# Identification and instruments

**Ari Hyytinen**

Hanken and Helsinki GSE

March 2024

(This lecture: Part 3 and 4)

# Table of contents

- **Part 1**: Econometric identification
  - Introduction
  - Basic concepts
- **Part 2**: Normalizations and special regressors in identification
  - Scale and location normalizations
  - Special regressors
- **Part 3**: Normalizations in discrete choice models
  - Scale of utility
  - Outside good
- **Part 4**: Instruments for and identification discrete choice demand models
  - Importance and intuition for non-parametric identification in BLP
  - Challenges
  - Logic of formal proof: Index – Inversion – Instruments

# Literature for part 3 and 4

- Nonparametric foundations and instruments: Demand estimation with discrete choice models
  - Berry, S. and P. Haile (2016). Identification in Differentiated Products Markets, ***Annual Review of Economics***.
  - Berry, S. and P. Haile (2021). Foundations of demand estimation, ***Handbook of Industrial Organization.***
    - Berry, S. and P. Haile (2014). Identification in differentiated products markets using market level fata, ***Econometrica***, 82, 1749
    - Berry, S. A. Gandhi and P. Haile (2013). Connected substitutes and invertibility of demand, ***Econometrica***, 81, 2087-2111.

# Part 3

Discrete choice models

Scale of utility and outside good

# Scale of utility

- In (many) discrete choice models, the scale of utility is normalized by normalizing the variance of the indirect utility's error component.
  - Let i's utility from choice j be $U_{ij} = V_{ij} + \varepsilon_{ij}$, for j = {0, 1, ..., J}.
    - Multiplying $U_{ij}$ by a constant, $\sigma > 0$, changes the *scale* of utility but *does not affect choice*, since $U_{ij} > U_{ik} \Leftrightarrow \sigma U_{ij} > \sigma U_{ik}$
  - By setting the variance of the error term to a constant, the scale of utility is normalized.

# Example: Multinomial logit

- Consider non-normalized variance: $\sigma\pi^2/6$
  - $\pi^2/6$ is related to a constant of integration.

$$U^*_{ij} = V_{ij} + \varepsilon^*_{ij} \text{ with } Var(\varepsilon^*) = \sigma^2\pi^2/6.$$

- Utility can divided by $\sigma$ without changing behaviour.

# ...

- Utility can divided by σ without changing behaviour.

$$U_{ij} = V_{ij}/\sigma + \varepsilon_{ij}$$

$$Var(\varepsilon^*/\sigma) = Var(\varepsilon) = \pi^2/6$$

$$s_{ij} = \frac{e^{V_{ij}/\sigma}}{\sum_k e^{V_{ik}/\sigma}} \approx \frac{e^{\beta^*/\sigma \cdot x_{ij}}}{\sum_k e^{\beta^*/\sigma \cdot x_{ik}}}$$

# …

- Every *coefficient* is rescaled by $\sigma$
- Only the ratio $\beta = \beta^*/\sigma$ is identified (*up-to-scale*)
  - Interpretation of the coefficients:
    - *They are relative to the variance of the unobserved factors*
  - Greater unobserved variance, the smaller $\beta$.
  - Ratio $\beta_j$ / $\beta_k$ is invariant to scaling

# Outside good

- Is outside good a free normalization?
  - Let i's utility from choice j be $U_{ij} = V_{ij} + \varepsilon_{ij}$, for $j = \{0, 1, \ldots, J\}$
  - Utility maximization $\rightarrow$ j chosen if $U_{ij} > U_{ik}$ $k \neq j$
  - There is a set of unobservable "taste parameters", $\varepsilon_{ij}$, that result in the purchase of good j.

# …

- Good j chosen with probability $\Pr(\varepsilon_i > V_i)$,

  where $V_i = V_{ik} - V_{ij}$ and $\varepsilon_i = \varepsilon_{ij} - \varepsilon_{ik}$, $k \neq j$.

$$s_{ij} = \int I(\varepsilon_{ij} - \varepsilon_{ik} > V_{ik} - V_{ij}) f(\varepsilon_i) \partial \varepsilon_i$$

## ...

- Example: Multinomial logit

$$s_{ij} = \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}}$$

$$\frac{s_{ij}(\theta)}{s_{ik}(\theta)} = \frac{e^{V_{ij}}}{e^{V_{ik}}} = e^{V_{ij}-V_{ik}}$$

- This shows that variations in market shares (LHS) are informative of difference $V_{ij}$ - $V_{ik}$

- Only differences in the indirect utilities are identified

# …

- Side-note: This also shows how IIA is at work:

$$\frac{s_{ij}(\theta)}{s_{ik}(\theta)} = \frac{e^{V_{ij}}}{e^{V_{ik}}} = e^{V_{ij}-V_{ik}}$$

- Ratio of choice probabilites depends only on j and k.
- It does not depend on any other alternative h.

## …

- A key consequence of the fact that only differences in the utility matter for choice:
  - Adding constants to utility irrelevant for choice
    - Only *differences* in *alternative specific* constants can be identified.
  - Also: Effects of individual specific factors, such as income, not identified
    - They do not vary between goods j and k

# Adding constant to utility

- Example: Multinomial logit with $V_{ij} = X_i'\beta_j$

- Add a constant C to each $\beta_j \rightarrow$ no effect on choice

$$s_{ij} = \frac{\exp[\mathbf{x_i}(\beta_j + C)]}{\sum_k \exp[\mathbf{x_i}(\beta_k + C)]} = \frac{\exp[\mathbf{x_i}C]\exp[\mathbf{x_i}\beta_j]}{\exp[\mathbf{x_i}C]\sum_k \exp[\mathbf{x_i}\beta_k]}$$

# Adding constant to utility

- Because adding constant has no effect on choice, C has to be normalized.

$$s_{ij} = \frac{\exp[\mathbf{x_i}(\beta_j + C)]}{\sum_k \exp[\mathbf{x_i}(\beta_k + C)]}$$

- Consider setting $C = -\beta_0$

- Interpretation: Good j = 0 produces ”no utility in expectation”

$$e^{V_{i0}} = e^0 = 1$$

# …

- This normalizes the *level* of indirect utility that agents get from the outside good.

- This is a *location* normalization and implies that utility from the outside option is: $U_{i0} = \varepsilon_{i0}$

  - We have to interpret $U_{ij}$ as utility from j (j > 0) relative to the outside option.

## …

- In static discrete choice models:
  - This is a free normalization (*almost*; see fn 22 in Berry and Haile -21)
- In dynamic discrete choice models:
  - This is not a free normalization
  - Implies restrictions on preferences and thus behaviour
    - see, e.g., Rust (1994) and Magnac and Thesmar (2002).

# What is outside good?

- One of the goods in the choice set is the outside good and this good produces no utility in expectation.

  - What is this “j = 0” good?

  - It is the good (whatever it is) *for which price is not set in response to the prices of the inside goods*

# …

- *Existence* of outside good important:
  - If there was no outside good, consumers would be *forced* to choose one inside good.
    - Demand would depend only on differences in prices.
  - General increase in price level would not decrease the (total) amount bought (*implausible*)

## …

- Outside good

  - Typical assumption: Market size equals the size of population in a market × constant

    - E.g., in a soda market: constant = max amount i can potentially consume

  - The constant is not a free normalization: It affects estimates of preferences and counterfactuals

  - Can market size identified and estimated? (Yes, see L. Zhang, JMP)

# Part 4

This part is about identification of BLP model(s)

# Introduction

- Why should one be interested in nonparametric identification of BLP type models?

- To understand better
  - sources (drivers) of identification
  - role of parametric assumptions
  - working of parametric estimators
    - must be used in practise when working with finite samples $\rightarrow$

## …

- Are functional form and distributional assumptions
  - essential for identification (key features of the model); or
  - useful practical tools when working with finite samples?
- Identification also provides guidance for applications & empirical work
  - E.g., what types of instruments are needed?
    - What are the implications of distributional restrictions,  better data (e.g., micro data), or functional form assumptions for the kinds & number of IVs needed?

*Challenges of demand estimation*

# Demand estimation

- Goal: Measure responses of quantities demanded to *ceteris paribus* changes in prices or other factors

  - What is needed: sufficiently flexible functional forms, valid sources of exogenous variation, and sufficient account for unobserved heterogeneity

  - *Challenge #1*: Unobservable demand shocks $\rightarrow$ price endogeneity

  - *Challenge #2*: Unobservable demand shocks must be held fixed to measure e.g. demand elasticities

# Challenges 1# and #2

- *Challenge #1*: Endogeneity of prices
  - Statistical dependence between prices and latent demand shocks
  - Results from unobservables, U, that affect demand

$$Q = D(X, P, U)$$
$$P = C(W, Q, V).$$

# …

- *Challenge #2*: Demand of good j depends on more than one latent demand shock
  - Demand for good j cannot be considered in isolation from $j \neq k$
    - E.g., change in the price (or quality) of a substitute or complement will cause demand to shift
  - Demand for good j changes if the prices or characteristics of *any* of the related goods change

## …

- Demand shocks ($U_1$, ..., $U_J$) are associated with *all* related goods.

- Prices and characteristics of related goods $j \neq k$ cannot, in general, be *excluded* from the demand for good j.

- Demand is not a regression $\rightarrow$

# …

Consider a market with J interrelated goods, with the demand for good j = (1,…, J) given by:

$$Q_j = D_j\,(X, P, U) \qquad \text{(general demand)}$$

where

$$X = (X_1, \ldots, X_J)$$
$$P = (P_1, \ldots, P_J)$$
$$U = (U_1, \ldots, U_J).$$

# …

- There are J structural errors that enter on the right-hand side of the demand equation
  - The presence of demand shocks ($U_1$, ..., $U_J$) implies that this is **not** a standard regression equation.
- Econometric models with *multiple structural errors* are harder to identify and estimate than regression models
  - see also Matzkin (2013)

# …

- To estimate the *level* and *slope* of demand at specific points is *different from* estimating e.g. some weighted average responses (such as LATE).
  - see Berry and Haile (-21, section 2.5.3)
  - *Averaging over latent variables ≠ holding them fixed*
- In IO, we are rarely interested in average responses.
  - Instead, the interest is in the ceteris paribus effects of *counterfactual* (price) changes or *ex ante* analyses of proposed policy changes

## ...

- To generically identify demand elasticities requires that a price can be varied while *holding all else constant*
  - *This includes U* $\rightarrow$ demand shocks ($U_1$, ..., $U_J$) also need to be held fixed when defining a ceteris paribus effect
- Having J instruments for J endogenous prices is not enough for non-parametric identification of $D_j$ (entire function)
  - Or, more specifically, not without further functional form restrictions

# Example: Randomized prices

- Observed variation in quantities with *randomized prices*:
  - Randomized prices remove the dependence of prices on demand shocks
  - Randomization of prices does *not* keep the demand shocks, U, constant

→ What can one learn?

# …

- When P randomized, observed variation in Q allow identifying certain averages of demand responses, based on integration over $U = (U_1, \ldots, U_J)$
    - I.e., certain types of LATEs can be obtained by integrating over the vector of demand shocks.
- Such averages are not informative of any elasticity of demand at e.g. observed prices or quantities

# Example: Functional form

- Consider the following demand restriction:

$$D_j(X, P, U) = D_j(X, P, \varepsilon_j(U)) \qquad (**)$$

where $\varepsilon_j(U)$ is a scalar and $D_j$ is increasing in $\varepsilon_j(U)$

- This could be the case *if* demand for good j is assumed to be linear in the demand shocks U.

# …

- Here randomized prices (P _|_ U) or J instruments for prices would allow identifying the demand function (**)
  - As in Matzkin (-03): Quantile $\tau$ of distribution of $Q_j$|X, P allows tracing $D_j(X, P, \varepsilon_j(U))$ for $\varepsilon_j(U)$ fixed at its $\tau$ quantile.
- However: $\varepsilon_j(U)$ is a strong functional form assumption.
  - Ruled out by many common parametric demand specifications, such as multinomial logit.
  - See Berry and Haile (-21, section 2.4) for further discussion.

*BLP setup and role of $\xi_{jt}$*

# Set-up: BLP demand model

- Preference heterogeneity + product or market-specific unobservable ($\rightarrow$ endogeneity)

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

$$\beta_{it}^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} d_{it} + \beta_2^{(k)} \nu_{it}^{(k)}$$

$$(\epsilon_{ijt}, \nu_{it}) \perp\!\!\!\perp (x_t, p_t)$$

$$\epsilon_{ijt} \text{ i.i.d. extreme value, } \nu_{it}^{(k)} \text{ i.i.d. normal}$$

# Role of $\xi_{jt}$

- Notion of a market (*t*) is important:
  - Markets = natural combinations of geography (e.g., metropolitan areas) and time (e.g., years, quarters)
  - But: what is e.g. a market for a given digital good?
- Demand shocks associated with good j and market t, giving rise to unobserved $\xi_{jt}$

# …

- What are unobserved $\xi_{jt}$?
- *Narrow interpretation:*
  - Good j's unobserved characteristics
- *Broader interpretation*:
  - Any combination of latent product characteristics and latent taste variation *that is common to consumers in market t*
  - High $\xi_{jt} \rightarrow$ consumers have a "high mean taste" for j in market t

# …

- $\xi_{jt}$ is observed by firms when prices are set but not by us econometricians

- Potentially correlated with price Corr($\xi_{jt}$, $p_{jt}$) $\neq$ 0, just like "demand shocks".

  - Typical assumption: Not correlated with other characteristics $E[\xi_{jt} \mid x_{jt}] = 0$.

- Unobserved product characteristics allow product j to be better than product k in a way that is not explained by differences in $x_j$ and $x_k$.

  - Vertical in nature: Consumers agree on their value

# Example: Automarket

- Narrow interpretation: $\xi_{jt}$ mirrors anything
  - … that makes Volvo better than Skoda that is not fully captured by the observable characteristics in the data
  - … and that affect demand for different products and that leads higher sales (and/or higher prices).
- Difficult-to-quantify aspects: style, prestige, reputation, past experience, etc
- Quantifiable characteristics, but not in the data

*Insights from parametric models*

# How is the BLP model identified?

- Standard intuition:
  - Exogenous changes in choice sets via exclusion restrictions (instruments)
  - Functional form / distributional assumptions
  - Supply side $\rightarrow$ cross equation restrictions $\rightarrow$ overidentifying restrictions for parameters
    - This is why imposing supply side can be “informative of demand”

# …

- Preview: Key lessons from non-param. identification results of Berry and Haile (2014, 2021):
  - Main requirement for non-parametric identification in BLP-type demand models:
    - **i+i+i -requirement: index restriction + invertability + instruments**
    - In practise: Trade-off between functional form restrictions vs IV needs (what is available)
  - Functional form assumptions mainly in the "standard role"
    - Approximation in finite samples / interpolation / extrapolation

# Building intuition for i+i+i

- Intuition for

  i. Index: How the index structure links $\xi_{jt}$ to observables,

  ii. Inversion: How inversion yields equations that can be estimated using standard econometric tools

  iii. Instruments: Need for instruments for the endogenous variables in appering in those equations

## …

- Rewrite

1. Multinomial logit
2. Nested logit
3. BLP

# 1. Multinomial logit

$$v_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

Indirect utlity

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

Linear index

$$s_{jt} = \frac{e^{\delta_{jt}}}{1 + \sum_k e^{\delta_{kt}}}$$

Mkt share (choice probability)

$$\delta_{jt} = ln(s_{jt}) - ln(s_{0t})$$

Inversion using $s_{0t}$ (see next page)

$$ln(s_{jt}) - ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

Like regressing quantity on price, need instruments for price

# …

How inversion using $s_{0t}$ works:

$$\ln s_{0t} = -\log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right)$$

$$\ln s_{jt} = [x_{jt}\beta - \alpha p_{jt} + \xi_{jt}] - \log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right)$$

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{Data!} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

# ...

$$ln(s_{jt}) - ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

Rewrite this to get an index on LHS

$$x_{jt} = \left(x_{jt}^{(1)}, x_{jt}^{(2)}\right)$$

Decompose $x_{jt}$ and set scale by dividing $\beta_1$ ($\xi_{jt}$ have no natural scale) + fix $x^{(2)}_{jt}$

$$x_{jt}^{(1)} + \tilde{\xi}_{jt} = \frac{1}{\beta_1}\left(ln(s_{jt}) - ln(s_{0t})\right) + \frac{\alpha}{\beta} p_{jt}$$

LHS: Index

RHS: Tightly parameterized function of shares and price.

where: $\tilde{\xi}_{jt} = \frac{\xi_{jt}}{\beta^{(1)}}$

# ...

$$x_{jt}^{(1)} = \frac{1}{\beta^{(1)}}\left(\ln(s_{jt}) - \ln(s_{0t})\right) + \frac{\alpha}{\beta^{(1)}} p_{jt} - \tilde{\xi}_{jt},$$

- resembles a regression equation, with an additively separable error on RHS

- forms a connection to the more complicated models we discuss

## …

- In spite of the two endogenous variables on RHS, only one *excluded* instrument $z_{jt}$ needed to identify this equation
    - Note role of $x^{(1)}_{jt}$: It can be interpreted as a type of special regressor
        - (in BLP: $x^{(1)}_{jt}$ does not have a random coefficient)
- Implication:

Bivariate moment condition: $E[\xi_{jt} | x^{(1)}_{jt} , z_{jt}] = 0$

# 2. Nested logit

$$\ln(s_{jt}) - \ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + (1-\lambda)\ln(s_{j/g,t}) + \xi_{jt}$$

$$x_{jt}^{(1)} + \tilde{\xi}_{jt} = \frac{1}{\beta^{(1)}}\left(ln(s_{jt}) - ln(s_{0t}) - (1-\lambda)\ln(s_{j/g,t})\right) + \frac{\alpha}{\beta^{(1)}} p_{jt}$$

Like a regression equation: Same LHS as for mlogit, RHS now a more complicated function of the mkt shares and price.

Instruments needed for the price and for $\ln(s_{j/g,t})$, which is a specific function of the share vector = $(s_1, \ldots, s_J)$ ("quantities")

# 3. BLP

$$x_{jt}^{(1)} + \tilde{\xi}_{jt} = \frac{1}{\beta^{(1)}} \tilde{\delta}_j \left( s_t, p_t, x_t^{(2)}, \theta \right)$$

Inverse market share function

Has to be evaluated numerically

Depends nonlinearly on parameters of the random coefficients.

Like non-linear regression:

Same LHS as before, but RHS a complicated function of prices ($p_t$) and market shares ($s_t$), all of which are correlated with $\xi_{jt}$ i.e., endogenous:

Need more IVs. How many? $\rightarrow$

$$x_{jt}^{(1)} + \tilde{\xi}_{jt} = \frac{1}{\beta^{(1)}} \tilde{\delta}_j \left( s_t, p_t, x_t^{(2)}, \theta \right)$$

# …

- $\xi_{jt}$ vary across products and markets and complicate the identification of BLP-demand:
  - Each shock affects the quantity demanded and price of *all* related goods, implying that *all* components of $s_t=(s_{1t},\ldots, s_{Jt})$ and $p_t=(p_{1t},\ldots, p_{Jt})$ depend on $\xi_{jt}$
  - RHS: With J products, 2J endogenous variables:
    - J prices ($p_t$)
    - J quantities or market shares ($s_t$).

*Non-parametric identification of the BLP model*

# Identification using market level data

- Berry-Haile (2014, 2021): Nonparametric generalization of the BLP model, with the following three key elements:
  - **Index**: Index restriction
  - **Inversion**: Generalized multivariate inversion of choice probabilities:
    - Express each index as function of endogenous variables
  - **Instruments**: IVs can identify the inverse market share functions and thereby the model's structural errors $\xi$.

# General demand model

- Demand for good j in market t:

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

- **Goal of non-parametric identification**: Learn *function* $\sigma_j(x_t, p_t, \xi_t)$, including all its partial derivatives
  - This demand system can be derived from a general random utlity discrete choice model; see e.g. Berry and Haile (2021)

# Three assumptions

- Inversion (i.e., existence of J inverse share equations)
  - **Assumption 1** (non-parametric functional form assumption): *Index restriction*
  - **Assumption 2** (invertibility of the demand system): *Connected substitutes*
- Non-parametric identification of the inverted share functions (non-parametric IV regression)
  - **Assumption 3** (non-parametric IV-assumptions): Mean independence & completeness

# A1: The index restriction

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

Partition $x_t = (x_t^{(1)}, x_t^{(2)})$

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

Demand shocks have no natural location or scale: wlog, $E[\xi_{jt}] = 0$, $|\beta_j|=1$ (special regressor)

**Assumption (index):**

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

This is a non-parametric functional form assumption. *It restricts how $x_{jt}^{(1)}$ and $\xi_{jt}$ can affect the demand.*

They can enter the non-parametric function, $\sigma_j$, only through the index $\delta_{jt}$

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

# …

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

- **Note #1**: $x_{jt}^{(1)}$ and $\xi_{jt}$ are "perfect substitutes" in the index.
- **Note #2**: The index restriction is what leaves $x_t^{(1)}$ out of from the inversion, making them available as instrument for shares
  - (i.e., BLP-instruments can be excluded)
- **Note #3**: Since $\delta_t = (\delta_{1t}, \ldots, \delta_{tJ})$, the demand shock to, say, good *k* can still affect the demand for good j, through a fully non-parametric function $\sigma_j$

# A2: Connected substitutes (CS)

- Goods are
  - (i) *weak substitutes* (w.r.t. $\delta_t$); and
  - (ii) *connected* to each other in the demand system
- Violations of CS assumption unlikely in a discrete choice model (Berry and Haile 2014)

## ...

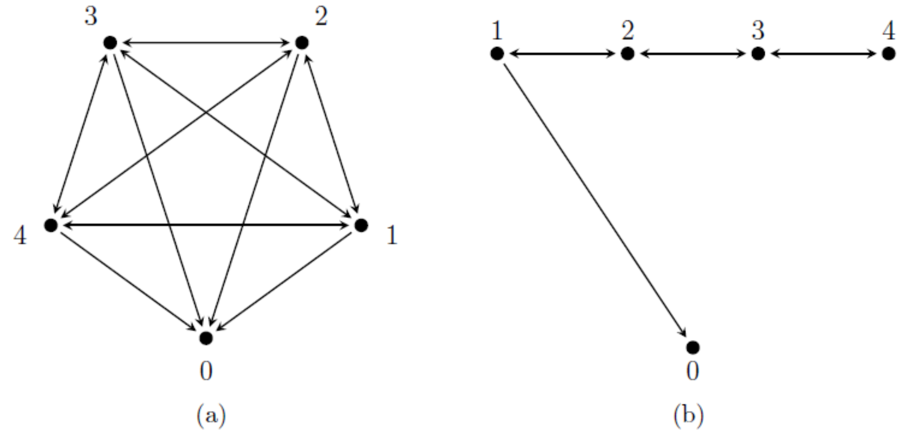
**Assumption 5.2** (Connected substitutes).

**(i)** $\sigma_k(\delta_t, p_t)$ *is nonincreasing in* $\delta_{jt}$ *for all* $j > 0$, $k \neq j$, *and any* $(\delta_t, p_t) \in \mathbb{R}^{2J}$;

**(ii)** *for each* $(\delta_t, p_t) \in supp(\delta_t, p_t)$ *and any nonempty* $\mathcal{K} \subseteq \{1, \ldots, J\}$, *there exist* $k \in \mathcal{K}$ *and* $\ell \notin \mathcal{K}$ *such that* $\sigma_\ell(\delta_t, p_t)$ *is strictly decreasing in* $\delta_{kt}$.

- Part (i): *Weak* substitution
  - I.e., greater $\delta_{jt}$ must weakly reduce the demand for other goods
- Part (ii): *Strict* substitution among at least some goods
  - I.e., goods are *connected*: No strict subset of goods substitute only among themselves; all goods belong in one demand system

Figure 2: Substitution in Standard Discrete Choice Models




Directed graphs of the substitution matrix for standard discrete choice models, with $J = 4$ inside goods. Panel (a): standard random utility models of horizontal differentiation, such as the multinomial logit, multinomial probit, nested logit, mixed logit/probit. Panel (b): the pure vertical model with an outside good. From each vertex associated with an inside good there is a directed path to the vertex associated with the outside good.

From Berry and Haile -21, illustrating existence of a directed path from any good j > 0 to the outside good (j = 0)

# A1 + A2 => inversion

- Berry, Gandhi and Haile (2013) generalize the Berry (1994) invertibility result

**Lemma (BGH).** Under Assumptions 1-2, for each $j$ there exists a function $\sigma_j^{-1}$ such that $\delta_{jt} = \sigma_j^{-1}(s_t, p_t)$ for all $(s_t, p_t)$ in their support.

Existence lemma:

For all demand vectors, there exist an inverse demand system of form $\delta_{jt} = \sigma^{-1}_j(s_t, p_t)$ for j = 1,..., J

# …

- Inverting demand system = tool to obtain a representation with *one structural error per equation*

- Inverted demand equations similar to regression equations:

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

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

# A1-A3: Identification of demand

- A3: Instruments satisfy a mean independence $E[\xi_{jt} | x^{(1)}_{jt}, z_{jt}] = 0$ and a relevance ("completeness") condition (Newey-Powell 2003)

**Lemma 1.** *Under Assumptions 5.1–5.3, for all $j = 1, \ldots, J, \sigma_j^{-1}$ is identified on the support of $(s_t, p_t)$.*

- Newey-Powell identification argument for non-parametric IV regression can be extended → identification of each $\sigma^{-1}_j()$ → plug in $(s_t, p_t)$ → recover each $\xi_{jt}$ → identification of demand ($\sigma_j$) for all j.

**Theorem 5.1** (Berry and Haile (2014)). *Suppose $(s_t, x_t, p_t, z_t)$ are observable and that Assumptions 5.1–5.3 hold. Then for all $j$, the demand function $\sigma_j$ is identified.*

*Instruments*

# Instruments: Which and why?

- Starting point:
  - Learning about demand requires instruments for price, such as e.g. cost shifters.
  - When products are differentiated, *we also have to learn about ”substitution patterns”*
    - … in the dimension of observed product characteristics, $x_j$
    - … in the unobserved (vertical) dimension as captured by the product-specic demand error $\xi_{jt}$

  $\rightarrow$ *variation in the shares are informative of these*

$$\delta_{jt} \equiv x_{jt} + \xi_{jt}$$

# …

- Besides generating variation in prices, we need to move shares at any given price vector.
  - **Prices**: The need of IVs for prices:
    - Prices endogenous, because likely to correlate with $\xi_{jt}$
    - We need changes in each price, holding all others fixed and in a way that isn't confounded by changes in $\xi$.
  - **Shares:** The need of IVs for shares:
    - Use excluded instruments $x_{-jt}$ to hold shares $s_{-jt}$ fixed while prices change

- Intuition

  - To identify substitution patterns, we need observed exogenous changes *in the choice set*

  - Learning about substitution: *Exogenous shifters of own and rival-product demand* to handle vertical substitution

    - Exogenous shifters of rival-product demand $\rightarrow$ instrument for the vector of market shares in inverse demand

# How many instruments? (2J)

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

$\sigma^{-1}_j(s_t, p_t)$ strictly increasing in $s_{jt}$

$$s_{jt} = h_j\left(s_{-jt}, p_t, x_{jt}^{(1)}, \xi_{jt}\right)$$

for some function $h_j$

This shows why 2J instruments are needed: Each inverse equation, $\sigma^{-1}_j(s_t, p_t)$ has 2J *endogenous* variables. The latter equation shows that:

- $x^{(1)}_{jt}$ can act as an instrument for itself
- Need J instruments for prices, $p_t$
- J - 1 instruments for the endogenous quantities $s_{-jt}$.

# Instrument menu for IV-regression

1. Cost shifters
2. BLP-instruments
3. Waldfogel-Fan instruments
4. Exogenous market structure changes
5. Differentiation instruments

- *Separate question*: optimal functions of the instruments to use in the conditional moment conditions $\rightarrow$ *optimal instruments*

# Cost shifters

- Shifts in marginal costs (materials, tax, tariffs, etc) have no direct effect on quantities and can be used as an instrument for prices to identify demand
  - In BLP: Need variation in the costs *across alternatives*.
  - Proxies for cost shifters: Local wage levels
  - “Hausman IVs” = prices of the same product in other markets since such prices mirror variation in costs (valid if demand shocks are not correlated across markets)

# BLP-instruments

- Single-product firms: IVs = average (exogenous) characteristics of competing products in the same market

  - **Relevance**: In oligopoly, firm j sets the price as a function of characteristics of products produced by competing firms, suggesting their relevance

  - **Exclusion**: Characteristics of competing products should not depend on $\xi_{jt}$ (i.e., consumers' valuation of focal firm j's product).

- Multiproduct firms: IVs = characteristics of all other products produced by same firm

# Waldfogel-Fan instruments

- **Mark-up shifters** -- e.g., characteristics of “nearby" markets (”Waldfogel instruments")
  - e.g., sometimes firms use of the same price for all markets in a region (“zone-pricing”)
  - e.g., demographics, such as age, in Helsinki may affect prices (markups) in Lahti, but may be independent of Lahti preferences (including Lahti demand shocks):
    - Conditional on Lahti observables they act through price (i.e., costs or mark-ups).

# Exogenous shifters of mkt structure

- Exogenous shifters of market structure:
  - Something that affects prices through mark-ups
    - (changes in the intensity of competition  overall or locally in product space)
  - Exogenous entry and exit
  - Changes in firm ownership, mergers

# Differentiation instruments

- **Differentiation instruments refers to** the proximity (distance) in product characteristics (Gandhi and Houde 2020)

  - IVs should mirror the exogenous degree of differentiation of each product in a market

  - E.g., counts of "close" rival and non-rival products in each market

  - E.g., sums over squared differences between rival and non-rival products in each market.

  - May help with weak identification of BLP

## …

- Berry and Haile (2014):
- When 2J instruments needed, one can use:
  - J BLP instruments ("for shares")
    - BLP instruments unique in the sense that they affect shares both through prices *and* directly through choice problem.
  - J cost shifters ("for prices")

# Concluding remarks

# Implementation / PyBLP

- Conlon and Gortmaker (RJE, -20)

**Algorithm 1 Nested Fixed Point**

For each guess of $\theta_2$:

(a) For each market $t$, solve $\mathcal{S}_{jt} = s_{jt}(\delta_t, \theta_2)$ for $\hat{\delta}_t(\mathcal{S}_t, \theta_2) \equiv \hat{\delta}_t(\theta_2)$.

(b) For each market $t$, use the $J_t \times 1$ vector $\hat{\delta}_t(\theta_2)$ to construct the $J_t \times J_t$ matrix $\Delta_t(p_t, \hat{\delta}_t(\theta_2), \theta_2)$

(c) For each market $t$, recover $\hat{\eta}_t(\theta_2) = \Delta_t(\hat{\delta}_t(\theta_2), \theta_2)^{-1}\mathcal{S}_t$ by solving the $J_t \times J_t$ linear system.

(d) Stack up $\hat{\delta}_t(\mathcal{S}_{jt}, \theta_2)$ and $\hat{c}_{jt}(\hat{\delta}_t(\theta_2), \theta_2) = f_{MC}(p_{jt} - \hat{\eta}_{jt}(\hat{\delta}_t(\theta_2), \theta_2))$ and use linear IV-GMM to recover $[\hat{\theta}_1(\theta_2), \hat{\theta}_3(\theta_2)]$ following the recipe in Appendix A.1. The following is our somewhat different formulation:

$$\begin{aligned} \hat{\delta}_{jt}(\mathcal{S}_t, \theta_2) + \alpha p_{jt} &= [x_{jt}, v_{jt}]\beta + \xi_{jt}, \\ f_{MC}(p_{jt} - \hat{\eta}_{jt}(\theta_2)) &= [x_{jt}, w_{jt}]\gamma + \omega_{jt}. \end{aligned} \tag{11}$$

(e) Construct the residuals:

$$\begin{aligned} \hat{\xi}_{jt}(\theta_2) &= \hat{\delta}_{jt}(\theta_2) - [x_{jt}, v_{jt}]\hat{\beta}(\theta_2) + \alpha p_{jt}, \\ \hat{\omega}_{jt}(\theta_2) &= \hat{c}_{jt}(\theta_2) - [x_{jt}, w_{jt}]\hat{\gamma}(\theta_2). \end{aligned} \tag{12}$$

(f) Stack the sample moments:

$$g(\theta_2) = \begin{bmatrix} \frac{1}{N}\sum_{jt} \hat{\xi}_{jt}(\theta_2) Z_{jt}^D \\ \frac{1}{N}\sum_{jt} \hat{\omega}_{jt}(\theta_2) Z_{jt}^S \end{bmatrix}. \tag{13}$$

(g) Construct the GMM objective: $q(\theta_2) = g(\theta_2)'W g(\theta_2)$.

# Microdata

- Berry and Haile, 2022, Nonparametric Identification of Differentiated Products Demand Using Micro Data (arXiv:2204.06637)
  - Micro-data (consumer characteristic) provide variation in consumers' choice problems within a market ("a panel structure") i.e., without contamination from variation in the un-observables (fixed within a market)
  - Use this variation to learn about substitution patterns
  - Micro-data $\rightarrow$ richer demand specifications $\rightarrow$ reduces requirements on the number and types of IVs
- Conlon & Gortmaker, 2023, Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP, NBER WP 31605

# Appendix

# Generic random utility model

$$x_t = (x_{1t}, \ldots, x_{Jt}), p_t = (p_{1t}, \ldots, p_{Jt}), \xi_t = (\xi_{1t}, \ldots, \xi_{Jt})$$

$$\chi_t = (x_t, p_t, \xi_t)$$

matrix of all product & mkt characteristics

$$(v_{i1t}, \ldots, v_{iJt}) \sim F_v(\cdot|\chi_t)$$

conditional indirect utilities, $v_{iot} = 0$

- Very general random utility model:
  - $F_v(|)$ not derived from specification of utility
  - Restriction: $\xi_{jt}$ is scalar (jt-level unobservable)

# A1: The index restriction

$$\chi_t = (x_t, p_t, \xi_t)$$

$$x_{jt} = \left(x_{jt}^{(1)}, x_{jt}^{(2)}\right), \; x_{jt}^{(1)} \in \mathbb{R}$$

$$\delta_{jt} = x_{jt}^{(1)}\beta_j + \xi_{jt}, \quad \delta_t = (\delta_{1t}, \ldots, \delta_{Jt})$$

**Assumption 1** ("index") $F_v(\cdot|\chi_t) = F_v\left(\cdot|\delta_t, x_t^{(2)}, p_t\right)$

Restricts how $x_{jt}^{(1)}$ and $\xi_{jt}$ can affect the distribution of conditional indirect utilities. Note: $x_{jt}^{(1)}$ and $\xi_{jt}$ are "perfect substitutes".

Note also: The index restriction is what leaves $x^{(1)}$ out of from the inversion ($\sigma^{-1}{}_j$), making them available as instrument for shares (i.e., BLP-instruments can be excluded)

## …

$$s_{jt} = \sigma_j(\chi_t) = \sigma_j(\delta_t, p_t) = \Pr\left(\arg\max_{j \in \mathcal{J}} v_{ijt} = j \middle| \delta_t, p_t\right)$$ Utility maximization

$$s_{jt} = \sigma_j(\chi_t) \qquad j = 1, \ldots, J \qquad \chi_t = (x_t, p_t, \xi_t)$$ Demand system

- Note that ($s_t$ , $x_t$ , $p_t$) are observed: If each $\xi_{jt}$ were also observable, the functions ($\sigma_1$, . . . , $\sigma_J$) would be observable (trivial identification).
- Endogeneity because of $\xi_t \rightarrow$ challenge to nonparametric identification.
- More difficult than the usual case: Each $s_{jt}$ and $p_{jt}$ is a function of all J unobservables ($\xi_{1t}$ , . . . , $\xi_{Jt}$).
- Handling endogeneity harder with *multiple* structural errors in each equation