ECN 594: Logit Demand and Identification

Nicholas Vreugdenhil

January 4, 2026

Announcements

- **Homework 1 released today**
- Due: Feb 4 (before Lecture 6)
- Demand estimation using Python and `pyblp`
- Start early!

Plan for today

1. Logit model derivation
2. Berry (1994) inversion
3. Elasticity formulas
4. IIA problem (preview)

---

5. The identification problem
6. Direction of bias
7. Instrumental variables
8. Uber example

Part 1: The Logit Demand Model

Recap: Random utility

- Consumer i chooses among J products
- Utility:

$$u_{ij} = x_j \beta - \alpha p_j + \xi_j + \varepsilon_{ij}$$

- Consumer chooses the product with highest utility
- Last time: we left ε_{ij} unspecified
- Today: we assume ε_{ij} has a specific distribution

The logit assumption

- Assume ε_{ij} is i.i.d. **Type I Extreme Value**
- Also called Gumbel distribution
- CDF: $F(\varepsilon) = \exp(-\exp(-\varepsilon))$
- Why this assumption?
  - Gives us **closed-form** choice probabilities!
  - Computationally tractable

Logit choice probabilities

- Define **mean utility**:

$$\delta_j = x_j\beta - \alpha p_j + \xi_j$$

- So utility is: $u_{ij} = \delta_j + \varepsilon_{ij}$

- With Type I Extreme Value errors, the probability that consumer chooses j:

$$P(\text{choose } j) = \frac{\exp(\delta_j)}{\sum_{k=1}^{J} \exp(\delta_k)}$$

- This is the **logit** formula

The outside option

- Problem: Our formula doesn't allow consumers to "not buy"
- We need an **outside option** (product $j = 0$)
- Utility of outside option:

$$u_{i0} = \varepsilon_{i0}$$

- We normalize: $\delta_0 = 0$
- All other utilities are *relative* to this outside option

Logit with outside option

- With the outside option, the share of product j is:

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$

- And the share of the outside option is:

$$s_0 = \frac{1}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$

- Note: $s_0 + \sum_{j=1}^{J} s_j = 1$ (shares sum to 1)

Berry (1994) inversion: the key insight

- We observe: market shares s_j
- We want: mean utilities δ_j (to estimate β, α)
- **Problem:** How do we get δ_j from s_j?
- **Berry's insight:** Take the log of shares!

Berry (1994) inversion

- Start with:

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)}, \quad s_0 = \frac{1}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$

- Take logs:

$$\ln(s_j) = \delta_j - \ln\left(1 + \sum_{k=1}^{J} \exp(\delta_k)\right)$$
$$\ln(s_0) = -\ln\left(1 + \sum_{k=1}^{J} \exp(\delta_k)\right)$$

- Subtract:

$$\ln(s_j) - \ln(s_0) = \delta_j$$

Berry (1994) inversion: the estimating equation

- We have: $\ln(s_j) - \ln(s_0) = \delta_j$
- Substitute $\delta_j = x_j\beta - \alpha p_j + \xi_j$:

$$\ln(s_j) - \ln(s_0) = x_j\beta - \alpha p_j + \xi_j$$

- This is a **linear regression**!
- LHS: can compute from observed shares
- RHS: product characteristics, price, and an error term

Logit elasticities

- Given shares $s_j = \frac{\exp(\delta_j)}{1+\sum_k \exp(\delta_k)}$
- We can derive price elasticities:

$$\eta_{jj} = \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = -\alpha p_j(1 - s_j) \quad \text{(own-price)}$$

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \alpha p_k s_k \quad \text{(cross-price)}$$

- Note: $\alpha > 0$, so own-price elasticity is **negative** (as expected)

Worked example: Logit elasticities

- **Question:**
- Suppose $\alpha = 0.5$, product j has price $p_j = 20$ and market share $s_j = 0.1$.
- Compute the own-price elasticity for product j.

Take 2 minutes to solve this.

Worked example: Logit elasticities (solution)

- Own-price elasticity formula:

$$\eta_{jj} = -\alpha p_j (1 - s_j)$$

- Plug in: $\alpha = 0.5$, $p_j = 20$, $s_j = 0.1$

$$\begin{aligned} \eta_{jj} &= -0.5 \times 20 \times (1 - 0.1) \\ &= -0.5 \times 20 \times 0.9 \\ &= -9 \end{aligned}$$

- Interpretation: A 1% price increase $\Rightarrow$ 9% decrease in quantity

Worked example: Cross-price elasticity

- Now compute the cross-price elasticity with product k

- Given: $\alpha = 0.5$, $p_k = 25$, $s_k = 0.05$

- Cross-price elasticity formula:

$$\eta_{jk} = \alpha p_k s_k$$

- Plug in:

$$\begin{aligned}\eta_{jk} &= 0.5 \times 25 \times 0.05 \\ &= 0.625\end{aligned}$$

- Interpretation: A 1% increase in $p_k \Rightarrow 0.625\%$ increase in s_j

The IIA problem (preview)

- Look at the cross-price elasticity again:

$$\eta_{jk} = \alpha p_k s_k$$

- This doesn't depend on product j at all!
- Implication: All products have the **same** cross-elasticity with product k
- Is this realistic?
- Suppose BMW raises its price. Logit says: same fraction go to Mercedes as to Honda Civic!
- This is the **IIA** (Independence of Irrelevant Alternatives) property
- We'll discuss this in detail in Lecture 4

Part 2: Identification and Instrumental Variables

The estimating equation (reminder)

- From Berry inversion:

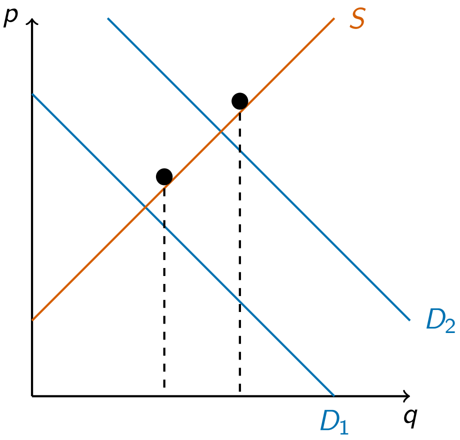
$$\ln(s_j) - \ln(s_0) = x_j\beta - \alpha p_j + \xi_j$$

- This looks like a regression we can run with OLS
- **But there's a problem...**

The classic identification problem

- We observe equilibrium prices and quantities

- Problem: can't tell if demand shifted or supply shifted



The classic identification problem

- If we just regress q on p, what do we get?
- Neither demand nor supply!
- **Key insight:** We need variation that shifts ONE curve but not the other
  - To identify demand: need **supply shifters**
  - To identify supply: need **demand shifters**

Price endogeneity in demand estimation

- Our estimating equation:

$$\ln(s_j) - \ln(s_0) = x_j\beta - \alpha p_j + \xi_j$$

- ξ_j = unobserved product quality
- **Problem:** Firms observe ξ_j when setting prices!
- High quality products (ξ_j high) tend to have high prices
- $\Rightarrow \text{Cov}(p_j, \xi_j) > 0$
- OLS gives biased estimates

Direction of bias

- If high-quality products have high prices...
- OLS sees: high price, but demand still high (because of ξ)
- OLS concludes: price doesn't hurt demand much
- **Result:** $\hat{\alpha}$ biased toward zero (less negative than truth)

Worked example: Bias direction

- **Question:** You estimate a logit demand model using OLS and get $\hat{\alpha} = -0.3$. A colleague says the true α is likely -0.5.

- Is this consistent with endogeneity bias? Why?

Worked example: Bias direction

- **Question:** You estimate a logit demand model using OLS and get $\hat{\alpha} = -0.3$. A colleague says the true α is likely -0.5.
- Is this consistent with endogeneity bias? Why?
- **Answer:** Yes!
  - OLS overstates how much consumers like expensive products
  - So OLS finds a smaller (less negative) price coefficient
  - $-0.3 > -0.5$, so this is exactly what we'd expect

The gold standard: Experiments

- Best solution: randomize prices!
- **Uber's price "wiggles"** (Cohen et al. 2016)
  - Uber experimentally varies surge multipliers up and down
  - Same time, same location $\rightarrow$ different riders see different prices
  - This randomization creates exogenous price variation
- Key insight: demand conditions are identical, only price differs
- Result: demand elasticity ≈ -0.5 (inelastic!)

Why experiments are powerful

- No confounding: price variation is independent of demand shocks
- Clean identification of the demand curve
- **But:**
  - Tech companies can run experiments
  - Traditional industries can't randomize prices
  - Most IO settings require **instrumental variables**

Instrumental variables: the solution

- Need variables z that are:
    1. **Relevant:** Correlated with price ($\text{Cov}(z, p) \neq 0$)
    2. **Exogenous:** Uncorrelated with ξ ($\text{Cov}(z, \xi) = 0$)
- These are cost shifters or other supply-side variables

Common IVs in demand estimation

1. **Hausman IVs:** Prices in other markets
   - Same product in different cities has similar costs
   - But demand shocks may differ across markets
2. **BLP IVs:** Characteristics of competing products
   - More/different competitors $\rightarrow$ lower prices
   - Competitors' characteristics don't affect YOUR ξ
3. **Cost shifters:** Input prices, exchange rates
   - Affect production costs, hence prices
   - No direct effect on demand

Worked example: IV intuition

- **Question:** Why do competitor characteristics work as IVs?

Worked example: IV intuition

- **Question:** Why do competitor characteristics work as IVs?
- **Answer:**
- **Relevance:** More competitors nearby $\rightarrow$ more competition $\rightarrow$ lower price ✓
- **Exogeneity:** Competitor characteristics don't affect YOUR unobserved quality ξ_j ✓
- Example: If Toyota enters with a new Camry, this affects Civic's price but not Civic's unobserved quality

Worked example: Evaluating an IV

- **Question:** A researcher proposes using gasoline prices as an IV for car prices. Is this valid?

Worked example: Evaluating an IV

- **Question:** A researcher proposes using gasoline prices as an IV for car prices. Is this valid?
- **Relevance:** Higher gas prices $\rightarrow$ higher operating costs $\rightarrow$ might affect car prices? Maybe weakly.
- **Exogeneity:** Do gas prices affect car quality ξ_j?
  - Gas prices affect *demand* for fuel-efficient cars
  - This might shift which cars look "good" to consumers
  - Potentially problematic!
- **Verdict:** Probably not a great IV

Summary: IV conditions

- For z to be a valid IV:
  1. **Relevant:** z must predict prices
     - Can test this! Run first-stage regression
  2. **Exogenous:** z must not affect demand directly
     - Cannot test this directly (requires economic reasoning)
- This is the standard IV framework from econometrics
- Applied to demand estimation: use supply-side variation

Key Points

1. **Logit model**: ε_{ij} is Type I Extreme Value $\rightarrow$ closed-form shares
2. **Share equation**: $s_j = \frac{\exp(\delta_j)}{1+\sum_k \exp(\delta_k)}$
3. **Berry inversion**: $\ln(s_j) - \ln(s_0) = \delta_j$ turns demand estimation into a regression
4. **Elasticities**: Own $= -\alpha p_j(1 - s_j)$; Cross $= \alpha p_k s_k$
5. **IIA**: Cross-elasticities don't depend on product similarity (a limitation)
6. **Price is endogenous**: Firms observe ξ_j when pricing $\rightarrow$ $\text{Cov}(p, \xi) > 0$
7. **Bias direction**: OLS gives $\hat{\alpha}$ biased toward zero
8. **Solution**: Instrumental variables (cost shifters, BLP IVs, Hausman IVs)

Next time

- **Lecture 3:** Demographic Interactions and pyblp
  - Extending logit to allow preference heterogeneity
  - Estimation using pyblp package
  - Worked example with car data