Nonparametric Identification of Differentiated Products Demand Using Micro Data

Steven Berry and Phil Haile
Yale

Spring 2020

Early versions of this work were presented in the working paper "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," which is superseded by the present paper.

Idea of This Paper

Nonparametric identification of demand when one has micro data linking the characteristics of individual consumers to their choices.

Micro data creates a kind of panel data structure of consumers-within-markets.

Our main insight: micro data can substantially reduce the reliance on instrumental variable for identification. Still need instruments for prices.

Very Broad Background

- Empirical IO has been criticized in the past for estimating models that are rich & complicated, but maybe not well-identified (Angrist and Pischke, 2010).
- There has been a lot of work on formal identification in IO.
- Nonparametric identification can help us understand basic sources of identification and also understand the role of parametric assumptions, which we are likely to continue to use.

Demand Estimation

- Unobserved product characteristics ("demand shocks") that vary across products and markets present a challenge to identification of demand.
- Each shock affects the quantity demanded, and price, of all related goods.
- In the case of "market level data," Berry & Haile (2014) obtain identification with instruments for all endogenous variables: the prices and quantities of all goods in the demand system.

Market Level Instruments

- With J products, $2J$ endogenous variables: J prices, p, and J quantities or "market shares," s.
- Berry and Haile (2014) then require $2J$ instruments: J "BLP instruments" (exogenous characteristics of other products) plus J cost shifters
- Functional form restrictions can reduce this number.

Index, Invert, IV

In the market-level context:

$$\underbrace{x_{jt}\beta + \xi_{jt}}_{\text{index}} = \underbrace{\sigma_j^{-1}}_{\text{inverse}} \underbrace{(s, p)}_{\text{IVs}}$$

(β normalized to one to choose units). E.g. for the logit:

$$\sigma^{-1}(s, p) = \ln(s_j) - \ln(s_0) + \alpha p_{jt}.$$

With micro data:

- use an index in consumer characteristics, z
- add the panel-data like possibility of within-market movement in the space of z, exploiting variation choice probabilities while holding ξ_{jt} fixed.

Micro Data

Micro data discrete choice is much older than market level, but did not originally address endogeneity. There are many identification papers on discrete choice, but much not on our context of consumers-within-markets and market-level endogeneity.

McFadden, et al, (1977): household level transport demand for auto v. bus shifted by wage-time-distance to work, auto ownership, etc. Distance is also used in modern studies of schools, hospitals, jobs, etc. In autos: family size $\times$ size, rural $\times$ pickup, education $\times$ import. Many others!

Voting example: income, education, age race, urban/rural shift voting preferences, but not just the preference for one candidate. Not a clear exclusive assignment of, say, education to the taste for one party or the other.

Example: Job Vacancies as Differentiated Products

Azar, Berry, Marinescu (2019)

Simplified version: i is applicant, j is job listing, t is commuting zone:

$$u_{ijt} = x_{jt}\beta + \alpha w_{jt} + \underbrace{z_{ijt}\gamma + \xi_{jt}}_{\text{index}} + \nu_{igt}(\lambda) + \epsilon_{ijt}$$

where z_{ijt} is distance to job, the errors are nested logit (on, say, occupation), x_{jt} are unobserved job characteristics and w_{jt} is the endogenous wage (need IVs)

The current paper can provide a nonparametric identification foundation for frameworks like this, using an index that generalizes

$$z_{ijt}\gamma + \xi_{jt}$$

and otherwise uses much weaker assumptions overall.

Our Results

- allow demand to depend on consumer observables
- are nonparametric
- do not require a "special regressor," or exclusive assignment of some z's to the utility of a particular produce,
- do not require a full support assumption
- do not restrict the way that product characteristics or prices shift demand,
- works for discrete choice demand as well as continuous demand or mixed discrete/continuous
- Allow for "Waldfogel" instruments (market level distribution of z that shifts prices) and do not require BLP instruments.

Our Results

however

- require an index in z and invertibility assumptions,
- in lieu of full support, require a "common choice probability" that is found in all markets (for different z's in different markets),
- still require J instruments for prices,
- require J elements of z (to get fully flexible demand)

The requirements of J price IVs and J elements of z can be relaxed under stronger assumptions.

Questions?

Next: Model, Normalizations, Example

Setup

We consider multinomial choice among J goods (or "products") and an outside option, $j = 0$, by consumers i in "markets" t. A market is defined formally by:

- 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 product-market unobservables (Ξ is capital ξ, sorry);
- a distribution $F_Z(\cdot; t)$ of consumer-specific observables $Z_{it} = (Z_{i1t}, \ldots, Z_{iJt})$, with support $\mathcal{Z}(X_t)$.

Uppercase denotes a random variable.

Notes

- We need at least J elements of Z_{it},
- additional elements could be included in the X_t, and so we assume $Z_{it} \in \mathbb{R}^J$, but
- we don't necessarily assign Z_{ijt} to the "utility of product j."
- We assume J elements of Ξ_t (one for each product).
- We don't assume independence between Z_{it} and Ξ_{it}
- We don't assume independence between X_t and Ξ_t, although this might be required to answer some counterfactuals.

Choice

The choice environment of consumer i is

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

The choice of consumer i is

$$Q_{it} = (Q_{i1t}, \ldots, Q_{iJt})$$

In the discrete choice setting, the joint distribution of decision rules is given by the choice probabilities:

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

Core Assumptions

Assumption (Index)

$$\sigma\left(C_{it}\right) = \sigma(\underbrace{\gamma(Z_{it}, X_t, \Xi_t)}_{index}, X_t, P_t)$$

with the index $\gamma\left(Z_{it}, X_t, \Xi_t\right) \in \mathbb{R}^J$.

Assumption (Invertible Demand)

$\sigma\left(\cdot, X_t, P_t\right)$ *is injective on* $\operatorname{supp} \gamma(Z_{it}, X_t, \Xi_t) | (X_t, P_t)$ *for all* X_t, P_t.

Assumption (Injective Index)

$\gamma\left(\cdot, X_t, \Xi_t\right)$ *is injective on* $\mathcal{Z}\left(X_t\right)$ *for all* X_t, Ξ_t.

Additional Assumptions

Index is linear in Ξ_{jt} but not in Z_{it}.

Assumption (Separable Index)

For all j,

$$\gamma_j\left(Z_{it}, \Xi_t\right) = g_j\left(Z_{it}\right) + \Xi_{jt}.$$

Assumption (Regularity)

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

Transformations of the Index

A linear transformation of our additively separable index can be reversed by a modification of the choice probability function σ. For example, for a J-vector A and a nonsingular $J \times J$ matrix B, define

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

and also define

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

Then (σ, γ) and $(\tilde{\sigma}, \tilde{\gamma})$ are two representations of the same decision rules, the latter satisfying our assumptions whenever the former does.

Discussion of the Linear Transformation

This illustrates an inherent ambiguity. 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 the nonparametric discrete choice identification literature, 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.

Similarly, a fully parametric discrete choice models will (often) not retain the same parametric structure under the transformation and so the issue will not arise.

Transformation Normalization

We set $E\left[\Xi_t\right] = 0$ and for some z^0 we set

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

In our transformation example, start from the original model and then let $B = \left[\partial g(z^0)/\partial z\right]^{-1}$ and $A = -B\,E\left[\Xi_t\right]$. We then drop the tildes from the notation.

As a local normalization in the index, we also set $g(z^0) = 0$.

Random Coefficients Discrete Choice Example

A classic model is:

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

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

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

In our model, we could simply condition on income y_{it}, treating it fully flexibly. We can rewrite the model to illustrate our index:

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

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

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

Verifying Assumptions for the RCL Model

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 the index assumptions. It is easy to show that the resulting discrete choice probabilities satisfy Berry, Gandhi and Haile's (2013) "connected substitutes" condition with respect to the vector of indices and therefore satisfy the assumed injectivity of demand.

We require at least J non-trivial elements of the vector z_{it} and the injectivity of $g(z)$ depends on the invertibility of the implied matrix of linear coefficients in the index.

Example: Ho (2009)

How to get a "rich" set of z's? Ho builds them from an auxiliary dataset.

Consumer i's demand for insurance depends on her preferences and the likelihood of having of each diagnosis. Ho uses data on hospital choice to derive the expected utility of each hospital network:

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

where n_{jt} are measures of the insurer's network and d_{it} are demographics. This then enters a classic logit:

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

Injectivity of the index requires only that $\lambda \neq 0$ and the other assumptions are also easy to verify.

Questions?

Next: Identification Results and Discussion (with questions and examples)

The Consumer with Choice Probability s

Our assumptions imply that for each $s \in \mathcal{S}(\xi, p)$ there must be a unique $z^* \in \mathcal{Z}$ such that

$$\sigma\left(g\left(z^*\right) + \xi, p\right) = s.$$

Here, z^* is the vector of consumer characteristics whose associated choice probability vector is s.

We can then write the inverted model as:

$$g\left(z^*\left(s; \xi, p\right)\right) + \xi = \sigma^{-1}\left(s; p\right)$$

Note that for each $s \in \mathcal{S}(\xi, p)$, $z^*(s; \xi, p)$ is observed even though ξ is not observed.

Identification of the Index Function

$$g\left(z^{*}\left(s;\xi_{t},p\right)\right)+\xi_{t}=\sigma^{-1}\left(s;p\right).$$

Differentiating and setting $z=z^{*}\left(s;\xi_{t},p\right)$,

$$\frac{\partial g\left(z\right)}{\partial z}\frac{\partial z^{*}\left(s;\xi_{t},p\right)}{\partial s}=\frac{\partial\sigma^{-1}\left(s;p\right)}{\partial s}.$$

Do the same in another market t' with the same p and the same $s\in\mathcal{S}(\xi_{t'},p)$. Setting the two RHS equal,

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

with $z'=z^{*}\left(s;\xi_{t'},p\right)$.

We normalized $\partial g(z^{0})/\partial z=I$ and so we can start with $z=z^{0}$ and then move through a series of markets with the same p and "overlapping" $\mathcal{S}(\xi,p)$ to identify $g(z)\ \forall z\in\mathcal{Z}$.

A Common Choice Probability

We assume that is some choice probability vector s^* that is common to all markets.

Assumption (Common Choice Probability)

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

This requires that the intersection of the supports $\mathcal{S}(\xi, p)$ be nonempty. The supports of s are driven by variation in Z_{it} across its support and so the assumption requires "sufficient" variation in Z_{it}.

The common choice probability assumption is verifiable in the data.

Contrast to Large Support

A large support assumption would imply that *every* interior choice probability vector s is a common choice probability.

The common choice probability assumption 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 and p_t on choice behavior.

An IV Equation to Identify Demand

Under the common choice probability assumption, we can write, in every market,

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

The LHS is known and we are using variation in z to keep the RHS choice probabilities fixed at s^*. This leaves only the variation in p and we can identify the function $\sigma^{-1}(s^*,p)$ given instruments, W_t, for prices. This identifies ξ_t, which in turn identifies demand.

The remaining necessary assumption is therefore a classic nonparametric "completeness" condition

Assumption (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* $\Psi\left(P_t\right)$ *with finite expectation,* $E\left[\Psi\left(P_t\right)|W_t\right]=0$ *almost surely implies* $\Psi\left(P_t\right)=0$ *almost surely.*

Discussion

As compared to the market-level results, we replace IVs for quantities with micro data z variables.

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 results have some intuitive connection to "within estimation" of slope parameters in panel data models with fixed effects.

Next: discussion of particular questions.

What Are Appropriate Instruments?

Depending on the application, IVs for price might include

- cost shifters excluded from the demand system, or proxies for these.
- exogenous shifters of market structure
- Waldfogel IVs: market-level demographics such as the distribution of income and ethnicity that alter equilibrium markups (recall that we condition on consumer-level demographics).

But not (without further restrictions) the BLP instruments, because we have not even specified which elements of X are associated with what product.

What About Stronger Functional Forms?

For our nonparametric model, we need adequate exogenous variation of dimension equal to the dimension of the endogenous variables.

In practice, functional form assumptions enable interpolation, extrapolation, and the bridging of gaps between the exogenous variation in the sample and variation needed for nonparametric identification.

We can also ask how imposing additional structure on the demand model might allow relaxation of our identification requirements. We consider three example directions.

Strengthening the Index Structure

functional form example 1

Say that there are some elements $X_t^{(1)}$ that we can assign to specific products. Condition only on $x_t \backslash x_t^{(1)}$ and assume

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

with

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

We assume that z_{ijt} is exclusive to $g_j(\cdot)$.

The IV regression equation becomes

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

The BLP instruments are now potentially available.

A Special Regressor

functional form example 2

Following one approach of Berry and Haile (2010), consider

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

with μ_{ijt} a random scalar whose nonparametric distribution depends on x_{jt} and p_{jt}. In this case, our results imply identification of $g_j(\cdot)$ up to units, turning $g_j(Z_{ijt})$ into a known special regressor. With a (restrictive) full support assumption on $g_j(Z_j)$, a standard argument then identifies

$$m_{jt} \equiv E(\mu_{ijt}|x_{jt}, p_{jt}) + \xi_{jt}.$$

This defines a nonparametric IV regression equation for each choice j in which the prices and characteristics of other choices are excluded. Thus, one needs only one instrument for price, and the BLP IVs are again available.

A Semiparametric Model

functional form example 3

As one example, consider a semiparametric nested logit model, conditional on X_t, where inverse demand is

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

Take any market t and any $z \in \mathcal{Z}$. Differentiating,

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

The only unknowns are $\partial g_j(z)/\partial z_1$ and θ. Moving to another market t', we get a second equation of the same form with an identical LHS. Equating the right-hand sides allows us to solve for θ. With θ known, we then identify (over-identify) all derivatives of $g_j(z)$ and identification of α requires only a single excluded instrument.

What about Continuous Demand Systems?

There is nothing in our proofs that requires the consumer-level quantities, Q_{ijt} to binary outcomes from a discrete choice model.

Berry, Gandhi and Haile describe a broad class of continuous choice models that can satisfy the key demand invertibility property.

As an example, the paper considers a "mixed CES" model where consumers have random coefficients in a CES model of continuous demand.