Foundations of Demand Estimation

Nonparametric Identification

ZHANG, Jianhao

Hong Kong University of Science and Technology

April 24, 2025

Outline

Insights from Parametric Models

How standard parametric models are identified brings out three recurring themes:

- Demand shocks that enter through indices for each good.
- The presence of a one-to-one mapping between the indices and market shares, allowing inversion of the demand system.
- The application of instrumental variables to identify the components of the inverse demand.

Nonparametric Demand Model

Without loss, condition on a fixed number of inside goods J, the demand for each product j in market t can be given by:

$$s_{jt} = \sigma_j(x_t, p_t, \xi_t) \qquad j = 1, \ldots, J. \tag{1}$$

- s_{jt}, measure of demand at market-level.
- x_t, *all* observed exogenous characteristics of the market and goods.
- p_t, prices of all goods.
- ξ_t, the J-vector of demand shocks.

To demonstrate identification, Berry and Haile(2014) require three main assumptions.

A Nonparametric Index

Assumption 1 (Index)

For all j, $\sigma_j(x_t, p_t, \xi_t) = \sigma_j\left(x_t^{(2)}, \delta_t, p_t\right)$.

- Partition x_t as $(x_t^{(1)}, x_t^{(2)})$ where $x_t^{(1)} = (x_{1t}^{(1)}, \ldots, x_{Jt}^{(1)}) \in \mathbb{R}^J$.
- For each market t, define a vector of indices $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})$ where

$$\delta_{jt} = x_{jt}^{(1)}\beta_j + \xi_{jt} \tag{2}$$

A Nonparametric Index

We may assume without loss of generality that $E[\xi_{jt}] = 0$ and $|\beta_j| = 1$ for all j:

$$\delta_{jt} = x_{jt}^{(1)} + \xi_{jt} \tag{3}$$

Furthermore, because variables $x_t^{(2)}$ is exogenous and play no role in the identification argument, we will henceforth condition on an arbitrary value of $x_t^{(2)}$ without loss of generality and suppress $x_t^{(2)}$ in the notation.

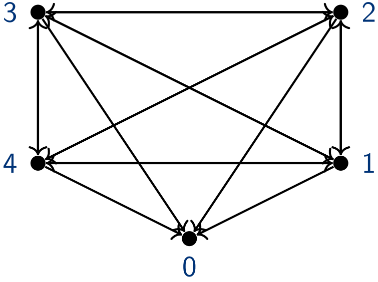
Inverting Demand

Assumption 2 ((Connected substitutes))

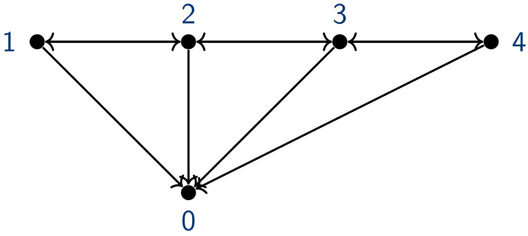
(i) $\sigma_k(\delta_t, p_t)$ is nonincreasing in δ_{jt} for all $j > 0$, $k \neq j$, and any $(\delta_t, p_t) \in \mathbb{R}^{2J}$;
(ii) For each $(\delta_t, p_t) \in \text{supp}(\delta_t, p_t)$ and any $\emptyset \neq \mathcal{K} \subseteq \{1, \ldots, J\}$, there exist $k \in \mathcal{K}$ and $\ell \notin \mathcal{K}$ such that $\sigma_\ell(\delta_t, p_t)$ is strictly decreasing in δ_{kt}.

Part (i) of Assumption 2 requires that goods be weak substitutes with respect to the indices: an improvement in the index δ_{jt} must weakly reduce the demand for other goods.
Part (ii) requires at least some strict substitution among goods $j = 0, 1, \ldots, J$, that there is no strict subset of goods that substitute only among themselves.

Inverting Demand



(a)



(b)

Inverting Demand

- Berry et al. (2013) demonstrate in a wide range of demand models that invertibility of demand is ensured whenever the connected substitutes conditions hold for some injective transformation of the demand system.
- Key implication is that for all demand vectors s_t such that $s_{jt} > 0$ for all j, there exists an inverse demand system taking the form:

$$\delta_{jt} = \sigma_j^{-1}(s_t; p_t) \qquad j = 1, \ldots, J. \tag{4}$$

Identification via instruments

- In the case of nonparametric regression we are interested in an equation of the form:
$$y = \Gamma(x) + \epsilon \tag{5}$$
where $x \in \mathbb{R}^K$.
- Newey and Powell (2003) showed that given instruments z satisfying the mean independence condition $E[\epsilon|z] = 0$, a necessary and sufficient condition for identification of the regression function Γ is a standard "completeness" condition.

Identification via instruments

Standard "completeness" condition(Newey and Powell, 2003)

In the class of functions $B(\cdot)$ on $\mathbb{R}^K$ such that $E[B(x)|z]$ is finite, the only function B such that $E[B(x)|z] = 0$ almost surely is a function that maps to zero almost surely on its domain.

Identification via instruments

To connect this to demand, observe that we may re-arrange each equation of (3) as:

$$x_{jt}^{(1)} = \sigma_j^{-1}(s_t; p_t) - \xi_{jt} \tag{6}$$

yielding a form similar to (5).

Identification via instruments

Assumption 3 (Instruments)

(i) For all $j = 1, \ldots, J$, $E[\xi_{jt} | z_t, x_t^{(1)}] = 0$ almost surely

(ii) For all functions $B(s_t, p_t)$ with finite expectation, if $E[B(s_t, p_t) | z_t, x_t^{(1)}] = 0$ almost surely, then $B(s_t, p_t) = 0$ almost surely.

Lemma 1. Under Assumptions 1–3, for all $j = 1, \ldots, J$, σ_j^{-1} is identified on the support of (s_t, p_t).

Identification via instruments

Theorem(Berry and Haile, 2014).

Suppose (s_t, x_t, p_t, z_t) are observable and that Assumptions 1–3 hold. Then for all j, the demand function σ_j is identified.

Micro Data

- "Micro data" refers to a setting where we can observe individual consumer characteristics d_{it} matched with the choices q_{it} of each consumer.
- Micro data can also provide a panel structure of observed outcomes for many individual consumers within each market.

Micro Data: Estimation

- One significant advantage of micro data: the potential for within-market variation to lessen (but not eliminate) reliance on instrumental variables.
- Suppose we have a mixed logit specification, with conditional indirect utilities of the form

$$u_{ijt} = x_{jt}\beta_{it} - \alpha_0 p_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{7}$$

  where

$$\beta_{it}^{(k)} = \beta_0^{(k)} + \sum_{\ell=1}^{L} \beta_d^{(\ell,k)} d_{i\ell t} + \beta_\nu^{(k)} \nu_{it}^{(k)}. \tag{8}$$

Micro Data: Estimation

We can rewrite this as:

$$u_{ijt} = \delta_{jt} + \mu_{ijt}(\nu_{it}; \beta_d, \beta_\nu) + \epsilon_{ijt}, \tag{9}$$

with

$$\mu_{ijt}(\nu_{it}; \beta_d, \beta_\nu) = \sum_{k=1}^{K} x_{jt}^{(k)} \left(\beta_\nu^{(k)} \nu_{it}^{(k)} + \sum_{\ell=1}^{L} \beta_d^{(\ell,k)} d_{i\ell t} \right) \tag{10}$$

and

$$\delta_{jt} = x_{jt}\beta_0 - \alpha_0 p_{jt} + \xi_{jt}. \tag{11}$$

McFadden et al.(1977) referred to δ_{jt} as the "alternative-specific" constant.

Micro Data: Estimation

- With the specification above, choice probabilities for each consumer i take the form:

$$s_{ijt} = \int \frac{\exp\{\delta_{jt} + \mu_{ijt}(\nu_{it}; \beta_d, \beta_\nu)\}}{\sum_{k=0}^{J_t} \exp\{\delta_{kt} + \mu_{ikt}(\nu_{it}; \beta_d, \beta_\nu)\}} dF_\nu(\nu_{it}) \tag{12}$$

- If j denote the good selected by consumer i in market t. The (15) gives the likelihood contribution of consumer i's choice as a function of parameters $(\delta, \alpha_y, \beta_d, \beta_\nu)$.

Micro Data: Estimation

- When the number of observed consumers per good is large in each market. We could estimate $(\delta, \beta_d, \beta_\nu)$ by maximizing the product of these (simulated) likelihoods over all consumers:

$$\mathcal{L}(\delta, \beta_d, \beta_\nu) = \prod_{i,t} \int \frac{\exp\{\delta_{j(i)t} + \mu_{ij(i)t}(\nu_{it}; \beta_d, \beta_\nu)\}}{\sum_{k=0}^{J_t} \exp\{\delta_{kt} + \mu_{ikt}(\nu_{it}; \beta_d, \beta_\nu)\}} dF_\nu(\nu_{it}) \quad (13)$$

- We can run a second-step linear IV regression(different from 2SLS) to estimate the parameters α_0 and β_0.
- With micro data, we now require only one excluded instrument.

Micro Data: Estimation

- It is more preferable to estimate all parameters at once exploiting both with-market and cross-market variation.
- Exploiting all variation can often lead to much more precise estimates.
- Estimation using simulated likelihood can be computationally demanding if we want sufficient precision

Micro Data: Moment Condition

To estimate all parameters jointly, one could use moment conditions reflecting the score of the likelihood (16) with respect to $(\delta, \beta_d, \beta_\nu)$ together with orthogonality conditions of the form:

$$E\left[\left(\delta_{jt} - x_{jt}\beta_0 - \alpha_0 p_{jt}\right) z_{jt}\right] = 0, \tag{14}$$

where z_{jt} represents the exogenous x_{jt} combined with excluded instruments for p_{jt}.

Micro Data: Moment Condition

- Following Berry et al. (2004a) We can combine moments reflecting market shares with "micro moments" characterizing key features of the joint distribution of consumer i's characteristics and the characteristics of her choice $j(i)$.
- Typical micro moments include covariances, or conditional expectations of consumer characteristics given characteristics of the chosen product (or vice versa).
- When we have micro data, we will have more limited reliance on orthogonality conditions: aggregate moment condition can be replaced by micro moments that are sufficient to identify $(\delta,\ \beta_d,\ \beta_\nu)$.

Examples of Estimation from Micro Data

- Demand for hospital(e.g., Capps er al.(2003), Ho(2009)).
- Retail outlets(e.g., Burda et al. (2015))
- Residential locations (e.g., Bayer et al. (2007), Diamond(2016))
- Automobiles (e.g., Goldberg (1995), Petrin (2002))

Consumer Panels

- Typically, consumer panel data refers to observation of each consumer on multiple choice occasions.
- One advantage of a consumer panel is that it provide more information about the role of individual characteristics in determing substitution patterns.

Consumer Panels: Estimation

- We can write likelihood contribution for consumer i, as a function of the parameters $(\delta, \beta_d, \beta_\nu)$.

$$s_{ijkm} = \int \left(\frac{e^{\delta_{jm0} + \mu_{ijm0}(v_{im};\beta_d,\beta_\nu)}}{1 + \sum_\ell e^{\delta_{\ell m0} + \mu_{i\ell m0}(v_{im};\beta_d,\beta_\nu)}} \right) \left(\frac{e^{\delta_{km1} + \mu_{ikm1}(v_{im};\beta_d,\beta_\nu)}}{1 + \sum_\ell e^{\delta_{\ell m1} + \mu_{i\ell m1}(v_{im};\beta_d,\beta_\nu)}} \right) dF_\nu(\nu_{im}). \tag{15}$$

- Another typical approach would start from the types of aggregate moments and micro moments.

Ranked Choice Data

- Data on each consumer's rank ordering of products.
- The absence of temporal separation can avoid any question about which stochastic components of the model should be viewed as fixed.
- Variation in ranked choice data is ideal for assessing the closest substitutes.
- Estimation can proceed along the lines suggested previously: likelihood approach (see Train (2009)) and moment method(see Berry et al. (2004a)).

Nonparametric Demand Model

Following Berry and Haile(2024), consider a nonparametric model of demand characterized by equations

$$s_{ijt} = \sigma_j(d_{it}, y_{it}, x_t, p_t, \xi_t) \qquad j = 1, \ldots, J. \tag{16}$$

Compared to the model considered in the case of market-level data, here we have added observed individual-specific measures (d_{it}, y_{it}) as determinants of demand.

Identification Assumptions

In addition to the required degree of variation in d_{it}, choice sets and price instruments, the identification results in Berry and Haile (2024) rely on a set of core assumptions on demand. The four main assumptions are:

(i) For all j, $\sigma_j(d_{it}, y_{it}, p_t, \xi_t) = \sigma_j(\gamma(d_{it}, y_{it}, \xi_t), y_{it}, p_t)$, with $\gamma(d_{it}, y_{it}, \xi_t) \in \mathbb{R}^J$.

(ii) $\sigma(\cdot, y_{it}, p_t)$ is injective on the support of $\gamma(d_{it}, y_{it}, \xi_t)$ conditional on (y_{it}, p_t).

(iii) $\gamma(\cdot, y_{it}, \xi_t)$ is injective on the support of $d_{it}|y_{it}$.

(iv) For all j, $\gamma_j(d_{it}, y_{it}, \xi_t) = g_j(d_{it}, y_{it}) + \xi_{jt}$.

Connection to Parametric Model

Consider the mixed-logit random utility specification:

$$u_{ijt} = x_{jt}\beta_{ijt} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{17}$$

- where $\beta_{ijt}^{(k)} = \beta_{0j}^{(k)} + \sum_{\ell=1}^{L} \beta_{dj}^{(\ell,k)} d_{i\ell t} + \beta_{\nu j}^{(k)} \nu_{it}^{(k)}$;
- $\ln(\alpha_{it}) = \alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}$.

Connection to Parametric Model

- We can rewrite (20) as

$$u_{ijt} = g_j(d_{it}, x_t) + \xi_{jt} + \mu_{ijt}, \tag{18}$$

where

$$g_j(d_{it}, x_t) = \sum_k x_{jt}^{(k)} \sum_{\ell=1}^{L} \beta_{dj}^{(\ell,k)} d_{i\ell t} = \sum_{\ell=1}^{L} d_{i\ell t} \sum_k x_{jt}^{(k)} \beta_{dj}^{(\ell,k)} \tag{19}$$

and

$$\mu_{ijt} = \sum_k x_{jt}^{(k)} \left(\beta_{0j}^{(k)} + \beta_{vj}^{(k)} v_{it}^{(k)} \right) - p_{jt} \exp(\alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}) + \epsilon_{ijt} \tag{20}$$

- Notice that if $L = J$, our key assumptions hold as long as the $J \times J$ matrix of coefficients on d_{it} (whose elements are $\sum_k x_{jt}^{(k)} \beta_{dj}^{(\ell,k)}$) is full rank.

Identification: A Sketch

- First, a combination of within-market and cross-market variation is exploited to uncover the index function $g : \mathbb{R}^J \to \mathbb{R}^J$.
- Then cross-market variation—including that produced by excluded instruments for prices—allows identification of the demand shocks ξ_{jt} for all goods and markets in the same way that residuals in a nonparametric regression model are identified
- Finally, with the demand shocks known, identification of demand is immediate from the definition of demand in (19).

Identification of Index Function

- Let $\mathcal{S}(\xi, p)$ denote the support of the share vector when the random variables (ξ_t, p_t) take the values (ξ, p). Because d_{it} varies within each market, the set $\mathcal{S}(\xi, p)$ is not a singleton: each d_{it} in market t is associated with a different observed conditional choice probability vector s_{it}.
- Given the assumptions on demand, for each vector of market shares $s \in \mathcal{S}(\xi, p)$ there will be a unique d^* in the support of d_{it} such that

$$\sigma(g(d^*) + \xi, p) = s. \tag{21}$$

Identification of Index Function

- This d^* is the vector of consumer characteristics that generate the choice probability vector s (given $(\xi_t, p_t) = (\xi, p)$). So we may write

$$d^*(s; \xi, p). \tag{22}$$

- Furthermore, the inverted demand system at this point is

$$g(d^*(s; \xi, p)) + \xi = \sigma^{-1}(s; p). \tag{23}$$

- Choice probabilities conditional on d_{it} in each market t are observed, $d^*(s; \xi_t, p_t)$ is observed for all t and $s \in \mathcal{S}(\xi_t, p_t)$ even though no ξ_t is observed or known at this point.

Identification of Index Function

- If we differentiate (26) within a market t where $p_t = p$ and $d^*(s; \xi_t, p) = d$, we obtain
$$\frac{\partial g(d)}{\partial d} \frac{\partial d^*(s; \xi_t, p)}{\partial s} = \frac{\partial \sigma^{-1}(s; p)}{\partial s}. \tag{24}$$
- If we do the same within another market t' with the same p and same $s \in \mathcal{S}(\xi_{t'}, p)$, we get a similar expression with an identical right-hand side. Setting the two left-hand sides equal and letting $d' = d^*(s; \xi_{t'}, p)$, we see that

$$\frac{\partial g(d')}{\partial d} = \left[\frac{\partial g(d)}{\partial d}\right] \frac{\partial d^*(s; \xi_t, p)}{\partial s} \left[\frac{\partial d^*(s; \xi_{t'}, p)}{\partial s}\right]^{-1}. \tag{25}$$

Identification of Demand

- Berry and Haile (2024) require that there exist some "common choice probability" vector s^* that is reached in every market by a consumer with the "right" characteristics d_{it} for that market.
- Specifically, to our earlier assumptions (i)-(iv) we add:
  (v) There exists s^* such that $s^* \in \mathcal{S}(\xi, p)$ for all $(\xi, p) \in \text{supp}(\xi_t, p_t)$.
- This assumption requires existence of at least one vector of choice probabilities for the inside goods is reached in every market t.

Identification of Demand

- With a common choice probability vector s^*, in every market t we have J inverse demand equations of the form

$$g_j(d^*(s^*;\xi_t, p_t)) = \sigma_j^{-1}(s^*;p_t) - \xi_{jt}. \tag{26}$$

- Identification of $\sigma_j^{-1}(s^*;p_t)$ will be done given instruments for endogenous varialbles p_t.

THANK YOU