Demand Estimation

Mixtape Session

Jeff Gortmaker and Ariel Pakes

Last Class

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

- In each market $t \in \mathcal{T}$, individuals with types $i \in \mathcal{I}_t$ choose a $j \in \mathcal{J}_t \cup \{0\}$.

Last Class

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})}$$

- In each market $t \in \mathcal{T}$, individuals with types $i \in \mathcal{I}_t$ choose a $j \in \mathcal{J}_t \cup \{0\}$.
- Logit shocks ε_{ijt} give mixed (over individual types) logit market shares.

Last Class

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})}$$

- In each market $t \in \mathcal{T}$, individuals with types $i \in \mathcal{I}_t$ choose a $j \in \mathcal{J}_t \cup \{0\}$.
- Logit shocks ε_{ijt} give mixed (over individual types) logit market shares.
- On day 1, we set $\mu_{ijt} = 0$ to get a conveniently linear estimating equation:

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

Last Class

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})}$$

- In each market $t \in \mathcal{T}$, individuals with types $i \in \mathcal{I}_t$ choose a $j \in \mathcal{J}_t \cup \{0\}$.
- Logit shocks ε_{ijt} give mixed (over individual types) logit market shares.
- On day 1, we set $\mu_{ijt} = 0$ to get a conveniently linear estimating equation:

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

- Let's go over your first coding exercise.

Unrealistic Substitution Patterns

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

- In the price cut exercise, the pure logit model didn't perform well. Why?

Unrealistic Substitution Patterns

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

- In the price cut exercise, the pure logit model didn't perform well. Why?
- Last week we derived the own-price elasticity. What about the cross-price one?

$$\eta_{jkt} = \frac{\partial \log q_{jt}}{\partial \log p_{kt}} = \frac{\partial q_{jt}}{\partial p_{kt}} \frac{p_{kt}}{q_{jt}} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = -\alpha \cdot p_{kt} \cdot s_{kt}$$

Unrealistic Substitution Patterns

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

- In the price cut exercise, the pure logit model didn't perform well. Why?
- Last week we derived the own-price elasticity. What about the cross-price one?

$$\eta_{jkt} = \frac{\partial \log q_{jt}}{\partial \log p_{kt}} = \frac{\partial q_{jt}}{\partial p_{kt}} \frac{p_{kt}}{q_{jt}} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = -\alpha \cdot p_{kt} \cdot s_{kt}$$

- Doesn't depend on the characteristics of j!

  $\rightarrow$ Independence of Irrelevant Alternatives (IIA) property.

Red Bus/Blue Bus Problem

- Most industrial organization examples are about cereals or automobiles.

Red Bus/Blue Bus Problem

- Most industrial organization examples are about cereals or automobiles.
- There are two options: buying a car or a blue bus. Each has a 50% market share.

Red Bus/Blue Bus Problem

- Most industrial organization examples are about cereals or automobiles.
- There are two options: buying a car or a blue bus. Each has a 50% market share.
- Introduce a second bus, but it's red. Pure logit (IIA) predicts 33% market shares.
  - $\rightarrow$ In your exercise, consumers substituted *proportionally* from each cereal.

Red Bus/Blue Bus Problem

- Most industrial organization examples are about cereals or automobiles.
- There are two options: buying a car or a blue bus. Each has a 50% market share.
- Introduce a second bus, but it's red. Pure logit (IIA) predicts 33% market shares.
  - $\rightarrow$ In your exercise, consumers substituted *proportionally* from each cereal.
- In reality, we'd expect the car to still have 50% and each bus to have 25%.
  - $\rightarrow$ In your exercise, we'd hope for more substitution from more similar cereals.

Roadmap

- Preference Heterogeneity
- Mixed Logit Estimation
- Numerical Best Practices
- Differentiation Instruments
- Coding Exercise 2

Red Bus/Blue Bus Solution

- Our solution will be to re-introduce non-logit preference heterogeneity.

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

Red Bus/Blue Bus Solution

- Our solution will be to re-introduce non-logit preference heterogeneity.

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- This will allow 50% of consumers to really like cars and 50% to really like buses.
  $\rightarrow$ When a new bus is introduced, this doesn't really affect the car-lovers' choice.

Red Bus/Blue Bus Solution

- Our solution will be to re-introduce non-logit preference heterogeneity.

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}$$

- This will allow 50% of consumers to really like cars and 50% to really like buses.
  - → When a new bus is introduced, this doesn't really affect the car-lovers' choice.
- Want μ_{ijt} to dominate logit substitution from convenient but unrealistic ε_{ijt}.
  - → Want to add multiple dimensions of heterogeneity that really matter in our setting.

Random Coefficients

$$u_{ijt} = x_{jt}'\beta + \xi_{jt} + \varepsilon_{ijt}$$

- How to add preference heterogeneity to our pure logit model?
  - $\rightarrow$ For simplicity, I'll just let x_{jt} denote all characteristics, including prices p_{jt}.

Random Coefficients

$$u_{ijt} = x'_{jt}\beta_{it} + \xi_{jt} + \varepsilon_{ijt}$$

- How to add preference heterogeneity to our pure logit model?
  - $\rightarrow$ For simplicity, I'll just let x_{jt} denote all characteristics, including prices p_{jt}.
- Intuitively, we want to replace β with *random coefficients* β_{it}.
  - $\rightarrow$ *Random* in that they're drawn from a distribution of consumer types $i \in \mathcal{I}_t$.
  - $\rightarrow$ For $x_{jt} = \text{car}_{jt}$ and $\mathcal{I}_t = \{\text{car-lovers}, \text{bus-lovers}\}$, want $\beta_{it} \gg 0$ for car-lovers.

Random Coefficients

$$u_{ijt} = x'_{jt} \underbrace{(\beta + \Pi y_{it} + \Sigma \nu_{it})}_{\beta_{it}} + \xi_{jt} + \varepsilon_{ijt}$$

- How to add preference heterogeneity to our pure logit model?
  - → For simplicity, I'll just let x_{jt} denote all characteristics, including prices p_{jt}.
- Intuitively, we want to replace β with *random coefficients* β_{it}.
  - → *Random* in that they're drawn from a distribution of consumer types $i \in \mathcal{I}_t$.
  - → For $x_{jt} = \text{car}_{jt}$ and $\mathcal{I}_t = \{\text{car-lovers}, \text{bus-lovers}\}$, want $\beta_{it} \gg 0$ for car-lovers.
- Most common specification is $\beta_{it} \sim N(\beta + \Pi y_{it}, \Sigma\Sigma')$.
  - → Π shifts preferences according to "observed" demographics $y_{it} \sim$ census.
  - → Σ shifts preferences according to "unobserved" preferences $\nu_{it} \sim N(0, I)$.
  - → Σ is the *Cholesky root* of the variance matrix. Usually diagonal with standard deviations.

Random Coefficients

$$u_{ijt} = \underbrace{x'_{jt}\beta + \xi_{jt}}_{\delta_{jt}} + \underbrace{x'_{jt}(\Sigma\nu_{it} + \Pi y_{it})}_{\mu_{ijt}} + \varepsilon_{ijt}$$

- How to add preference heterogeneity to our pure logit model?
  - → For simplicity, I'll just let x_{jt} denote all characteristics, including prices p_{jt}.
- Intuitively, we want to replace β with *random coefficients* β_{it}.
  - → *Random* in that they're drawn from a distribution of consumer types $i \in \mathcal{I}_t$.
  - → For $x_{jt} = \text{car}_{jt}$ and $\mathcal{I}_t = \{\text{car-lovers}, \text{bus-lovers}\}$, want $\beta_{it} \gg 0$ for car-lovers.
- Most common specification is $\beta_{it} \sim N(\beta + \Pi y_{it}, \Sigma\Sigma')$.
  - → Π shifts preferences according to "observed" demographics $y_{it} \sim$ census.
  - → Σ shifts preferences according to "unobserved" preferences $\nu_{it} \sim N(0, I)$.
  - → Σ is the *Cholesky root* of the variance matrix. Usually diagonal with standard deviations.

Random Coefficients in Practice

$$u_{ijt} = \underbrace{x_{jt}'\beta + \xi_{jt}}_{\delta_{jt}} + \underbrace{x_{jt}'(\Sigma\nu_{it} + \Pi y_{it})}_{\mu_{ijt}} + \varepsilon_{ijt}$$

- In practice, we implement random coefficients by making a new dataset.
  - $\rightarrow$ In PyBLP lingo, “product data” rows are (j, t)’s, and new “agent data” rows are (i, t)’s.

Random Coefficients in Practice

$$u_{ijt} = \underbrace{x_{jt}'\beta + \xi_{jt}}_{\delta_{jt}} + \underbrace{x_{jt}'(\Sigma\nu_{it} + \Pi y_{it})}_{\mu_{ijt}} + \varepsilon_{ijt}$$

- In practice, we implement random coefficients by making a new dataset.
  - → In PyBLP lingo, "product data" rows are (j, t)'s, and new "agent data" rows are (i, t)'s.
- In your coding exercise, you'll just draw $|\mathcal{I}_t| = 100$ types per market.
  - → Draw $\nu_{it} \sim N(0, I)$ from a random number generator.
  - → Draw y_{it} from census data on demographics: income, etc.
  - → Each type is equally-likely, so use equal sampling weights $w_{it} = 1/|\mathcal{I}_t|$.

Random Coefficients in Practice

$$u_{ijt} = \underbrace{x_{jt}'\beta + \xi_{jt}}_{\delta_{jt}} + \underbrace{x_{jt}'(\Sigma\nu_{it} + \Pi y_{it})}_{\mu_{ijt}} + \varepsilon_{ijt}$$

- In practice, we implement random coefficients by making a new dataset.
  - → In PyBLP lingo, "product data" rows are (j, t)'s, and new "agent data" rows are (i, t)'s.
- In your coding exercise, you'll just draw $|\mathcal{I}_t| = 100$ types per market.
  - → Draw $\nu_{it} \sim N(0, I)$ from a random number generator.
  - → Draw y_{it} from census data on demographics: income, etc.
  - → Each type is equally-likely, so use equal sampling weights $w_{it} = 1/|\mathcal{I}_t|$.
- The goal is to have a dataset that reflects the *distribution* of individuals.
  - → Realism aside, this allows us to address distributional concerns.
  - → E.g. will a tax or price change affect high- or low-income individuals differently?

Roadmap

Preference Heterogeneity

Mixed Logit Estimation

Numerical Best Practices

Differentiation Instruments

Coding Exercise 2

From Linear Regression to GMM

$$\log \frac{s_{jt}}{s_{0t}} = \delta_{jt} = x_{jt}'\beta + \xi_{jt}$$

- In your exercise, you estimated β by running the above regression.
  - $\rightarrow$ Again, let x_{jt} include price, a constant, any other characteristics.
  - $\rightarrow$ Let z_{jt} include our price IV and exogenous characteristics in x_{jt}.

From Linear Regression to GMM

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

- In your exercise, you estimated β by running the above regression.
  - $\rightarrow$ Again, let x_{jt} include price, a constant, any other characteristics.
  - $\rightarrow$ Let z_{jt} include our price IV and exogenous characteristics in x_{jt}.

- Our exclusion restriction implies the moment condition $\mathbb{E}[\xi_{jt} \cdot z_{jt}] = 0$.

From Linear Regression to GMM

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

- In your exercise, you estimated β by running the above regression.
  - $\rightarrow$ Again, let x_{jt} include price, a constant, any other characteristics.
  - $\rightarrow$ Let z_{jt} include our price IV and exogenous characteristics in x_{jt}.

- Our exclusion restriction implies the moment condition $\mathbb{E}[\xi_{jt} \cdot z_{jt}] = 0$.

- We'd get the exact same $\hat{\beta}$ by optimizing the following GMM objective:

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

The BLP Contraction

- With preference heterogeneity, $\delta_{jt} = \log \frac{s_{jt}}{s_{0t}}$ no longer holds.

The BLP Contraction

- With preference heterogeneity, $\delta_{jt} = \log \frac{s_{jt}}{s_{0t}}$ no longer holds.
- Instead, given a guess of (Σ, Π), we numerically find the δ_{jt}'s that solve:

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + \mu_{ijt}(\Sigma, \Pi)]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + \mu_{ikt}(\Sigma, \Pi)]} \quad \text{for all} \quad j \in \mathcal{J}_t$$

The BLP Contraction

- With preference heterogeneity, $\delta_{jt} = \log \frac{s_{jt}}{s_{0t}}$ no longer holds.
- Instead, given a guess of (Σ, Π), we numerically find the δ_{jt}'s that solve:

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + \mu_{ijt}(\Sigma, \Pi)]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + \mu_{ikt}(\Sigma, \Pi)]} \quad \text{for all} \quad j \in \mathcal{J}_t$$

- Many ways to solve and speed up BLP's (1995) contraction.
  → PyBLP will take care of this, but see Conlon and Gortmaker (2020) if interested.

The BLP Contraction

- With preference heterogeneity, $\delta_{jt} = \log \frac{s_{jt}}{s_{0t}}$ no longer holds.
- Instead, given a guess of (Σ, Π), we numerically find the δ_{jt}'s that solve:

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp[\delta_{jt} + \mu_{ijt}(\Sigma, \Pi)]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\delta_{kt} + \mu_{ikt}(\Sigma, \Pi)]} \quad \text{for all} \quad j \in \mathcal{J}_t$$

- Many ways to solve and speed up BLP's (1995) contraction.
  - $\rightarrow$ PyBLP will take care of this, but see Conlon and Gortmaker (2020) if interested.
- BLP's (1995) big advancement was how to incorporate flexible preference heterogeneity.
  - $\rightarrow$ Built on simulation estimator advancements (Pakes and Pollard, 1989; McFadden, 1989).

The BLP Estimator

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

- BLP estimation consists of two nested loops.
  1. In the "outer" loop, we optimize over $\theta = (\beta, \Sigma, \Pi)$.
  2. In the "inner" loop, we solve the BLP contraction for $\delta_{jt}(\Sigma, \Pi)$.

The BLP Estimator

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

- BLP estimation consists of two nested loops.
  1. In the "outer" loop, we optimize over $\theta = (\beta, \Sigma, \Pi)$.
  2. In the "inner" loop, we solve the BLP contraction for $\delta_{jt}(\Sigma, \Pi)$.

- Actually, since $g(\theta)$ is linear in x_{jt}, we can "concentrate out" β and optimize (Σ, Π).
  - $\rightarrow$ Get $\hat{\beta}$ by running an IV regression of $\delta_{jt}(\Sigma, \Pi)$ on x_{jt}, like in the pure logit exercise.

The BLP Estimator

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

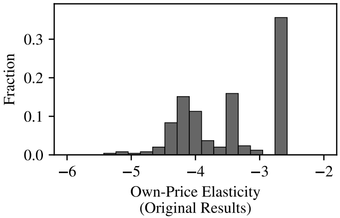
- BLP estimation consists of two nested loops.
  1. In the "outer" loop, we optimize over $\theta = (\beta, \Sigma, \Pi)$.
  2. In the "inner" loop, we solve the BLP contraction for $\delta_{jt}(\Sigma, \Pi)$.

- Actually, since $g(\theta)$ is linear in x_{jt}, we can "concentrate out" β and optimize (Σ, Π).
  $\rightarrow$ Get $\hat{\beta}$ by running an IV regression of $\delta_{jt}(\Sigma, \Pi)$ on x_{jt}, like in the pure logit exercise.

- What about the GMM weighting matrix W?
  $\rightarrow$ If you're just-identified ($\dim z_{jt} = \dim \theta$), it doesn't matter. You'll get a zero objective.
  $\rightarrow$ Otherwise, you may want to repeat optimization with optimal the two-step GMM $\hat{W}$.

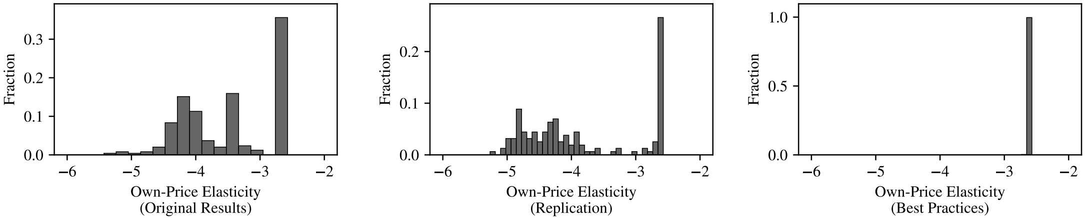
Roadmap

Motivation for Numerical Best Practices



- Variation in BLP estimates across different optimization algorithms and starting values has disillusioned some researchers (Knittel and Metaxoglou, 2014).

Motivation for Numerical Best Practices



- Variation in BLP estimates across different optimization algorithms and starting values has disillusioned some researchers (Knittel and Metaxoglou, 2014).
- But there are some numerical best practices that you can follow to avoid these kinds of issues (Conlon and Gortmaker, 2020).
  - → They're likely to be useful for most computation-heavy structural estimation, not just BLP!

Nonlinear Optimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q(\theta)$$

Nonlinear Optimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q(\theta)$$

- Set box constraints $\theta \in [\underline{\theta}, \overline{\theta}]$ to preclude unrealistic and unstable guesses of θ.
  - $\rightarrow$ E.g. huge Σ values can make the inner loop unstable.
  - $\rightarrow$ Economic intuition and initial estimates will give a sense for reasonable bounds.

Nonlinear Optimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q(\theta)$$

- Set box constraints $\theta \in [\underline{\theta}, \overline{\theta}]$ to preclude unrealistic and unstable guesses of θ.
- Check that 3-5 different starting values $\theta \sim U(\underline{\theta}, \overline{\theta})$ give the same $\hat{\theta}$.
  - $\rightarrow$ For 2-step GMM, do this twice, once for each step (6-10 jobs total).
  - $\rightarrow$ If you have access to a cluster, each can be a separate job, run in parallel.

Nonlinear Optimization

$$\hat{\theta} = \operatorname*{argmin}_{\theta} Q(\theta)$$

- Set box constraints $\theta \in [\underline{\theta}, \overline{\theta}]$ to preclude unrealistic and unstable guesses of θ.
- Check that 3-5 different starting values $\theta \sim U(\underline{\theta}, \overline{\theta})$ give the same $\hat{\theta}$.
- Prefer using gradient-based algorithms for "smooth" problems like BLP.
  - $\rightarrow$ Avoid derivative-free methods like Nelder-Mead/simplex, which tend to work worse.
  - $\rightarrow$ I prefer trust-region algorithms, e.g. SciPy's `trust-constr` or Knitro if you have it.

Nonlinear Optimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q(\theta)$$

- Set box constraints $\theta \in [\underline{\theta}, \overline{\theta}]$ to preclude unrealistic and unstable guesses of θ.
- Check that 3-5 different starting values $\theta \sim U(\underline{\theta}, \overline{\theta})$ give the same $\hat{\theta}$.
- Prefer using gradient-based algorithms for "smooth" problems like BLP.
- Try to terminate on strict first-order conditions, e.g. $||\text{gradient}||_\infty <$ `1e-8`.
  - → Inner loop should be tighter to prevent error "bubbling up." PyBLP default is very tight.
  - → Can also check second-order conditions, i.e. Hessian eigenvalues are positive.

Nonlinear Optimization

$$\hat{\theta} = \operatorname*{argmin}_{\theta} Q(\theta)$$

- Set box constraints $\theta \in [\underline{\theta}, \overline{\theta}]$ to preclude unrealistic and unstable guesses of θ.
- Check that 3-5 different starting values $\theta \sim U(\underline{\theta}, \overline{\theta})$ give the same $\hat{\theta}$.
- Prefer using gradient-based algorithms for "smooth" problems like BLP.
- Try to terminate on strict first-order conditions, e.g. $||\text{gradient}||_\infty < \texttt{1e-8}$.
- Configure your optimizer! Defaults may not work for your setting.

Numerical Integration

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

Numerical Integration

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \approx \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \,\mathrm{d}F(\mu_{it})$$

- Individual types i are typically an *approximation* to a population distribution.

Numerical Integration

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \approx \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \, \mathrm{d}F(\mu_{it})$$

- Individual types i are typically an *approximation* to a population distribution.
- Sometimes there are only a few types that we can integrate exactly.
  - $\rightarrow$ E.g. high- and low-income types $i \in \{1, 2\}$ with known shares w_{1t} and $w_{2t} = 1 - w_{1t}$.

Numerical Integration

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \approx \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \mathrm{d}F(\mu_{it})$$

- Individual types i are typically an *approximation* to a population distribution.
- Sometimes there are only a few types that we can integrate exactly.
- But usually we approximate the distribution with Monte Carlo integration.
  - → Use a random number generator (RNG) to draw $|\mathcal{I}_t| \approx$ 1,000 of (ν_{it}, y_{it})'s per market.
  - → Even better than your default RNG are quasi-Monte Carlo sequences.
  - → I recommend scrambled Halton sequences. R: Owen (2017). Python: SciPy or PyBLP.

Numerical Integration

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \approx \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \, \mathrm{d}F(\mu_{it})$$

- Individual types i are typically an *approximation* to a population distribution.
- Sometimes there are only a few types that we can integrate exactly.
- But usually we approximate the distribution with Monte Carlo integration.
- If you just need a few $\nu_{it} \sim N(0, I)$'s, try out Gauss-Hermite quadrature.
  - → 10-100× fewer carefully-chosen (w_{it}, ν_{it})'s that do just as well as Monte Carlo.
  - → Chosen to exactly integrate a polynomial expansion of the integrand.

Numerical Integration

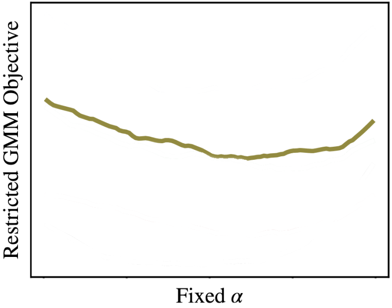
$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \cdot \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \approx \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} \, \mathrm{d}F(\mu_{it})$$

- Individual types i are typically an *approximation* to a population distribution.
- Sometimes there are only a few types that we can integrate exactly.
- But usually we approximate the distribution with Monte Carlo integration.
- If you just need a few $\nu_{it} \sim N(0, I)$'s, try out Gauss-Hermite quadrature.
- Keep increasing $|\mathcal{I}_t|$ until your estimates stabilize across draws/starting values.

What Typically Goes Wrong

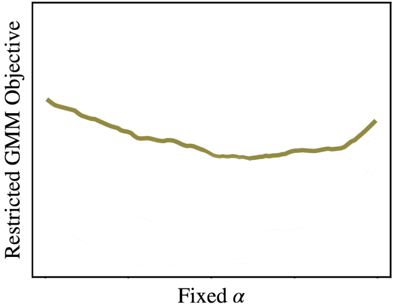
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)W g(\theta)'$ varies with θ.



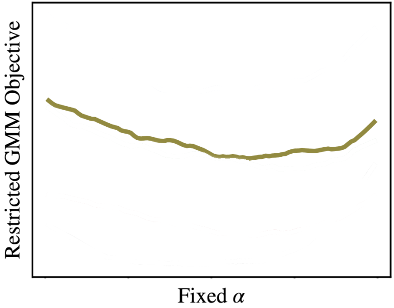
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)Wg(\theta)'$ varies with θ.
- Here, there's a minimum but also some challenges.



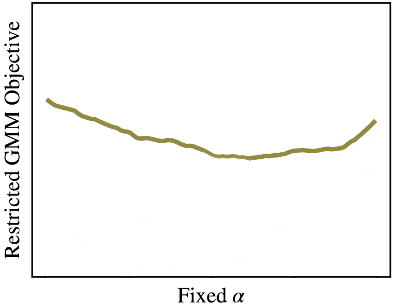
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)Wg(\theta)'$ varies with θ.
- Here, there's a minimum but also some challenges.
  - $\rightarrow$ Too few draws $|\mathcal{I}_t|$ makes the objective "choppy."



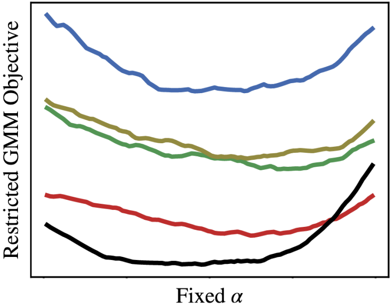
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)W g(\theta)'$ varies with θ.
- Here, there's a minimum but also some challenges.
  - $\rightarrow$ Too few draws $|\mathcal{I}_t|$ makes the objective "choppy."
  - $\rightarrow$ Poorly-configured optimizers can stop too early.



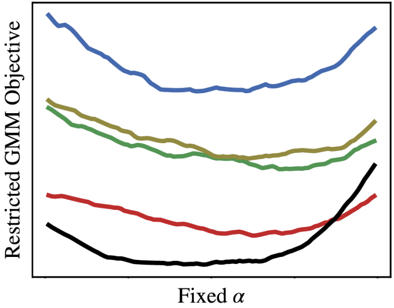
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta) W g(\theta)'$ varies with θ.
- Here, there's a minimum but also some challenges.
  - $\rightarrow$ Too few draws $|\mathcal{I}_t|$ makes the objective "choppy."
  - $\rightarrow$ Poorly-configured optimizers can stop too early.
- Different instruments give different objectives.



What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta) W g(\theta)'$ varies with θ.
- Here, there's a minimum but also some challenges.
  - $\rightarrow$ Too few draws $|\mathcal{I}_t|$ makes the objective "choppy."
  - $\rightarrow$ Poorly-configured optimizers can stop too early.
- Different instruments give different objectives.
  - $\rightarrow$ Even if they're all valid, some may be weaker.
  - $\rightarrow$ Weaker means flatter and harder to optimize.



Roadmap

Preference Heterogeneity

Mixed Logit Estimation

Numerical Best Practices

Differentiation Instruments

Coding Exercise 2

Adding Instruments

- For each new parameter in (Σ, Π), we need another instrument in z_{jt}.
  - $\rightarrow$ If you have fewer moments than parameters, you're *under-identified*.

Adding Instruments

- For each new parameter in (Σ, Π), we need another instrument in z_{jt}.
  - $\rightarrow$ If you have fewer moments than parameters, you're *under-identified*.

- In general, I recommend starting with one instrument per parameter.
  - $\rightarrow$ Try to choose an instrument that "targets" that parameter.
  - $\rightarrow$ For example, a single strong cost-shifter that "targets" α on p_{jt}.

Adding Instruments

- For each new parameter in (Σ, Π), we need another instrument in z_{jt}.
  - $\rightarrow$ If you have fewer moments than parameters, you're *under-identified*.
- In general, I recommend starting with one instrument per parameter.
  - $\rightarrow$ Try to choose an instrument that "targets" that parameter.
  - $\rightarrow$ For example, a single strong cost-shifter that "targets" α on p_{jt}.
- This makes your estimation strategy clear, and makes optimization easier.
  - $\rightarrow$ *Just-identified* models typically give $Q(\hat{\theta}) \approx 0$ at the optimum.
  - $\rightarrow$ This is regardless of your weighting matrix W, so you typically don't need 2-step GMM.

Adding Instruments

- For each new parameter in (Σ, Π), we need another instrument in z_{jt}.
  - $\rightarrow$ If you have fewer moments than parameters, you're *under-identified*.
- In general, I recommend starting with one instrument per parameter.
  - $\rightarrow$ Try to choose an instrument that "targets" that parameter.
  - $\rightarrow$ For example, a single strong cost-shifter that "targets" α on p_{jt}.
- This makes your estimation strategy clear, and makes optimization easier.
  - $\rightarrow$ *Just-identified* models typically give $Q(\hat{\theta}) \approx 0$ at the optimum.
  - $\rightarrow$ This is regardless of your weighting matrix W, so you typically don't need 2-step GMM.
- Later, adding more can help with weakness and testing exclusion restrictions.

Linear Regression Approximation

- There's a lot of confusion about what instruments are needed for BLP estimation.
  - → Identification of nonlinear models like BLP can be challenging.
  - → See Berry and Haile (2014, 2023) for a more formal, nonparametric framework.

Linear Regression Approximation

- There's a lot of confusion about what instruments are needed for BLP estimation.
  - → Identification of nonlinear models like BLP can be challenging.
  - → See Berry and Haile (2014, 2023) for a more formal, nonparametric framework.

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

$$u_{ijt} = (\beta + \sigma\nu_{it} + \pi y_{it})x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

Linear Regression Approximation

- There's a lot of confusion about what instruments are needed for BLP estimation.
  - → Identification of nonlinear models like BLP can be challenging.
  - → See Berry and Haile (2014, 2023) for a more formal, nonparametric framework.

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

$$u_{ijt} = (\beta + \sigma \nu_{it} + \pi y_{it}) x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- Salanié and Wolak (2022) approximate the BLP model around $\sigma, \pi \approx 0$:

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \underbrace{\sigma^2 d_{jt}^x + \pi m_t^y x_{jt} + \pi^2 v_t^y d_{jt}^x}_{\text{Defined on the next slide.}} + \xi_{jt}$$

Linear Regression Approximation

- There's a lot of confusion about what instruments are needed for BLP estimation.
  - → Identification of nonlinear models like BLP can be challenging.
  - → See Berry and Haile (2014, 2023) for a more formal, nonparametric framework.

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

$$u_{ijt} = (\beta + \sigma\nu_{it} + \pi y_{it})x_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

- Salanié and Wolak (2022) approximate the BLP model around $\sigma, \pi \approx 0$:

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \underbrace{\sigma^2 d_{jt}^x + \pi m_t^y x_{jt} + \pi^2 v_t^y d_{jt}^x}_{\text{Defined on the next slide.}} + \xi_{jt}$$

- Let's use our stronger intuition about linear regression to think about instruments!

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} = \beta x_{jt} + \xi_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} = \beta x_{jt} + \xi_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
  - → Use the same IV as before to target β: if $x_{jt} = p_{jt}$, a price IV; if exogenous, x_{jt} itself.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \Big(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\Big) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
  - $\rightarrow$ Can't use d_{jt}^x itself because it depends on endogenous market shares s_{kt}.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \left(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\right) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
  - → Can't use d_{jt}^x itself because it depends on endogenous market shares s_{kt}.
  - → Conventional choice was $\sum_{k \neq j} x_{kt}$, the BLP instruments from day 1.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \Big(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\Big) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
  - → Can't use d_{jt}^x itself because it depends on endogenous market shares s_{kt}.
  - → Conventional choice was $\sum_{k \neq j} x_{kt}$, the BLP instruments from day 1.
  - → A stronger choice is $\sum_{k \neq j} (x_{jt} - x_{kt})^2$ or similar from Gandhi and Houde (2020).

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \left(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\right) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
  - → Can't use d_{jt}^x itself because it depends on endogenous market shares s_{kt}.
  - → Conventional choice was $\sum_{k \neq j} x_{kt}$, the BLP instruments from day 1.
  - → A stronger choice is $\sum_{k \neq j} (x_{jt} - x_{kt})^2$ or similar from Gandhi and Houde (2020).
  - → We want cross-market choice set variation, otherwise d_{jt}^x is collinear with x_{jt}^2.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \pi m_t^y x_{jt} + \pi^2 v_t^y d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \Big(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\Big) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
- To target $\pi \neq 0$, we can interact x_{jt} with mean within-market income m_t^y.
  - → We want cross-market demographic variation, otherwise $m_t^y x_{jt}$ is collinear with x_{jt}.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \pi m_t^y x_{jt} + \pi^2 v_t^y d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \Big(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\Big) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
- To target $\pi \neq 0$, we can interact x_{jt} with mean within-market income m_t^y.
  - → We want cross-market demographic variation, otherwise $m_t^y x_{jt}$ is collinear with x_{jt}.
  - → Can technically identify π from higher-order variation, e.g. in variance v_t^y.

Linear Regression Intuition

$$\log \frac{s_{jt}}{s_{0t}} \approx \beta x_{jt} + \sigma^2 d_{jt}^x + \pi m_t^y x_{jt} + \pi^2 v_t^y d_{jt}^x + \xi_{jt} \quad \text{where} \quad d_{jt}^x = \Big(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} s_{kt} x_{kt}\Big) x_{jt}$$

- If we set $\sigma = \pi = 0$ like on day 1, we get our familiar pure logit regression.
- To target $\sigma \neq 0$, need a measure of how "differentiated" j is in terms of x_{jt} within t.
- To target $\pi \neq 0$, we can interact x_{jt} with mean within-market income m_t^y.
- In your exercise, you'll target (β, σ, π) with $z_{jt} = (x_{jt}, \sum_{k \neq j}(x_{jt} - x_{kt})^2, m_t^y x_{jt})$.
  - $\rightarrow$ If $x_{jt} = p_{jt}$, can replace x_{jt} with fitted values $\hat{p}_{jt}$ from the price IV's first stage.

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
  - $\rightarrow$ E.g. z_{jt} itself, z_{jt}^2, z_{jt}^3, or any function $f(z_{jt})$ of z_{jt}.

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
  - $\rightarrow$ E.g. z_{jt} itself, z_{jt}^2, z_{jt}^3, or any function $f(z_{jt})$ of z_{jt}.
- But adding a ton of instruments will bias your estimator.
  - $\rightarrow$ "Many weak IVs" problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
  - $\rightarrow$ Similar for nonlinear GMM (Han and Phillips, 2006; Newey and Windmeijer, 2009).

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
  - $\rightarrow$ E.g. z_{jt} itself, z_{jt}^2, z_{jt}^3, or any function $f(z_{jt})$ of z_{jt}.
- But adding a ton of instruments will bias your estimator.
  - $\rightarrow$ "Many weak IVs" problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
  - $\rightarrow$ Similar for nonlinear GMM (Han and Phillips, 2006; Newey and Windmeijer, 2009).
- Optimal IVs overweight observations with ξ_{jt} very sensitive to θ (Chamberlain, 1987):

$$f^*(z_{jt}) = \mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \theta'} \middle| z_{jt}\right]$$

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
  - $\rightarrow$ E.g. z_{jt} itself, z_{jt}^2, z_{jt}^3, or any function $f(z_{jt})$ of z_{jt}.
- But adding a ton of instruments will bias your estimator.
  - $\rightarrow$ “Many weak IVs” problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
  - $\rightarrow$ Similar for nonlinear GMM (Han and Phillips, 2006; Newey and Windmeijer, 2009).
- Optimal IVs overweight observations with ξ_{jt} very sensitive to θ (Chamberlain, 1987):

$$f^*(z_{jt}) = \mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \theta'} \middle| z_{jt}\right]$$

- Can be a bit tricky to compute, but with PyBLP it’s just one line of code.
  - $\rightarrow$ In practice, can update your IVs along with your weighting matrix for a second GMM step.

Roadmap

Preference Heterogeneity

Mixed Logit Estimation

Numerical Best Practices

Differentiation Instruments

Coding Exercise 2

Coding Exercise 2

- Try to do the second exercise before day 3's class, when I'll do it live.
    1. Incorporating preference heterogeneity.
    2. Mixed logit estimation.
    3. Evaluating improvements to the price cut counterfactual.

Coding Exercise 2

- Try to do the second exercise before day 3's class, when I'll do it live.
    1. Incorporating preference heterogeneity.
    2. Mixed logit estimation.
    3. Evaluating improvements to the price cut counterfactual.

- Think critically about the limitations of the model you estimate.
    → What dimensions of preference heterogeneity are missing?

Coding Exercise 2

- Try to do the second exercise before day 3's class, when I'll do it live.
    1. Incorporating preference heterogeneity.
    2. Mixed logit estimation.
    3. Evaluating improvements to the price cut counterfactual.

- Think critically about the limitations of the model you estimate.
    - $\rightarrow$ What dimensions of preference heterogeneity are missing?

- If you have time, try the supplemental exercises.
    - $\rightarrow$ Numerical integration alternatives.
    - $\rightarrow$ Optimal weights and instruments.
    - $\rightarrow$ Supply-side restrictions.

References I

Angrist, Joshua D, Guido W Imbens, and Alan B Krueger, “Jackknife instrumental variables estimation,” *Journal of Applied Econometrics*, 1999, *14* (1), 57–67.

Berry, Steven, James Levinsohn, and Ariel Pakes, “Automobile prices in market equilibrium,” *Econometrica*, 1995, *63* (4), 841–890.

Berry, Steven T and Philip A Haile, “Identification in differentiated products markets using market level data,” *Econometrica*, 2014, *82* (5), 1749–1797.

_ and _, “Nonparametric identification of differentiated products demand using micro data,” 2023.

Chamberlain, Gary, “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 1987, *34* (3), 305–334.

References II

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

Gandhi, Amit and Jean-François Houde, "Measuring substitution patterns in differentiated-products industries," 2020.

Han, Chirok and Peter CB Phillips, "GMM with many moment conditions," *Econometrica*, 2006, *74* (1), 147–192.

Knittel, Christopher R and Konstantinos Metaxoglou, "Estimation of random-coefficient demand models: Two empiricists' perspective," *Review of Economics and Statistics*, 2014, *96* (1), 34–59.

McFadden, Daniel, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 1989, pp. 995–1026.

References III

Newey, Whitney K and Frank Windmeijer, "Generalized method of moments with many weak moment conditions," *Econometrica*, 2009, *77* (3), 687–719.

Owen, Art B, "A randomized Halton algorithm in R," 2017.

Pakes, Ariel and David Pollard, "Simulation and the asymptotics of optimization estimators," *Econometrica*, 1989, pp. 1027–1057.

Salanié, Bernard and Frank A Wolak, "Fast, detail-free, and approximately correct: Estimating mixed demand systems," 2022.