

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 guides:
 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 guides:
 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.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

Model Overview

- BLP is a **full equilibrium model** for both demand *and* supply of discrete goods
 - **We'll focus on demand side** , but recall that prices are endogenous.
 - You will see the full model in Micro B with Anders (Bertrand price competition of multi-product producers).

Model Overview

- BLP is a **full equilibrium model** for both demand *and* supply of discrete goods
 - **We'll focus on demand side**, but recall that prices are endogenous.
 - You will see the full model in Micro B with Anders (Bertrand price competition of multi-product producers).
- Choices are made in **markets** denoted by $t \in \mathcal{T}$.
 - Time periods, geographic regions, etc.

Model Overview

- BLP is a **full equilibrium model** for both demand *and* supply of discrete goods
 - **We'll focus on demand side**, but recall that prices are endogenous.
 - You will see the full model in Micro B with Anders (Bertrand price competition of multi-product producers).
- Choices are made in **markets** denoted by $t \in \mathcal{T}$.
 - Time periods, geographic regions, etc.
- Each market has **individuals** with types denoted by $i \in \mathcal{I}_t$.
 - Different demographics and preferences.

Model Overview

- BLP is a **full equilibrium model** for both demand *and* supply of discrete goods
 - **We'll focus on demand side**, but recall that prices are endogenous.
 - You will see the full model in Micro B with Anders (Bertrand price competition of multi-product producers).
- Choices are made in **markets** denoted by $t \in \mathcal{T}$.
 - Time periods, geographic regions, etc.
- Each market has **individuals** with types denoted by $i \in \mathcal{I}_t$.
 - Different demographics and preferences.
- Individuals are faced with **choices** denoted by $j \in \mathcal{J}_t$.
 - Products, hospitals, candidates, etc.
 - 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} .
 - 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} .
 - 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} .
 - 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} .
 - 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

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

- Individuals choose an alternative to maximize (indirect) utility u_{ijt} .
 - 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} .
 - 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

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

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

Aggregate Market Shares

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

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

Aggregate Market Shares

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

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

Aggregate Market Shares

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

- Assume a convenient distribution for ε_{ijt} : iid type I extreme value.
→ “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.
- Problem: μ_{it} and s_{ijt} are typically unobserved.
Assume $\mu_{it} \sim F(\mu_{it} | \Sigma, \Pi)$ and aggregate over μ_{it}

$$s_{jt} = E(s_{ijt}) = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t$$

Aggregate Market Shares

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

- Assume a convenient distribution for ε_{ijt} : iid type I extreme value.
→ “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.
- Problem: μ_{it} and s_{ijt} are typically unobserved.
Assume $\mu_{it} \sim F(\mu_{it} | \Sigma, \Pi)$ and aggregate over μ_{it}

$$s_{jt} = E(s_{ijt}) = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t$$

- We'll match these to observed quantities \hat{s}_{jt} in our data.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

Pure Logit Model

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

- Start with the simplest case: no heterogenous utility. This is the model we worked with in Econometrics A. We will put μ_{ijt} back soon.

Pure Logit Model

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

- Start with the simplest case: no heterogenous utility. This is the model we worked with in Econometrics A. We will put μ_{ijt} back soon.
- Market shares simplify. No aggregation over individual types.

Pure Logit Model

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

- Start with the simplest case: no heterogenous utility. This is the model we worked with in Econometrics A. We will put μ_{ijt} back soon.
- Market shares simplify. No aggregation over individual types.
 - The 1 in the denominator is from our level normalization $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{0t} = 0$.

Pure Logit Model

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

- Start with the simplest case: no heterogeneous utility. This is the model we worked with in Econometrics A. We will put μ_{ijt} back soon.
- Market shares simplify. No aggregation over individual types.
 - The 1 in the denominator is from our level normalization $u_{i0t} = \varepsilon_{i0t}$, i.e. $\delta_{0t} = 0$.
- We can recover mean utilities from observed market shares (Berry, 1994).
 - 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: Grieco et al (2024)?
 - In Grieco et al (2024), products j are combinations of makes and models of cars; markets t are simply time.
 - 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: Grieco et al (2024)?
 - In Grieco et al (2024), products j are combinations of makes and models of cars; markets t are simply time.
 - 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} .
 - So p_{jt} is price of a car; x_{jt} includes a constant, a "horse power", "fuel efficiency", make dummies, etc.

Pure Logit Estimating Equation

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

- Running example: Grieco et al (2024)?
 - In Grieco et al (2024), products j are combinations of makes and models of cars; markets t are simply time.
 - 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} .
 - So p_{jt} is price of a car; x_{jt} includes a constant, a "horse power", "fuel efficiency", make dummies, etc.
- Interpret the regression error ξ_{jt} as unobserved product quality not in our data.
 - Unobserved characteristics, advertising, average taste variation, "demand shocks," etc.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

Endogeneity Concerns

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

- In Econometrics A, you ran an OLS regressions of δ_{jt} on p_{jt} and x_{jt} .

Endogeneity Concerns

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

- In Econometrics A, you ran an OLS regressions of δ_{jt} on p_{jt} and x_{jt} .
- Typically, we expect price to be strongly correlated with unobserved quality, ξ_{jt}
 - Firms know more than us about demand when setting prices.
 - Supply side lead to $\mathbb{C}(p_{jt}, \xi_{jt}) > 0$, so $\hat{\alpha} < 0$ is biased towards zero.

Endogeneity Concerns

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

- In Econometrics A, you ran an OLS regressions of δ_{jt} on p_{jt} and x_{jt} .
- Typically, we expect price to be strongly correlated with unobserved quality, ξ_{jt}
 - Firms know more than us about demand when setting prices.
 - Supply side lead to $\mathbb{C}(p_{jt}, \xi_{jt}) > 0$, so $\hat{\alpha} < 0$ is biased towards zero.
- Potential solutions IV or Fixed Effects

Fixed Effects

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

- Adding product and market fixed effects to x_{jt} can eliminate a lot of bias.
 - E.g. if p_{jt} is correlated with fixed effects ξ_j and/or ξ_t in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
 - We 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 eliminate a lot of bias.
 - E.g. if p_{jt} is correlated with fixed effects ξ_j and/or ξ_t in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
 - We need multiple observations per product and market to add ξ_j and ξ_t .
- Modern grocery scanner datasets have many thousands of products/markets.
 - Dummies take too much memory, so we “absorb” them, i.e. de-mean using within transformation.

Fixed Effects

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

- Adding product and market fixed effects to x_{jt} can eliminate a lot of bias.
 - E.g. if p_{jt} is correlated with fixed effects ξ_j and/or ξ_t in $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$.
 - We need multiple observations per product and market to add ξ_j and ξ_t .
- Modern grocery scanner datasets have many thousands of products/markets.
 - Dummies take too much memory, so we “absorb” them, i.e. de-mean using within transformation.
- Helpful but insufficient: ξ_{jt} typically varies by product *and* market, e.g. $\mathbb{C}(p_{jt}, \Delta\xi_{jt}) > 0$.

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.
→ 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.
 - 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} .
 - Does the sign of the coefficient on z_{jt} make sense?
 - 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.
 - 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} .
 - Does the sign of the coefficient on z_{jt} make sense?
 - 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.
→ 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.
 - 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.

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.
- **BLP**: Average characteristics x_{kt} of *competing* products $k \neq j$.

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.
- **BLP**: Average characteristics x_{kt} of *competing* products $k \neq j$.
 - Characteristics of competing products affect markups.
 - We'll come back to these later, since they can also serve a different purpose.
- I recommend starting with just one. A straightforward cost-shifter if you have it.

Empirical Example: Demand for cars (Grieco et al. QJE, 2024)

- Assume $\mu_{ijt} = 0$ and estimate **pure logit demand model for cars** with OLS and IV:

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

- x_{jt} : Observed product characteristics
 - Variables in logs (height, hp, mpg, weight, footprint, number of trims)
 - Variables in levels (release year, years since design)

Empirical Example: Demand for cars (Grieco et al. QJE, 2024)

- Assume $\mu_{ijt} = 0$ and estimate **pure logit demand model for cars** with OLS and IV:

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

- x_{jt} : Observed product characteristics
 - Variables in logs (height, hp, mpg, weight, footprint, number of trims)
 - Variables in levels (release year, years since design)
- **Regression specifications:**
 - Baseline: $\delta_{jt} = \alpha p_{jt} + \text{const} + \xi_{jt}$
 - Controls: Adding x_{jt} , car type, year, and make fixed effects (dummies)

Empirical Example: Demand for cars (Grieco et al. QJE, 2024)

- Assume $\mu_{ijt} = 0$ and estimate **pure logit demand model for cars** with OLS and IV:

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

- x_{jt} : Observed product characteristics
 - Variables in logs (height, hp, mpg, weight, footprint, number of trims)
 - Variables in levels (release year, years since design)
- **Regression specifications:**
 - Baseline: $\delta_{jt} = \alpha p_{jt} + \text{const} + \xi_{jt}$
 - Controls: Adding x_{jt} , car type, year, and make fixed effects (dummies)
- **Endogeneity problem:**
 - Price p_{jt} may be correlated with ξ_{jt} , biasing OLS estimates.
 - Use **IV regression** with exchange rate instrument (RXR) for price.

Let's Code!



- **Jupyter Notebook:** [15_blp.ipynb](#)
- **Part 1:** IV estimation with $\mu_{ijt} = 0$ and real exchange rate as instrument for price

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

Random utility model for panel data

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

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

Random utility model for panel data

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Longrightarrow \quad s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi)$$

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

Random utility model for panel data

$$\max_{j \in \mathcal{J}_t \cup \{0\}} u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \quad \Rightarrow \quad s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi)$$

- In each market $t \in \mathcal{T}$, individuals with types $i \in \mathcal{I}_t$ choose a $j \in \mathcal{J}_t \cup \{0\}$.
- Extreme value shocks ε_{ijt} give logit market shares (conditional on individual types)
→ Mixed Logit aggregate market shares
- Before we set $\mu_{ijt} = 0$ to get a conveniently linear estimating equation:

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

but now, δ_{jt} is implicitly given by s_{jt} as a fixed point on a non-linear equation.

Random Coefficients

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

- 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} .

Random Coefficients

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

- 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.

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 $\mathcal{I}_t = \{\text{car-lovers, bus-lovers}\}$, 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 $y_{it} \sim \text{census}$.
 - 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 $\mathcal{I}_t = \{\text{car-lovers, bus-lovers}\}$, 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 $y_{it} \sim \text{census}$.
 - 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.

Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

From Linear Regression to GMM

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

- For μ_{ijt} it is easy to estimate β by running the above regression.
 - Again, let x_{jt} include price, a constant, any other characteristics.
 - Let z_{jt} include our price IV and exogenous characteristics in x_{jt} .

From Linear Regression to GMM

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

- For μ_{ijt} it is easy to estimate β by running the above regression.
 - Again, let x_{jt} include price, a constant, any other characteristics.
 - 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

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

- For μ_{ijt} it is easy to estimate β by running the above regression.
 - Again, let x_{jt} include price, a constant, any other characteristics.
 - 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:

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

Aggregate market shares

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

Aggregate market shares

- 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:

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t$$

Aggregate market shares

- 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:

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t$$

- This effectively makes $\delta_{jt} = \delta_{jt}(\Sigma, \Pi)$ an *implicit function* of Σ and Π
- Need to approximate integral (Monte Carlo or Quadrature) and solve for δ_{jt}
- BLP's (1995) big advancement was how to incorporate flexible preference heterogeneity.
 - Built on simulation estimator advancements (Pakes and Pollard, 1989; McFadden, 1989).

Approximating the Integral using Monte Carlo

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi)$$

1. Draw individual preferences from $F(\mu_{it} | \Sigma, \Pi)$:

$$\mu_{ijt}^{(m)} = x'_{jt}(\Sigma \nu_{it}^{(m)} + \Pi y_{it}^{(m)}), \quad m = 1, \dots, M$$

where $\nu_{it}^{(m)} \sim N(0, I)$ and $y_{it}^{(m)}$ comes from census data.

Approximating the Integral using Monte Carlo

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi)$$

1. Draw individual preferences from $F(\mu_{it} | \Sigma, \Pi)$:

$$\mu_{ijt}^{(m)} = x'_{jt}(\Sigma \nu_{it}^{(m)} + \Pi y_{it}^{(m)}), \quad m = 1, \dots, M$$

where $\nu_{it}^{(m)} \sim N(0, I)$ and $y_{it}^{(m)}$ comes from census data.

2. Approximate the integral with the **sample average**:

$$s_{jt} \approx \frac{1}{M} \sum_{m=1}^M \frac{\exp(\delta_{jt} + \mu_{ijt}^{(m)})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt}^{(m)})}$$

Approximating the Integral using Monte Carlo

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi)$$

1. Draw individual preferences from $F(\mu_{it} | \Sigma, \Pi)$:

$$\mu_{ijt}^{(m)} = x'_{jt}(\Sigma \nu_{it}^{(m)} + \Pi y_{it}^{(m)}), \quad m = 1, \dots, M$$

where $\nu_{it}^{(m)} \sim N(0, I)$ and $y_{it}^{(m)}$ comes from census data.

2. Approximate the integral with the **sample average**:

$$s_{jt} \approx \frac{1}{M} \sum_{m=1}^M \frac{\exp(\delta_{jt} + \mu_{ijt}^{(m)})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt}^{(m)})}$$

- Better alternatives: **Quasi-Monte Carlo (QMC)** methods like **Sobol** or **Halton** sequences.
 - Converge faster than standard random sampling.
 - Reduce variance of integral approximation.

The BLP Contraction

- Given an estimate \hat{s}_{jt} of s_{jt} and a guess of (Σ, Π) we could use Newtons Method to numerically solve the system of equations

$$\hat{s}_{jt} = s_{jt}(\Sigma, \Pi) \equiv \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t \text{ and } t \in \mathcal{T}$$

The BLP Contraction

- Given an estimate \hat{s}_{jt} of s_{jt} and a guess of (Σ, Π) we could use Newtons Method to numerically solve the system of equations

$$\hat{s}_{jt} = s_{jt}(\Sigma, \Pi) \equiv \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it} | \Sigma, \Pi) \quad \text{for all } j \in \mathcal{J}_t \text{ and } t \in \mathcal{T}$$

- Alternatively we can use the method of **successive approximations** to solve for the fixed point on the **BLP contraction mapping**, $\Gamma_{(\Sigma, \Pi)}$:

$$\delta_{jt}^{k+1} = \Gamma_{(\Sigma, \Pi)}(\delta_{jt}^k) = \delta_{jt}^k + \log(\hat{s}_{jt}) - \log(s_{jt}(\Sigma, \Pi)) \quad \text{for all } j \in \mathcal{J}_t \text{ and } t \in \mathcal{T}$$

- Since $\Gamma_{(\Sigma, \Pi)}$ is a contraction, successive approximation will always find the *implicit function* $\delta_{jt}(\Sigma, \Pi)$ as the **unique fixed point** on the BLP contraction operator $\Gamma_{(\Sigma, \Pi)}$, i.e. where $\delta_{jt} = \Gamma_{(\Sigma, \Pi)}(\delta_{jt})$

The BLP Estimator

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

- 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

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

- 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) .
 - 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

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

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

Nonlinear Optimization

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

Nonlinear Optimization

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

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

Nonlinear Optimization

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

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

Nonlinear Optimization

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

- Set **box constraints** $\theta \in [\underline{\theta}, \bar{\theta}]$ to preclude unrealistic and unstable guesses of θ .
- Check that 3-5 **different starting values** $\theta \sim U(\underline{\theta}, \bar{\theta})$ 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. $\|\text{gradient}\|_{\infty} < 1\text{e-}8$.
 - Inner loop should be tighter to prevent error “bubbling up.”
 - Can also check second-order conditions, i.e. Hessian eigenvalues are positive.

Nonlinear Optimization

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

- Set **box constraints** $\theta \in [\underline{\theta}, \bar{\theta}]$ to preclude unrealistic and unstable guesses of θ .
- Check that 3-5 **different starting values** $\theta \sim U(\underline{\theta}, \bar{\theta})$ 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. $\|\text{gradient}\|_{\infty} < 1\text{e-}8$.
- **Configure your optimizer!** Defaults may not work for your setting.

Numerical Integration

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

Numerical Integration

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

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

Numerical Integration

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

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

Numerical Integration

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

- 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 $M \approx 1,000$ of (ν_{it}, y_{it}) 's per market.
 - Even better than your default RNG are quasi-Monte Carlo sequences.
 - I recommend scrambled Sobol sequences. Python: Chaospy.

Numerical Integration

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

- 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

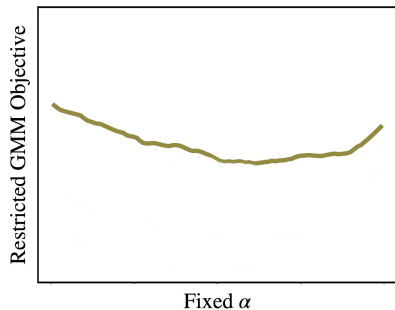
$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k \in \mathcal{J}_t} \exp(\delta_{kt} + \mu_{ikt})} dF(\mu_{it})$$

- 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 M** until your estimates stabilize across draws/starting values.

What Typically Goes Wrong

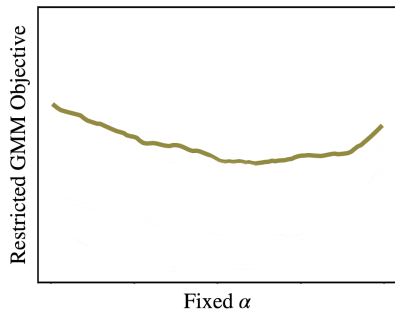
What Typically Goes Wrong

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



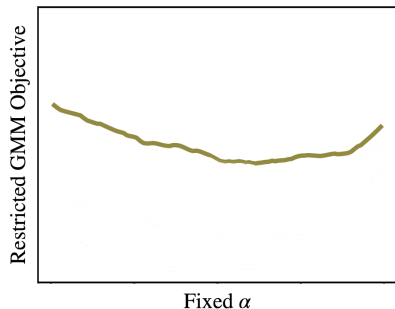
What Typically Goes Wrong

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



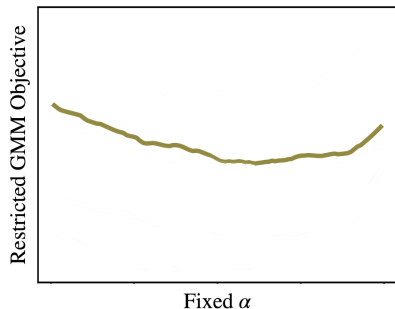
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)Wg(\theta)'$ varies with θ .
- Here, there's a minimum but also some challenges.
 - Too few draws M makes the objective “choppy.”



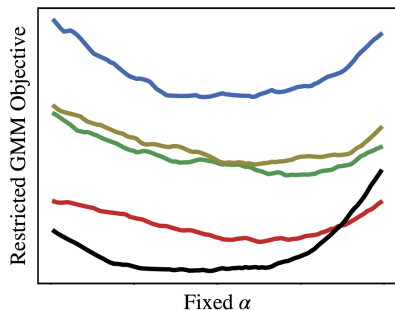
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)Wg(\theta)'$ varies with θ .
- Here, there's a minimum but also some challenges.
 - Too few draws M makes the objective “choppy.”
 - Poorly-configured optimizers can stop too early.



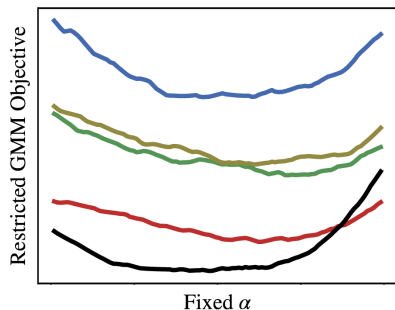
What Typically Goes Wrong

- Can see how $Q(\theta) = g(\theta)Wg(\theta)'$ varies with θ .
- Here, there's a minimum but also some challenges.
 - Too few draws M makes the objective “choppy.”
 - Poorly-configured optimizers can stop too early.
- Different instruments give different objectives.



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.
 - Too few draws M makes the objective “choppy.”
 - Poorly-configured optimizers can stop too early.
- Different instruments give different objectives.
 - Even if they're all valid, some may be weaker.
 - Weaker means flatter and harder to optimize.



Roadmap

The BLP Model

Pure Logit Estimation

Price Endogeneity

Preference Heterogeneity

Mixed Logit Estimation

Differentiation Instruments

Adding Instruments

- For each new parameter in (Σ, Π) , we need another instrument in z_{jt} .
 - 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} .
 - If you have fewer moments than parameters, you're *under-identified*.
- In general, I recommend starting with one instrument per parameter.
 - Try to choose an instrument that “targets” that parameter.
 - 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} .
 - If you have fewer moments than parameters, you're *under-identified*.
- In general, I recommend starting with one instrument per parameter.
 - Try to choose an instrument that “targets” that parameter.
 - For example, a single strong cost-shifter that “targets” α on p_{jt} .
- This makes your estimation strategy clear, and makes optimization easier.
 - *Just-identified* models typically give $Q(\hat{\theta}) \approx 0$ at the optimum.
 - 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} .
 - If you have fewer moments than parameters, you're *under-identified*.
- In general, I recommend starting with one instrument per parameter.
 - Try to choose an instrument that “targets” that parameter.
 - For example, a single strong cost-shifter that “targets” α on p_{jt} .
- This makes your estimation strategy clear, and makes optimization easier.
 - *Just-identified* models typically give $Q(\hat{\theta}) \approx 0$ at the optimum.
 - 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

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

- 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} .

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

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

- 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).

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

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

- 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 .
- 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

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

- 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 .
- 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

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

- 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 .
- 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 $z_{jt} = (x_{jt}, \sum_{k \neq j} (x_{jt} - x_{kt})^2, m_t^y x_{jt})$.
→ 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 $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
→ 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 $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
 - 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.
 - “Many weak IVs” problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
 - Similar for nonlinear GMM (Han and Phillips, 2006; Newey and Windmeijer, 2009).

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
 - 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.
 - “Many weak IVs” problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
 - 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):

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

Optimal Instruments

- There are many valid instruments that satisfy exclusion restrictions $\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$.
 - 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.
 - “Many weak IVs” problem is well-known for 2SLS (Angrist, Imbens and Krueger, 1999).
 - 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):

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

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

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**, “Estimating discrete-choice models of product differentiation,” *RAND Journal of Economics*, 1994, pp. 242–262.
- , **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.
- Berry, Steven T and Philip A Haile**, “Identification in differentiated products markets using market level data,” *Econometrica*, 2014, 82 (5), 1749–1797.

References II

- **and** — , “Foundations of demand estimation,” in “Handbook of industrial organization,” Vol. 4 2021, pp. 1–62.
- **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.
- Conlon, Christopher and Jeff Gortmaker**, “Best practices for differentiated products demand estimation with PyBLP,” *RAND Journal of Economics*, 2020, 51 (4), 1108–1161.
- **and** — , “Incorporating micro data into differentiated products demand estimation with PyBLP,” 2023.

References III

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

- Newey, Whitney K and Frank Windmeijer**, “Generalized method of moments with many weak moment conditions,” *Econometrica*, 2009, 77 (3), 687–719.
- Pakes, Ariel and David Pollard**, “Simulation and the asymptotics of optimization estimators,” *Econometrica*, 1989, pp. 1027–1057.
- Petrin, Amil**, “Quantifying the benefits of new products: The case of the minivan,” *Journal of Political Economy*, 2002, 110 (4), 705–729.
- Salanié, Bernard and Frank A Wolak**, “Fast, detail-free, and approximately correct: Estimating mixed demand systems,” 2022.