Demand Estimation

Mixtape Session

Jeff Gortmaker and Ariel Pakes

Who Am I?

- A fifth-year Economics PhD candidate at Harvard University.

Who Am I?

- A fifth-year Economics PhD candidate at Harvard University.
- Making BLP-style estimation more accessible to researchers.
  - → Best practices papers (Conlon and Gortmaker, 2020, 2023).
  - → Open-source Python package (PyBLP).
  - → This course!

This Course

- Three days, 6pm-9pm.
  1. Today: BLP model, pure logit, price endogeneity.
  2. Wednesday: Mixed logit, identification, numerical best practices.
  3. Friday: Micro BLP, consumer survey data, other extensions.

This Course

- Three days, 6pm-9pm.
    1. Today: BLP model, pure logit, price endogeneity.
    2. Wednesday: Mixed logit, identification, numerical best practices.
    3. Friday: Micro BLP, consumer survey data, other extensions.

- Please participate and ask questions!
    $\rightarrow$ Ask questions in the Discord chat or raise your hand on Zoom.

This Course

- Three days, 6pm-9pm.
  1. Today: BLP model, pure logit, price endogeneity.
  2. Wednesday: Mixed logit, identification, numerical best practices.
  3. Friday: Micro BLP, consumer survey data, other extensions.

- Please participate and ask questions!
  → Ask questions in the Discord chat or raise your hand on Zoom.

- Three coding exercises, one after each day.
  → Try these on your own or with your classmates' help. Use Discord rooms!
  → I'll do the first two exercises live at the start of days 2 and 3. We'll post solutions.

Readings

- There are a lot of possible references for how to do BLP-style estimation.

Readings

- There are a lot of possible references for how to do BLP-style estimation.
- Modern guides:
  1. Berry and Haile (2021)
  2. Conlon and Gortmaker (2020)
  3. Conlon and Gortmaker (2023)

Readings

- There are a lot of possible references for how to do BLP-style estimation.
- Modern guides:
  1. Berry and Haile (2021)
  2. Conlon and Gortmaker (2020)
  3. Conlon and Gortmaker (2023)
- Foundational papers:
  1. Berry, Levinsohn and Pakes (1995)
  2. Nevo (2000)
  3. Petrin (2002)
  4. Berry, Levinsohn and Pakes (2004)

Readings

- There are a lot of possible references for how to do BLP-style estimation.
- Modern guides:
  1. Berry and Haile (2021)
  2. Conlon and Gortmaker (2020)
  3. Conlon and Gortmaker (2023)
- Foundational papers:
  1. Berry, Levinsohn and Pakes (1995)
  2. Nevo (2000)
  3. Petrin (2002)
  4. Berry, Levinsohn and Pakes (2004)
- None of these are required for the course, but I recommend taking a look afterwards.

Running Example

- BLP can be used to better understand all sorts of decisions.
  - → Product purchases, hospital visits, school choice, voting behavior, etc.

Running Example

- BLP can be used to better understand all sorts of decisions.
  - $\rightarrow$ Product purchases, hospital visits, school choice, voting behavior, etc.
- Typically used for counterfactual analysis of something that hasn't happened.
  - $\rightarrow$ Need a structural model when we can't just estimate a treatment effect.

Running Example

- BLP can be used to better understand all sorts of decisions.
  - → Product purchases, hospital visits, school choice, voting behavior, etc.
- Typically used for counterfactual analysis of something that hasn't happened.
  - → Need a structural model when we can't just estimate a treatment effect.
- Running example: What if we halved an important product's price?
  - → Practitioners: Increased sales vs. cannibalization?
  - → Regulators: Revenue loss from eliminating a tax?
  - → Academics: Welfare consequences?

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Coding Exercise 1

Model Overview

- Model of individuals making a discrete choice from different alternatives.
  - $\rightarrow$ Original BLP (1995) also modeled firm price-setting. We'll focus on demand.

Model Overview

- Model of individuals making a discrete choice from different alternatives.
  - $\rightarrow$ Original BLP (1995) also modeled firm price-setting. We'll focus on demand.
- Choices are made in markets denoted by $t \in \mathcal{T}$.
  - $\rightarrow$ Time periods, geographic regions, etc.

Model Overview

- Model of individuals making a discrete choice from different alternatives.
  - $\rightarrow$ Original BLP (1995) also modeled firm price-setting. We'll focus on demand.
- Choices are made in markets denoted by $t \in \mathcal{T}$.
  - $\rightarrow$ Time periods, geographic regions, etc.
- Each market has individuals with types denoted by $i \in \mathcal{I}_t$.
  - $\rightarrow$ Different demographics and preferences.

Model Overview

- Model of individuals making a discrete choice from different alternatives.
  - $\rightarrow$ Original BLP (1995) also modeled firm price-setting. We'll focus on demand.
- Choices are made in markets denoted by $t \in \mathcal{T}$.
  - $\rightarrow$ Time periods, geographic regions, etc.
- Each market has individuals with types denoted by $i \in \mathcal{I}_t$.
  - $\rightarrow$ Different demographics and preferences.
- Individuals are faced with choices denoted by $j \in \mathcal{J}_t$.
  - $\rightarrow$ Products, hospitals, candidates, etc.
  - $\rightarrow$ Outside option $j = 0$: no purchase, no treatment, no vote, etc.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  - $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.
- Will help to decompose utility into three parts.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  - $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.

- Will help to decompose utility into three parts.
  1. Mean utility δ_{jt}: Average preference across all individuals in the market.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  - $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.

- Will help to decompose utility into three parts.
  1. Mean utility δ_{jt}: Average preference across all individuals in the market.
  2. Systematic heterogeneity μ_{ijt}: Different preferences, e.g. due to different demographics.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.

- Will help to decompose utility into three parts.
  1. Mean utility δ_{jt}: Average preference across all individuals in the market.
  2. Systematic heterogeneity μ_{ijt}: Different preferences, e.g. due to different demographics.
  3. Idiosyncratic heterogeneity ε_{ijt}: Superimposed noise that accommodates estimation.

Utility Maximization

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Individuals choose an alternative to maximize (indirect) utility u_{ijt}.
  - $\rightarrow$ We will specify a function for u_{ijt} and use revealed preference to estimate it.

- Will help to decompose utility into three parts.
  1. Mean utility δ_{jt}: Average preference across all individuals in the market.
  2. Systematic heterogeneity μ_{ijt}: Different preferences, e.g. due to different demographics.
  3. Idiosyncratic heterogeneity ε_{ijt}: Superimposed noise that accommodates estimation.

- We will parameterize δ_{jt} and μ_{ijt} and make a convenient assumption about ε_{ijt}.

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \qquad \Longrightarrow \qquad s_{ijt} = \mathbb{P}_{\varepsilon_{it}}\Big(u_{ijt} \geq u_{ikt} \text{ for all } k \in \mathcal{J}_t \cup \{0\}\Big)$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.
  - $\rightarrow$ "Logit shocks" are convenient because they give multinomial logit choice probabilities s_{ijt}.

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \qquad \Longrightarrow \qquad s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{k \in \mathcal{J}_t \cup \{0\}} \exp(\delta_{kt} + \mu_{ikt})}$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.
  - $\rightarrow$ "Logit shocks" are convenient because they give multinomial logit choice probabilities s_{ijt}.

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{k \in \mathcal{J}_t \cup \{0\}} \exp(\delta_{kt} + \mu_{ikt})}$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.
  - $\rightarrow$ "Logit shocks" are convenient because they give multinomial logit choice probabilities s_{ijt}.
- Want μ_{ijt} to be sufficiently flexible that this convenient assumption matters little.
  - $\rightarrow$ Possible to eliminate ε_{ijt} but computation gets difficult (Berry and Pakes, 2007).

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{k \in \mathcal{J}_t \cup \{0\}} \exp(\delta_{kt} + \mu_{ikt})}$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.
  - $\rightarrow$ "Logit shocks" are convenient because they give multinomial logit choice probabilities s_{ijt}.
- Want μ_{ijt} to be sufficiently flexible that this convenient assumption matters little.
  - $\rightarrow$ Possible to eliminate ε_{ijt} but computation gets difficult (Berry and Pakes, 2007).
- Each type i is a share w_{it} of the population. Aggregating over them gives market shares.

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot s_{ijt}$$

Aggregate Market Shares

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{k \in \mathcal{J}_t \cup \{0\}} \exp(\delta_{kt} + \mu_{ikt})}$$

- Assume a convenient distribution for ε_{ijt}: iid type I extreme value.
  - $\rightarrow$ "Logit shocks" are convenient because they give multinomial logit choice probabilities s_{ijt}.
- Want μ_{ijt} to be sufficiently flexible that this convenient assumption matters little.
  - $\rightarrow$ Possible to eliminate ε_{ijt} but computation gets difficult (Berry and Pakes, 2007).
- Each type i is a share w_{it} of the population. Aggregating over them gives market shares.

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot s_{ijt}$$

- We'll match these to observed quantities $q_{jt} = s_{jt} \cdot M_t$ in our data.

Choosing a Market Size

- In our data, we observe quantities $q_{jt} = s_{jt} \cdot M_t$.
  - $\rightarrow$ Need to divide by some market size M_t to get our model's market shares s_{jt}.
  - $\rightarrow$ Issue here is that we often don't observe the quantity of outside choices q_{0t}.

Choosing a Market Size

- In our data, we observe quantities $q_{jt} = s_{jt} \cdot M_t$.
  - $\rightarrow$ Need to divide by some market size M_t to get our model's market shares s_{jt}.
  - $\rightarrow$ Issue here is that we often don't observe the quantity of outside choices q_{0t}.
- Sometimes the choice of market size is straightforward.
  - $\rightarrow$ Market size for drugs to treat a condition is how many people have that condition.

Choosing a Market Size

- In our data, we observe quantities $q_{jt} = s_{jt} \cdot M_t$.
  - $\rightarrow$ Need to divide by some market size M_t to get our model's market shares s_{jt}.
  - $\rightarrow$ Issue here is that we often don't observe the quantity of outside choices q_{0t}.
- Sometimes the choice of market size is straightforward.
  - $\rightarrow$ Market size for drugs to treat a condition is how many people have that condition.
- But typically, the choice of market size is neither easy nor innocuous.
  - $\rightarrow$ E.g. how many choices of which cereal to buy are made every day in a specific city?
  - $\rightarrow$ Population $\times$ max cereals per day? Foot traffic estimate $\times$ max cereals per trip?

Choosing a Market Size

- In our data, we observe quantities $q_{jt} = s_{jt} \cdot M_t$.
  - $\rightarrow$ Need to divide by some market size M_t to get our model's market shares s_{jt}.
  - $\rightarrow$ Issue here is that we often don't observe the quantity of outside choices q_{0t}.
- Sometimes the choice of market size is straightforward.
  - $\rightarrow$ Market size for drugs to treat a condition is how many people have that condition.
- But typically, the choice of market size is neither easy nor innocuous.
  - $\rightarrow$ E.g. how many choices of which cereal to buy are made every day in a specific city?
  - $\rightarrow$ Population $\times$ max cereals per day? Foot traffic estimate $\times$ max cereals per trip?
- You should try different assumptions and see how they change your results.
  - $\rightarrow$ In general, the bigger the market size, the more substitution to the outside good.
  - $\rightarrow$ We'll learn how to discipline these assumptions with data on day 3.

Identification and Normalizations

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- We will estimate our utility function with revealed preference.
  - $\rightarrow$ Holding μ_{ijt} fixed, a higher quantity $q_{jt} > q_{kt}$ implies a higher mean utility $\delta_{jt} > \delta_{kt}$.

Identification and Normalizations

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- We will estimate our utility function with revealed preference.
  - $\rightarrow$ Holding μ_{ijt} fixed, a higher quantity $q_{jt} > q_{kt}$ implies a higher mean utility $\delta_{jt} > \delta_{kt}$.

- Utility is invariant to positive affine transformations. Need two normalizations.

$$u_{ijt} > u_{ikt} \quad \overset{b>0}{\Longleftrightarrow} \quad a + b \cdot u_{ijt} > a + b \cdot u_{ikt}$$

Identification and Normalizations

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- We will estimate our utility function with revealed preference.
  - $\rightarrow$ Holding μ_{ijt} fixed, a higher quantity $q_{jt} > q_{kt}$ implies a higher mean utility $\delta_{jt} > \delta_{kt}$.

- Utility is invariant to positive affine transformations. Need two normalizations.
  a. Level: We will normalize $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{jt} = \mu_{ijt} = 0$
     $\implies$ Estimates are relative to outside option utility.

Identification and Normalizations

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- We will estimate our utility function with revealed preference.
  - $\rightarrow$ Holding μ_{ijt} fixed, a higher quantity $q_{jt} > q_{kt}$ implies a higher mean utility $\delta_{jt} > \delta_{kt}$.
- Utility is invariant to positive affine transformations. Need two normalizations.
  a. Level: We will normalize $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{jt} = \mu_{ijt} = 0$

     $\implies$ Estimates are relative to outside option utility.

  b. Scale: We already normalized $\mathbb{V}(\varepsilon_{ijt}) = \pi^2/6$ when deriving choice probabilities.

     $\implies$ Estimates are relative to scale of noise.

Identification and Normalizations

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- We will estimate our utility function with revealed preference.
  - $\rightarrow$ Holding μ_{ijt} fixed, a higher quantity $q_{jt} > q_{kt}$ implies a higher mean utility $\delta_{jt} > \delta_{kt}$.

- Utility is invariant to positive affine transformations. Need two normalizations.
  a. Level: We will normalize $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{jt} = \mu_{ijt} = 0$

     $\implies$ Estimates are relative to outside option utility.

  b. Scale: We already normalized $\mathbb{V}(\varepsilon_{ijt}) = \pi^2/6$ when deriving choice probabilities.

     $\implies$ Estimates are relative to scale of noise.

- Now that our model can in theory be identified, how do we estimate it?

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Coding Exercise 1

Pure Logit Model

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \cancelto{0}{\mu_{ijt}} + \varepsilon_{ijt}$$

- Start with the simplest case: no heterogenous utility. We'll add μ_{ijt} back on day 2.

Pure Logit Model

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \varepsilon_{ijt} \qquad \Longrightarrow \qquad s_{jt} = \frac{\exp \delta_{jt}}{\sum_{k \in \mathcal{J}_t \cup \{0\}} \exp \delta_{kt}}$$

- Start with the simplest case: no heterogenous utility. We'll add μ_{ijt} back on day 2.
- Market shares simplify. No aggregation over individual types.

Pure Logit Model

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \varepsilon_{ijt} \qquad \Longrightarrow \qquad s_{jt} = \frac{\exp \delta_{jt}}{1 + \sum_{k \in \mathcal{J}_t} \exp \delta_{kt}}$$

- Start with the simplest case: no heterogenous utility. We'll add μ_{ijt} back on day 2.
- Market shares simplify. No aggregation over individual types.
  - $\rightarrow$ The 1 in the denominator is from our level normalization $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{jt} = 0$.

Pure Logit Model

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{jt} = \frac{\exp \delta_{jt}}{1 + \sum_{k \in \mathcal{J}_t} \exp \delta_{kt}} \quad \Longrightarrow \quad \log \frac{s_{jt}}{s_{0t}} = \delta_{jt}$$

- Start with the simplest case: no heterogenous utility. We'll add μ_{ijt} back on day 2.
- Market shares simplify. No aggregation over individual types.
  - $\rightarrow$ The 1 in the denominator is from our level normalization $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{jt} = 0$.
- We can recover mean utilities from observed market shares (Berry, 1994).
  - $\rightarrow$ If we specify a function for δ_{jt}, we'll have a linear regression!

Pure Logit Estimating Equation

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt}$$

- Running example: What if we halved an important product's price?
  - $\rightarrow$ In your exercise, products j are breakfast cereals; markets t are city-quarters.
  - $\rightarrow$ If we estimate the model, we can change p_{jt} and estimate how consumers react.

Pure Logit Estimating Equation

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Running example: What if we halved an important product's price?
  - $\rightarrow$ In your exercise, products j are breakfast cereals; markets t are city-quarters.
  - $\rightarrow$ If we estimate the model, we can change p_{jt} and estimate how consumers react.
- Specify δ_{jt} as a function of price p_{jt} and other product characteristics x_{jt}.
  - $\rightarrow$ In your exercise, p_{jt} is per serving; x_{jt} includes a constant, a "mushy" dummy, etc.

Pure Logit Estimating Equation

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Running example: What if we halved an important product's price?
  - $\rightarrow$ In your exercise, products j are breakfast cereals; markets t are city-quarters.
  - $\rightarrow$ If we estimate the model, we can change p_{jt} and estimate how consumers react.
- Specify δ_{jt} as a function of price p_{jt} and other product characteristics x_{jt}.
  - $\rightarrow$ In your exercise, p_{jt} is per serving; x_{jt} includes a constant, a "mushy" dummy, etc.
- Interpret the regression error ξ_{jt} as unobserved product quality not in our data.
  - $\rightarrow$ Unobserved characteristics, advertising, average taste variation, "demand shocks," etc.

Interpreting Parameters

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Let's say we estimate this equation. How to interpret our parameter estimates?

Interpreting Parameters

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Let's say we estimate this equation. How to interpret our parameter estimates?
- Prices are in dollars, so the units of α are "utils" per dollar. Not very helpful.

Interpreting Parameters

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Let's say we estimate this equation. How to interpret our parameter estimates?
- Prices are in dollars, so the units of α are "utils" per dollar. Not very helpful.
  - $\rightarrow$ Instead, report own-price elasticities, or a quantity-weighted average/median.
  - $\rightarrow$ You can derive elasticities by differentiating the multinomial logit expression for s_{jt}.

$$\eta_{jjt} = \frac{\partial \log q_{jt}}{\partial \log p_{jt}} = \frac{\partial q_{jt}}{\partial p_{jt}} \frac{p_{jt}}{q_{jt}} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \alpha \cdot p_{jt} \cdot (1 - s_{jt})$$

Interpreting Parameters

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Let's say we estimate this equation. How to interpret our parameter estimates?
- Prices are in dollars, so the units of α are "utils" per dollar. Not very helpful.
  - $\rightarrow$ Instead, report own-price elasticities, or a quantity-weighted average/median.
  - $\rightarrow$ You can derive elasticities by differentiating the multinomial logit expression for s_{jt}.

$$\eta_{jjt} = \frac{\partial \log q_{jt}}{\partial \log p_{jt}} = \frac{\partial q_{jt}}{\partial p_{jt}} \frac{p_{jt}}{q_{jt}} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \alpha \cdot p_{jt} \cdot (1 - s_{jt})$$

- If x_{jt} is a "mushy" cereal dummy, β is "utils" from mushyness. Again, not helpful.

Interpreting Parameters

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Let's say we estimate this equation. How to interpret our parameter estimates?
- Prices are in dollars, so the units of α are "utils" per dollar. Not very helpful.
  - $\rightarrow$ Instead, report own-price elasticities, or a quantity-weighted average/median.
  - $\rightarrow$ You can derive elasticities by differentiating the multinomial logit expression for s_{jt}.

$$\eta_{jjt} = \frac{\partial \log q_{jt}}{\partial \log p_{jt}} = \frac{\partial q_{jt}}{\partial p_{jt}} \frac{p_{jt}}{q_{jt}} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \alpha \cdot p_{jt} \cdot (1 - s_{jt})$$

- If x_{jt} is a "mushy" cereal dummy, β is "utils" from mushyness. Again, not helpful.
  - $\rightarrow$ Instead, report β/α, the dollar willingness to pay for mushyness.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Coding Exercise 1

Endogeneity Concerns

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- In your coding exercise, you'll run an OLS regression of δ_{jt} on p_{jt} and x_{jt}.

Endogeneity Concerns

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- In your coding exercise, you'll run an OLS regression of δ_{jt} on p_{jt} and x_{jt}.
- As usual, if a regressor is correlated with the error, then its coefficient is biased.

Endogeneity Concerns

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- In your coding exercise, you'll run an OLS regression of δ_{jt} on p_{jt} and x_{jt}.
- As usual, if a regressor is correlated with the error, then its coefficient is biased.
- Typically, we expect price to be strongly correlated with unobserved quality.
  - $\rightarrow$ Firms know more than us about demand when setting prices.
  - $\rightarrow$ Often, $\mathbb{C}(p_{jt}, \xi_{jt}) > 0$, so $\hat{\alpha} < 0$ is biased towards zero.

Endogeneity Concerns

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

- In your coding exercise, you'll run an OLS regression of δ_{jt} on p_{jt} and x_{jt}.
- As usual, if a regressor is correlated with the error, then its coefficient is biased.
- Typically, we expect price to be strongly correlated with unobserved quality.
  - $\rightarrow$ Firms know more than us about demand when setting prices.
  - $\rightarrow$ Often, $\mathbb{C}(p_{jt}, \xi_{jt}) > 0$, so $\hat{\alpha} < 0$ is biased towards zero.
- Today we'll focus on handling just price endogeneity for simplicity.

Fixed Effects

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

- Adding product and market fixed effects to x_{jt} can help a lot.
  - $\rightarrow$ E.g. if p_{jt} is correlated with product-specific ξ_j in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
  - $\rightarrow$ But do need multiple observations per product and market to add ξ_j and ξ_t.

Fixed Effects

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Adding product and market fixed effects to x_{jt} can help a lot.
  - $\rightarrow$ E.g. if p_{jt} is correlated with product-specific ξ_j in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
  - $\rightarrow$ But do need multiple observations per product and market to add ξ_j and ξ_t.
  - $\rightarrow$ Aside: Related to dynamic panel approach. Let $\xi_{jt} = \rho\xi_{jt-1} + \Delta\xi_{jt}$, estimate ρ.

Fixed Effects

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Adding product and market fixed effects to x_{jt} can help a lot.
  - $\rightarrow$ E.g. if p_{jt} is correlated with product-specific ξ_j in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
  - $\rightarrow$ But do need multiple observations per product and market to add ξ_j and ξ_t.
  - $\rightarrow$ Aside: Related to dynamic panel approach. Let $\xi_{jt} = \rho\xi_{jt-1} + \Delta\xi_{jt}$, estimate ρ.
- Modern grocery scanner datasets have many thousands of products/markets.
  - $\rightarrow$ Dummies take too much memory, so we "absorb" them, i.e. iteratively de-mean.
  - $\rightarrow$ Stata: Reghdfe. R: Fixest. Python: PyFixest. Coding exercise: PyBLP via PyHDFE.

Fixed Effects

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Adding product and market fixed effects to x_{jt} can help a lot.
  - $\rightarrow$ E.g. if p_{jt} is correlated with product-specific ξ_j in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
  - $\rightarrow$ But do need multiple observations per product and market to add ξ_j and ξ_t.
  - $\rightarrow$ Aside: Related to dynamic panel approach. Let $\xi_{jt} = \rho\xi_{jt-1} + \Delta\xi_{jt}$, estimate ρ.
- Modern grocery scanner datasets have many thousands of products/markets.
  - $\rightarrow$ Dummies take too much memory, so we “absorb” them, i.e. iteratively de-mean.
  - $\rightarrow$ Stata: Reghdfe. R: Fixest. Python: PyFixest. Coding exercise: PyBLP via PyHDFE.
- Helpful but insufficient: ξ_{jt} typically varies by product *and* market.
  - $\rightarrow$ And if prices don’t, then we don’t have any variation left to estimate α.

Instrumental Variables

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- With or without fixed effects, a carefully-chosen IV can be a good solution.
  - $\rightarrow$ Relevance: $\mathbb{C}(p_{jt}, z_{jt}) \neq 0$. Exclusion: $\mathbb{C}(\xi_{jt}, z_{jt}) = 0$.

Instrumental Variables

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- With or without fixed effects, a carefully-chosen IV can be a good solution.
  - $\rightarrow$ Relevance: $\mathbb{C}(p_{jt}, z_{jt}) \neq 0$. Exclusion: $\mathbb{C}(\xi_{jt}, z_{jt}) = 0$.
- Always run a first-stage regression of p_{jt} on z_{jt} and x_{jt}.
  - $\rightarrow$ Does the sign of the coefficient on z_{jt} make sense?
  - $\rightarrow$ Is the instrument strong, or should you worry about weak instruments?

Instrumental Variables

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- With or without fixed effects, a carefully-chosen IV can be a good solution.
  - $\rightarrow$ Relevance: $\mathbb{C}(p_{jt}, z_{jt}) \neq 0$. Exclusion: $\mathbb{C}(\xi_{jt}, z_{jt}) = 0$.
- Always run a first-stage regression of p_{jt} on z_{jt} and x_{jt}.
  - $\rightarrow$ Does the sign of the coefficient on z_{jt} make sense?
  - $\rightarrow$ Is the instrument strong, or should you worry about weak instruments?
- Many places to look. I'll discuss the most common ones.

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
  - $\rightarrow$ We want valid instruments that shift costs and/or markups.

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
- Cost-shifters: Measures of input prices, tariffs, etc.

  $\rightarrow$ Consumers should only care about them through their effect on prices.

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
- Cost-shifters: Measures of input prices, tariffs, etc.
- Hausman: Current price of the same product averaged across *other* locations.
  - → Need costs to be correlated across locations, but not unobserved quality.
  - → Helpful that retailers tend to do “uniform pricing” (DellaVigna and Gentzkow, 2019).

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
- Cost-shifters: Measures of input prices, tariffs, etc.
- Hausman: Current price of the same product averaged across *other* locations.
- Waldfogel: Average consumer characteristics in *nearby* locations.
  $\rightarrow$ With uniform pricing, your neighbors' demographics will affect your prices.

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
- Cost-shifters: Measures of input prices, tariffs, etc.
- Hausman: Current price of the same product averaged across *other* locations.
- Waldfogel: Average consumer characteristics in *nearby* locations.
- BLP: Average characteristics x_{kt} of *competing* products $k \neq j$.
  - $\rightarrow$ Characteristics of competing products affect markups.
  - $\rightarrow$ We'll come back to these later, since they can also serve a different purpose.

Typical Instruments for Price

$$\delta_{jt} = \alpha p_{jt} + x'_{jt}\beta + \xi_{jt}$$

- Typically, prices are marginal costs plus a markup term.
- Cost-shifters: Measures of input prices, tariffs, etc.
- Hausman: Current price of the same product averaged across *other* locations.
- Waldfogel: Average consumer characteristics in *nearby* locations.
- BLP: Average characteristics x_{kt} of *competing* products $k \neq j$.
- I recommend starting with just one. A straightforward cost-shifter if you have it.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Coding Exercise 1

Coding Exercise 1

- Try to do the first exercise before day 2's class, when I'll do it live.
  1. Getting set up with Python and PyBLP.
  2. Pure logit estimation.
  3. Running the price cut counterfactual.

Coding Exercise 1

- Try to do the first exercise before day 2's class, when I'll do it live.
  1. Getting set up with Python and PyBLP.
  2. Pure logit estimation.
  3. Running the price cut counterfactual.

- When doing the exercise, think critically about the pure logit model's limitations.
  - $\rightarrow$ Do the substitution patterns you estimate seem reasonable?

Coding Exercise 1

- Try to do the first exercise before day 2's class, when I'll do it live.
  1. Getting set up with Python and PyBLP.
  2. Pure logit estimation.
  3. Running the price cut counterfactual.

- When doing the exercise, think critically about the pure logit model's limitations.
  $\rightarrow$ Do the substitution patterns you estimate seem reasonable?

- If you have time, try the supplemental exercises.
  $\rightarrow$ Statistical inference.
  $\rightarrow$ Modeling the supply side.
  $\rightarrow$ Checking your code by simulating data.

References I

Bergé, Laurent, "Fixest: Fast fixed-effects estimations." Available at https://github.com/lrberge/fixest.

Berry, Steven, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

__ **and Ariel Pakes**, "The pure characteristics demand model," *International Economic Review*, 2007, *48* (4), 1193–1225.

__ **, James Levinsohn, and Ariel Pakes**, "Automobile prices in market equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

__ **,** __ **, and** __ , "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

References II

Berry, Steven T and Philip A Haile, "Foundations of demand estimation," in "Handbook of industrial organization," Vol. 4 2021, pp. 1–62.

Conlon, Christopher and Jeff Gortmaker, "Best practices for differentiated products demand estimation with PyBLP," *The RAND Journal of Economics*, 2020, *51* (4), 1108–1161.

__ **and** __ , "Incorporating micro data into differentiated products demand estimation with PyBLP," 2023.

__ **and** __ , "PyBLP: BLP Demand Estimation with Python." Available at `https://github.com/jeffgortmaker/pyblp`.

Correia, Sergio, "Reghdfe: Linear regressions with multiple fixed effects." Available at `https://github.com/sergiocorreia/reghdfe`.

References III

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in US retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Fischer, Alexander, "PyFixest: Fast high-dimensional fixed effects regression in Python following fixest-syntax." Available at `https://github.com/s3alfisc`.

Gortmaker, Jeff and Anya Tarascina, "PyHDFE: High dimensional fixed effect absorption with Python." Available at `https://github.com/jeffgortmaker/pyhdfe`.

Hausman, Jerry A, "Valuation of new goods under perfect and imperfect competition," in "The economics of new goods," University of Chicago Press, 1996, pp. 207–248.

Nevo, Aviv, "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 2000, *9* (4), 513–548.

References IV

Petrin, Amil, "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 2002, *110* (4), 705–729.

Waldfogel, Joel, "Preference externalities: An empirical study of who benefits whom in differentiated-product markets," *The RAND Journal of Economics*, 2003, *34* (3), 557.