Nonparametric Identification of Differentiated Products Demand Using Micro Data*

Steven T. Berry
Yale University

Philip A. Haile
Yale University

May 17, 2020

Abstract

We consider nonparametric identification of demand when one has "micro data" linking the characteristics of individual consumers to their choices. A primary challenge to identification of demand is the presence of unobserved product characteristics (or other "demand shocks") that vary across markets. Each of these latent factors directly affects the quantity demanded of all related goods and typically affects all prices as well. In the case of "market level data," Berry and Haile (2014) showed that the resulting simultaneity/endogenetiy challenges can be overcome with instruments for all endogenous variables, i.e., for the prices and quantities of all goods in the demand system. Here we show that micro data not only permits richer demand specifications but also can substantially soften the reliance on instrumental variables, reducing both the number and types of instruments required. We demonstrate identification of a nonparametric model of demand nesting common empirical specifications and requiring neither the structure of a "special regressor" nor a "full support" assumption on consumer-level observables. A key insight is that heterogeneity in consumer characteristics creates variation in consumers' decision problems within a single market, where the latent demand shocks are fixed. As a result, observed differences across markets in the consumer observables required to match certain demand vectors can reveal the latent vectors of demand shocks, using instruments only for prices.

*Early versions of this work were presented in the working paper "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," first circulated in 2007 and superseded by the present paper. We thank Suk Joon Son for helpful comments.

1 Introduction

Empirical models of differentiated products demand are an important part of the applied econometrics toolkit, underlying influential empirical work in many fields of economics. Although practical considerations typically dictate reliance on parsimonious functional forms for estimation of demand, an important question concerns the nonparametric foundation for demand estimation. In this paper we consider the identification of nonparametric differentiated products demand models, focusing on the case in which one has access to "micro data" matching attributes of individual consumers to their purchase decisions. We show that the availability of micro data not only allows a more richly specified model, but also can substantially relax the both the number and types of instrumental variables relied upon for identification.

Micro data linking consumer characteristics to consumer choices are not always available in applications. In Berry and Haile (2014) we explored identification of demand when, as in Bresnahan (1981) or Berry, Levinsohn, and Pakes (1995), one observes only market-level data on product prices, product/market characteristics, market shares, and the distribution of consumer characteristics. However many applications do offer micro data. A classic example is McFadden's (McFadden, Talvitie, and Associates (1977)) study of transportation demand, where each consumer's preferences over different modes of transport are altered by the distance from her location to each mode. This example illustrates the defining characteristic of the type of micro data considered here: consumer-specific observables that alter the relative attractiveness of different options. Consumers' distances to different options have been used in a number of other applications as well, including those involving demand for hospitals, retail outlets, residential locations, or schools, as in the examples of Capps, Dranove, and Satterthwaite (2003), Burda, Harding, and Hausman (2015), Bayer, Keohane, and Timmins (2009), and Neilson (2019). More broadly, observable consumer-level attributes that shift tastes for products might include a household's income or other demographics. Family size has been modeled as shifting preferences for cars (Goldberg (1995)), and for neighborhoods (Bayer, Ferreira, and McMillan (2007)). Other examples include product-specific advertising exposure (Ackerberg (2003)), consumer-newspaper ideological match (Gentzkow and Shapiro (2010)), the match between household demographics and those of a school or neighborhood (Bayer, Ferreira, and McMillan (2007), Hom (2018)), and the match between voter demographics and candidate characteristics (Kawai, Toyama, and Watanabe (2020)).[1]

It is unsurprising that micro data can allow a richer specification of the empirical demand model. Our main insight, however, is that such data can also substantially reduce the reliance on instrumental variables for identification. A fundamental challenge to identification of demand arises from the elementary observation that the quantity demanded of any one good depends on all characteristics of that good and all related goods (complements or substitutes). Such characteristics include not only the prices of

[1] Here we cite only a small representative handful of papers out of a selection that spans many topics and many years.

each good in the relevant market but also unobserved characteristics (or, more generally, latent "demand shocks") that vary across markets. Likewise, the equilibrium price of any one good typically depends on the all other characteristics (observed and unobserved) of that good and all related goods.[2] Thus, in a market with J goods, equilibrium prices and quantities are determined in system of $2J$ fully simultaneous equations. In such systems, identification generally cannot be obtained using strategies (e.g., control function) familiar from triangular models (see Blundell and Matzkin (2014) and Berry and Haile (2016)). In Berry and Haile (2014), however, we show that nonparametric identification of demand can be obtained from standard IV conditions, given instruments for all prices and quantities. The need to instrument for quantities—indeed, to consider a system of equations for both supply and demand to identify demand alone—may be surprising. But this need is tightly connected to identification results for other simultaneous models (see, e.g., Matzkin (2015), Berry and Haile (2014) and Berry and Haile (2018)) and is easily recognized in the IV requirements of parametric models used for estimation in practice.[3] At an intuitive level, to measure any own- or cross-price elasticity one needs to isolate the change in quantity demanded that results from shifting one price while holding fixed $J-1$ other prices and J latent demand shocks; $2J$ excluded instruments can provide the independent variation needed to isolate this response. One important finding in Berry and Haile (2014) is the essential role of competing goods' exogenous characteristics—sometimes called "BLP instruments"—in providing instruments for quantities.

In this paper, we develop conditions under which the availability of micro data cuts the number of required instruments in half. In particular, variation in micro data can eliminate the need to instrument for quantities and, therefore, the necessary reliance on BLP instruments. The use of micro data also makes it possible to specify a more flexible demand model and makes new kinds of instruments available. The reduction in IV requirements is obtained because micro data provides a form of observable variation in the choice problem faced by different consumers within the same market. This creates a panel structure, where we can exploit both "within market" and "between market" variation. Critically, within a given market the latent market-level demand shocks are fixed; thus the observed responses to variation in choice problems within a given market cannot be confounded by variation in these shocks. Of course, prices are also fixed within a market. But "clean" within-market variation can allow us to pin down the latent demand shocks by instrumenting only for prices in the cross-section of markets. Once the latent demand shocks are known, identification of demand becomes trivial.

Our model of demand is nonparametric and, although we focus on discrete choice demand, our results generalize to continuous demand systems with representations satis-

[2]We emphasize prices as the leading case of endogenous product characteristics. However our results generalize directly to cases with additional endogenous product characteristics, although additional instruments will be required.

[3]This is easiest to see in the linear-in-parameters nested logit model of Berry (1994), but is also clear in the nonlinear models of Berry, Levinsohn, and Pakes (1995) and Nevo (2001). See the discussion in Berry and Haile (2016).

fying the “connected substitutes” condition of Berry, Gandhi, and Haile (2013) or other conditions ensuring “invertibility” of demand. We allow all consumer attributes to shift preferences for all products, avoiding any *a priori* exclusivity assumptions on these observables.[4] However, in addition to standard IV conditions, our results rely on three important assumptions. First, we require a nonparametric index restriction—formally a weak separability assumption—on the way the market-level demand shocks and some observed consumer attributes enter the model.[5] Second, we require injectivity of the mappings that link consumer attributes to choice probabilities. Our injectivity assumptions generalize standard specifications, but include an important requirement that the number of consumer attributes be as large as the number of products.

Finally, we require sufficient variation in the consumer observables to satisfy a “common choice probability” condition that we believe is new to the literature. Given the number of products available, this condition requires that there be some point s^* in the probability simplex such that in every market one can obtain s^* as the conditional choice probability by conditioning on the “right” set of consumer observables for that market. This requirement contrasts with a standard “full support” condition, which would imply that every *every* point s within the simplex is a common choice probability. Our condition allows for a broad range of cases where choice probabilities are never close to one or zero. It is also verifiable.

Our insights build on strategies used in the parametric applied literature by, e.g., Berry, Levinsohn, and Pakes (2004) and Bayer, Ferreira, and McMillan (2007), who pointed out the potential for certain types of panel data to pin down “substitution patterns” without instruments beyond those for prices. We also connect to a substantial econometrics literature on the use of micro data to identify discrete choice models. Indeed, the traditional individual-level discrete choice literature exploits micro data almost by definition. However, our approach generalizes earlier work in several directions. As in the empirical literature using market-level data following on Berry (1994) and Berry, Levinsohn, and Pakes (1995), we emphasize the role of market-level unobserved product characteristics that result in the econometric endogeneity of some product-level characteristics, particularly price. Accounting for this endogeneity is key to the identification of policy-relevant market demand elasticities. Our focus on market-level endogeneity differentiates our work from much of the classic work on the identification of discrete choice models. In addition, many existing nonparametric and semiparametric identification results for discrete choice models require a consumer attribute for each choice with at least

[4]This contrasts with the frequent reliance on exclusion restrictions in the nonparametric simultaneous equations literature, as in Matzkin (2015) and Berry and Haile (2018).

[5]Despite some superficial similarity, both the form and role of this index restriction differ from those in our earlier work (Berry and Haile (2014) and Berry and Haile (2018)). In each case the index restriction helps to deal with the issue of a large vector of unobservables that nonlinearly affect the demand for each product. But the indices in this paper are tied to consumer attributes rather than product characteristics, and this index structure is employed in a different way to obtain identification.

some features of a "special regressor."[6] Such a variable is typically specified as entering the discrete choice utility function linearly and, in the multinomial context, with each such attribute restricted to enter the (conditional indirect) utility of only one choice—i.e., excluded from the utilities of other choices. These functional form and exclusion restrictions are then combined with a "full support" assumption. We relax the functional form restrictions of this approach, avoid the full support assumption, and remove the exclusion restriction altogether.

Although we focus on demand for differentiated products, our results apply to other choice settings as well. One example is a discrete choice model of voting in a two-party election (e.g., Gordon and Hartmann (2013)) applied to data on that matches individual votes to voter demographic information, along with data on candidate characteristics and market-level (e.g., metro area level) variation in campaign advertising. There are no prices, but advertising now plays the role of the endogenous choice characteristic.[7] The "market-level" unobservables capture the effects of unmeasured candidate characteristics and local political preferences. Observed demographic characteristics, say education and income, create variation in voter preferences for the two candidates within a given market. We could think of the discrete choices as D, R and Not Voting. Our common choice probability requirement then requires the existence of some fixed vote share (choice probability) vector—say 0.4 for D and 0.4 for R—such that in each metro area there is some level of education and income that generates that conditional vote share. The level of education required to match the given vote share might be higher (and income lower) in a conservative as opposed to a liberal area. Note that there is no exclusion restriction here: both voter demographics will effect the vote share of both candidates.

Throughout the paper, we maintain an exclusive focus on identification. Nonparametric identification results do not eliminate concerns about the impact of parametric assumptions relied on in practice. But they address the important question of whether such assumptions can be viewed properly as finite-sample approximations rather than essential maintained hypotheses. Formal identification results can also clarify which maintained assumptions may be most difficult to relax, can reveal the essential sources of exogenous variation in the data, can offer assurance that robustness analysis is possible, and can pave the way to development of alternative (parametric or nonparametric) estimation approaches.

In what follows section 2 sets up the model we consider. Section 3 connects our nonparametric model to parametric examples from the existing empirical literature. Section 4 then establishes identification in two steps, reflecting the panel structure of the micro data setting. We first demonstrate identification of the index function using within-market variation in consumer attributes and consumer choice probabilities. Intuitively,

[6]See the review by Lewbel (2014) and references therein. Our earlier work, Berry and Haile (2010), featured an example of this sort.

[7]See, e.g., Gerber (1998) and Gordon and Hartmann (2013). Possible instruments include candidate wealth, market-specific measures of advertising cost, and combinations of statewide characteristic and features of the electoral college system that alter the returns to advertising in different metros.

up to normalizations, the variation in within-market behavior reveals the effect of consumer attributes on choices. It also reveals the vector of consumer characteristics for each market that generates the common choice probability. Plugging this vector into the index function, we obtain for each product and market an index that is a nonparametric function of observable product characteristics (including prices) and a single additively separable demand shock. These nonparametric functions are then identified following standard results from nonparametric IV regression, where the only endogenous variables are prices. Identification of these functions reveals the values of the demand shocks, and identification of demand follows directly. We discuss implications and extensions in section 5.

2 Multinomial Choice Model

2.1 Setup

We consider multinomial choice among J goods (or "products") and an outside option ("good 0") by consumers i in "markets" t. A market is defined formally by:[8]

- a vector X_t of observable characteristics of the market and products;
- a price vector $P_t = (P_{1t}, \ldots, P_{Jt})$;
- a vector $\Xi_t = (\Xi_{1t}, \ldots, \Xi_{Jt})$ of unobservables at the product$\times$market level;[9] and
- a distribution $F_Z(\cdot; t)$ of consumer-specific observables $Z_{it} = (Z_{i1t}, \ldots, Z_{iJt})$, with support $\mathcal{Z}(X_t)$.

Although X_t will typically include observable product characteristics, it may also include any number factors defining the market, including consumer observables. For example, the population of consumers may be partitioned into "markets" based on a combination of geography, time, and a vector of consumer demographics included in X_t.

In contrast, Z_{it} is distinguished by the indexing of its elements by both i and j. Our results will require that Z_{it} have dimension at least J, and that changes in Z_{it} alter the relative attractiveness of different goods. Because all other consumer-level observables may be absorbed by X_t, we henceforth assume $Z_{it} \in \mathbb{R}^J$. Although this permits the case in which each component Z_{ijt} is a consumer-specific factor assumed to alter only the attractiveness of good j, we will not require any such exclusion assumption. We also will not require independence (or even conditional independence) of Z_{it} and Ξ_t.

[8] More generally, the definition of a market could also include the number of goods available, J_t. We condition on a fixed number of products J without loss.

[9] For clarity we write random variables in uppercase and their realizations lowercase. Note that Ξ is the uppercase form of the standard notation ξ for product$\times$market unobservables.

The choice environment of consumer i is then represented by

$$C_{it} = (Z_{it}, X_t, P_t, \Xi_t)$$

Let $\mathcal{C}$ denote the support of C_{it}. The most basic primitive characterizing consumer behavior in such a setting is a distribution of decision rules for each $c_{it} \in \mathcal{C}$.[10] As usual, heterogeneity in decision rules (i.e., nondegeneracy of the distribution) within a given choice environment may reflect latent preference heterogeneity, stochastic elements of individual preferences, or stochastic elements of choice (e.g., optimization error).

The choice made by consumer i can be represented by $Q_{it} = (Q_{i1t}, \ldots, Q_{iJt})$, where Q_{ijt} denotes the quantity (here, 0 or 1) of good j purchased. A distribution of decision rules is characterized by the conditional cumulative distribution functions $F_Q(q|C_{it}) = E[1\{Q_{it} \leq q\}|C_{it}]$. In the case of discrete choice, this joint distribution of can be represented without loss by the multinomial choice probability function

$$\sigma(C_{it}) = (\sigma_1(C_{it}), \ldots, \sigma_J(C_{it})) = E[Q_{it}|C_{it}].$$

Given the total measure of consumers in each choice environment, the choice probability function σ fully characterizes consumer demand.

Thus far we have implicitly made two significant assumptions: that the unobservables Ξ_t have dimension J, and that, conditional on X_t, the support of Z_{it} does not depend on the realization of Ξ_t. Our results will also rely on the following key structure.

Assumption 1 (Index). *$\sigma(C_{it}) = \sigma(\gamma(Z_{it}, X_t, \Xi_t), X_t, P_t)$, with $\gamma(Z_{it}, X_t, \Xi_t) \in \mathbb{R}^J$.*

Assumption 2 (Invertible Demand). *$\sigma(\cdot, X_t, P_t)$ is injective on $supp\,\gamma(Z_{it}, X_t, \Xi_t)|(X_t, P_t)$ for all X_t, P_t.*

Assumption 3 (Injective Index). *$\gamma(\cdot, X_t, \Xi_t)$ is injective on $\mathcal{Z}(X_t)$ for all X_t, Ξ_t.*

Assumption 1 is a weak separability condition requiring that, given (X_t, P_t), Z_{it} and Ξ_t affect choices only through a vector of indices $(\gamma_1(Z_{it}, X_t, \Xi_t), \ldots, \gamma_J(Z_{it}, X_t, \Xi_t))$. This structure is satisfied by standard specifications used in practice. Assumption 2 further requires that the choice probability function be "invertible" with respect to the index vector—that, holding (X_t, P_t) fixed, distinct index vectors map to distinct choice probabilities. This is not without loss, and in general injectivity requires that σ map to interior values, i.e., that $\sigma_j(C_{it}) > 0$ for all j. A sufficient condition is the "connected substitutes" property introduced by Berry, Gandhi, and Haile (2013), who point out that this property is natural in a discrete choice setting in which each $\gamma_j(Z_{it}, X_t, \Xi_t)$ can be interpreted as quality index for good j. In that case, the connected substitutes condition requires that there be no strict subset of goods that substitute only among themselves in

[10] As is well known, under additional conditions a distribution of decision rules can be represented as the result of utility maximization. See, e.g., Mas-Colell, Whinston, and Green (1995), Block and Marschak (1960), Falmagne (1978), and McFadden (2005). We do not require such conditions and will not consider a utility-based representation of choice behavior.

response to variation in the index vector $\gamma(Z_{it}, X_t, \Xi_t)$. Assumption 3 requires injectivity of the index function γ with respect to the vector Z_{it}. This generalizes common utility-based specifications in which each Z_{ijt} is assumed to affect only the conditional indirect utility of good j and to do so monotonically. For example, if each index were a linear function of the J components of Z_{it}, Assumption 3 would require the matrix of coefficients to be full rank.

We henceforth condition on an arbitrary value of X_t and suppress it in the notation. The remaining assumptions and results should be interpreted to hold conditional on X_t.[11] Although not essential, we will focus on the case in which the indices $\gamma_j(Z_{it}, \Xi_t)$ are additively separable in Ξ_t (Assumption 4).[12] In Assumption 5 we assume sufficient smoothness (as well as openness of $\mathcal{Z}$) to permit our applications of calculus below.[13] Part (iv) of Assumption 5 also strengthens the injectivity requirements of Assumptions 2 and 3 by requiring that the Jacobian matrices $\partial g(z)/\partial z$ and $\partial\sigma(\gamma, p)/\partial\gamma$ be nonsingular almost surely.[14].

Assumption 4 (Separable Index). *For all j, $\gamma_j(Z_{it}, \Xi_t) = g_j(Z_{it}) + \Xi_{jt}$.*

Assumption 5 (Regularity). *(i) $\mathcal{Z}$ is open and connected; (ii) $g(z)$ is continuously differentiable on $\mathcal{Z}$; (iii) $\sigma(\gamma, \xi)$ is continuously differentiable with respect to γ for all $(\gamma, \xi) \in supp\ (\gamma(Z_{it}, \Xi_t), \Xi_t)$; (iv) $\partial g(z)/\partial z$ and $\partial\sigma(\gamma, p)/\partial\gamma$ ar nonsingular almost surely on $\mathcal{Z}$ and $supp\,(\gamma(Z_{it}, \Xi_t), P_t)$, respectively.*

2.2 Normalization

The model requires two types of normalizations for the identification question to be properly posed. The first requirement reflects the fact that the unobservables have no natural location; therefore, adding a constant vector to Ξ_t and subtracting the same

[11] We have not assumed that X_t and Ξ_t are independent. Indeed, we could have indexed Ξ_t by X_t to emphasize that we can allow arbitrary dependence. Below we demonstrate identification of the index function g and choice probability function σ conditional on each value of X_t. Identification of the effects of X_t holding unobservables fixed could also be obtained under an additional assumption, such as the typical $X_t \perp\!\!\!\perp \Xi_t$. Identification of many functionals (e.g., price elasticities) and counterfactuals of interest will not require such an assumption.

[12] We can instead permit a fully nonlinear index under a strict monotonicity condition and reliance on the completeness condition of Chernozhukov and Hansen (2005) instead of the standard completeness condition.

[13] Given parts (i) and (ii) of Assumption 5, the injectivity of g required by Assumption 3 implies (by invariance of domain) that the image $g(\Omega)$ of any open set Ω (e.g., $\Omega = \mathcal{Z}$) is open.

[14] Even without part (iv) there could be no open set Ω on which $\partial g(z)/\partial z$ was singular, as (see footnote 13) $g(\Omega)$ or would then be an open subset of $\mathbb{R}^J$, contradicting Sard's theorem. A similar observation applies to $\partial\sigma(\gamma, p)/\partial\gamma$. Thus, part (iv) rules out injective continuously differentiable functions g or σ with (uncountably many) critical points on a set Ω containing no open subset of $\mathbb{R}^J$ but having positive measure nonetheless—e.g., for $J = 1$, a fat Cantor set.

vector from g yields the same distribution of consumer choice at every (z_{it}, ξ_t, p_t). Thus, we take an arbitrary point $z^0 \in \mathcal{Z}$ and set

$$g\left(z^0\right) = 0, \tag{1}$$

where the right-hand side is the zero J-vector.

The second normalization requirement arises from the fact that any injective transformation of the index vector $\gamma\left(Z_{it}, \Xi_t\right)$ can be reversed by a modification of the function σ. For example, let A be any J-vector of constants, let B be any nonsingular $J \times J$ matrix, and define

$$\tilde{\gamma}\left(Z_{it}, \Xi_t\right) = A + B\gamma\left(Z_{it}, \Xi_t\right). \tag{2}$$

If we then define $\tilde{\sigma}$ by

$$\tilde{\sigma}\left(\tilde{\gamma}, P_t\right) = \sigma\left(B^{-1}\left(\tilde{\gamma} - A\right), P_t\right), \tag{3}$$

then (σ, γ) and $(\tilde{\sigma}, \tilde{\gamma})$ are two representations of the same distribution of decision rules, the latter satisfying our assumptions whenever the former does.[15] We must choose a single representation before exploring whether the observables allow identification. We do this by taking the representation in which the index $\gamma\left(z^0, \Xi_t\right) = g\left(z^0\right) + \Xi_t$ has expectation zero, i.e.,

$$E\left[\Xi_t\right] = 0,$$

and is such that

$$\left[\frac{\partial g(z^0)}{\partial z}\right] = I. \tag{4}$$

For example, this representation is obtained from (2) and (3) by letting $B = \left[\partial g(z^0)/\partial z\right]^{-1}$, $A = -B\,E\left[\Xi_t\right]$, and then dropping the tildes from the notation for the transformed model.

3 Parametric Examples from the Literature

The empirical literature in economics includes many examples of parametric specifications that are special cases of our model. Discrete choice models are frequently formulated using a random utility specification of the form

$$u_{ijt} = x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{5}$$

where u_{ijt} represents individual i's conditional indirect utility from choice j in market t. As in our model, x_{jt}, p_{jt} and ξ_{jt} are, respectively, observed product characteristics, prices, and latent demand shocks such as unobserved product characteristics.

---

[15] This illustrates an inherent ambiguity in the interpretation of how variation in a given component of the vector Z_{it} alters preferences. For example, in terms of behavior, there is no difference between a change in Z_{ijt} that makes good j more desirable and a change in Z_{ijt} that makes all other goods (including the outside good) less desirable. In practice, this ambiguity is often resolved with *a priori* exclusion assumptions—e.g., an assumption that Z_{ijt} affects only the utility obtained from good j. Some of the examples discussed below utilize this structure. Such assumptions could only aid identification, and our choice of normalization remains valid in that case.

The additive ϵ_{ijt} is typically specified as a draw from a type-1 extreme value distribution or a normal distribution, yielding a mixed multinomial logit or probit model. Components k of the random coefficient vector β_{it} are often specified as

$$\beta_{it}^{(k)} = \lambda_0^{(k)} + \sum_{\ell=1}^{L} \lambda_\ell^{(k)} z_{i\ell t} + \lambda_\nu^{(k)} \nu_{it}^{(k)},$$

where $z_{i\ell t}$ represents one of L observable characteristics of individual i, and each $\nu_{it}^{(k)}$ is a random variable with a pre-specified distribution. Often, the coefficient on price is specified as varying with some additional observed characteristics y_{it} such as income. A typical specification of α_{it} takes the form

$$ln(\alpha_{it}) = \lambda_0^{(0)} + \lambda_1^{(0)} y_{it} + \lambda_\nu^{(0)} \nu_{i0}.$$

To connect this to our model, we simply condition on y_{it}, treating it fully flexibly.

We can then rewrite (5) as

$$u_{ijt} = g_j\left(z_{it}\right) + \xi_{jt} + x_{jt}\lambda_0 - \alpha_{it} p_{jt} + \mu_{ijt}, \tag{6}$$

where $\mu_{ijt} = \sum_k x_{jt}^{(k)} \lambda_\nu^{(k)} \nu_{it}^{(k)} + \epsilon_{it}$ and

$$g_j\left(z_{it}\right) = \sum_k x_{jt}^{(k)} \sum_{\ell=1}^{L} \lambda_\ell^{(k)} z_{i\ell t}.$$

Observe that all effects of z_{it} and ξ_t operate though indices

$$\gamma_j\left(z_{it}, \xi_t\right) = g_j\left(z_{it}\right) + \xi_{jt} \qquad j = 1, \ldots, J,$$

satisfying our Assumptions 1 and 4. It is easy to show that the resulting choice probabilities satisfy Berry, Gandhi and Haile's (2013) "connected substitutes" condition with respect to the vector of indices $\gamma\left(z_{it}, \xi_t\right) = \left(\gamma_1\left(z_{it}, \xi_t\right), \ldots, \gamma_J\left(z_{it}, \xi_t\right)\right)$; therefore, the injectivity of demand required by Assumption 2 holds. We assume that (after conditioning on y_{it}) there are at least J non-trivial elements of the vector z_{it}.[16] Injectivity of $g(z_{it}) = (g_1(z_{it}), \ldots, g_J(z_{it}))$ might then be assumed as a primitive condition of the model or else derived from other conditions, as in the example we discuss below.

Of course, our model does not rely on the linear structure of this example, nor on any parametric distributional assumptions. But this example connects our model to a large number of applications and shows one way that the observables z_{it} can interact with product characteristics to generate preference heterogeneity across individuals facing the same choice set (i.e., where all x_{jt} and ξ_{jt} are fixed). Note also that the example lacks features that are sometimes relied on in results showing identification of discrete choice

[16] This allows for individual characteristics that shift both the index and the distribution of the price coefficient. However, there must be at least J elements of z_{it} that vary after conditioning on y_{it}.

models: in addition to the absence of individual characteristics that exclusively affect the utility from one choice j, this model does not exhibit independence between the "error term" μ_{ijt} and any of the variables that appear on the right hand side of (6).[17]

To see another way that our index structure arises in practice, consider Ho's (2009) model of demand for health insurance. Each consumer i in market t considers J insurance plans as well as the outside option of remaining uninsured. Each consumer has a vector of observable characteristics d_{it}.[18] Let n_{jt} denote the set of hospitals in plan j's network, along with their characteristics (e.g. location and the availability of speciality services like cardiac care). Each insurance plan is associated with its network n_{jt}, an annual premium p_{jt}, additional observed plan characteristics x_{jt} (e.g., the size of its physician network), and an unobservable ξ_{jt}.

A consumer's insurance plan demand depends on her particular likelihood of having of each type of hospital need (diagnosis) as well as how her preferences over hospital characteristics will vary with the type of need. This gives each consumer i an expected utility $EU\left(n_{jt}, d_{it}\right)$ for the option to use plan j's hospital network. Ho derives this expected utility from auxiliary data on hospital choice.[19] which yields, from the perspective of identification, a known functional form for the consumer/choice measures

$$z_{ijt} \equiv EU\left(n_{jt}, d_{it}\right).$$

Similar to (5), consumer i's conditional indirect utility for plan j then takes the form[20]

$$u_{ijt} = \lambda z_{ijt} + x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}. \tag{7}$$

Ho assumes each ϵ_{ijt} is an independent draw from a type-1 extreme value distribution, yielding a multinomial logit model.

Observe that in this example Ho combines data on the characteristics of consumers and choices with a model to derive a scalar z_{ijt} that exclusively affects only the utility of choice j. In this case, the injectivity of the index vector $\gamma\left(z_{it}, \xi_t\right)$, as required by our Assumption 3, holds under an assumption that $\lambda \neq 0$. Given this condition, Assumption 5 is also satisfied as long as the support of Z_{it} is an open and connected subset of $\mathbb{R}^J$. Satisfaction of our remaining assumptions follows as in the previous example.

---

[17] In the traditional parametric model, the ν_{it} are assumed to be independent across i and t, but the x_{jt} variables enter the composite error μ_{ijt}. The x_{jt} are also typically correlated with p_{jt} and, in our framework, are allowed to be correlated with changes in the distribution of z_{it} across markets.

[18] Ho's data include measures of individual age, gender, income, home location, employment status, and industry of employment.

[19] See also Ho (2006). Here d_{it} affects both diagnosis probabilities and preferences over hospitals condition on diagnosis. Ho and Lee (2016) extend the model to treat insurance choice at the household level, with households anticipating diagnosis probabilities for each household member.

[20] Ho (2009) allows the premium coefficient to differ by income level. As discussed above, we can condition on income to treat it fully flexibly. Ho (2009) uses excluded plan-level cost shifters as instruments for premiums.

4 Identification

We examine the identifiability of g and σ from observation of the choice decisions of the population of consumers i in a population of markets t. In addition to the suppressed X_t, the observables consist of P_t, Z_{it}, Q_{it}, and a vector of instruments W_t discussed further below. Given any $(\xi, p) \in \text{supp}\,(\Xi_t, P_t)$, let

$$\mathcal{S}\,(\xi, p) = \sigma\,(g\,(\mathcal{Z}) + \xi, p)$$

denote the support of choice probabilities in markets t for which $\Xi_t = \xi, P_t = p$. Because $\mathcal{Z}$ is open, continuity and injectivity of σ with respect to the index and of the index with respect to Z_{it} imply (by invariance of domain) that $\mathcal{S}\,(\xi, p)$ is open.

A key observation for what follows is that, by Assumptions 2 and 3, for each $s \in \mathcal{S}\,(\xi, p)$ there must be a unique $z^* \in \mathcal{Z}$ such that $\sigma\,(g\,(z^*) + \xi, p) = s$. Thus, for $(\xi, p) \in \text{supp}\,(\Xi_t, P_t)$ and $s \in \mathcal{S}\,(\xi, p)$, we define the function $z^*\,(s; \xi, p)$ implicitly by

$$\sigma\,(g\,(z^*\,(s; \xi, p)) + \xi, p) = s. \tag{8}$$

We then have

$$g\,(z^*\,(s; \xi, p)) + \xi = \sigma^{-1}\,(s; p) \quad \forall\,(\xi, p) \in \text{supp}\,(\Xi_t, P_t)\,, s \in \mathcal{S}\,(\xi, p)\,. \tag{9}$$

Note that in each market t, the set $\mathcal{S}(\xi_t, p_t)$ and the values of $z^*\,(s; \xi_t, p_t)$ for all $s \in \mathcal{S}\,(\xi_t, p_t)$ are observed, even though the value of the argument ξ_t in each market t is unknown.

4.1 Identification of the Index Function

Let $||\cdot||$ denote the Euclidean norm and let $\mathcal{B}\,(b, \Delta)$ denote an open ball in $\mathbb{R}^J$ of radius $\Delta > 0$, centered at b. We demonstrate identification of the index function $g = (g_1, \ldots, g_J)$ under the following condition.

Assumption 6 (Nondegeneracy)**.** *For some $\Delta > 0$ and $p \in \text{supp}\,P_t$, $\text{supp}\,\Xi_t|\{P_t = p\}$ contains an open ball of radius Δ.*

Assumption 6 requires continuously distributed Ξ_t but is otherwise mild. It can be derived as an implication of standard models of supply in which cost shifters (which need not be observed) allow the same equilibrium price vector p to arise under different demand conditions (different ξ_t). The key implication, exploited in the following result, is that there exist $p \in \text{supp}\,P_t$ and $\Delta > 0$ such that for any $d \in \mathbb{R}^J$ satisfying $||d|| < \Delta$, $\text{supp}\,\Xi_t|\{P_t = p\}$ contains points ξ and ξ' satisfying $\xi - \xi' = d$.

Lemma 1. *Let Assumptions 1–6 hold and take (p, Δ) as defined by Assumption 6. Then for every z and z' in $\mathcal{Z}$ such that $||g\,(z') - g\,(z)|| < \Delta$ there exist ξ and ξ' in $\text{supp}\,\Xi_t|\{P_t = p\}$ such that, for some s, $z = z^*\,(s; \xi, p)$ and $z' = z^*\,(s; \xi', p)$.*

Proof. Take any z and z' in $\mathcal{Z}$ such that $||g(z') - g(z)|| < \Delta$. By Assumption 6 and the choice of (p, Δ), there exist ξ and ξ' in $\operatorname{supp} \Xi_t | \{P_t = p\}$ such that $\xi - \xi' = g(z') - g(z)$, i.e., $\gamma(z', \xi') = \gamma(z, \xi)$. Taking $s = \sigma(\gamma(z', \xi'), p) = \sigma(\gamma(z, \xi), p)$, the result follows from the definition (8). □

Lemma 2. *Let Assumptions 1–6 hold. Then there exists $\Delta > 0$ such that for almost all $z, z' \in \mathcal{Z}$ satisfying $||g(z') - g(z)|| < \Delta$ the matrix $\left[\frac{\partial g(z)}{\partial z}\right]^{-1} \left[\frac{\partial g(z')}{\partial z}\right]$ is identified.*

Proof. Take p and Δ as in Assumption 6. Consider markets t and t' in which $P_t = P_{t'} = p$ but, for some choice probability vector s,

$$z = z^*(s; \xi_t, p) \neq z' = z^*(s; \xi_{t'}, p), \tag{10}$$

revealing that $\xi_t \neq \xi_{t'}$. Lemma 1 ensures that such t, t', and s exist for all $z, z' \in \mathcal{Z}$ satisfying $||g(z') - g(z)|| < \Delta$.[21] And although ξ_t and $\xi_{t'}$ are latent, the identities of markets t and t' satisfying (10) is observed. Differentiating (9) with respect to the vector s within these two markets, we obtain

$$\frac{\partial g(z)}{\partial z} \frac{\partial z^*(s; \xi_t, p)}{\partial s} = \frac{\partial \sigma^{-1}(s; p)}{\partial s} \tag{11}$$

and

$$\frac{\partial g(z')}{\partial z} \frac{\partial z^*(s; \xi_{t'}, p)}{\partial s} = \frac{\partial \sigma^{-1}(s; p)}{\partial s}.$$

Thus, recalling Assumption 5, for almost all such z, z' we have

$$\left[\frac{\partial g(z')}{\partial z}\right]^{-1} \frac{\partial g(z)}{\partial z} = \frac{\partial z^*(s; \xi_{t'}, p)}{\partial s} \left[\frac{\partial z^*(s; \xi_t, p)}{\partial s}\right]^{-1}.$$

Because the matrices on the right-hand side are observed, the result follows. □

Theorem 1. *Under Assumptions 1–6, g is identified on $\mathcal{Z}$.*

Proof. Take $\Delta > 0$ as in Lemma 2. For each vector of integers $\tau \in \mathbb{Z}^J$, define the set

$$\mathcal{B}_\tau = g(\mathcal{Z}) \cap \mathcal{B}\left(g\left(z^0\right) + \Delta\tau, \Delta\right),$$

and let $\mathcal{I}_\tau$ denote the pre-image of $\mathcal{B}_\tau$ under g. By construction, all z and z' in any given set $\mathcal{I}_\tau$ satisfy $||g(z') - g(z)|| < \Delta$. So by Lemma 2, $[\partial g(z)/\partial z]^{-1} [\partial g(z')/\partial z]$ is known for almost all z and z' in any set $\mathcal{I}_\tau$. Because $\cup_{\tau \in \mathbb{Z}^J} \mathcal{B}_\tau$ forms an open cover of $g(\mathcal{Z})$, $\cup_{\tau \in \mathbb{Z}^J} \mathcal{I}_\tau$ forms an open cover of $\mathcal{Z}$. Thus, given any $z \in \mathcal{Z}$ there exists a simple chain of

[21] Absent Assumption 6 and Lemma 1, the argument here still demonstrates identification of $[\partial g(z)/\partial z]^{-1} [\partial g(z')/\partial z]$ for every pair (z, z') such that for some p and s, we have $z = z^*(s; \xi_t, p) \neq z' = z^*(s; \xi_{t'}, p)$. Existence of (p, s) satisfying this condition for a given pair (z, z') is observable.

open sets $\mathcal{I}_\tau$ in $\mathcal{Z}$ linking the point z^0 to z.[22] Thus, $\left[\partial g(z^0)/\partial z\right]^{-1}\left[\partial g(z)/\partial z\right]$ is known for almost all $z \in \mathcal{Z}$. With the normalization (4) (and the continuity of $\partial g(z)/\partial z$), the result then follows from the fundamental theorem of calculus for line integrals and the boundary condition (1). □

Before moving to identification of the choice probability function, we pause to point out that our constructive identification of $g(\cdot)$ used only a single price vector p—that required by Assumption 6. In typical models of supply this condition would hold for almost all price vectors in the support of P_t. In addition to providing falsifiable restrictions, this indicates a form of redundancy that would typically be exploited by estimators used in practice. Similarly, our proof of Theorem 1 used, for each $z \in \mathcal{Z}$, only one of infinitely many paths between z^0 and z; integrating along any such path must yield the same function $g(\cdot)$.

4.2 Identification of the Choice Probability Function

We demonstrate identification of the choice probability function σ under the following additional conditions.

Assumption 7 (Common Choice Probability)**.** *There exists a choice probability vector s^* such that $s^* \in \mathcal{S}(\xi, p)$ for all $(\xi, p) \in$ supp (Ξ_t, P_t).*

Assumption 8 (Instruments for Prices)**.** *(i) For all $j = 1, \ldots, J$, $E\left[\Xi_{jt}|W_t\right] = 0$ almost surely; (ii) In the class of functions $\Psi(P_t)$ with finite expectation, $E\left[\Psi(P_t)|W_t\right] = 0$ almost surely implies $\Psi(P_t) = 0$ almost surely.*

Assumption 7 is a requirement that there exist some choice probability vector s^* that is common to all markets—that $\cap_{(\xi,p)\in \text{supp}\,(\Xi_t,P_t)}\mathcal{S}(\xi, p)$ be nonempty. The nondegeneracy of each set $\mathcal{S}(\xi_t, p_t)$ reflects variation in Z_{it} across its support. Assumption 7 requires sufficient variation in Z_{it} that for some s^* we have $s^* \in \mathcal{S}(\xi_t, p_t)$ for all realizations of (ξ_t, p_t). The strength of this assumption depends on the joint support of (Ξ_t, P_t) and the relative impacts of (Z_{it}, Ξ_t, P_t) on choice behavior. Observe that P_{jt} and Ξ_{jt} typically will have opposing impacts and will be positively correlated under equilibrium pricing behavior; thus, large support for $g(Z_{it})$ may not be required even if Ξ_t has large support. Indeed, we can contrast our assumption of a single common choice probability vector with a requirement of a special regressor with large support: the latter would imply that *every* interior choice probability vector s is a common choice probability.[23] Note also that, because choice probabilities conditional on Z_{it} are observable in all markets (i.e., for all realizations of (Ξ_t, P_t)), Assumption 7 is verifiable—i.e., its satisfaction or failure is identified.[24]

[22] See, e.g., van Mill (2002, Lemma 1.5.21).

[23] Identification arguments exploiting special regressors also commonly rely on linearity, exclusion, and independence conditions that we have not required.

[24] See Berry and Haile (2018) for a formal definition of verifiability.

Assumption 8 requires instruments for prices satisfying standard nonparametric IV conditions. Part (i) is the exclusion restriction, ensuring that variation in W_t not alter the mean of the unobservables Ξ_t. Part (ii) is a standard completeness condition—the nonparametric analog of the classic rank condition for linear regression. For example, Newey and Powell (2003) have shown that under mean independence, completeness is necessary and sufficient for identification in separable nonparametric regression. The following result demonstrates that, given knowledge of the index function g and existence of a common choice probability vector s^*, the same instrumental variables conditions suffice here.

Theorem 2. *Under Assumptions 1–8, σ is identified.*

Proof. Taking $s = s^*$ in equation (9) we have $g\left(z^*\left(s^*; \xi_t, p_t\right)\right) = \sigma^{-1}\left(s^*; p_t\right) - \xi_t$ for all tl; i.e., for all t and each $j = 1, \ldots, J$,

$$g_j\left(z^*\left(s^*; \xi_t, p_t\right)\right) = \sigma_j^{-1}\left(s^*; p_t\right) - \xi_{jt}. \tag{12}$$

By Theorem 1 the left side of (12) is known (recall that the values of each $z^*\left(s^*; \xi_t, p_t\right)$ are observable). Thus, for each j this equation takes the form of a separable nonparametric regression model. Identification of each function $\sigma_j^{-1}\left(s^*; \cdot\right)$ follows immediately from the identification result of Newey and Powell (2003). This implies identification of each ξ_{jt} as well. With Theorem 1, this implies that the value of each $\gamma\left(z_{it}, \xi_t\right)$ is identified for all t and $z_{it} \in \mathcal{Z}$. Identification of σ is then immediate from the observability of the choice vectors Q_{it}, since

$$\sigma\left(\gamma\left(z_{it}, \xi_t\right), p_t\right) = E\left[Q_{it} | Z_{it} = z_{it}, \Xi_t = \xi_t, P_t = p_t\right].$$

□

5 Discussion

We have shown that availability of micro data not only permits demand specifications that condition on consumer-level observables, but also can substantially reduce the reliance on instrumental variables to address the key challenge to identification of demand: the presence of unobserved product characteristics or other latent demand shocks that affect the prices and quantities of all goods in the demand system. This softening of instrumental variables requirements is achieved because consumer-level observables create within-market variation in choice problems. Such variation is similar to that which can be created by instruments for quantities; however, the exogeneity of the micro-data variation arises not from an exclusion restriction in the cross-section of markets but from the fact that within a single market the market-level demand shocks simply do not vary. Thus, our insights also have some connection to those underlying "within estimation" of slope parameters in panel data models with fixed effects.

Our results lead to several natural questions, which we discuss in the remainder of this concluding section.

5.1 What Are Appropriate Instruments?

Candidate instruments for prices include those typically relied upon in the case of market-level data (see, e.g., Berry and Haile (2016)). Classic instruments for prices are cost shifters excluded from the demand system. When cost shifters are not observed, proxies for these shifters may be available and excludable.[25] Exogenous shifters of market structure (e.g., firm ownership) that affect prices through equilibrium markups can also serve as instruments. Micro data can also result in availability of a related category of candidate instruments: market-level demographics such as the distribution of income and ethnicity that alter equilibrium markups. Berry and Haile (2014, 2016) refer to these as "Waldfogel" instruments, after Waldfogel (2003).[26] When micro-data are available, we can directly account for the impacts of individual-specific demographics, so it may be reasonable to assume that market-level demographics are excluded from the conditional demands we seek to identify. The requirement that these market-level measures be mean independent of the market-level demand shocks is a significant assumption, ruling out certain kinds of geographic sorting or peer effects. But in many applications such an assumption may be natural.

5.2 What About Stronger Functional Forms?

In practice, estimation is almost always influenced by functional form assumptions—e.g., the choice of parametric structure, kernel functions, or sieve basis. Such functional forms enable interpolation, extrapolation, and bridging of gaps between the exogenous variation present in the sample and that needed for nonparametric identification. A study of nonparametric identification can reveal whether such functional form assumptions play a larger role in one precise sense. One interpretation of our results is that only limited nonparametric structure is essential: for our nonparametric model, the main requirement for identification is adequate exogenous variation of dimension equal to the dimension of the endogenous variables.

But one can also ask how imposing additional structure on the demand model might allow relaxation of our identification requirements. Answers to this question may be of direct interest and can also suggest the sensitivity of identification to particular conditions. We may feel more comfortable when we know that relaxation of one condition for identification can be offset by strengthening another. A full exploration of these potential trade-offs describes an entire research agenda. But the examples below illustrate three directions one can go to enlarge the set of potential instruments, further reduce the number of required instruments, or reduce the required dimensionality of the micro data.

---

[25] An example, plausibly excludable in some applications, are so-called "Hausman instruments": prices of the same good in other markets (e.g., Hausman, Leonard, and Zona (1994), Hausman (1996), or Nevo (2000, 2001)).

[26] See also Gentzkow and Shapiro (2010) and Fan (2013).

5.2.1 Strengthening the Index Structure

Our model made no assumption on the way the characteristics X_t enter demand. For example, we have not assumed that there are certain elements X_{jt} of X_t that in some sense only affect good j. With such a restriction, however, another class of instruments—the exogenous characteristics of competing goods (i.e., BLP instruments) can become available.[27] One way to re-introduce the BLP instruments is to assume that for at least some component $X_t^{(1)}$ of X_t, choice probabilities can be written as (now conditioning out and suppressing only $x_t \backslash x_t^{(1)}$)

$$\sigma(z_{it}, x_t^{(1)}, \xi_t, p_t) = \sigma(\gamma_{it}, p_t),$$

with

$$\gamma_{ijt} = g_j(z_{jt}) + \xi_{jt} + h_j\left(x_{jt}^{(1)}\right).$$

Here we have strengthened our index structure by including $X_{jt}^{(1)}$ in the index and assuming that Z_{jt} is exclusive to the index for good j.[28]

In this case, the IV regression equation (12) becomes

$$g_j\left(z_j^*\left(s^*; \xi_t, p_t\right)\right) = \sigma_j^{-1}\left(s^*, p_t\right) - h_j\left(x_{jt}^{(1)}\right) + \xi_{jt}.$$

Identification of $\sigma_j^{-1}\left(s^*, \cdot\right)$ and $h_j(\cdot)$ then follows with instruments for P_t when $X_{jt}^{(1)}$ is mean independent of Ξ_{jt}. With the additional assumption that $X_{-jt}^{(1)}$ is mean independent of Ξ_{jt}, $X_{-jt}^{(1)}$ are available as instruments for P_{jt}.

5.2.2 A Special Regressor

Following Berry and Haile (2010), a different approach is to assume that the demand system of interest is generated by a random utility model with conditional indirect utilities of the form

$$u_{ijt} = g_j(z_{ijt}) + \xi_{jt} + \mu_{ijt},$$

with μ_{ijt} a scalar random term whose nonparametric distribution depends on x_{jt} and p_{jt}. In this case, our Theorem 1 demonstrates identification of each function $g_j(\cdot)$ up to a units normalization on the utility associated with product j. This turns $g_j(Z_{ijt})$ into a known special regressor. With a (typically very restrictive) full support assumption on $g_j(Z_j)$, a standard argument demonstrates identification of the marginal distribution of each $\mu_{ijt}|(x_{jt}, p_{jt})$. Berry and Haile (2010) show that one can use this marginal distribution to define a nonparametric IV regression equation for each choice j. In each such equation, the prices and characteristics of other choices are excluded. Thus, in this framework one needs only one instrument for price, and exogenous characteristics of competing goods (BLP instruments) are again available.

---

[27] In standard oligopoly models, each good's markup is dependent on the characteristics of related goods.

[28] Regarding the latter, recall the discussion in footnote 15.

5.2.3 A Semiparametric Model

Another way to add structure is to consider semiparametric models. As one example, consider semiparametric nested logit model where inverse demand, given z_t, is[29]

$$g_j(z_t) + \xi_{jt} = \ln(s_{jt}(z_t)/s_{0t}(z_t)) - \theta \ln(s_{j/g,t}(z_t)) + \alpha p_{jt}. \tag{13}$$

Here $s_{j/g,t}(z_t)$ denotes the within-group share and θ denotes the usual "nesting parameter."

Take any market t and any $z \in \mathcal{Z}$. Differentiating (13) with respect to one (possibly, the only) element of z_t—say z_{1t}—at the point z yields

$$\frac{\partial g_j(z)}{\partial z_1} = \frac{\partial \ln s_{jt}(z)}{\partial z_1} - \frac{\partial \ln s_{0t}(z)}{\partial z_1} - \theta \frac{\partial \ln s_{j/gt}(z)}{\partial z_1}. \tag{14}$$

In this equation, $\frac{\partial g_j(z)}{\partial z_1}$ and θ are the only unknowns. Moving to another market t', we can obtain a second equation of the same form in which the LHS is identical to that in (14). Equating the right-hand sides yields

$$\frac{\partial \ln s_{jt}(z)}{\partial z_1} - \frac{\partial \ln s_{0t}(z)}{\partial z_1} - \theta \frac{\partial \ln s_{j/gt}(z)}{\partial z_1} = \frac{\partial \ln s_{jt'}(z)}{\partial z_1} - \frac{\partial \ln s_{0t'}(z)}{\partial z_1} - \theta \frac{\partial \ln s_{j/gt'}(z)}{\partial z_1}.$$

Thus, we can solve for θ as long as

$$\frac{\partial \ln s_{j/gt}(z)}{\partial z_1} \neq \frac{\partial \ln s_{j/gt'}(z)}{\partial z_1},$$

a condition that will typically hold when $\xi_{t'} \neq \xi_t$ or $p_{t'} \neq p_t$, and which is directly observed. With θ known, we then identify (indeed, over-identify) all derivatives of $g_j(z)$ from (14). Identification of the remaining parameter α can then be obtained from (13) with a single excluded instrument—e.g., an excluded market-level cost shifter or markup shifter that affects all prices.

Although this example involves a model that is more flexible than nested logit models typically estimated in practice, it moves a considerable distance from our fully nonparametric model. But this example makes clear that additional structure can further reduce the dimension of the required exogenous variation. Indeed, here we can obtain identification with a single instrument (vs. the usual requirement of two instruments for the fully parametric nested logit (see Berry (1994))) and a scalar individual level observable z_{it}. Other semiparametric models may offer more intermediate points in set of feasible trade-offs between the flexibility of the model and the dimension of exogenous variation needed for identification.

[29] Recall that we have conditioned on x_t, permitting it to enter the model flexibly. For example, conditional indirect utilities might take the form

$$u_{ijt} = h\left(x_t, g_j\left(z_t, x_t\right) + \xi_{jt} - \alpha\left(x_t\right) p_{jt} + \mu_{ijt}\left(x_t\right)\right),$$

where h is strictly increasing in its second argument, $\alpha(x_t)$ is arbitrary, and $\mu_{ijt}(x_t)$ is a stochastic component taking the standard composite nested-logit form at each x_t.

5.3 What about Continuous Demand Systems?

Although we have focused on the case in which the consumer-level quantities Q_{ijt} are binary outcomes arising from a discrete choice model, there is nothing in our proofs requiring this. Applying our results to continuous demand is therefore just a matter of verifying the suitability of our assumptions.

As an example, consider a "mixed CES" model of continuous choice, similar to the model in Adao et al. (2017), with $J+1$ products. Each consumer i has utility over consumption vectors $q \in R_{+}^{J+1}$ given by

$$u\left(q; z_{it}, x_t, p_t, \xi_t\right) = \left(\sum_{j=0}^{J} \phi_{ijt} q_j^{\rho}\right)^{1/\rho},$$

where $\rho \in (0,1)$ is a parameter and each ϕ_{ijt} represents idiosyncratic preferences of consumer i for the product characteristics x. We set $\phi_{i0t} = 0$ and let

$$\phi_{ijt} = \exp\left[(1-\rho)\left(g_j\left(z_{it}\right) + \xi_{jt} + x_{jt}\beta_{it}\right)\right], \; j = 1, \ldots, J,$$

where β_{it} is a random vector with distribution F representing consumer-level preferences for product characteristics. With $p_{0t} = 1$ and consumer income of y_{it}, familiar CES algebra shows that Marshallian demands are

$$q_{ijt} = \frac{y_{it} \exp\left(g_j\left(z_{it}\right) + \xi_{jt} + x_{jt}\beta_{it} - \alpha \ln(p_{jt})\right)}{1 + \left[\sum_{k=1}^{J} \exp\left(g_k\left(z_{it}\right) + \xi_{kt} + x_{kt}\beta_{it} - \alpha\rho \ln(p_{kt})\right)\right]}.$$

This equation resembles a choice probability for a random coefficients logit model, although the quantities q_{it} here take on continuous values and do not sum to one. Conditioning on y_{it} (formally, using income level as one factor defining markets), it is easy to show that our Assumptions 1–4 are satisfied for the expected CES demand functions, which take the form

$$\sigma_t(g(z_{it}) + \xi_t, x_t, p_t) = E\left[Q_{it} | z_{it}, x_t, p_t, \xi_t\right],$$

where the jth component of $E[Q_{ijt}|z_{it}, x_t, p_t, \xi_t]$ is

$$\int \frac{y_{it} \exp\left(g_j\left(z_{it}\right) + \xi_{jt} + x_{jt}\beta_{it} - \alpha \ln(p_{jt})\right)}{1 + \left[\sum_{k=1}^{J} \exp\left(g_k\left(z_{it}\right) + \xi_{kt} + x_{kt}\beta_{it} - \alpha\rho \ln(p_{kt})\right)\right]} dF(\beta_{it}).$$

Berry, Gandhi, and Haile (2013) also describe a broad class of continuous choice models that can satisfy the key injectivity property of Assumption 2. These models can include mixed continuous/discrete settings, where individual consumers may purchase zero or any positive quantity of each good.

References

Ackerberg, D. (2003): "Advertising Learning and Customer Choice in Experience Good Markets: A Structural Empirical Examination," *International Economic Review*, 44, 1007–1040.

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," *American Economic Review*, 107, 633–89.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115, 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and Hedonic valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, 58, 1 – 14.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 23, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60, 889–917.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T., A. Gandhi, and P. A. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81, 2087–2111.

Berry, S. T. and P. A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Discussion Paper 1718, Cowles Foundation, Yale University.

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

——— (2016): "Identification in Differentiated Products Markets," *Annual Review of Economics*, 8, 27–52.

——— (2018): "Nonparametric Identification of Simultaneous Equations Models with a Residual Index Structure," *Econometrica*, 86, 289–315.

Block, H. and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. by I. Olkin, S. Ghurye, W. Hoeffding, W. G. Mado, and H. B. Mann, Stanford University Press, 97–132.

Blundell, R. and R. L. Matzkin (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, 5, 271–295.

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

Burda, M., M. Harding, and J. Hausman (2015): "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," *Journal of Applied Econometrics*, 30, 353–376.

Capps, C., D. Dranove, and M. Satterthwaite (2003): "Competition and Market Power in Option Demand Markets," *RAND Journal of Economics*, 34, 737–763.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Falmagne, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the U.S. Daily Newspaper Market," *American Economic Review*, 103, 1598–1628.

Gentzkow, M. and J. Shapiro (2010): "What Drives Media Slant? Evidence from U.S. Newspapers," *Econometrica*, 78, 35–71.

Gerber, A. (1998): "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables," *The American Political Science Review*, 92, 401–411.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951.

Gordon, B. and W. R. Hartmann (2013): "Advertising Effects in Presidential Elections," *Marketing Science*, 32, 19–35.

Hausman, J., G. Leonard, and J. Zona (1994): "Competitive Analysis with Differentiated Products," *Annales d'Economie et de Statistique*, 34, 159–180.

Hausman, J. A. (1996): "Valuation of New Goods under Perfect and Imperfect Competitioin," in *The Economics of New Goods*, ed. by T. F. Bresnahan and R. J. Gordon, Chicago: University of Chicago Press, chap. 5, 209–248.

Ho, K. (2006): "The Welfare Effects of Restricted Hospital Choice in the U.S. Medical Care Market," *Journal of Applied Econometrics*, 21, 1039–1079.

——— (2009): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 99, 393–430.

Ho, K. and R. Lee (2016): "Insurer Competition in Health Care Markets," *Econometrica*, forthcoming.

Hom, M. (2018): "School Choice, Segregation and Access to Quality Schools: Evidence from Arizona," Tech. rep., Yale University.

Kawai, K., Y. Toyama, and Y. Watanabe (2020): "Voter Turnout and Preference Aggregation," *American Economic Journals: Microeconomics*, forthcoming.

Lewbel, A. (2014): "An Overview of the Special Regressor Method," in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, ed. by J. S. Racine, L. Su, and A. Ullah, Oxford University Press, 38–62.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, Oxford University Press.

Matzkin, R. L. (2015): "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1–66.

McFadden, D. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26, 245–264.

McFadden, D., A. Talvitie, and Associates (1977): *Demand Model Estimation and Validation*, Berkeley CA: Institute of Transportation Studies.

Neilson, C. (2019): "Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students," Tech. rep., Princeton University.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9, 513–548.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–42.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation in Nonparametric Models," *Econometrica*, 71, 1565–1578.

van Mill, J. (2002): *The Infinite-Dimensional Topology of Function Spaces*, Elsevier.

Waldfogel, J. (2003): "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *RAND Journal of Economics*, 34, 557–568.