Demand Estimation

Huan Deng

April 29, 2025

Outline

Introduction & Motivation

Why Estimate Demand?

- Demand elasticities are crucial for analyzing:
  - Market power & markups
  - Welfare effects of policies (taxes, subsidies, tariffs)
  - Mergers, new product introductions
  - Consumer surplus, deadweight loss, etc.
- Many IO questions fundamentally require *quantitative* measures of:
  - How do changes in price or product characteristics affect quantity demanded?
  - How do we interpret these results in counterfactual simulations?
- Today, we discuss identification in differentiated products markets and estimation methods, following Berry & Haile (2016), Berry & Haile (2021), and NBER Summer Institute 2012 Methods Lectures by Nevo and Pakes.

Two Fundamental Challenges

1) Price Endogeneity:

- Equilibrium prices typically correlate with unobserved demand factors ξ.
- Example: If a product is especially appealing (high ξ), the producer may set a higher price.

2) Multiple Demand Shocks:

- Demand for a given good j depends on *all* goods' unobserved shocks.
- Need to hold $\xi_1, \ldots, \xi_J$ fixed to measure own-price elasticity for one good j.
- Hence, standard IV for *just* p_j is insufficient: we also need exogenous variation for the other goods' shocks or for the quantities themselves.

Demand Is Not Regression

- Multiple goods: $Q_j = D_j(X, p_1, \ldots, p_J,\ U_1, \ldots, U_J)$.
- Each Q_j depends on *all* the U_k.
- In typical regression $Y = f(X) + \varepsilon$, we only have one unobservable ε.
- Here, Y_j (or Q_j) must *somehow* hold $(U_1, \ldots, U_J)$ fixed for partial derivatives.
- Tools like standard OLS or basic 2SLS with a single instrument won't suffice.

Even With Exogenous Prices...

Hypothetical: Suppose we could randomize each product's price p_j across markets (like an experiment).

- This fixes correlation: $p_j \perp U_j$ (in expectation).
- But demand depends on $\mathbf{U} = (U_1, \ldots, U_J)$.
- Observed variation in Q_j with random p_j is *averaging over* $\mathbf{U}$, not holding them fixed.
- We only identify *average treatment effects* of price, not the *ceteris paribus* effect essential for elasticity or surplus.
- Must do more (function form or more exogeneity) to pin down actual *demand curves*.

Key insight: In a system with multiple structural errors, pure randomization of price is *not* enough for standard demand identification.

The Characteristic Space Approach

Motivation

- In Industrial Organization, modeling demand for *differentiated products* is critical.
- Two main conceptual frameworks:
  1. **Product space** (each product is a unique point)
  2. **Characteristic space** (products as bundles of attributes)
- Goal: compare these approaches and see why modern IO favors characteristic space.

Product Space: Key Features

Definition:

- Each product is treated as its own entity (e.g., brand dummies).
- Often leads to large parameter counts (one parameter per product or brand).

Challenges:

- **Parameter explosion:** With J products, we can end up with $\mathcal{O}(J)$ or even $\mathcal{O}(J^2)$ parameters.
- **Poor extrapolation:** Hard to predict demand for a *new* product not in the original set.
- **Interpretation:** Brand-level effects can mask *which attributes* matter to consumers.

Characteristic Space: Key Features

Definition:

- Each product is described by a vector of observable attributes, e.g. size, horsepower, flavor, etc.
- Consumer utility is a function of these attributes (and price).

Advantages:

- **Dimensionality reduction:** Instead of J product dummies, we estimate preferences for K attributes.
- **Predictive power:** Easy to analyze what happens when a new product is introduced (with known attribute vector).
- **Economic meaning:** Directly measures how consumers value each characteristic (fuel efficiency, screen size, etc.).
- **Flexible substitution:** Products with similar characteristics are closer substitutes.

Why the Shift? Historical Context

- **Lancaster (1966)**: Pioneered the idea that consumers value *attributes*, not the product itself.
- **Discrete choice demand** (e.g., Berry, Levinsohn, & Pakes 1995): Empirically implements characteristic-based utility in IO.
- Better data on product attributes made it possible to move away from simple "brand fixed effects."

Key result: Modern IO typically uses random-coefficients logit or similar characteristic-based approaches to capture rich substitution patterns and better interpret consumer preferences.

Core Differences in a Nutshell

Product Space

- High dimensional (product-level).
- Often one dummy or parameter *per* brand.
- Hard to handle large J.
- Poor fit for forecasting new products.
- Interpretation: brand "fixed effect" but no direct view of *why*.

Characteristic Space

- Lower dimensional (K attributes).
- Flexible substitution among similar products.
- Easier to predict new product's demand (given attributes).
- Parameters map onto real product features.
- More intuitive for welfare or policy analysis (attribute values can drive consumer surplus).

Example: Introducing a New Car Model

- **Product space approach:**
  - New car brand is a *completely new* dummy.
  - No direct way to link it with existing brands; might require a full re-estimation or ad hoc assumptions.
- **Characteristic space approach:**
  - Plug in new car's attributes (e.g. engine size, MPG, safety).
  - Predict how consumers switch from existing cars based on similarity in attributes.
  - More realistic, especially if new product overlaps with existing attribute distributions.

Common Critiques and Caveats

Characteristic Space Is Not a Panacea:

- Must observe (or instrument for) the key attributes driving utility.
- Attributes could be very high-dimensional (still a challenge).
- Unobserved product attributes create endogeneity problems (BLP solution uses instruments).
- For truly novel products (outside the range of existing characteristics), extrapolation remains hard.

Key Takeaways

1. **Product space** was a natural first approach but struggles with high dimensionality and poor out-of-sample predictions.
2. **Characteristic space** aligns better with consumer theory (utility from attributes).
3. Allows more parsimonious models, richer substitution patterns, and better forecasts for new products.
4. Has become the *standard* approach in modern IO demand estimation (e.g. Berry–Levinsohn–Pakes).

Canonical Model

Setup: Random Utility Model

Consumer i's utility for product j in market t:

$$u_{i,j,t} = v_{i,j}(x_{j,t}, p_{j,t}, \xi_{j,t},\ \varepsilon_{i,j,t}).$$

- Each consumer chooses the product with highest $u_{i,j,t}$.
- If $j = 0$, "outside good."
- Heterogeneity from:
  - $\varepsilon_{i,j,t}$: i.i.d. idiosyncratic (Type I EV often).
  - random coeffs in $v_{i,j}$: taste variation for x and p.
  - $\xi_{j,t}$: unobserved quality in the market-product dimension.
- $\implies$ *Aggregate shares* $s_{j,t}$ by integrating over consumer distribution.

Canonical Mixed Logit

$$u_{i,j,t} = \underbrace{x_{j,t}\,\beta_0 - \alpha_0\,p_{j,t} + \xi_{j,t}}_{\delta_{j,t}} + \underbrace{x_{j,t}\,\tilde{\beta}_i - \tilde{\alpha}_i \cdot p_{j,t}}_{\mu_{i,j,t}} + \varepsilon_{i,j,t}, \quad j > 0,$$

$$u_{i,0,t} = \varepsilon_{i,0,t}.$$

- $\delta_{j,t}$: mean utility, includes unknown $\xi_{j,t}$.
- $\mu_{i,j,t}$: random coefficient deviations (consumer-level).
- $\varepsilon_{i,j,t}$: Gumbel $\implies$ logit structure, but with a "mixed" layer from $\mu_{i,j,t}$.
- Market shares:

$$s_{j,t}(\theta) = \int \frac{\exp(\delta_{j,t} + \mu_{i,j,t})}{1 + \sum_{k=1}^{J} \exp(\delta_{k,t} + \mu_{i,k,t})}\, dF_\mu(\mu).$$

- Variation in β_i and α_i crucial for flexible substitution patterns.

Random Coefficients and Substitution

- Simple logit (no RC) imposes "Independence of Irrelevant Alternatives" (IIA).
- In reality: goods with similar x are stronger substitutes.
- Random coefficients let β_i vary so that:
  - Consumers who highly value "hybrid engine" also prefer other hybrids more strongly.
  - Cross-price elasticities can differ widely, even if shares are the same.
- Typically specify $\beta_i = \beta_0 + \Pi D_i + \Sigma \nu_i$ and $\ln(\alpha_i) = \alpha_0 + \alpha_y y_i + \alpha_\nu \nu_i^{(0)}$:
  - D_i = demographics, ν_i = taste shocks.
  - $\implies$ more flexible patterns of consumer choice and substitution.

Market-Level Data & Inversion

- Observed:

$$(s_{j,t},\ p_{j,t},\ x_{j,t}),\quad j = 1,\ldots,J_t,\ t = 1,\ldots,T.$$

- **BLP Key Insight (1995)**:
  - We denote coefficients in $\delta_{j,t}$ (θ_1) as "linear parameters", and coefficients governing individual heterogeneity (θ_2) as "nonlinear parameters"
  - There is a unique $\delta_{j,t}$ for each product so that

$$s_{j,t} = \sigma_j(\delta_t,\ x_t,\ \theta_2).$$

  - where $\delta_{j,t} = x_{j,t}\beta_0 - \alpha_0 p_{j,t} + \xi_{j,t}$.
  - Solve this *inversion* to recover $\xi_{j,t}(\theta)$ from observed shares $s_{j,t}$.
- If $\xi_{j,t}$ is correlated with $p_{j,t}$, we can't just regress $\delta_{j,t}$ on $(x_{j,t}, p_{j,t})$.
- **Hence GMM approach**: find $\hat{\theta}$ so that $\xi_{j,t}(\hat{\theta})$ is orthogonal to instruments $z_{j,t}$.

BLP GMM Steps (High-Level)

1. **Guess** a parameter vector θ.
2. **Invert**:
$$s_{j,t} = \sigma_j(\delta_t,\ x_t,\ \theta_2).$$
Solve for each $\delta_{j,t}$ s.t. model shares = observed shares.
3. **Compute** residual: $\hat{\xi}_{j,t} = \delta_{j,t} - x_{j,t}\beta_0 + \alpha_0\, p_{j,t}$.
4. **Moment Conditions**:
$$E[\hat{\xi}_{j,t}(\theta)\, z_{j,t}] = 0 \quad \Rightarrow \quad \text{GMM objective minimized.}$$
5. **Update guess** θ; repeat until GMM converges.

Result: $\hat{\theta}_{GMM}$ plus standard errors from typical GMM formula.

Identification

Identification

Index, Inversion and Instruments

Three Is

- To discuss the identification the BLP model, we need to discuss
  - Index: each demand shocks enters through an index for each good
  - Inversion: the presence of a one-to-one mapping between the indices and market shares, allowing inversion of the demand system
  - Instruments: provide exogenous variation in prices *and* shares, ensuring identification.

Partitioning $\mathbf{x}_t$ and Defining Indices δ_t

- Let
$$x_t = \left(x_t^{(1)},\ x_t^{(2)}\right), \quad x_t^{(1)} = \left(x_{1t}^{(1)}, \dots, x_{Jt}^{(1)}\right) \in \mathbb{R}^J.$$
- For each market t, define a vector of indices
$$\delta_t = \left(\delta_{1t}, \dots, \delta_{Jt}\right) \quad \text{where} \quad \delta_{jt} = x_{jt}^{(1)}\,\beta_j \ + \ \xi_{jt}.$$
- ξ_{jt} is the unobserved "demand shock" or unobserved characteristic for good j in market t.

Index

Index

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

- $\sigma_j(\cdot)$ is the nonparametric demand function for good j.
- **Index requirement**: $x_{jt}^{(1)}$ and ξ_{jt} affect demand *only* through the scalar δ_{jt}.
- This is a nonparametric functional form restriction:

$$(x_{jt}^{(1)}, \xi_{jt}) \to \delta_{jt} \quad \Rightarrow \quad \sigma_j \text{ depends on } \delta_{jt} \text{ plus } x_t^{(2)}, p_t.$$

- Meanwhile, $x_t^{(2)}$ and p_t can freely enter σ_j in a fully flexible (nonparametric) way.

Connected Substitutes

Connected Substitutes condition is key to demand inversion.

Part (i): Weak Substitutes

- For all $j > 0$, $k \neq j$, and any (δ, p) in the support:

  if δ_j increases, $\sigma_k(\delta, p)$ does not increase (weakly decreasing).

- **Interpretation**: an improvement in good j's index does not raise demand for other goods k.
- Automatic in standard discrete choice: higher δ_j means higher (j's "quality"), so k's share typically falls.

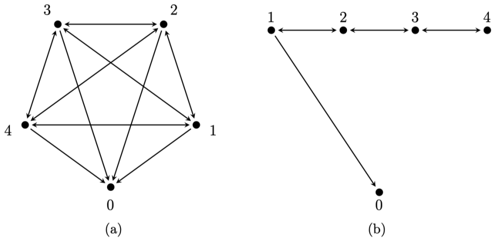
Part (ii): Connected Strict Substitution

- For each (δ, p) and any nonempty $K \subset \{1, \ldots, J\}$, there exists $k \in K$ and $\ell \notin K$ s.t. $\sigma_\ell(\delta, p)$ is strictly decreasing in δ_k.
- **Meaning**: from each good k, there's *some* strict substitution path to the outside good 0.

Interpretation of Connected Substitutes

- **Weak Substitutes**: ensures that δ_j acts as a "quality index" for j, so raising it lowers others' demands (or leaves them unchanged).
- **Connectedness**: ensures no strict subset of goods that substitute *only* among themselves, excluding the outside good.
- **Graph Perspective**:
    - Represent each good $j > 0$ as a vertex.
    - Draw a directed edge $j \to k$ if an increase in δ_j strictly lowers demand for k.
    - Condition: from every good $j > 0$, there is a directed path eventually leading to the outside good 0.

Directed Graphs for Substitution (from Berry and Haile (2021))



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.

Demand Inversion Under Connected Substitutes

- **Invertibility**:
  - For any vector of market shares $s = (s_1, \ldots, s_J)$ with $s_j > 0$, there is a unique δ s.t.

$$s_j = \sigma_j(\delta, p) \quad \forall j.$$

  - Not just local invertibility but a *global* one-to-one mapping, thanks to connected structure.
- **Implication**: we can define

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

- **Berry, Gandhi, & Haile (2013)**:
  - This invertibility extends to a wide range of demand forms, even if goods have complementarity in some local sense, as long as the net effect satisfies connected substitutes.

Why Does Inversion Matter?

- **Identifying δ_j:**

$$\delta_j \; = \; x_j + \xi_j \quad \Longrightarrow \quad \xi_j = \delta_j - x_j.$$

- **Unobserved quality** ξ_j is thus recovered from data (s_j, p_j, x_j) if demand is invertible.
- After that, estimation uses *instrumental variables* to solve endogeneity: $E[\xi_j \mid z_j] = 0$.
- **Crucial Step**: "One error per equation" $\Longrightarrow$ standard IV arguments become valid.

Why Instrument for Both Price *and* Shares?

- Typical question: "Isn't it enough to instrument for $p_{j,t}$?"
- But the *random coefficient* part requires identifying substitution across products.
- Variation in *shares* $s_{j,t}$ is also endogenous (depends on $\xi_{k,t}$ for all k).
- We need exogenous variation to uncover how changes in (p, x) re-allocate market shares across goods.
- BLP instruments handle that: x_{-j} affects good j's share even if p_j is held fixed, letting us learn about cross-substitution.
- Hence, dimension of valid IVs typically $\geq 2J$ (for J inside goods).
- Below we discuss four sets instruments used in BLP models

Cost Shifters & Proxies

- **Exogeneity assumption**: cost variables uncorrelated with demand shock $\xi_{j,t}$.
- **Examples**:
  - Input prices (e.g. wages, raw materials).
  - Taxes, tariffs, exchange rates.
- **Proxies**: noisy measures correlated with firm-level cost.
  - E.g. average local wage vs. precise labor cost measure.
  - Valid if proxy error is uncorrelated with $\xi_{j,t}$.
- **Relevance**: cost changes must pass through to equilibrium prices.

BLP Instruments (Competitors' Characteristics)

- Key **insight** (Berry, Levinsohn, Pakes, 1995):

$$\text{Exogenous characteristics of rival products} \;\; (x_{-j,t})$$

  shift good j's markup *and* share.

- **Logic**:
  - Rival products' $x_{k,t}$ affect j's residual demand $\rightarrow$ markup.
  - As long as $x_{k,t}$ are mean-independent of $\xi_{j,t}$, they serve as valid instruments.

- **Dimension**:
  - Often need many instruments to handle multi-product environment.
  - BLP instruments can strongly shift both prices *and* quantity shares.

Hausman Instruments

- **Idea**: same good's price in another market $p_{j,t'}$ can reflect cost variation if:
  - Common marginal cost shocks across markets.
  - Demand shocks $\xi_{j,t}$ are local and uncorrelated with $p_{j,t'}$.
- **Validity** depends on:
  - No correlation between $\xi_{j,t}$ and $\xi_{j,t'}$ across markets t, t'.
  - Or that exogenous cost shocks are principal driver of $p_{j,t'}$ variation.
- Potential **concern**: if demand shocks vary systematically across markets, $p_{j,t'}$ may correlate with $\xi_{j,t}$.

Waldfogel-Fan Instruments

- Another approach for exogenous variation:
  - **Demographics** in other "linked" markets or regions that share a pricing zone.
- **Logic**:
  - If pricing is done at regional or zone level:

$$p_{j,t} \text{ depends on } \sum_{\ell \in \text{zone}} \text{market} \ell\text{-level demand variables.}$$

  - Variation in *other* markets' demographics can shift $p_{j,t}$ but not $\xi_{j,t}$.
- **Same excludability caution**: cross-market correlation in unobservables can compromise validity.

Identification

Micro Data and Identification Gains

Market- vs. Consumer-Level Data

Market-Level Data:

- Observations consist of market shares or total quantity sold for each product j in each market t.
- Also observe prices $p_{j,t}$, product characteristics $x_{j,t}$, possibly distribution of demographics.
- Typical advantage: Easier to collect large cross-sections of markets.
- Typical disadvantage: Less clarity on individual substitution patterns; potential for "endogeneity" of price.

Consumer-Level Data:

- Observe each consumer i's chosen product j, along with consumer attributes D_i (income, etc.).
- Potential advantage: Clearer identification of heterogeneity in tastes and better control of endogeneity.
- Potential disadvantage: Costly data collection, possible sample-selection issues.

Why Micro Data Helps

- With *consumer-level* choice data:
  - Within a single market t, $p_{j,t}$, $\xi_{j,t}$ are fixed for all i.
  - Variation in consumer attributes $z_{i,t}$ reveals heterogeneity in $\mu_{i,j,t}$.
- This internal variation helps identify random coefficients without needing as many external instruments for the share side.
- E.g. if we see how individuals with different income or family size shift choices among goods with different x_j, we learn about substitution patterns.
- $\implies$ fewer external instruments needed, often only for *price* $p_{j,t}$.

Estimation

Mixed Logit Setup

- We have briefly discussed estimation in our previous discussion of identification. Here we offer details on implementing a Mixed Logit model using micro data.
- Let the *utility* for person i choosing product j in market t be:

$$U_{ijt} = V(p_{jt}, x_{jt}, D_i, \beta_i) + \xi_{jt} + \varepsilon_{ijt}.$$

- Separate $V(\cdot)$ into a *common* part $\bar{V}(\cdot)$ and an *individual-specific* part $\tilde{V}(\cdot)$:

$$U_{ijt} = \underbrace{\bar{V}(p_{jt}, x_{jt}, \bar{\beta}) + \xi_{jt}}_{\delta_{jt}} + \underbrace{\tilde{V}(p_{jt}, x_{jt}, D_i, \tilde{\beta}_i) + \varepsilon_{ijt}}_{\text{varies over } i}.$$

- Define $\delta_{jt} = \bar{V}(p_{jt}, x_{jt}, \bar{\beta}) + \xi_{jt}$.

Mixed Logit Form

- If ε_{ijt} is i.i.d. extreme value, then choice probabilities are Mixed Logit:

$$P_{ijt} = \int \left[\frac{\exp\left(\delta_{jt} + \tilde{V}(p_{jt}, x_{jt}, D_i, \tilde{\beta}_i\right)}{\sum_{j'} \exp\left(\delta_{j't} + \tilde{V}(p_{j't}, x_{j't}, D_i, \tilde{\beta}_i)\right)} \right] f(\tilde{\beta}_i \mid \theta_2)\, d\tilde{\beta}_i.$$

- We can write out the likelihood function and estimate $\{\theta_1, \delta\}$ altogether
- However, if the number of δ_{jt} is too many, then optimization would be hard, we can use the below contraction mapping for dimension reduction.

The Contraction Mapping

- For each trial value of parameters θ:

  Find $\delta(\theta)$ such that the predicted shares match actual shares.

- We update:

$$\delta_{jt}^{(t+1)} = \delta_{jt}^{(t)} + \ln\left(\frac{S_{jt}}{\hat{S}_{jt}(\delta^{(t)}, \theta_2)}\right),$$

- This is called *the contraction* and guarantees a unique δ that solves $\hat{S}_{jt}(\delta, \theta_2) = S_{jt}$.
- After convergence, $\delta(\theta)$ is fully determined by θ_2.

Instrumental Variables Regression for δ_{jm}

- The constants δ_{jt} contain price p_{jt} and the unobserved term ξ_{jt}:
$$\delta_{jm} \;=\; \bar{\beta}' v(p_{jt}, x_{jt}) \;+\; \xi_{jm}.$$
- This is a linear regression in δ_{jt}
- Price p_{jm} is endogenous (p_{jt} correlated with ξ_{jt}), so estimate by *instrumental variables* (IV).

More on Estimation

- After inverting the demand system, estimation is straightforward: using moment conditions generated by IVs.
- GMM is a natural choice.
- In the inner loop, we solve for mean utilities for trial value of nonlinear parameters. This nested approach might be time consuming, can consider MPEC by Dube, Fox and Su (2012)
- The devil is in the details. For best practice in implemention, please see Conlon and Gortmaker (2020) and their PyBLP package.