Demand Estimation

Mixtape Session

Jeff Gortmaker and Ariel Pakes

Last Class

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}_t} \overbrace{(\delta_{jt} - x_{jt}'\beta)}^{\xi_{jt}(\theta)} \cdot z_{jt}$$

$$\text{subject to} \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + \mu_{ijt}(\theta)]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + \mu_{ikt}(\theta)]}$$

- On day 2, adding preference heterogeneity μ_{ijt} gave more realistic substitution patterns.

Last Class

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}_t} \overbrace{(\delta_{jt} - x_{jt}'\beta)}^{\xi_{jt}(\theta)} \cdot z_{jt}$$

$$\text{subject to} \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + x_{jt}'(\Sigma\nu_{it} + \Pi y_{it})]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + x_{kt}'(\Sigma\nu_{it} + \Pi y_{it})]}$$

- On day 2, adding preference heterogeneity μ_{ijt} gave more realistic substitution patterns.
  - $\rightarrow$ Most common form is $\mu_{ijt} = x_{jt}'(\Sigma\nu_{it} + \Pi y_{it})$ for $\nu_{it} \sim N(0, I)$ and y_{it} from census data.
  - $\rightarrow$ Implements random coefficients $\beta_{it} \sim N(\beta + \Pi y_{it}, \Sigma\Sigma')$ on characteristics x_{jt} in utility.

Last Class

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}_t} \overbrace{(\delta_{jt} - x'_{jt}\beta)}^{\xi_{jt}(\theta)} \cdot z_{jt}$$

$$\text{subject to} \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + x'_{jt}(\Sigma\nu_{it} + \Pi y_{it})]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + x'_{kt}(\Sigma\nu_{it} + \Pi y_{it})]}$$

- On day 2, adding preference heterogeneity μ_{ijt} gave more realistic substitution patterns.
  - $\rightarrow$ Most common form is $\mu_{ijt} = x'_{jt}(\Sigma\nu_{it} + \Pi y_{it})$ for $\nu_{it} \sim N(0, I)$ and y_{it} from census data.
  - $\rightarrow$ Implements random coefficients $\beta_{it} \sim N(\beta + \Pi y_{it}, \Sigma\Sigma')$ on characteristics x_{jt} in utility.
- This required adding consumer type i data to supplement our product j data from day 1.

Last Class

$$\hat{\theta} = \arg\min_{\theta} g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}_t} \overbrace{(\delta_{jt} - x'_{jt}\beta)}^{\xi_{jt}(\theta)} \cdot z_{jt}$$

$$\text{subject to} \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + x'_{jt}(\Sigma \nu_{it} + \Pi y_{it})]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + x'_{kt}(\Sigma \nu_{it} + \Pi y_{it})]}$$

- On day 2, adding preference heterogeneity μ_{ijt} gave more realistic substitution patterns.
  - $\rightarrow$ Most common form is $\mu_{ijt} = x'_{jt}(\Sigma \nu_{it} + \Pi y_{it})$ for $\nu_{it} \sim N(0, I)$ and y_{it} from census data.
  - $\rightarrow$ Implements random coefficients $\beta_{it} \sim N(\beta + \Pi y_{it}, \Sigma\Sigma')$ on characteristics x_{jt} in utility.
- This required adding consumer type i data to supplement our product j data from day 1.
- Let's go over your second coding exercise.

Limited Cross-Market Variation

- The price cut exercise seems more reasonable with random coefficients.
  - $\rightarrow$ Consumers substitute from similar products, particularly along the price dimension.

Limited Cross-Market Variation

- The price cut exercise seems more reasonable with random coefficients.
  - $\rightarrow$ Consumers substitute from similar products, particularly along the price dimension.
- But which random coefficients we can add is limited by variation in our data.
  - $\rightarrow$ Recall the linear regression intuition from day 2.

Limited Cross-Market Variation

- The price cut exercise seems more reasonable with random coefficients.
  - $\rightarrow$ Consumers substitute from similar products, particularly along the price dimension.
- But which random coefficients we can add is limited by variation in our data.
  - $\rightarrow$ Recall the linear regression intuition from day 2.
- Can't credibly estimate a standard deviation in Σ on the mushy dummy.
  - $\rightarrow$ Same cereals in each market, so no choice set variation along mushy dimension.
  - $\rightarrow$ Results in unrealistically limited substitution between similar cereals.

Limited Cross-Market Variation

- The price cut exercise seems more reasonable with random coefficients.
  - $\rightarrow$ Consumers substitute from similar products, particularly along the price dimension.
- But which random coefficients we can add is limited by variation in our data.
  - $\rightarrow$ Recall the linear regression intuition from day 2.
- Can't credibly estimate a standard deviation in Σ on the mushy dummy.
  - $\rightarrow$ Same cereals in each market, so no choice set variation along mushy dimension.
  - $\rightarrow$ Results in unrealistically limited substitution between similar cereals.
- Also can't estimate a parameter in Π on log income alone.
  - $\rightarrow$ Market fixed effects are collinear with market-level income means.
  - $\rightarrow$ Unrealistic that overall cereal preference doesn't vary with income.

Within-Market Variation

- Without much cross-market variation, what we really want is within-market variation.

Within-Market Variation

- Without much cross-market variation, what we really want is within-market variation.
- “Micro data” has information about individual choices, not just market-level quantities.

Within-Market Variation

- Without much cross-market variation, what we really want is within-market variation.
- "Micro data" has information about individual choices, not just market-level quantities.
- Typical example is consumer survey data.
  - → Internal surveys conducted by firms.
  - → Ad-hoc surveys conducted by academics.
  - → Marketing research datasets (e.g. NielsenIQ's Consumer Panel).
  - → Regulatory agencies like the UK's antitrust authority (Reynolds and Walters, 2008).

Intuition for the Running Example

- Imagine surveying people at the supermarket who purchased cereal.

Intuition for the Running Example

- Imagine surveying people at the supermarket who purchased cereal.
- "What was your annual income last year?"
  - $\rightarrow$ Should be informative about a parameter in Π on log income alone.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.

Intuition for the Running Example

- Imagine surveying people at the supermarket who purchased cereal.
- "What was your annual income last year?"
  - $\rightarrow$ Should be informative about a parameter in Π on log income alone.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.
- "Would you have purchased another mushy cereal if your first choice wasn't available?"
  - $\rightarrow$ Should be informative about a standard deviation in Σ on the mushy dummy.
  - $\rightarrow$ Within-mushy substitution is precisely what we hope this parameter will increase!

Intuition for the Running Example

- Imagine surveying people at the supermarket who purchased cereal.
- "What was your annual income last year?"
  - $\rightarrow$ Should be informative about a parameter in Π on log income alone.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.
- "Would you have purchased another mushy cereal if your first choice wasn't available?"
  - $\rightarrow$ Should be informative about a standard deviation in Σ on the mushy dummy.
  - $\rightarrow$ Within-mushy substitution is precisely what we hope this parameter will increase!
- Let's incorporate answers to these questions into estimation.
  - $\rightarrow$ We'll set up a general framework and come back to these when we have notation to do so.

Roadmap

Micro BLP Estimation

Choosing Micro Moments

Using More Information

Coding Exercise 3

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}_t} (\delta_{jt}(\theta) - x_{jt}' \beta) \cdot z_{jt}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N} \sum_{j,t} (\delta_{jt}(\theta) - x'_{jt}\beta) \cdot z_{jt} \\ \overline{v}_1 - v_1(\theta) \end{bmatrix}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.
    1. $\overline{v}_1$: Mean income of cereal purchasers. Matched to the model's prediction $v_1(\theta)$.

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N}\sum_{j,t}(\delta_{jt}(\theta) - x'_{jt}\beta)\cdot z_{jt} \\ \overline{v}_1 - v_1(\theta) \\ \overline{v}_2 - v_2(\theta) \end{bmatrix}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.
    1. $\overline{v}_1$: Mean income of cereal purchasers. Matched to the model's prediction $v_1(\theta)$.
    2. $\overline{v}_2$: Share of cereal purchasers who chose a mushy cereal and would choose another mushy cereal if their first choice wasn't available. Model analogue is $v_2(\theta)$.

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N} \sum_{j,t} (\delta_{jt}(\theta) - x'_{jt}\beta) \cdot z_{jt} \\ \overline{v} - v(\theta) \end{bmatrix}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.
    1. $\overline{v}_1$: Mean income of cereal purchasers. Matched to the model's prediction $v_1(\theta)$.
    2. $\overline{v}_2$: Share of cereal purchasers who chose a mushy cereal and would choose another mushy cereal if their first choice wasn't available. Model analogue is $v_2(\theta)$.

- Denote the vector of matched statistics by $\overline{v} = [\overline{v}_1, \overline{v}_2]'$.

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N} \sum_{j,t} (\delta_{jt}(\theta) - x'_{jt}\beta) \cdot z_{jt} \\ f(\overline{v}) - f(v(\theta)) \end{bmatrix}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.
  1. $\overline{v}_1$: Mean income of cereal purchasers. Matched to the model's prediction $v_1(\theta)$.
  2. $\overline{v}_2$: Share of cereal purchasers who chose a mushy cereal and would choose another mushy cereal if their first choice wasn't available. Model analogue is $v_2(\theta)$.
- Denote the vector of matched statistics by $\overline{v} = [\overline{v}_1, \overline{v}_2]'$.
- Don't have to be averages $\overline{v}$. Can match any smooth function $f(\overline{v})$.
  - → Ratios (e.g. mean income given mushy), correlations (e.g. between income and price), etc.

Micro BLP Estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N}\sum_{j,t}(\delta_{jt}(\theta) - x_{jt}'\beta) \cdot z_{jt} \\ f(\overline{v}) - f(v(\theta)) \end{bmatrix}$$

- We can extend our BLP estimator by matching statistics from consumer surveys.
    1. $\overline{v}_1$: Mean income of cereal purchasers. Matched to the model's prediction $v_1(\theta)$.
    2. $\overline{v}_2$: Share of cereal purchasers who chose a mushy cereal and would choose another mushy cereal if their first choice wasn't available. Model analogue is $v_2(\theta)$.

- Denote the vector of matched statistics by $\overline{v} = [\overline{v}_1, \overline{v}_2]'$.

- Don't have to be averages $\overline{v}$. Can match any smooth function $f(\overline{v})$.
    - $\rightarrow$ Ratios (e.g. mean income given mushy), correlations (e.g. between income and price), etc.

- The resulting "micro BLP" estimator is used a lot in industrial organization.

Micro BLP Popularity

- First popularized by Petrin (2002) and BLP (2004).

Paper	Demand Estimation: Industry	Country	Years
Petrin (2002)	Automobiles	United States	1981–1993
Berry, Levinsohn, and Pakes (2004)	Automobiles	United States	1993
Thomadsen (2005)	Fast Food	United States	1999
Goeree (2008)	Personal Computers	United States	1996–1998
Ciliberto and Kuminoff (2010)	Cigarettes	United States	1993–2002
Nakamura and Zerom (2010)	Coffee	United States	2000–2004
Beresteanu and Li (2011)	Automobiles	United States	1999–2006
Li (2012)	Automobiles	United States	1999–2006
Copeland (2014)	Automobiles	United States	1999–2008
Starc (2014)	Health Insurance	United States	2004–2008
Ching, Hayashi, and Wang (2015)	Nursing Homes	United States	1999
Li, Xiao, and Liu (2015)	Automobiles	China	2004–2009
Nurski and Verboven (2016)	Automobiles	Belgium	2010–2011
Barwick, Cao, and Li (2017)	Automobiles	China	2009–2011
Murry (2017)	Automobiles	United States	2007–2011
Wollmann (2018)	Commercial Vehicles	United States	1986–2012
Li (2018)	Automobiles	China	2008–2012
Li, Gordon, and Netzer (2018)	Digital Cameras	United States	2007–2010
Backus, Conlon, and Sinkinson (2021)	Cereal	United States	2007–2016
Grieco, Murry, and Yurukoglu (2021)	Automobiles	United States	1980–2018
Neilson (2021)	Primary Schools	Chile	2005–2016
Armitage and Pinter (2022)	Automobiles	United States	2009–2017
Döpper, MacKay, Miller, and Stiebale (2022)	Retail	United States	2006–2019
Durrmeyer (2022)	Automobiles	France	2003–2008
Weber (2022)	Trucks	United States	2010–2018
Bodéré (2023)	Preschools	United States	2010–2018
Montag (2023)	Laundry Machines	United States	2005–2015
Conlon and Rao (2023)	Distilled Spirits	United States	2007–2013
Calder-Wang and Kim (2024)	Rental Housing	United States	2011–2018

Micro BLP Popularity

- First popularized by Petrin (2002) and BLP (2004).
- Used a lot. But each paper has different notation.

Paper	Demand Estimation		
	Industry	Country	Years
Petrin (2002)	Automobiles	United States	1981–1993
Berry, Levinsohn, and Pakes (2004)	Automobiles	United States	1993
Thomadsen (2005)	Fast Food	United States	1999
Goeree (2008)	Personal Computers	United States	1996–1998
Ciliberto and Kuminoff (2010)	Cigarettes	United States	1993–2002
Nakamura and Zerom (2010)	Coffee	United States	2000–2004
Beresteanu and Li (2011)	Automobiles	United States	1999–2006
Li (2012)	Automobiles	United States	1999–2006
Copeland (2014)	Automobiles	United States	1999–2008
Starc (2014)	Health Insurance	United States	2004–2008
Ching, Hayashi, and Wang (2015)	Nursing Homes	United States	1999
Li, Xiao, and Liu (2015)	Automobiles	China	2004–2009
Nurski and Verboven (2016)	Automobiles	Belgium	2010–2011
Barwick, Cao, and Li (2017)	Automobiles	China	2009–2011
Murry (2017)	Automobiles	United States	2007–2011
Wollmann (2018)	Commercial Vehicles	United States	1986–2012
Li (2018)	Automobiles	China	2008–2012
Li, Gordon, and Netzer (2018)	Digital Cameras	United States	2007–2010
Backus, Conlon, and Sinkinson (2021)	Cereal	United States	2007–2016
Grieco, Murry, and Yurukoglu (2021)	Automobiles	United States	1980–2018
Neilson (2021)	Primary Schools	Chile	2005–2016
Armitage and Pinter (2022)	Automobiles	United States	2009–2017
Döpper, MacKay, Miller, and Stiebale (2022)	Retail	United States	2006–2019
Durrmeyer (2022)	Automobiles	France	2003–2008
Weber (2022)	Trucks	United States	2010–2018
Bodéré (2023)	Preschools	United States	2010–2018
Montag (2023)	Laundry Machines	United States	2005–2015
Conlon and Rao (2023)	Distilled Spirits	United States	2007–2013
Calder-Wang and Kim (2024)	Rental Housing	United States	2011–2018

Micro BLP Popularity

- First popularized by Petrin (2002) and BLP (2004).
- Used a lot. But each paper has different notation.
- We'll use the standardized framework for PyBLP from Conlon and Gortmaker (2025).

	Demand Estimation		
Paper	Industry	Country	Years
Petrin (2002)	Automobiles	United States	1981–1993
Berry, Levinsohn, and Pakes (2004)	Automobiles	United States	1993
Thomadsen (2005)	Fast Food	United States	1999
Goeree (2008)	Personal Computers	United States	1996–1998
Ciliberto and Kuminoff (2010)	Cigarettes	United States	1993–2002
Nakamura and Zerom (2010)	Coffee	United States	2000–2004
Beresteanu and Li (2011)	Automobiles	United States	1999–2006
Li (2012)	Automobiles	United States	1999–2006
Copeland (2014)	Automobiles	United States	1999–2008
Starc (2014)	Health Insurance	United States	2004–2008
Ching, Hayashi, and Wang (2015)	Nursing Homes	United States	1999
Li, Xiao, and Liu (2015)	Automobiles	China	2004–2009
Nurski and Verboven (2016)	Automobiles	Belgium	2010–2011
Barwick, Cao, and Li (2017)	Automobiles	China	2009–2011
Murry (2017)	Automobiles	United States	2007–2011
Wollmann (2018)	Commercial Vehicles	United States	1986–2012
Li (2018)	Automobiles	China	2008–2012
Li, Gordon, and Netzer (2018)	Digital Cameras	United States	2007–2010
Backus, Conlon, and Sinkinson (2021)	Cereal	United States	2007–2016
Grieco, Murry, and Yurukoglu (2021)	Automobiles	United States	1980–2018
Neilson (2021)	Primary Schools	Chile	2005–2016
Armitage and Pinter (2022)	Automobiles	United States	2009–2017
Döpper, MacKay, Miller, and Stiebale (2022)	Retail	United States	2006–2019
Durrmeyer (2022)	Automobiles	France	2003–2008
Weber (2022)	Trucks	United States	2010–2018
Bodéré (2023)	Preschools	United States	2010–2018
Montag (2023)	Laundry Machines	United States	2005–2015
Conlon and Rao (2023)	Distilled Spirits	United States	2007–2013
Calder-Wang and Kim (2024)	Rental Housing	United States	2011–2018

Micro Moments

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N}\sum_{j,t}(\delta_{jt}(\theta) - x'_{jt}\beta)\cdot z_{jt} \\ f(\overline{v}) - f(v(\theta)) \end{bmatrix}$$

- There are two new components.
  1. Micro statistics $f(\overline{v}) = [f_1(\overline{v}), \ldots, f_M(\overline{v})]'$.
  2. Their model analogues $f(v(\theta)) = [f_1(v(\theta)), \ldots, f_M(v(\theta))]'$.

Micro Moments

$$\hat{\theta} = \operatorname*{argmin}_{\theta} g(\theta) W g(\theta)' \quad \text{where} \quad g(\theta) = \begin{bmatrix} \frac{1}{N} \sum_{j,t} (\delta_{jt}(\theta) - x_{jt}'\beta) \cdot z_{jt} \\ f(\overline{v}) - f(v(\theta)) \end{bmatrix}$$

- There are two new components.
  1. Micro statistics $f(\overline{v}) = [f_1(\overline{v}), \ldots, f_M(\overline{v})]'$.
  2. Their model analogues $f(v(\theta)) = [f_1(v(\theta)), \ldots, f_M(v(\theta))]'$.

- Statistically, we need $f(\overline{v}) \to f(v(\theta_0))$ as the micro dataset expands.
  - $\to$ This gives what we'll call $m = 1, \ldots, M$ different "micro moments."
  - $\to$ A bit different from our "aggregate moments" $\mathbb{E}[\xi_{jt} \cdot z_{jt}] = 0$.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.
- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.
- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
  1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.
- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
    1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.
    2. ... be of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_n t_n}$. Same individual type weight as before.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.
- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
  1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.
  2. ... be of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_n t_n}$. Same individual type weight as before.
  3. ... choose $j_n \in \mathcal{J}_{t_n} \cup \{0\}$ with probability $s_{i_n j_n t_n}$. Same logit choice probability as before.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.
- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
  1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.
  2. ... be of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_n t_n}$. Same individual type weight as before.
  3. ... choose $j_n \in \mathcal{J}_{t_n} \cup \{0\}$ with probability $s_{i_n j_n t_n}$. Same logit choice probability as before.
  4. ... be selected into the survey with known probability $w_{d i_n j_n t_n}$. Often choice-based.

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.

- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
  1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.
  2. ... be of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_n t_n}$. Same individual type weight as before.
  3. ... choose $j_n \in \mathcal{J}_{t_n} \cup \{0\}$ with probability $s_{i_n j_n t_n}$. Same logit choice probability as before.
  4. ... be selected into the survey with known probability $w_{d i_n j_n t_n}$. Often choice-based.

- Micro statistics $f_m(\overline{v})$ are smooth functions of $p = 1, \ldots, P$ "micro part" averages:

$$\overline{v}_p = \frac{1}{N_d} \sum_{n \in \mathcal{N}_d} v_{p i_n j_n t_n}$$

Micro Statistics

- Each micro statistic $f_m(\overline{v})$ is a summary statistic computed on a micro dataset.

- Micro datasets $d \in \mathcal{D}$ consist of surveyed consumers $n \in \mathcal{N}_d$ who could ...
  1. ... be from each market $t_n \in \mathcal{T}$ with equal probability.
  2. ... be of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_n t_n}$. Same individual type weight as before.
  3. ... choose $j_n \in \mathcal{J}_{t_n} \cup \{0\}$ with probability $s_{i_n j_n t_n}$. Same logit choice probability as before.
  4. ... be selected into the survey with known probability $w_{d i_n j_n t_n}$. Often choice-based.

- Micro statistics $f_m(\overline{v})$ are smooth functions of $p = 1, \dots, P$ “micro part” averages:

$$\overline{v}_p = \frac{1}{N_d} \sum_{n \in \mathcal{N}_d} v_{p i_n j_n t_n}$$

- Different weights w_{dijt}, values v_{pijt}, and functions $f_m(\cdot)$ support most summary stats.

Micro Statistic Example

- To match $\overline{v}_1$, the mean income of cereal purchasers, PyBLP needs some info.

Micro Statistic Example

- To match $\overline{v}_1$, the mean income of cereal purchasers, PyBLP needs some info.
- First, you need to define your micro dataset d.
  - $\rightarrow$ Sampling weights $w_{dijt} = 1\{j \neq 0\}$ imply only cereal purchasers were surveyed.
  - $\rightarrow$ In practice, you specify a function to compute a matrix of weights for each market t.
  - $\rightarrow$ You also need to specify the number of survey observations $N_d = |\mathcal{N}_d|$.

Micro Statistic Example

- To match $\overline{v}_1$, the mean income of cereal purchasers, PyBLP needs some info.
- First, you need to define your micro dataset d.
  - $\rightarrow$ Sampling weights $w_{dijt} = 1\{j \neq 0\}$ imply only cereal purchasers were surveyed.
  - $\rightarrow$ In practice, you specify a function to compute a matrix of weights for each market t.
  - $\rightarrow$ You also need to specify the number of survey observations $N_d = |\mathcal{N}_d|$.
- Second, you need to define your micro part p.
  - $\rightarrow$ Micro values $v_{pijt} = \text{income}_{it}$ implies $\overline{v}_p$ is mean surveyed income.
  - $\rightarrow$ In practice, you specify a second function to compute a matrix of values for each market t.

Micro Statistic Example

- To match $\overline{v}_1$, the mean income of cereal purchasers, PyBLP needs some info.

- First, you need to define your micro dataset d.
  - $\rightarrow$ Sampling weights $w_{dijt} = 1\{j \neq 0\}$ imply only cereal purchasers were surveyed.
  - $\rightarrow$ In practice, you specify a function to compute a matrix of weights for each market t.
  - $\rightarrow$ You also need to specify the number of survey observations $N_d = |\mathcal{N}_d|$.

- Second, you need to define your micro part p.
  - $\rightarrow$ Micro values $v_{pijt} = \text{income}_{it}$ implies $\overline{v}_p$ is mean surveyed income.
  - $\rightarrow$ In practice, you specify a second function to compute a matrix of values for each market t.

- Lastly, you need to define your micro moment m.
  - $\rightarrow$ The identity function $f_m(\overline{v}_p) = \overline{v}_p$ just matches the mean surveyed income.
  - $\rightarrow$ You also need to specify the actual value of the micro statistic $\overline{v}_1$.

Model Analogues

$$f_m(\overline{v}_p) \to f_m(v_p(\theta_0))$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\cdots v_{pijt}}{\cdots}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.

  $\to$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\cdots v_{pijt}}{\cdots}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.
  - $\to$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.
- Recall the data generating process for surveyed consumers $n \in \mathcal{N}_d$.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\sum_{t\in\mathcal{T}}\cdots v_{pijt}}{\sum_{t\in\mathcal{T}}\cdots}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.
  - $\rightarrow$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.
- Recall the data generating process for surveyed consumers $n \in \mathcal{N}_d$.
  1. In market $t_n \in \mathcal{T}$ with equal probability.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t} w_{it}\cdots v_{pijt}}{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t} w_{it}\cdots}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.
  - $\to$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.
- Recall the data generating process for surveyed consumers $n \in \mathcal{N}_d$.
  1. In market $t_n \in \mathcal{T}$ with equal probability.
  2. Of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_nt_n}$.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t}\sum_{j\in\mathcal{J}_t\cup\{0\}} w_{it}\cdot s_{ijt}(\theta_0)\cdots v_{pijt}}{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t}\sum_{j\in\mathcal{J}_t\cup\{0\}} w_{it}\cdot s_{ijt}(\theta_0)\cdots}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.
  - $\to$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.
- Recall the data generating process for surveyed consumers $n \in \mathcal{N}_d$.
  1. In market $t_n \in \mathcal{T}$ with equal probability.
  2. Of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_nt_n}$.
  3. Chooses $j_n \in \mathcal{J}_{t_n}$ with probability $s_{i_nj_nt_n}$.

Model Analogues

$$f_m\left(\frac{1}{N_d}\sum_{n\in\mathcal{N}_d} v_{pi_nj_nt_n}\right) \to f_m\left(\frac{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t}\sum_{j\in\mathcal{J}_t\cup\{0\}} w_{it}\cdot s_{ijt}(\theta_0)\cdot w_{dijt}\cdot v_{pijt}}{\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}_t}\sum_{j\in\mathcal{J}_t\cup\{0\}} w_{it}\cdot s_{ijt}(\theta_0)\cdot w_{dijt}}\right)$$

- For each guess of θ, PyBLP will compute the model analogue $f_m(v_p(\theta))$.
- The model analogue $v_p(\theta)$ of a micro part $\overline{v}_p$ is a conditional expectation.
  - $\to$ Expected micro value $v_{pi_nj_nt_n}$ divided by the probability of being selected.
- Recall the data generating process for surveyed consumers $n \in \mathcal{N}_d$.
  1. In market $t_n \in \mathcal{T}$ with equal probability.
  2. Of type $i_n \in \mathcal{I}_{t_n}$ with probability $w_{i_nt_n}$.
  3. Chooses $j_n \in \mathcal{J}_{t_n}$ with probability $s_{i_nj_nt_n}$.
  4. Selected to be in the survey with known probability $w_{di_nj_nt_n}$.

Roadmap

Micro BLP Estimation

Choosing Micro Moments

Using More Information

Coding Exercise 3

Adding Micro Moments

- Recall my advice from last class about adding instruments.
  - $\rightarrow$ For each new parameter, we need another instrument.
  - $\rightarrow$ Start by choosing a single instrument that “targets” that parameter.

Adding Micro Moments

- Recall my advice from last class about adding instruments.
  - → For each new parameter, we need another instrument.
  - → Start by choosing a single instrument that "targets" that parameter.
- My advice for adding micro moments is similar.
  - → For each new parameter without an instrument, we need a micro moment.
  - → Start by choosing a single micro moment that "targets" the parameter.

Adding Micro Moments

- Recall my advice from last class about adding instruments.
  - $\rightarrow$ For each new parameter, we need another instrument.
  - $\rightarrow$ Start by choosing a single instrument that "targets" that parameter.
- My advice for adding micro moments is similar.
  - $\rightarrow$ For each new parameter without an instrument, we need a micro moment.
  - $\rightarrow$ Start by choosing a single micro moment that "targets" the parameter.
- What if you could estimate a parameter with either aggregate or micro variation?
  - $\rightarrow$ Could just choose the variation that seems more "credible." Often the micro moment.
  - $\rightarrow$ Can use both. Micro moments can reduce large SEs from limited aggregate variation.

Targeting Micro Moments

- Simplest case: 1 characteristic x_{jt} (e.g. price) and 1 demographic y_{it} (e.g. income).

$$u_{ijt} = \beta_1 + \pi_1 y_{it} + (\beta_x + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

Targeting Micro Moments

- Simplest case: 1 characteristic x_{jt} (e.g. price) and 1 demographic y_{it} (e.g. income).

$$u_{ijt} = \beta_1 + \pi_1 y_{it} + (\beta_x + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- In our example, matching a "$\mathbb{E}[y_{it} \mid j \neq 0]$" micro moment targets π_1.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.

Targeting Micro Moments

- Simplest case: 1 characteristic x_{jt} (e.g. price) and 1 demographic y_{it} (e.g. income).

$$u_{ijt} = \beta_1 + \pi_1 y_{it} + (\beta_x + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- In our example, matching a "$\mathbb{E}[y_{it} \mid j \neq 0]$" micro moment targets π_1.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.
- To target π_x, you can match "$\mathbb{E}[y_{it} \cdot x_{jt} \mid j \neq 0]$" or similar, e.g. "$\mathbb{C}(y_{it}, x_{jt} \mid j \neq 0)$."
  - $\rightarrow$ Relationship between income and price targets how income shifts price sensitivity.
  - $\rightarrow$ Other common examples include "$\mathbb{E}[y_{it} \mid x_{jt} < \overline{x}]$" and "$\mathbb{E}[x_{jt} \mid y_{it} < \overline{y}]$."

Targeting Micro Moments

- Simplest case: 1 characteristic x_{jt} (e.g. price) and 1 demographic y_{it} (e.g. income).

$$u_{ijt} = \beta_1 + \pi_1 y_{it} + (\beta_x + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- In our example, matching a "$\mathbb{E}[y_{it} \mid j \neq 0]$" micro moment targets π_1.
  - $\rightarrow$ Mean income of cereal purchasers targets how income shifts cereal preference.
- To target π_x, you can match "$\mathbb{E}[y_{it} \cdot x_{jt} \mid j \neq 0]$" or similar, e.g. "$\mathbb{C}(y_{it}, x_{jt} \mid j \neq 0)$."
  - $\rightarrow$ Relationship between income and price targets how income shifts price sensitivity.
  - $\rightarrow$ Other common examples include "$\mathbb{E}[y_{it} \mid x_{jt} < \overline{x}]$" and "$\mathbb{E}[x_{jt} \mid y_{it} < \overline{y}]$."
- Micro data is not directly informative about "linear parameters" β_1 or β_x.
  - $\rightarrow$ Mean utility $\delta_{jt} = \beta_1 + \beta_x x_{jt} + \xi_{jt}$ is already pinned down by market shares s_{jt}.

Second Choice Data

- What about targeting unobserved preference heterogeneity parameters in Σ?

Second Choice Data

- What about targeting unobserved preference heterogeneity parameters in Σ?
- The literature has had success with incorporating second choice data (e.g. BLP, 2004).
  - $\rightarrow$ Survey asks consumers which $k_n \neq j_n$ they'd choose if their first choice weren't available.
  - $\rightarrow$ Micro weights and values now just have an extra index: w_{dijkt} and v_{pijkt}.

Second Choice Data

- What about targeting unobserved preference heterogeneity parameters in Σ?
- The literature has had success with incorporating second choice data (e.g. BLP, 2004).
  - $\rightarrow$ Survey asks consumers which $k_n \neq j_n$ they'd choose if their first choice weren't available.
  - $\rightarrow$ Micro weights and values now just have an extra index: w_{dijkt} and v_{pijkt}.
- Direct measures of substitution are very informative about Σ.
  - $\rightarrow$ Recall the red bus/blue bus example that motivated adding preference heterogeneity.
  - $\rightarrow$ Each second choice is like observing a new market with the first choice removed.

Second Choice Moments

- Extend the simple example from before with unobserved preference heterogeneity:

$$u_{ijt} = \beta_1 + \sigma_1 \nu_{1it} + \pi_1 y_{it} + (\beta_x + \sigma_x \nu_{2it} + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

Second Choice Moments

- Extend the simple example from before with unobserved preference heterogeneity:

$$u_{ijt} = \beta_1 + \sigma_1 \nu_{1it} + \pi_1 y_{it} + (\beta_x + \sigma_x \nu_{2it} + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- To target σ_x, we want a measure of how much people substitute within x_{jt}.
  - → In your exercise, you'll match the share "$\mathbb{P}(\text{mushy}_{jt} \text{ and } \text{mushy}_{kt} \mid j \neq 0)$."
  - → For a continuous x_{jt}, you could match "$\mathbb{C}(x_{jt}, x_{kt} \mid j, k \neq 0)$" or similar.

Second Choice Moments

- Extend the simple example from before with unobserved preference heterogeneity:

$$u_{ijt} = \beta_1 + \sigma_1 \nu_{1it} + \pi_1 y_{it} + (\beta_x + \sigma_x \nu_{2it} + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- To target σ_x, we want a measure of how much people substitute within x_{jt}.
  - $\rightarrow$ In your exercise, you'll match the share "$\mathbb{P}(\text{mushy}_{jt} \text{ and } \text{mushy}_{kt} \mid j \neq 0)$."
  - $\rightarrow$ For a continuous x_{jt}, you could match "$\mathbb{C}(x_{jt}, x_{kt} \mid j, k \neq 0)$" or similar.
- To target σ_1, we want a measure of how much people substitute to $k = 0$.
  - $\rightarrow$ In your exercise, you'll match another share "$\mathbb{P}(k = 0 \mid j \neq 0)$."

Second Choice Moments

- Extend the simple example from before with unobserved preference heterogeneity:

$$u_{ijt} = \beta_1 + \sigma_1 \nu_{1it} + \pi_1 y_{it} + (\beta_x + \sigma_x \nu_{2it} + \pi_x y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- To target σ_x, we want a measure of how much people substitute within x_{jt}.
  - $\rightarrow$ In your exercise, you'll match the share "$\mathbb{P}(\text{mushy}_{jt}$ and $\text{mushy}_{kt} \mid j \neq 0)$."
  - $\rightarrow$ For a continuous x_{jt}, you could match "$\mathbb{C}(x_{jt}, x_{kt} \mid j, k \neq 0)$" or similar.

- To target σ_1, we want a measure of how much people substitute to $k = 0$.
  - $\rightarrow$ In your exercise, you'll match another share "$\mathbb{P}(k = 0 \mid j \neq 0)$."

- Diversion ratios are straightforward to interpret and collect.

Outside Substitution and Market Size

- Estimating a σ_1 is important if you're interested in inside-outside substitution.
  - $\rightarrow$ How many consumers will stop purchasing soda if all sodas are taxed?
  - $\rightarrow$ How much larger will a market grow if innovation reduces production costs?

Outside Substitution and Market Size

- Estimating a σ_1 is important if you're interested in inside-outside substitution.
  - $\rightarrow$ How many consumers will stop purchasing soda if all sodas are taxed?
  - $\rightarrow$ How much larger will a market grow if innovation reduces production costs?

- On day 1, we discussed how choosing a market size is neither easy nor innocuous.
  - $\rightarrow$ Assuming a small market size M_t means assuming a small outside share s_{0t}.
  - $\rightarrow$ Typically, this results in relatively high inside qualities ξ_{jt}.
  - $\rightarrow$ Implies little substitution to the outside good in counterfactuals.

Outside Substitution and Market Size

- Estimating a σ_1 is important if you're interested in inside-outside substitution.
  - $\rightarrow$ How many consumers will stop purchasing soda if all sodas are taxed?
  - $\rightarrow$ How much larger will a market grow if innovation reduces production costs?
- On day 1, we discussed how choosing a market size is neither easy nor innocuous.
  - $\rightarrow$ Assuming a small market size M_t means assuming a small outside share s_{0t}.
  - $\rightarrow$ Typically, this results in relatively high inside qualities ξ_{jt}.
  - $\rightarrow$ Implies little substitution to the outside good in counterfactuals.
- Directly matching an outside diversion ratio will help discipline outside substitution.
  - $\rightarrow$ If $\hat{\sigma}_1$ is large, many people will dislike all inside goods and usually choose $j = 0$.
  - $\rightarrow$ This reduces the *effective* market size, helping to compensate for a too-large M_t.
  - $\rightarrow$ See Zhang (2024) for more on how σ_1 can help and other solutions.

Roadmap

Micro BLP Estimation

Choosing Micro Moments

Using More Information

Coding Exercise 3

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?
- May prefer to use summary stats for a few reasons.

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?

- May prefer to use summary stats for a few reasons.
  - $\rightarrow$ Cost: More data costs more. Headline summary stats can be free.

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?
- May prefer to use summary stats for a few reasons.
  - $\rightarrow$ Cost: More data costs more. Headline summary stats can be free.
  - $\rightarrow$ Confidentiality: Data providers may want to protect respondent identities.

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?
- May prefer to use summary stats for a few reasons.
  - → Cost: More data costs more. Headline summary stats can be free.
  - → Confidentiality: Data providers may want to protect respondent identities.
  - → Compatibility: Aggregate and micro data may come from different sampling schemes.

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?

- May prefer to use summary stats for a few reasons.
  - → Cost: More data costs more. Headline summary stats can be free.
  - → Confidentiality: Data providers may want to protect respondent identities.
  - → Compatibility: Aggregate and micro data may come from different sampling schemes.
  - → Clarity: Matching a single statistic makes it clear where identification comes from.

Information Tradeoffs

- So far we have focused on matching summary statistics. Why not use the full survey?

- May prefer to use summary stats for a few reasons.
  - $\rightarrow$ Cost: More data costs more. Headline summary stats can be free.
  - $\rightarrow$ Confidentiality: Data providers may want to protect respondent identities.
  - $\rightarrow$ Compatibility: Aggregate and micro data may come from different sampling schemes.
  - $\rightarrow$ Clarity: Matching a single statistic makes it clear where identification comes from.

- But adding more info can greatly increase the precision of our estimates.
  - $\rightarrow$ Ideally we'd observe a complete micro dataset $\{t_n, j_n, k_n, y_{i_n t_n}\}_{n \in \mathcal{N}_d}$.

Maximum Likelihood

- If we only had micro data, we may want to just work with its log likelihood: (technically, this likelihood is conditional on the aggregate data)

$$\log \mathcal{L}(\theta, \delta) = \sum_{n \in \mathcal{N}_d} \log \mathbb{P}(t_n,\, j_n,\, k_n,\, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta, \delta)$$

Maximum Likelihood

- If we only had micro data, we may want to just work with its log likelihood: (technically, this likelihood is conditional on the aggregate data)

$$\log \mathcal{L}(\theta, \delta) = \sum_{n \in \mathcal{N}_d} \log \mathbb{P}(t_n,\, j_n,\, k_n,\, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta, \delta)$$

- This classic approach proceeds into two steps:

Maximum Likelihood

- If we only had micro data, we may want to just work with its log likelihood: (technically, this likelihood is conditional on the aggregate data)

$$\log \mathcal{L}(\theta, \delta) = \sum_{n \in \mathcal{N}_d} \log \mathbb{P}(t_n,\, j_n,\, k_n,\, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta, \delta)$$

- This classic approach proceeds into two steps:
  1. Find the $\hat{\theta} = (\hat{\Sigma}, \hat{\Pi})$ and mean utilities $\hat{\delta}$ that maximize $\log \mathcal{L}(\theta, \delta)$.

Maximum Likelihood

- If we only had micro data, we may want to just work with its log likelihood: (technically, this likelihood is conditional on the aggregate data)

$$\log \mathcal{L}(\theta, \delta) = \sum_{n \in \mathcal{N}_d} \log \mathbb{P}(t_n,\ j_n,\ k_n,\ y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta, \delta)$$

- This classic approach proceeds into two steps:
  1. Find the $\hat{\theta} = (\hat{\Sigma}, \hat{\Pi})$ and mean utilities $\hat{\delta}$ that maximize $\log \mathcal{L}(\theta, \delta)$.
  2. Run an IV regression of $\hat{\delta}_{jt}$ on x_{jt} to recover linear parameters $\hat{\beta}$ like in day 1.

Maximum Likelihood

- If we only had micro data, we may want to just work with its log likelihood: (technically, this likelihood is conditional on the aggregate data)

$$\log \mathcal{L}(\theta, \delta) = \sum_{n \in \mathcal{N}_d} \log \mathbb{P}(t_n,\, j_n,\, k_n,\, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta, \delta)$$

- This classic approach proceeds into two steps:
  1. Find the $\hat{\theta} = (\hat{\Sigma}, \hat{\Pi})$ and mean utilities $\hat{\delta}$ that maximize $\log \mathcal{L}(\theta, \delta)$.
  2. Run an IV regression of $\hat{\delta}_{jt}$ on x_{jt} to recover linear parameters $\hat{\beta}$ like in day 1.

- For a modern take on this "MLE" approach, see Grieco, Murry, Pinkse and Sagl (2025).
  - $\rightarrow$ Combine 1, 2, and the likelihood for aggregate market shares into a single objective.
  - $\rightarrow$ Their Julia package Grumps.jl efficiently handles the high-dimensional $\delta = \{\delta_{jt}\}_{j,t}$.

Optimal Micro Moments

- In micro BLP, optimal micro moments match the first-order conditions in MLE:

$$f^*(\overline{v}) = \frac{1}{N_d} \sum_{n \in \mathcal{N}_d} \frac{\partial \mathbb{P}(t_n,\, j_n,\, k_n,\, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta)}{\partial \theta}$$

Optimal Micro Moments

- In micro BLP, optimal micro moments match the first-order conditions in MLE:

$$f^*(\overline{v}) = \frac{1}{N_d} \sum_{n \in \mathcal{N}_d} \frac{\partial \mathbb{P}(t_n, j_n, k_n, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta)}{\partial \theta}$$

- These use all the information in a micro dataset (Conlon and Gortmaker, 2025).
  - $\rightarrow$ Intuition for statistical efficiency here is just that MLE is efficient.

Optimal Micro Moments

- In micro BLP, optimal micro moments match the first-order conditions in MLE:

$$f^*(\overline{v}) = \frac{1}{N_d} \sum_{n \in \mathcal{N}_d} \frac{\partial \mathbb{P}(t_n, j_n, k_n, y_{i_n t_n} \mid n \in \mathcal{N}_d; \theta)}{\partial \theta}$$

- These use all the information in a micro dataset (Conlon and Gortmaker, 2025).
  - $\rightarrow$ Intuition for statistical efficiency here is just that MLE is efficient.
- Can be a bit tricky to compute, but only a few lines of code with PyBLP.
  - $\rightarrow$ Like optimal IVs, can update along with the weighting matrix for a second GMM step.

Roadmap

Micro BLP Estimation

Choosing Micro Moments

Using More Information

Coding Exercise 3

Coding Exercise 3

- This is the end of the class, but try to do the third exercise on your own.
    1. Incorporating micro moments.
    2. Micro BLP estimation.
    3. Evaluating improvements to the price cut counterfactual.

Coding Exercise 3

- This is the end of the class, but try to do the third exercise on your own.
    1. Incorporating micro moments.
    2. Micro BLP estimation.
    3. Evaluating improvements to the price cut counterfactual.

- Also try to the supplemental exercises.
    - → Varying the market size.
    - → Optimal micro moments.
    - → Estimating a nesting parameter.

Coding Exercise 3

- This is the end of the class, but try to do the third exercise on your own.
  1. Incorporating micro moments.
  2. Micro BLP estimation.
  3. Evaluating improvements to the price cut counterfactual.

- Also try to the supplemental exercises.
  - $\rightarrow$ Varying the market size.
  - $\rightarrow$ Optimal micro moments.
  - $\rightarrow$ Estimating a nesting parameter.

- I'll post the remaining solutions today after questions.

Good luck with estimating your own demand systems!

j.gortmaker@nyu.edu

References I

Berry, Steven, James Levinsohn, and Ariel Pakes, “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.

Conlon, Christopher and Jeff Gortmaker, “Incorporating micro data into differentiated products demand estimation with PyBLP,” *Journal of Econometrics*, 2025, p. 105926.

Grieco, Paul LE, Charles Murry, Joris Pinkse, and Stephan Sagl, “Optimal estimation of discrete choice demand models with consumer and product data,” 2025.

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

References II

Pinkse, Joris and contributors, "Grumps.jl is a package for computing random coefficients demand models using consumer and product level data." Available at `https://github.com/NittanyLion/Grumps.jl`.

Reynolds, Graeme and Chris Walters, "The use of customer surveys for market definition and the competitive assessment of horizontal mergers," *Journal of Competition Law and Economics*, 2008, *4* (2), 411–431.

Zhang, Linqi, "Identification and estimation of market size in discrete choice demand models," 2024.