_0\underline{\kappa}^2_{D_F\frac{\partial}{\partial\theta}m}\int\mu\left(dp\right)\left|F_N-F\right|^2_\infty\int\left(\sqrt{N}m^\dagger\left(p,F_N\right)\right)^2\mu\left(dp\right),
\end{aligned}$$

where we first use Cauchy Schwarz inequality, then we take a pointwise mean value expansion at $\frac{\partial}{\partial\theta_k}m\left(p,\theta_0,F\right)$. Then remaining integrals in the second inequality are bounded and $\left|\left(J_2\right)_k\right|\xrightarrow{p}0$ since $\left|F_N-F\right|^2_\infty\xrightarrow{p}0$.

For J_3, take out the upper bounds of the integrand:

$$\left|\left(J_3\right)_k\right|\leq\underline{\kappa}_0\underline{\kappa}_{\frac{\partial}{\partial\theta}m}\sup_{p\in\mathcal{S}_P}\left|\sqrt{N}m^\dagger\left(p,F_N\right)\right|\int\left(\mu_N\left(dp\right)-\mu\left(dp\right)\right).$$

Since the supremum is a linear functional, we have $\sup_{p\in\mathcal{S}_P}\left|\sqrt{N}m^{\dagger}\left(p,F_N\right)\right| = O_p\left(1\right)$. Then by A2(ii) $\int\left(\mu_N\left(dp\right)-\mu\left(dp\right)\right)\xrightarrow{p}0$, and $\left|\left(J_3\right)_k\right|\xrightarrow{p}0$.

For J_4, applying similar arguments to J_2 and J_3, we have

$$\left|\left(J_4\right)_k\right| \leq \underline{\kappa}_0\underline{\kappa}_{D_F\frac{\partial}{\partial\theta}m}\left|F_N-F\right|_{\infty}\int\sqrt{N}m^{\dagger}\left(p,F_N\right)\mu\left(dp\right)\int\left(\mu_N\left(dp\right)-\mu\left(dp\right)\right).$$

So that $\left|\left(J_4\right)_k\right|\xrightarrow{p}0$ since $\int\sqrt{N}m^{\dagger}\left(p,F_N\right)\mu\left(dp\right)=O_p\left(1\right)$ and $\left|F_N-F\right|_{\infty}\int\left(\mu_N\left(dp\right)-\mu\left(dp\right)\right)\xrightarrow{p}0.\blacksquare$

Lemma 5. Under Assumptions A2(i), A2(ii) and A3, $\sup_{\theta\in\Theta}\left|M_N^*\left(\theta\right)-M\left(\theta\right)\right|\xrightarrow{a.s.^*}0$ for almost all samples $\{P_i\}_{i=1}^N$.

Proof of Lemma 5. Write

$$M_N^*\left(\theta\right)-M\left(\theta\right)=M_N^*\left(\theta\right)-M_N\left(\theta\right)+M_N\left(\theta\right)-M\left(\theta\right).$$

From Lemma 2, $\sup_{\theta\in\Theta}\left|M_N\left(\theta\right)-M\left(\theta\right)\right|\xrightarrow{a.s.}0$. Next,

$$\begin{aligned}M_N^*\left(\theta\right)-M_N\left(\theta\right) &= \int m\left(p,\theta,F_N^*\right)^2\left(\mu_N^*\left(dp\right)-\mu_N\left(dp\right)\right)+\int m\left(p,\theta,F_N^*\right)^2-m\left(p,\theta,F_N\right)^2d\mu_N\\ &= I_1^*\left(\theta\right)+I_2^*\left(\theta\right).\end{aligned}$$

We can use analogous arguments made in the proof of Lemma 2 to show $\sup_{\theta\in\Theta}\left|M_N^*\left(\theta\right)-M_N\left(\theta\right)\right|\xrightarrow{a.s.^*}0$. The result then follows from an application of the triangle inequality.$\blacksquare$

Lemma 6. Under Assumptions A2(i), A2(ii) and A3, $\widehat{\theta}^*\xrightarrow{a.s.^*}\theta_0$ for almost all samples $\{P_i\}_{i=1}^N$.

Proof of Lemma 6. Follows immediately from Lemmas 1 and 5.$\blacksquare$

Lemma 7. Under Assumptions A2(i), A2(ii) and A3, for any θ_N such that $\left\|\theta_N-\theta_0\right\|\xrightarrow{a.s.^*}0$, $\left\|H_N^*(\theta_N)-H\left(\theta_0\right)\right\|\xrightarrow{a.s.^*}0$ for almost all samples $\{P_i\}_{i=1}^N$.

Proof of Lemma 7. The same argument used in Lemma 3 can be applied to show that $\sup_{\theta\in\Theta}\left\|H_N^*(\theta)-H_N\left(\theta\right)\right\|\xrightarrow{a.s.^*}0$ by replacing the quantities defined using the original data by the bootstrap sample, and the limiting (population) objects by the sample counterparts using the original data. Then the triangle inequality we have:

$$\left\|H_N^*(\theta_N)-H\left(\theta_0\right)\right\|\leq\sup_{\theta\in\Theta}\left\|H_N^*(\theta)-H_N\left(\theta\right)\right\|+\left\|H_N(\theta_N)-H\left(\theta_0\right)\right\|.$$

Then by Lemma 3 $\left\|H_N^*(\theta_N)-H\left(\theta_0\right)\right\|\xrightarrow{a.s.^*}0.\blacksquare$

Lemma 8. Under Assumptions A2(i), A2(ii) and A3, $\sqrt{N}\left(\frac{\partial}{\partial\theta}M_N^*\left(\theta_0\right)-\frac{\partial}{\partial\theta}M_N\left(\theta_0\right)\right)$ converges in distribution to $\mathcal{N}\left(0,\Sigma\right)$ under $\mathbb{P}^*$ conditionally given $\{P_i\}_{i=1}^N$

Proof of Lemma 8. For notational simplicity we set μ_N^* and μ_N to be equal to μ for all N, otherwise the proof can be extended in the same manner as done in Lemma 4 with more algebra. Then

$$\begin{aligned}
& \sqrt{N}\frac{\partial}{\partial\theta}M_N^*(\theta_0) - \sqrt{N}\frac{\partial}{\partial\theta}M_N(\theta_0) \\
= \; & 2\int \frac{\partial}{\partial\theta}m(p,\theta_0,F)\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N)\right)\mu(dp) \\
& +2\int\left(\frac{\partial}{\partial\theta}m(p,\theta_0,F_N) - \frac{\partial}{\partial\theta}m(p,\theta_0,F)\right)\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N)\right)\mu(dp) \\
& +2\int\left(\frac{\partial}{\partial\theta}m(p,\theta_0,F_N^*) - \frac{\partial}{\partial\theta}m(p,\theta_0,F_N)\right)\sqrt{N}m(p,\theta_0,F_N)\mu(dp) \\
& +2\int\left(\frac{\partial}{\partial\theta}m(p,\theta_0,F_N^*) - \frac{\partial}{\partial\theta}m(p,\theta_0,F_N)\right)\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N^*)\right)\mu(dp) \\
= \; & J_1^* + J_2^* + J_3^* + J_4^*.
\end{aligned}$$

From Giné and Zinn (1990) we know the empirical distribution can be bootstrapped. So that $\sqrt{N}(F_N^* - F_N)$ has the same distribution as $\sqrt{N}(F_N - F)$ asymptotically, and similarly for their corresponding linear functionals. Thus J_1^* gives the desired distribution theory in the limit.

For the other terms, first consider J_2^*. Take the $k-$th component of J_2^* and apply Cauchy Schwarz inequality:

$$\begin{aligned}
|(J_2^*)_k| \; & \leq \; 2\int\left(\frac{\partial}{\partial\theta_k}m(p,\theta_0,F_N) - \frac{\partial}{\partial\theta_k}m(p,\theta_0,F)\right)^2\mu(dp)\int\left(\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N)\right)\right)^2\mu(dp \\
& \leq \; \underline{\kappa}_0\underline{\kappa}^2_{D_F\frac{\partial}{\partial\theta}m}\int\mu(dp)\left|F_N - F\right|_\infty^2\int\left(\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N)\right)\right)^2\mu(dp).
\end{aligned}$$

So $|(J_2^*)_k| \xrightarrow{p^*} 0$ since $|F_N - F|_\infty^2 \xrightarrow{p} 0$ and $\int\left(\sqrt{N}\left(m(p,\theta_0,F_N^*) - m(p,\theta_0,F_N)\right)\right)^2\mu(dp)$ is asymptotically tight under $\mathbb{P}^*$.

By an analogous reasoning, it is straightforward to show that $\|J_3^*\| \xrightarrow{p^*} 0$ and $\|J_4^*\| \xrightarrow{p^*} 0$.

The proof then follows from the triangle inequality.■

We define the following objects for the remaining lemmas. Let: $\widehat{Q}_T = \widehat{\mathbf{g}}^\top\widehat{\mathbf{g}}/T$, $Q_T = \mathbf{g}^\top\mathbf{g}/T$ and $Q = E[Q_T]$; $\widehat{1}_T = \mathbf{1}[\underline{\lambda}(\widehat{Q}_T) > 0]$ and $1_T = \mathbf{1}[\underline{\lambda}(Q_T) > 0]$; $\|\cdot\|_F$ denote the Frobenius norm for matrices, so that for any matrix A, $\|A\|_F = \mathrm{tr}\left(A^\top A\right)^{1/2}$ where $\mathrm{tr}(\cdot)$ is the trace operator. Note that $\|x\| = \|x\|_F$ for any column vector x.

Lemma 9. Under Assumptions B1, $\max_{1\leq t\leq T}\left\|\widehat{\mathbf{W}}^t - \mathbf{W}^t\right\| = O_p\left(1/\sqrt{N_T}\right)$.

Proof of Lemma 9. Under B1, the implications of Theorem 2 and Corollary 2 hold for all markets. Therefore for all t, $\left\|\widehat{\mathbf{W}}^t - \mathbf{W}^t\right\| = O_p\left(1/\sqrt{N^t}\right)$, and the proof follows since $N^t \geq N_T$.■

LEMMA 10. Under Assumptions B1 - B3, $\|Q_T - Q\|_F^2 = o_p(1)$.

PROOF OF LEMMA 10. For this it is sufficient to show $E\left[\|Q_T - Q\|_F^2\right] = o(1)$. First we write:

$$\begin{aligned} Q_T &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{g}^{t\top}\mathbf{g}^t, \text{ where} \\ \mathbf{g}^t &= \left(g^L\left(\Delta_1^t\right), \ldots, g^L\left(\Delta_{K^t-1}^t\right)\right)^\top. \end{aligned}$$

Under B2 $\left\{\mathbf{g}^{t\top}\mathbf{g}^t\right\}_{t=1}^{T}$ is an i.i.d. sequence of squared matrices of size L. Therefore $E[Q_T] = Q$ does not depend on T. Since $\|Q_T - Q\|_F^2$ is the sum of the squared of every element in $Q_T - Q$, we have:

$$\begin{aligned} E\left[\|Q_T - Q\|_F^2\right] &= \sum_{l,l'=1}^{L} E\left[(Q_T - Q)_{ll'}\right]^2 \\ &= \sum_{l,l'=1}^{L} \mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{g}^{t\top}\mathbf{g}^t\right)_{ll'}\right) \\ &= \frac{1}{T}\sum_{l,l'=1}^{L} \mathrm{Var}\left(\sum_{K=2}^{\overline{K}}\sum_{k=1}^{K-1} g_{lL}\left(\Delta_k^t\right) g_{l'L}\left(\Delta_k^t\right)\mathbf{1}\left[K^t = K\right]\right). \end{aligned}$$

The variance term can be bounded by using the law of total variance and, since $\overline{K} < \infty$, applying Cauchy Schwarz inequality together with B3(ii) repeatedly, so that

$$\begin{aligned} &\mathrm{Var}\left(\sum_{K=2}^{\overline{K}}\sum_{k=1}^{K-1} g_{lL}\left(\Delta_k^t\right) g_{l'L}\left(\Delta_k^t\right)\mathbf{1}\left[K^t = K\right]\right) \\ \leq\ & E\left[\mathrm{Var}\left(\left.\sum_{K=2}^{\overline{K}}\sum_{k=1}^{K-1} g_{lL}\left(\Delta_k^t\right) g_{l'L}\left(\Delta_k^t\right)\right| K^t = K\right)\right] \\ \leq\ & \underline{\kappa}_0\zeta(L)^4. \end{aligned}$$

Therefore $E\left[\|Q_T - Q\|_F^2\right] \leq \underline{\kappa}_0\zeta(L)^4 L^2/T$. By B3(ii) $E\left[\|Q_T - Q\|_F^2\right] = o(1)$, which implies $\|Q_T - Q\|_F^2 = o_p(1)$.■

LEMMA 11. Under Assumptions B1 - B3, $1_T = 1 + o_p(1)$.

PROOF OF LEMMA 11. Since $|\underline{\lambda}(Q_T - Q)| \leq \|Q_T - Q\|_F$, as the latter is the square root of the sum of all squared eigenvalues of $Q_T - Q$, we also have $\underline{\lambda}(Q_T - Q) = o_p(1)$ by Lemma 10. By the implication of B3(i), $\underline{\lambda}(Q) > 0$ therefore $\lim_{T\to\infty}\Pr[\underline{\lambda}(Q_T) > 0] = 1$ which completes the proof.■

LEMMA 12. Under Assumptions B1 - B3, $\left\|1_T(\widetilde{\beta} - \beta_L)\right\| = O_p\left(L^{-\alpha}\right)$.

PROOF OF LEMMA 12. First write $1_T(\widetilde{\beta} - \beta_L) = 1_T\left(\mathbf{g}^\top\mathbf{g}\right)^{-}\mathbf{g}^\top(\mathbf{y} - \mathbf{g}\beta_L)$. By Lemma 11, we have with probability approaching one (w.p.a. 1), $1_T\left(\mathbf{g}^\top\mathbf{g}\right)^{-} = 1_T Q_T^{-1}/T$ therefore,

$$\left\|1_T\left(\mathbf{g}^\top\mathbf{g}\right)^{-}\mathbf{g}^\top(\mathbf{y} - \mathbf{g}\beta_L)\right\| \leq \left\|1_T Q_T^{-1}\mathbf{g}^\top/\sqrt{T}\right\|\left\|(\mathbf{y} - \mathbf{g}\beta_L)/\sqrt{T}\right\|.$$

Next we show that $\left\|1_T Q_T^{-1}\mathbf{g}^\top/\sqrt{T}\right\| = O_p(1)$. Since $\underline{\lambda}(1_T Q_T)$ is bounded away from zero w.p.a. 1, as seen from the previous lemma, we have that $\overline{\lambda}\left(1_T Q_T^{-1}\right)$ is bounded from above w.p.a. 1. Then we have: $\left\|1_T Q_T^{-1}\mathbf{g}^\top/\sqrt{T}\right\|^2 = \overline{\lambda}\left(1_T Q_T^{-1}\mathbf{g}^\top\mathbf{g}Q_T^{-1}/T\right) = \overline{\lambda}\left(1_T Q_T^{-1}\right) = O_p(1)$, so that $\left\|1_T Q_T^{-1}\mathbf{g}^\top/\sqrt{T}\right\| = O_p(1)$. Note that $\mathbf{y}$ can be written as a vector of $\{G(\Delta_k^t)\}_{k=1,t=1}^{K^t,T}$, see equation (14). Then using B3(iii), we have

$$
\begin{aligned}
\left\|(\mathbf{y}-\mathbf{g}\beta_L)/\sqrt{T}\right\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K^t-1}\left(G(\Delta_k^t) - g^L(\Delta_k^t)^\top\beta_L\right)^2 \\
&\leq L^{-2\alpha}\left(\overline{K}-1\right)\overline{K}/2.
\end{aligned}
$$

So that $\left\|(\mathbf{y}-\mathbf{g}\beta_L)/\sqrt{T}\right\| = O(L^{-\alpha})$, which completes the proof.■

LEMMA 13. Under Assumptions B1 - B4, $\left\|(\widehat{\mathbf{y}}-\mathbf{y})/\sqrt{T}\right\| = O_p\left(1/\sqrt{N_T}\right)$.

PROOF OF LEMMA 13. Recall that $\widehat{\mathbf{y}}$ is a vector of $\{\widehat{q}_k^t\}_{k=1,t=1}^{K^t,T}$, so that

$$
\left\|(\widehat{\mathbf{y}}-\mathbf{y})/\sqrt{T}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K^t-1}\left(\sum_{k'=1}^{k}\left(q_{k'}^t-\widehat{q}_{k'}^t\right)\right)^2.
$$

The proof is an immediate consequence of Lemma 9 since we have: $\sum_{k=1}^{K^t-1}\left(\sum_{k'=1}^{k}\left(q_{k'}^t-\widehat{q}_{k'}^t\right)\right)^2 = O_p(1/N_T)$ for all t.■

LEMMA 14. Under Assumptions B1 - B4, $\left\|(\widehat{\mathbf{g}}-\mathbf{g})/\sqrt{T}\right\| = O_p\left(\zeta(L)/\sqrt{N_T}\right)$.

PROOF OF LEMMA 14. Recall that $\widehat{\mathbf{g}}$ is a matrix of $\left\{g_{iL}(\widehat{\Delta}_k^t)\right\}_{k=1,t=1}^{K^t,T}$. From Assumption B4(i), we can take a mean value expansion so that for any t,k $\left|g_{iL}(\widehat{\Delta}_k^t) - g_{iL}(\Delta_k^t)\right| \leq \zeta(L)\left|\widehat{\Delta}_k^t - \Delta_k^t\right|$, which is $O_p\left(\zeta(L)/\sqrt{N_T}\right)$ by Lemma 9. The proof then follows by the same argument as used in Lemma 13.■

LEMMA 15. Under Assumptions B1 - B4, $\left\|\widehat{Q}_T - Q_T\right\| = O_p\left(\zeta(L)/\sqrt{N_T}\right)$.

PROOF OF LEMMA 15. Since $\widehat{\mathbf{g}}^\top\widehat{\mathbf{g}} - \mathbf{g}^\top\mathbf{g} = 2(\widehat{\mathbf{g}}-\mathbf{g})^\top\mathbf{g} + (\widehat{\mathbf{g}}-\mathbf{g})^\top(\widehat{\mathbf{g}}-\mathbf{g})$, we have

$$
\left\|\widehat{Q}_T - Q_T\right\| \leq 2\left\|(\widehat{\mathbf{g}}-\mathbf{g})^\top\mathbf{g}/T\right\| + \left\|(\widehat{\mathbf{g}}-\mathbf{g})^\top(\widehat{\mathbf{g}}-\mathbf{g})/T\right\|.
$$

We can bound $\left\|(\widehat{\mathbf{g}}-\mathbf{g})^\top\mathbf{g}/T\right\|$ by:

$$
\begin{aligned}
\left\|(\widehat{\mathbf{g}}-\mathbf{g})^\top\mathbf{g}/T\right\| &\leq \left\|(\widehat{\mathbf{g}}-\mathbf{g})/\sqrt{T}\right\|\left\|\mathbf{g}/\sqrt{T}\right\| \\
&= O_p\left(\zeta(L)/\sqrt{N_T}\right),
\end{aligned}
$$

as we have shown $\left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/\sqrt{T}\right\| = O_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$ in Lemma 14 and using the fact that $\left\|\mathbf{g}/\sqrt{T}\right\|^2 = \overline{\lambda}\left(Q_T\right) = O_p\left(1\right)$. The latter follows from Lemma 10, which implies $\left|\overline{\lambda}\left(Q_T - Q\right)\right| \leq \left\|Q_T - Q\right\|_F = o_p\left(1\right)$, together with B3(i) they ensure $\overline{\lambda}\left(Q_T\right)$ is bounded w.p.a.1. Also, by Lemma 14 $\left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/T\right\| = o_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$ since $\left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/T\right\| = \left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/\sqrt{T}\right\|^2 = O_p\left(\zeta\left(L\right)^2/N_T\right)$ which is $o_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$ by B4(ii).■

Lemma 16. Under Assumptions B1 - B4, $\widehat{1}_T = 1 + o_p\left(1\right)$.

Proof of Lemma 16. From Lemma 14, $\widehat{1}_T = 1_T + o_p\left(1\right)$. The proof then follows from Lemma 11.■

Lemma 17. Under Assumptions B1 - B4, $\left\|\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}} - \mathbf{g}^\top\mathbf{y}\right)/\sqrt{T}\right\| = O_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$.

Proof of Lemma 17. We begin by writing: $\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}} - \mathbf{g}^\top\mathbf{y} = \left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\mathbf{y} + \mathbf{g}^\top\left(\widehat{\mathbf{y}}-\mathbf{y}\right) + \left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\left(\widehat{\mathbf{y}}-\mathbf{y}\right)$. We can bound $\left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\mathbf{y}/T$ by:

$$\begin{aligned}
\left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\mathbf{y}/T\right\| &\leq \left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/\sqrt{T}\right\|\left\|\mathbf{y}/\sqrt{T}\right\| \\
&= O_p\left(\zeta\left(L\right)/\sqrt{N_T}\right),
\end{aligned}$$

as we have $\left\|\left(\widehat{\mathbf{g}}-\mathbf{g}\right)/\sqrt{T}\right\| = O_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$ from Lemma 14 and $\left\|\mathbf{y}/\sqrt{T}\right\| = O_p\left(1\right)$. The latter holds since $\left\|\mathbf{y}/\sqrt{T}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K^t-1}\left(1-\sum_{k'=1}^{k}q_{k'}^t\right)^2 \leq \overline{K}\left(\overline{K}-1\right)/2 < \infty$. The same line of arguments can be used to show that $\mathbf{g}^\top\left(\widehat{\mathbf{y}}-\mathbf{y}\right)/T = O_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$ and $\left(\widehat{\mathbf{g}}-\mathbf{g}\right)^\top\left(\widehat{\mathbf{y}}-\mathbf{y}\right)/T = o_p\left(\zeta\left(L\right)/\sqrt{N_T}\right)$.■

Proofs of Theorems

Our proofs of Theorems 1 and 2 follow standard steps for an M-estimator (e.g. see van der Vaart (2000)). The proof of Theorem 3 follows the approach of Arcones and Giné (1992). We employ a similar strategy used in Newey (1997) to prove Theorem 4.

Proof of Theorem 1. Immediately holds from Lemmas 1 and 2, following the standard conditions for consistency of an M-estimator.■

Proof of Theorem 2. Our estimator satisfies the following first order condition, $0 = \frac{\partial}{\partial\theta}M_N\left(\widehat{\theta}\right)$. Applying a mean value expansion,

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}M_N\left(\theta_0\right) + H_N\left(\widetilde{\theta}\right)\left(\widehat{\theta}-\theta_0\right) \\
&= \frac{\partial}{\partial\theta}M_N\left(\theta_0\right) + H\left(\theta_0\right)\left(\widehat{\theta}-\theta_0\right) + o_p\left(\left\|\widehat{\theta}-\theta_0\right\|\right),
\end{aligned}$$

where $\widetilde{\theta}$ denotes some intermediate value between $\widehat{\theta}$ and θ_0, and the second equality follows from Lemma 3 and Theorem 1. Assumption A2(iii) ensures $H \equiv H(\theta_0)$ is invertible, by re-arranging and multiplying by $\sqrt{N}$, we have

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right) = H^{-1}\left(\sqrt{N}\frac{\partial}{\partial\theta}M_N(\theta_0)\right) + o_p(1).$$

The result then follows from applying Cramér theorem to Lemma 4.■

Proof of Theorem 3. Similar to the proof of Theorem 2, our bootstrap estimator satisfies the following first order condition,

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta}M_N^*(\theta_0) + H_N^*\left(\widetilde{\theta}^*\right)\left(\widehat{\theta}^*-\theta_0\right) \\ &= \frac{\partial}{\partial\theta}M_N^*(\theta_0) + H\left(\widehat{\theta}^*-\theta_0\right) + o_{p^*}\left(\left\|\widehat{\theta}^*-\theta_0\right\|\right). \end{aligned}$$

where $\widetilde{\theta}^*$ denotes some intermediate value between $\widehat{\theta}^*$ and θ_0, and the second equality follows from Lemmas 6 and Lemma 7. Using A2(iii), we have

$$\sqrt{N}\left(\widehat{\theta}^*-\theta_0\right) = H^{-1}\left(\sqrt{N}\frac{\partial}{\partial\theta}M_N^*(\theta_0)\right) + o_{p^*}(1).$$

Take the difference between $\sqrt{N}\left(\widehat{\theta}^*-\theta_0\right)$ and $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ (from the last equation in the previous proof), we have

$$\sqrt{N}\left(\widehat{\theta}^*-\widehat{\theta}\right) = H^{-1}\left(\sqrt{N}\frac{\partial}{\partial\theta}M_N^*(\theta_0) - \sqrt{N}\frac{\partial}{\partial\theta}M_N(\theta_0)\right) + o_{p^*}(1).$$

The proof is completed by applying Cramér theorem to Lemma 8.■

Proof of Theorem 4. For each c, we decompose:

$$\widehat{G}(c) - G(c) = \widehat{G}(c) - \widetilde{G}(c) + \widetilde{G}(c) - G(c).$$

First consider $\widetilde{G}(c) - G(c)$, which can be decomposed further into $g^L(c)^\top\left(\widetilde{\beta}-\beta_L\right) + \left(g^L(c)^\top\beta_L - G(c)\right)$. These terms are similar to the components of a series estimator of a regression function. We have,

$$\begin{aligned} \left|1_T\left(\widetilde{G}-G\right)\right|_\infty &\le \zeta(L)\left\|1_T(\widetilde{\beta}-\beta_L)\right\| + O\left(L^{-\alpha}\right) \\ &= O_p\left(\zeta(L)L^{-\alpha}\right). \end{aligned}$$

The rate above follows from Lemma 12 and Assumption B3(iii). And from Lemma 11, we have $1 - 1_T = o_p(1)$, therefore:

$$\begin{aligned} \left|\widetilde{G}-G\right|_\infty &= \left|1_T\left(\widetilde{G}-G\right)\right|_\infty + o_p\left(\left|\widetilde{G}-G\right|_\infty\right) \\ &= O_p\left(\zeta(L)L^{-\alpha}\right). \end{aligned}$$

Next consider $\widehat{G}(c)-\widetilde{G}(c)$, which accounts for the generated variables. We focus on $\widehat{1}_T 1_T\left(\widehat{G}(c)-\widetilde{G}(c)\right)$. In particular Lemmas 11 and 16 ensure that Q_T^{-1} and $\widehat{Q}_T^{-1}$ exist w.p.a. 1, and we have

$$\widehat{1}_T 1_T\left(\widehat{G}(c)-\widetilde{G}(c)\right)=\widehat{1}_T 1_T g^L(c)^\top\left(\widehat{Q}_T^{-1}\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}/T-Q_T^{-1}\mathbf{g}^\top\mathbf{y}/T\right).$$

We now show that $\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}/T-Q_T^{-1}\mathbf{g}^\top\mathbf{y}/T\right)=O_p\left(\zeta(L)/\sqrt{N_T}\right)$. To see this, consider:

$$\begin{aligned}\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}/T-Q_T^{-1}\mathbf{g}^\top\mathbf{y}/T\right) &= \widehat{1}_T 1_T Q_T^{-1}\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T\\ &\quad+\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\mathbf{g}^\top\mathbf{y}/T\\ &\quad+\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T.\end{aligned}$$

Lemma 10 ensures that $\widehat{Q}_T^{-1}$ and Q_T^{-1} converge in probability to Q^{-1}, which is known to be bounded by assumption B3(i). Therefore, using Lemma 17,

$$\begin{aligned}\widehat{1}_T 1_T Q_T^{-1}\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T &= Q^{-1}\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T+o_p\left(\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T\right)\\ &= O_p\left(\zeta(L)/\sqrt{N_T}\right).\end{aligned}$$

Note we can write $\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)=\widehat{1}_T 1_T\widehat{Q}_T^{-1}\left(Q_T-\widehat{Q}_T\right)Q_T^{-1}$. Then, in addition to the above, by Lemma 15: $\left\|\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\right\|=O_p\left(\zeta(L)/\sqrt{N_T}\right)$. We also have $\left\|\mathbf{g}^\top\mathbf{y}/T\right\|\leq\left\|\mathbf{g}/\sqrt{T}\right\|\left\|\mathbf{y}/\sqrt{T}\right\|$, which we know is bounded in probability since both $\left\|\mathbf{g}/\sqrt{T}\right\|$ and $\left\|\mathbf{y}/\sqrt{T}\right\|$ are $O_p(1)$ (we have shown these in the proofs of Lemmas 15 and 17 respectively). Hence,

$$\begin{aligned}\left\|\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\mathbf{g}^\top\mathbf{y}/T\right\| &\leq \left\|\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\right\|\left\|\mathbf{g}^\top\mathbf{y}/T\right\|\\ &= O_p\left(\zeta(L)/\sqrt{N_T}\right).\end{aligned}$$

Lastly, under B4(ii), we have

$$\begin{aligned}\left\|\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T\right\| &\leq \left\|\widehat{1}_T 1_T\left(\widehat{Q}_T^{-1}-Q_T^{-1}\right)\right\|\left\|\left(\widehat{\mathbf{g}}^\top\widehat{\mathbf{y}}-\mathbf{g}^\top\mathbf{y}\right)/T\right\|\\ &= o_p\left(\zeta(L)/\sqrt{N_T}\right).\end{aligned}$$

Therefore we have $\left|\widehat{1}_T 1_T\left(\widehat{G}-\widetilde{G}\right)\right|_\infty=O_p\left(\zeta^2(L)/\sqrt{N_T}\right)$, and by Lemmas 11 and 16 we know $\widehat{1}_T 1_T=1+o_p(1)$, so that

$$\begin{aligned}\left|\widehat{G}-\widetilde{G}\right|_\infty &= \left|\widehat{1}_T 1_T\left(\widehat{G}-\widetilde{G}\right)\right|_\infty+o_p\left(\left|\left(\widehat{G}-\widetilde{G}\right)\right|_\infty\right)\\ &= O_p\left(\zeta^2(L)/\sqrt{N_T}\right).\end{aligned}$$

Then $\left|\widehat{G}-G\right|_\infty$ can be bounded by using the triangle inequality, which completes the proof.■

Asymptotic Variances for Corollaries 1, 2 and 3

We take the asymptotic distribution of $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ derived in Theorem 2 as the starting point. The asymptotic variances for the estimators described in Corollaries 1, 2 and 3 can be obtained using the delta-method. In particular, given that $\sqrt{N}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0, H^{-1}\Sigma H^{-1}\right)$ and θ_0 belongs to the interior of $\Theta \subseteq [0,1]^{K-1}$, then for any $l-$vector value function $\mathbf{x}:\Theta \rightarrow \mathbb{R}^l$ that is continuously differentiable at θ_0: $\sqrt{N}\left(\mathbf{x}(\widehat{\theta})-\mathbf{x}\left(\theta_0\right)\right) \xrightarrow{d} N\left(0, D_{\mathbf{x}}^{\top} H^{-1}\Sigma H^{-1} D_{\mathbf{x}}\right)$ where $D_{\mathbf{x}}^{\top}$ is the Jacobian matrix, $\left(D_{\mathbf{x}}^{\top}\right)_{ij}=\frac{\partial}{\partial q_j}\mathbf{x}_i\left(\mathbf{q}\right)$. We provide $D_{\mathbf{x}}^{\top}$ for the three cases below.

Corollary 1. For $\widehat{\theta}_K$, $\mathbf{x}\left(\theta_0\right)=1-\sum_{k=1}^{K-1} q_k$. Here $D_{\mathbf{x}}^{\top}$ is simply $(-1,\ldots,-1)$.

Corollary 2. For $\widehat{\boldsymbol{\Delta}}$, $\mathbf{x}_i\left(\theta_0\right)=\int_{z=0}^{1} w\left(z;\mathbf{q}\right)\left[\left(i+1\right)z-1\right]\left(1-z\right)^{i-1} dz$ for $i=1,\ldots,K-1$. Using equations (3) and (12), and substituting in $q_K=1-\sum_{k=1}^{K} q_k$, we can write

$$
\begin{aligned}
w\left(z;\mathbf{q}\right) &= \overline{p}s\left(z;\mathbf{q}\right)-s\left(z;\mathbf{q}\right)r\left(\mathbf{q}\right)+r\left(\mathbf{q}\right), \text{ where}\\
s\left(z;\mathbf{q}\right) &= \frac{q_1}{\sum_{k=1}^{K} kq_k\left(1-z\right)^{k-1}}\\
&= \frac{q_1}{K\left(1-z\right)^{K-1}+\sum_{k=1}^{K-1} q_k\left(k\left(1-z\right)^{k-1}-K\left(1-z\right)^{K-1}\right)},\\
r\left(\mathbf{q}\right) &= \underline{p}-\frac{\left(\overline{p}-\underline{p}\right)q_1}{\sum_{k=2}^{K} kq_k}\\
&= \underline{p}-\frac{\left(\overline{p}-\underline{p}\right)q_1}{K\left(1-q_1\right)+\sum_{k=2}^{K}\left(k-K\right)q_k}.
\end{aligned}
$$

The integrand defined in $\mathbf{x}_i$ is continuously differentiable at any $\mathbf{q}$ that lies in the interior of Θ for all z. Therefore we can differentiate under the integral sign. In particular,

$$
\begin{aligned}
\frac{\partial \mathbf{x}_i\left(\theta_0\right)}{\partial q_j} &= \int_{z=0}^{1}\frac{\partial w\left(z;\mathbf{q}\right)}{\partial q_j}\left[\left(i+1\right)z-1\right]\left(1-z\right)^{i-1} dz, \text{ where}\\
\frac{\partial w\left(z;\mathbf{q}\right)}{\partial q_j} &= \left(\overline{p}-r\left(\mathbf{q}\right)\right)\frac{\partial}{\partial q_j}s\left(z;\mathbf{q}\right)+\left(1-s\left(z;\mathbf{q}\right)\right)\frac{\partial}{\partial q_j}r\left(\mathbf{q}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial q_j}s\left(z;\mathbf{q}\right)\\
&= \begin{cases}
\frac{1}{K(1-z)^{K-1}+\sum_{k=1}^{K-1} q_k\left(k(1-z)^{k-1}-K(1-z)^{K-1}\right)}-\frac{q_1\left((1-z)-K(1-z)^{K-1}\right)}{\left(K(1-z)^{K-1}+\sum_{k=1}^{K-1} q_k\left(k(1-z)^{k-1}-K(1-z)^{K-1}\right)\right)^2} & \text{for } j=1\\
\frac{-q_1\left(j(1-z)^{j-1}-K(1-z)^{K-1}\right)}{\left(K(1-z)^{K-1}+\sum_{k=1}^{K-1} q_k\left(k(1-z)^{k-1}-K(1-z)^{K-1}\right)\right)^2} & \text{for } j>1
\end{cases},
\end{aligned}
$$

$$
\frac{\partial}{\partial q_j}r\left(\mathbf{q}\right)=\begin{cases}
-\frac{\left(\overline{p}-\underline{p}\right)}{K(1-q_1)+\sum_{k=2}^{K}(k-K)q_k}-\frac{K\left(\overline{p}-\underline{p}\right)q_1}{\left(K(1-q_1)+\sum_{k=2}^{K}(k-K)q_k\right)^2} & \text{for } j=1\\
\frac{(j-K)\left(\overline{p}-\underline{p}\right)q_1}{\left(K(1-q_1)+\sum_{k=2}^{K}(k-K)q_k\right)^2} & \text{for } j>1
\end{cases}.
$$

Corollary 3. For $\widehat{\mathbf{G}}$, $\mathbf{x}_i(\theta_0) = 1 - \sum_{k=1}^{i} q_k$ for $i = 1, \ldots, K-1$. Here $D_T^\top$ is the following lower triangular matrix consisting of -1's, i.e.

$$D_T^\top = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & & \ddots & 0 \\ \vdots & & \ddots & & \vdots \\ -1 & \ddots & & -1 & 0 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}.$$

References

[1] Andrews, D.W.K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, **59**, 307–345.

[2] Andrews, D.W.K. (1994): "Empirical Processes Methods in Econometrics," *Handbook of Econometrics*, vol. 4, eds. R.F. Engle and D. McFadden. North-Holland.

[3] Arcones, M., and E. Giné (1992): "On the Bootstrap of M-estimators and Other Statistical Functionals," in *Exploring the Limits of the Bootstrap.*

[4] Blevins, J., and G.T. Senney (2014): "Dynamic Selection and Distributional Bounds on Search Costs in Dynamic Unit-Demand Models," *Working Paper, Ohio State University.*

[5] Brown, D., and M. Wegkamp (2002): "Weighted Minimum Mean–Square Distance from Independence Estimation," *Econometrica*, **70**, 2035-2051.

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

[7] Carrasco, M., and J.P. Florens (2000): "Generalization Of Gmm To A Continuum Of Moment Conditions," *Econometric Theory*, **16**, 797-834.

[8] Carrasco, M., and J.P. Florens (2002): "Efficient GMM Estimation using the Empirical Characteristic Function," *Working Paper, University of Montreal.*

[9] Chaussé, P. (2011): "Generalized Empirical Likelihood for a Continuum of Moment Conditions," *Working Paper, University of Waterloo.*

[10] Chen, X. (2007): "Large Sample Sieve Estimation of Semi-nonparametric Models," in *Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 6. North Holland, Amsterdam, 5549–5632.*

[11] De los Santos, B., A. Hortaçsu and M. Wildenbeest (2012): "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, **102**, 2955-2580.

[12] Domínguez, M., and I. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, **72**, 1601-1615.

[13] Escanciano, J., D. Jacho-Chávez, and A. Lewbel (2012): "Identification and Estimation of Semi-parametric Two Step Models," *Working Paper, University of Indiana.*

[14] Escanciano, J., D. Jacho-Chávez, and A. Lewbel (2014): “Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing,” *Journal of Econometrics*, **178**, 426-443.

[15] Giné, E. and J. Zinn (1990): “Bootstrapping General Empirical Measures,” *The Annals of Probability*, **18**, 851-869.

[16] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms*, Themes in Modern Econometrics, Cambridge University Press.

[17] Hong, H. and M. Shum (2006): “Using Price Distribution to Estimate Search Costs,” *RAND Journal of Economics*, **37**, 257-275.

[18] Hortaçsu, A., and C. Syverson (2004): “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds,”, *Quarterly Journal of Economics*, May 2004

[19] Kitamura, Y. (2007): “Empirical Likelihood Methods in Econometrics: Theory and Practice,” in *Advances in Economics and Econometrics Theory and Applications, Ninth World Congress, Volume 3.*

[20] Komunjer, I. (2012): “Global Identification in Nonlinear Models with Moment Restrictions,” *Econometric Theory*, **28**, 719-729.

[21] Kosorok, M. (2008): *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics.

[22] Lee, J., and P.M. Robinson (2014): “Series Estimation under Cross-sectional Dependence,” *Working Paper, LSE.*

[23] Lorentz, G.G. (1986): *Bernstein Polynomials.* Chelsea Publishing Company: New York.

[24] Mammen, E., C. Rothe, and M. Schienle (2012): “Nonparametric Regression with Nonparametrically Generated Covariates,” *Annals of Statistics*, **40**, 1132-1170.

[25] Mammen, E., C. Rothe, and M. Schienle (2014): “Semiparametric Estimation with Generated Covariates,” *Working Paper, University of Heidelberg.*

[26] Manski, C. F. (1983): “Closest Empirical Distribution Estimation,” *Econometrica*, **51**, 305–320.

[27] Moraga-González, J. and M. Wildenbeest (2007): "Maximum Likelihood Estimation of Search Costs," *European Economic Review*, **52**, 820-48.

[28] Moraga-González, J., Z. Sándor and M. Wildenbeest (2010): "Nonsequential Search Equilibrium with Search Cost Heterogeneity," *Working Paper, University of Indiana.*

[29] Moraga-González, J., Z. Sándor and M. Wildenbeest (2012): "Consumer Search and Prices in the Automobile Market," *Working Paper, University of Indiana.*

[30] Moraga-González, J., Z. Sándor and M. Wildenbeest (2013): "Semi-nonparametric Estimation of Consumer Search Costs," *Journal of Applied Econometrics*, **28**, 1205-1223.

[31] Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, **79**, 147–168.

[32] Newey, W., and R. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, **72**, 219–255.

[33] Owen, A. (2001): *Empirical Likelihood.* Chapman and Hall.

[34] Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimator for Dynamic Games," *Review of Economics Studies*, **75**, 901-928.

[35] Robinson, P.M. (1988): "The Stochastic Difference Between Econometric Statistics," *Econometrica*, **56**, 531-548.

[36] Sanches, F., D. Silva Junior and S. Srisuma (2016): "Ordinary Least Sqaures Estimation of a Dynamic Game Model," *International Economic Review*, **57**, 623-634.

[37] Su, C. and K. Judd (2012), "Constrained Optimization Approaches to Estimating Structural Models," *Econometrica*, **80**, 2213–2230.

[38] van der Vaart, A. (2000): *Asymptotic Statistics.* Cambridge University Press.