

# Discrete Choice Model

## DCM

Lecturer: LI, Mingyi

June 29, 2024

# Outline

- 1 Introduction
- 2 Logit
- 3 Generalized Extreme Value Models (GEV)  
Nested Logit
- 4 Random Utility Model (RUM)
- 5 BLP

# Outline

- 1 Introduction
- 2 Logit
- 3 Generalized Extreme Value Models (GEV)  
Nested Logit
- 4 Random Utility Model (RUM)
- 5 BLP

# Introduction

- This is a demand-side model, which we use to estimate the demand
  - We cannot get “demand” in the usual sense, because demand is a random variable. The “best” we can do is the conditional expectation of individual demand
- When and how to use a Discrete Choice Model?
  - case 1. Ignoring the supply side
  - case 2. Modelling consumer demand while simplifying the supply side using an IO model (such as perfectly competitive or a monopolistic competitive market setting)
- Some examples. Farronato and Fradkin (2022); Nevo (2001); Berry et al. (1995); Train (1998)

# Outline

## Logit

- Setup
  - A decision maker  $n$  faces  $J$  alternatives,

$$U_{nj} = V_{nj} + \epsilon_{nj}$$

- $\epsilon_{nj}$  follows type-I extreme value distribution ( $\epsilon_{nj} \sim \text{Gumbel}(0, 1)$ ), i.e.,  $F(\epsilon_{nj}) = \exp(-\exp(-\epsilon_{nj}))$ , of which the variance is  $\pi^2/6$
- Assuming variance equaling  $\pi^2/6$  implicitly normalizes the scale of utility (homothetic property)
- $\epsilon_{nj}$  is i.i.d. for any  $n$  and  $j$

# Type-I Extreme Value Distribution

- CDF

$$\text{Gumbel}(\mu, \beta) : F(x) = e^{-e^{-(x-\mu)/\beta}}$$

where  $x \in R$

- location parameter,  $\mu$ , and scale parameter  $\beta > 0$
- Graph exposition (see python codes)

# Choice Probabilities

- Consumer  $n$ 's demand for an alternative  $i \in J$

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} \quad (1)$$



# Choice Probabilities

- Consumer  $n$ 's demand for an alternative  $i \in J$

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} \quad (1)$$

- Proof hint

$$\begin{aligned} \Pr(n \text{ chooses } i) &= \Pr(U_{ni} > U_{nj}, \forall j \neq i) \\ &= \Pr(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}, \forall j \neq i) \\ &= \prod_{j \neq i} \Pr(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}) \\ &= \int_{-\infty}^{\infty} \prod_{j \neq i} \Pr(\epsilon_{nj} < V_{ni} - V_{nj} + \epsilon) dF(\epsilon) \end{aligned}$$

where  $F(\epsilon)$  is the CDF of Gumbel(0,1). In the equations above, we use following theorems

- i.i.d. property; total probability theorem  
 $(\Pr(A) = \int_{\Omega} \Pr(A|X=x) dF(x))$ . In our case, event  
 $A = \prod_{j \neq i} \Pr(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj})$



## Choice Probabilities (Cont'd)

- Properties

- the choice probabilities for all alternatives sum to one for one decision maker
- Independence from Irrelevant Alternatives (IIA)

$$\frac{P_{ni}}{P_{nk}} = \frac{\exp(V_{ni} / \sum_j \exp(V_{nj}))}{\exp(V_{nk} / \sum_j \exp(V_{nj}))} = \exp(V_{ni} - V_{nk})$$

where the ratio does not depend on any alternatives other than  $i$  and  $k$

- Proportional substitution

$$\frac{\partial P_{ni} / P_{ni}}{\partial z_{nj} / z_{nj}} = E_{iz_{nj}} = -\beta_z z_{nj} P_{nj}$$

where  $z_{nj}$  is the attribute of alternative  $j$  as faced by person  $n$  and  $\beta_z$  is its coefficient. This cross-elasticity is the same,  $\forall i$

- The log-likelihood function is globally concave in parameters  $\beta$ , when the utility function is linear in parameters:  $V_{nj} = \beta' x_{nj}$   
MCFADDEN (1974)

# Panel Data Application

- Utility function.  $U_{njt} = V_{njt} + \epsilon_{njt}, \forall j, t$
- Choice probabilities.  $P_{nit} = \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}}$
- Adding dynamic patterns
  - If representative utility for each period is specified to depend on variables for that period; for example,  $V_{njt} = \beta' x_{njt}$ , then there is essentially no difference between the logic model with panel data and with purely cross-sectional data
  - Dynamic aspects of behavior can be captured by specifying representative utility in each period to depend on observed variables from other period. For example, a lagged price response is represented by entering the price in period  $t - 1$  as an explanatory variable in the utility for period  $t$

# Consumer Surplus Calculation

- Exercise target. Estimate the change in consumer surplus that is associated with a particular policy
- Consumer surplus.  $CS_n = (1/\alpha_n) \times \max_j(U_{nj})$ , where  $\alpha_n$  is the marginal utility of income,  $\frac{dU_n}{dY_n} = \alpha_n$ , with  $Y_n$  the income of person  $n$ .
- The expected consumer surplus

$$\begin{aligned}
 E(CS_n) &= \frac{1}{\alpha_n} E[\max_j (V_{nj} + \epsilon_{nj})] \\
 &= \frac{1}{\alpha_n} \ln\left(\sum_j e^{V_{nj}}\right) + C.
 \end{aligned}$$

Prove the Emax term is a log-sum form.

## Consumer Surplus Calculation (Cont'd)

- For example, the change in consumer surplus that results from a change in the alternatives and/or the choice set

$$\Delta E(CS_n) = \frac{1}{\alpha_n} [\ln(\sum_j^{J^1} \exp(V_{nj}^1)) - \ln(\sum_j^{J^0} \exp(V_{nj}^0))]$$

where the superscripts 0 and 1 refer to before and after the change.

- Total consumer surplus
  - $E(CS_n)$  is the average consumer surplus in the subpopulation of people who have the same representative utilities as person  $n$ . The total consumer surplus in the population is calculated as the weighted sum of  $E(CS_n)$  over a sample of decision makers, with the weights reflecting the numbers of people in the population who face the same representative utilities as the sampled person

## Estimation: Maximum Likelihood Estimates (MLE)

- Intuition. Observed sample has the largest likelihood to happen given population distribution
- Likelihood function of independent sample (joint probability)

$$L(\beta) = \prod_{n=1}^N \prod_i (P_{ni})^{y_{ni}}$$

where  $\beta$  is a vector containing the parameters of the model.

- Log-likelihood function

$$LL(\beta) = \sum_{n=1}^N \sum_i y_{ni} \ln P_{ni}$$

and the estimator is the value of  $\beta$  that maximizes this function; its derivative with respect to each of the parameters is zero. The MLE are therefore the values of  $\beta$  that satisfy this first-order condition

# Optimization

- One-dimensional minimization

$$\min_{x \in R} f(x) \quad (2)$$

where  $f : R \rightarrow R$

- Multi-dimensional minimization

$$\min_{x \in R} f(x) \quad (3)$$

where  $f : R^n \rightarrow R$

- Gradient

$$\nabla f(x) = (\partial f(x)/\partial x_1, \partial f(x)/\partial x_2, \dots, \partial f(x)/\partial x_n)$$

- Hessian matrix

$$H(x) = \left( \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j=1}^n$$



# Optimization: Steepest Descent

- Steepest descent
  - the steepest ascent direction is gradient
  - in the steepest method the search direction is which  $f$  falls most rapidly per unit length
  - the steepest method chooses the search direction:

$$s^{k+1} = -(\nabla f(x^k))'$$

- Graph interpretation:  
one-dimensional  $y = f(x)$ ; two-dimensional  $z = f(x, y)$
- Gradient only points out the direction but ignores the step length

## Examples: Compare Steepest Descent and Newton's Method

- **Example 1:**  $y = x^2$ 
  - Gradient  $\nabla f(x) = 2x$  points out the direction.
  - Let  $x_0 = 3$  or  $x_0 = 5$ .
  - Gradient descent method, the step is  $-6$  and  $-10$ , respectively.
  - So  $x_1 = -3$  or  $-5$ , not closer to the solution. Gradient descent method fails.
  - 
  - Newton's method.  $H(\cdot) = 2$ ; the step size is  $-3$  and  $-5$ .
  - One iteration obtains the solution.
  - **Theorem:** Newton's method reaches the optimum in one iteration (since a 2-degree Taylor expansion of a quadratic form is an accurate approximation).
- **Example 2:**  $y = x^3$ 
  - What about cubic form?
  - Similarly,  $\nabla f(x) = 3x^2$ ,  $H(x) = 6x$ .
  - Let  $x_0 = 5$ .  $\nabla f(5) = 75$ ,  $H(5) = 30$ .
  - Newton's method: Step size  $= -75/30 = -2.5$ .
  - By iteration:  $x_1 = 2.5$ ,  $x_2 = 1.25$ ,  $x_3 = 0.625$ ,  $x_4 = 0.3125$ , ...

# Optimization: Newton's Method (One-Dimension)

- Taylor's expansion
  - for  $C^2$  functions  $f(x)$ , Newton's method is often used. We define:

$$p(x) \equiv f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2$$

- The idea behind Newton's method is to start at a point  $a$  and find quadratic polynomials  $p(x)$ , that approximate  $f(x)$  at  $a$  to the 2nd degree. We next approximately minimize  $f(\cdot)$  by finding a point  $x_m$  that minimizes  $p(x)$ .
- Newton's method iteration process

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

- Convergence: quadratic convergence when approaching  $x^*$ :

$$\lim_{n \rightarrow \infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^2} = \frac{1}{2} \left| \frac{f'''(x^*)}{f''(x^*)} \right| < \infty$$

- proof hint. Expand  $f'(x)$  at  $x_0$  and set  $x_0 = x_n$

# Optimization: Newton's Method (Multi-Dimension)

- Taylor's expansion

$$f(x) \doteq f(x^k) + \nabla f(x^k)(x - x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k)$$

- Newton's method iteration process

$$x^{k+1} = x^k - H^{-1}(x^k)(\nabla f(x^k))^T$$

- Add line search in Newton's method
  - find iterating step  $s = -H^{-1}(x^k)(\nabla f(x^k))^T$
  - solve  $\lambda_k = \arg \min_{\lambda} f(x^k + \lambda s^k)$
  - iterate  $x^{k+1} = x^k + \lambda_k s^k$
  - intuition:  $\lambda_k$  modifies the step length in each iteration, ensuring the largest decrease in the target function.
  - Newton's method (without line search) implicitly sets step length,  $\lambda_k = 1$

# Optimization: Quasi Newton's Method

- Drawback of Newton's method
  - Newton's method needs to calculate Hessian in every iteration
- BFGS method
  - initially set weight matrix to be a diagonal  $I$  matrix
  - update Hessian while iterating instead of calculating the accurate Hessian
- Stochastic gradient descent (SGD) method is utilized in deep learning
- for your references: chapter 4 in Judd (1998), and Le et al. (2011)

# BFGS Algorithm

---

**Algorithm 1:** BFGS Algorithm
 

---

**Input** : Choose initial guess  $x^0$ , initial Hessian guess  $H^0$ ,  
stopping parameters  $\delta$  and  $\epsilon > 0$

**Output:** Optimal point  $x^*$

```

1 for  $k = 0, 1, 2, \dots$  do
2   Solve  $H_k s^k = -(\nabla f(x^k))^T$  for the search direction  $s^k$ ;
3   Solve  $\lambda_k = \arg \min_{\lambda} f(x^k + \lambda s^k)$ ;
4    $x^{k+1} = x^k + \lambda_k s^k$ ;
5   Update  $H_k$ :
      
$$z_k = x^{k+1} - x^k,$$

      
$$y_k = (\nabla f(x^{k+1}))^T - (\nabla f(x^k))^T,$$

      
$$H_{k+1} = H_k - \frac{H_k z_k z_k^T H_k}{z_k^T H_k z_k} + \frac{y_k y_k^T}{y_k^T z_k}$$

      if  $\|x^k - x^{k+1}\| < \epsilon(1 + \|x^k\|)$  then
6     if  $\|\nabla f(x^{k+1})\| < \delta(1 + |f(x^{k+1})|)$  then
7       stop and report success;
8     else
9       stop and report convergence to non-optimal point;
```

---

# Limitations

- Plain Logit can represent systematic taste variation but not random taste variation (cannot be linked to observed characteristics)
- IIA property of the choice probability
- The cross-elasticity is the same for all  $i$ 
  - a change in an attribute of alternative  $j$  changes the probabilities for all other alternatives by the same percent

## Code Example

- See Python script



# Outline

- 1 Introduction
- 2 Logit
- 3 Generalized Extreme Value Models (GEV)**  
Nested Logit
- 4 Random Utility Model (RUM)
- 5 BLP

# Introduction

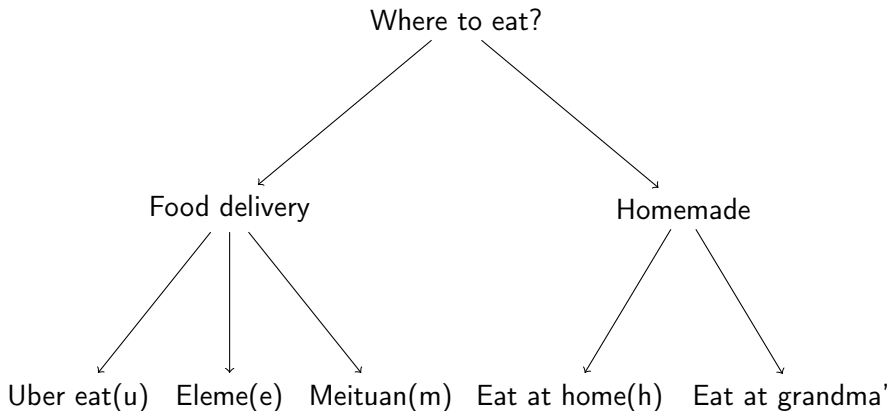
- The standard logit model exhibits independence from irrelevant alternatives (IIA), which implies proportional substitution across alternatives.
- Generalized Extreme Value (GEV) models constitute a broad class of models that accommodate a variety of substitution patterns.
- The unifying attribute of GEV models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value.
- This distribution allows for correlations among alternatives and is a generalization of the univariate extreme value distribution used in standard logit models.
- GEV models offer the advantage of providing choice probabilities that usually take a closed form, making them both flexible and practical for various applications.



# Food Delivery Choice Example

- **Choice set** = {Eleme (e), Meituan (m), Uber eat (u), Eating at home (h), Eating at grandma's (g)}
- If we use the plain Logit, when the price of eleme,  $p_e$ , increases, the probability of choosing Uber eat,  $P_{nu}$ , choosing Meituan,  $P_{nm}$ , choosing eating at home,  $P_{nh}$ , and choosing eating at grandma's,  $P_{ng}$ , would increase proportionately.
- **Is this realistic?**
- A more realistic scenario:  $P_{nu}$  and  $P_{nm}$  rise more relative to  $P_{nh}$  because the consumer may prefer placing a delivery order

## Food Delivery Choice Example (Cont'd)



## Choice Probabilities

- The utility that person  $n$  derives from alternative  $j$  in nest  $k$  is expressed as  $U_{nj} = V_{nj} + \epsilon_{nj}$ , where  $V_{nj}$  is observed by the researcher and  $\epsilon_{nj}$  is an unobserved random variable. The nested Logit model assumes that the vector of unobserved utilities  $\epsilon_n = \langle \epsilon_{n1}, \epsilon_{n2}, \epsilon_{n3}, \dots, \epsilon_{nJ} \rangle$  follows a cumulative distribution given by:

$$\exp \left( - \sum_{k=1}^K \left( \sum_{j \in B_k} e^{-\epsilon_{nj} / \lambda_k} \right)^{\lambda_k} \right) \quad (4)$$

- This distribution belongs to the Generalized Extreme Value (GEV) family and generalizes the distribution used in the plain Logit model

## Choice Probabilities (Cont'd)

- The choice probability for alternative  $i \in B_k$  is:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left( \sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^K \left( \sum_{j \in B_l} e^{V_{nj}/\lambda_l} \right)^{\lambda_l}}. \quad (5)$$

- For alternatives  $i \in B_k$  and  $m \in B_l$ , the ratio of their choice probabilities is:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k} \left( \sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l} \left( \sum_{j \in B_l} e^{V_{nj}/\lambda_l} \right)^{\lambda_l - 1}}. \quad (6)$$

- If  $k = l$  (i.e.,  $i$  and  $m$  are in the same nest), the terms in parentheses cancel out, yielding:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_k}}. \quad (7)$$

# Choice Probabilities (Cont'd)

- Proof hint for the choice probability
  - decompose utility function

$$U_{nj} = Y_{nj} + W_{nk} + \xi_{nk} + \epsilon_{nj}$$

where  $\epsilon \sim i.i.d. Gumbel(0, \lambda_k)$  in within nest  $k$

- Conditional probability
  - $\Pr_n(i) = \Pr_n(i|nest\ k) \Pr_n(nest\ k)$  where  $i$  is a choice in nest  $k$
  - log-sum term



## Remarks

- Independence from Irrelevant nests (IIN). The probability ratio does not depend on the attributes of alternatives in nests those NOT containing the two alternatives
- Loosely stated, the probability of choosing nest  $k$  depends on the expected utility that the person receives from that nest
- The log-likelihood function is NOT globally concave and is NOT close to a quadratic even in concave areas.
- Instead, nested Logit models can be estimated consistently, though not efficiently, in a sequential fashion

$$\begin{aligned} \text{For } i \in B_k, \quad \text{Prob}(i) &= \sum_{l=1}^L \text{Prob}(i \mid \text{nest } l) \times \text{Prob}(\text{nest } l) \\ &= \text{Prob}(i \mid \text{nest } k) \times \text{Prob}(\text{nest } k) \end{aligned}$$

since  $\text{Prob}(i \mid \text{nest } l) = 0, \quad l \neq k \text{ for } i \in B_k$

## Code Example

- See Python script

# Outline

- ① Introduction
- ② Logit
- ③ Generalized Extreme Value Models (GEV)  
Nested Logit
- ④ Random Utility Model (RUM)
- ⑤ BLP

# Introduction

- The Logit model is limited in three important ways:
  - It cannot represent random taste variation.
  - It exhibits restrictive substitution patterns due to the IIA property, even though GEV models relax this restriction.
  - It cannot be used with panel data when unobserved factors are correlated over time for each decision maker.
- In a RUM model, the explanatory variables did not vary over decision makers, but the coefficient is random over consumers
- The observed dependent variable was often market shares rather than individual customers' choices
- Two families:
  - Probit (multi-normal distributed unobserved factor)
  - Mixed logit (random coefficients)

## Mixed Logit (with individual-level data)

- The decision maker faces a choice among  $J$  alternatives.
- Utility of person  $n$  from alternative  $j$ ,

$$U_{nj} = \beta'_n x_{nj} + \epsilon_{nj}$$

where  $x_{nj}$  are observed variables that relate to the alternative and decision maker,  $\beta_n$  is a vector of coefficients of these variables for person  $n$  representing that person's tastes, and  $\epsilon_{nj}$  is a random term that is iid extreme value. The coefficients vary over decision makers in the population with density  $f(\beta; \theta)$ . This density is a function of parameters  $\theta$  that represent, for example, the mean and covariance of the  $\beta$ 's in the population

# Choice Probabilities

- This specification is the same as for plain Logit except that  $\beta$  varies over decision makers rather than being fixed.
- The decision maker knows the value of his own  $\beta_n$  and  $\epsilon_{nj}$ 's for all  $j$  and chooses alternative  $i$  iff  $U_{ni} > U_{nj}, \forall j \neq i$ .
- The researcher observes the  $x_{nj}$ 's but not  $\beta_n$  or the  $\epsilon_{nj}$ 's. That is, the probability conditional on  $\beta_n$  is,

$$L_{ni}(\beta_n) = \frac{e^{\beta'_n x_{ni}}}{\sum_j e^{\beta'_n x_{nj}}}$$

- If the researcher observed  $\beta_n$ , then the choice probability would be plain Logit

## Choice Probabilities (Cont'd)

- However, the researcher does not know  $\beta_n$  and therefore cannot condition on  $\beta$
- How to solve this?

## Choice Probabilities (Cont'd)

- However, the researcher does not know  $\beta_n$  and therefore cannot condition on  $\beta$
- How to solve this?
- Total Probability Theorem. The unconditional choice probability is the integral of  $L_{nj}(\beta)$  over all possible variables of  $\beta_n$ .

$$P_{ni} = \sum_{\beta \in \text{Support}\{\beta\}} \text{Prob}(i|\beta) \text{Prob}(\beta_n = \beta); \text{ discrete PMF}$$

$$\iff \int_{\beta \in \mathcal{B}} \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} f(\beta) d\beta; \text{ continuous PDF}$$

- Usually,  $f(\beta)$  has been specified to be Normal or Lognormal:  $\beta \sim N(b, W)$  or  $\ln \beta \sim N(b, W)$ , with parameters  $b$  and  $W$  that are estimated.



# Estimation

- Representative Utility

$$U_{nj} = \beta'_n x_{nj} + \epsilon_{nj},$$

- Choice probabilities

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta,$$

where

$$L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^J e^{\beta' x_{nj}}}.$$

- The coefficients  $\beta_n$  are distributed with density  $f(\beta|\theta)$ , where  $\theta$  refers collectively to the parameters of this distribution (such as the mean and covariance of  $\beta$ ). The researcher specifies the functional form  $f(\cdot)$  and wants to estimate the parameters  $\theta$ .
- The probabilities are approximated through simulation for any given value of  $\theta$ :

# Maximum Simulated Likelihood

---

## Algorithm 2: Mixed Logit Probability Simulation

---

**Input** : Choose a parameter vector  $\theta$  and set draw times  $R$

**Output**: Simulated probability  $\hat{P}_{ni}$

- 1 **for** *each* draw  $r$  from 1 to  $R$  **do**
- 2     Draw a value of  $\beta^r$  from  $f(\beta|\theta)$ . Calculate the Logit formula  $L_{ni}(\beta^r)$  with this draw.
- 3 Average the results to obtain the simulated probability:

$$\hat{P}_{ni} = \frac{1}{R} \sum_{r=1}^R L_{ni}(\beta^r).$$

---

# Maximum Simulated Likelihood (Cont'd)

- Two key properties of  $\hat{P}_{ni}$ .
  - $\hat{P}_{ni}$  is an unbiased estimator of  $P_{ni}$  by construction. Its variance decreases as  $R$  increases.
  - it is strictly positive, so that  $\ln \hat{P}_{ni}$  is defined, approximating the log-likelihood function
  - $\hat{P}_{ni}$  is smooth (twice differentiable) in the parameters  $\theta$  and the variables  $x$ , which facilitates the numerical search for the maximum likelihood function and the calculation of elasticities.
- The simulated probabilities are inserted into the log-likelihood function to give a simulated log likelihood:

$$SLL = \sum_{n=1}^N \sum_{j=1}^J d_{nj} \ln \hat{P}_{nj},$$

where  $d_{nj} = 1$  if person  $n$  chose  $j$  and zero otherwise. The maximum simulated likelihood (MSL) estimator is the value of  $\theta$  that maximizes SLL.

## Code Example

- See Python script

# Outline

# Introduction

- Question 1: What if we do not have individual-level data?
- Question 2: What if there are other observed characteristics correlated with price? (Endogeneity)
- Solution: Use market-level data and BLP method.

## BLP: Representative Utility Function

- The indirect utility of consumer  $i$  from consuming product  $j$  in market  $t$ ,  $U(x_{jt}, \xi_{jt}, p_{jt}, \tau_i; \theta)$ , is a function of observed and unobserved (by the researcher) product characteristics,  $x_{jt}$  and  $\xi_{jt}$ , respectively
- Consider a particular specification

$$u_{ijt} = \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \epsilon_{ijt},$$

$$i = 1, 2, \dots, I_t, \quad j = 1, \dots, J, \quad t = 1, \dots, T$$

where  $y_i$  is the income of consumer  $i$ ,  $p_{jt}$  is the price of product  $j$  in market  $t$ ,  $x_{jt}$  is a  $K$ -dimensional (row) vector of observable characteristics of product  $j$ ,  $\xi_{jt}$  is the unobserved product characteristic, and  $\epsilon_{ijt}$  is a mean-zero stochastic term

## Representative Utility Function (Cont'd)

- $\alpha_i$  is consumer  $i$ 's marginal utility from income, and  $\beta_i$  is a  $K$ -dimensional (column) vector of individual-specific taste coefficients
- Observed characteristics vary with the product. BLP (1995) examines the demand for cars and include as observed characteristics like horsepower, size, and air conditioning
- Depending on the structure of the data, some components of the unobserved characteristics can be captured by dummy variables. For example, we can model  $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$  and capture  $\xi_j$  and  $\xi_t$  by brand- and market-specific dummy variables
- Source of endogeneity.  $\xi_{jt}$  is unobserved and potentially correlated with  $p_{jt}$



## Representative Utility Function (Cont'd)

- Implicit assumption in the particular specification
  - The form of the indirect utility can be derived from a quasilinear utility function, which is free of wealth effects. Including wealth effects alters the way the term  $y_i - p_{jt}$  enters the equation above.
    - For instance, BLP(1995) builds on a Cobb-Douglas utility function to derive an indirect utility that is a function of  $\log(y_i - p_{jt})$ .
  - $\xi_{jt}$  is the same for all consumers.
  - All consumers face the same product characteristics.
  - In particular, all consumers are offered the same price.

# Decompose Utility Function: Consumer Preferences and Characteristics

- Consumer preferences  $\langle \alpha_i, \beta_i \rangle$  vary as a function of individual characteristics
  - The individual characteristics consist of two components: observed demographics,  $D_i$ , and additional unobserved characteristics,  $v_i$
  - Even though we do not observe individual data, we know something about the distribution of the demographics and additional characteristics
  - Demographics: i.e. income, age, family size, race, and education Information we might have includes large samples we can use to estimate some features of the distribution (i.e., Census data, Current Population Survey)

## Consumer Preferences and Characteristics (Cont'd)

- Formally, this will be modeled as:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \Pi D_i + \Sigma v_i,$$

$$v_i \sim P_v^*(v), \quad D_i \sim \hat{P}_D^*(D),$$

where  $D_i$  is a  $d \times 1$  vector,  $\Pi$  is a  $(K+1) \times d$  matrix of parameters, and  $\Sigma$  is a  $(K+1) \times (K+1)$  matrix of parameters

- If we assume that  $P_v^*(\cdot)$  is a standard multivariate normal distribution, then the matrix  $\Sigma$  allows each component of  $v_i$  to have a different variance and allows for correlation between these characteristics
- For simplicity, we assume that  $v_i$  and  $D_i$  are independent

## Decompose Utility Function (Cont'd)

- Let  $\theta = (\theta_1, \theta_2)$  be a vector containing all the parameters of the model
- The vector  $\theta_1 = (\alpha, \beta)$  contains the *linear* parameters, and the vector  $\theta_2 = (\Pi, \Sigma)$  contains the *nonlinear* parameters. We have:

$$u_{ijt} = \alpha_i y_i + \delta_{jt}(x_{jt}, p_{jt}, \xi_{jt}; \theta_1) + \mu_{ijt}(x_{jt}, p_{jt}, v_i, D_i; \theta_2) + \epsilon_{ijt}$$

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

$$\mu_{ijt} = [-p_{jt}, x_{jt}] \cdot (\Pi D_i + \Sigma v_i)$$

where  $[-p_{jt}, x_{jt}]$  is a  $1 \times (K + 1)$  vector

# Decompose Utility Function (Cont'd)

- Explanation
  - The first term,  $\alpha_i y_i$ , is given only for consistency with the indirect utility function, which will be cancelled out.
  - The second term,  $\delta_{jt}$ , which is referred to as the mean utility, is common to all consumers.
  - The last term,  $\mu_{ijt} + \epsilon_{ijt}$ , represents a mean-zero heteroskedastic deviation from the mean utility that captures the effects of the random coefficients.
- A comment: Is it realistic that consumers choose no more than one good, which is the assumption in BLP model
- Response:
  - even though many of us buy more than one brand at a time, less actually consume more than one at a time, so discreteness of choice can be sometimes defended by defining the choice period appropriately;
  - in some cases, the researcher has to model the choice of multiple products, or continuous quantities, explicitly

## Normalization: Outside Goods

- The specification of the demand system is completed with the introduction of an outside good: consumers may decide not to purchase any of the brands
- The indirect utility from this outside option is given by:

$$u_{i0t} = \alpha_i y_i + \xi_{0t} + \pi_0 D_i + \sigma_0 v_{i0} + \epsilon_{i0t}$$

- For simplicity, we can regard the outside option as “buy nothing.” Hence, the consumer would not get the characteristics bundle and does not need to pay
- The standard practice is to set  $\xi_{0t}$ ,  $\pi_0$ , and  $\sigma_0$  all to zero. Since the term  $\alpha_i y_i$  will eventually cancel out (common to all products), this is equivalent to normalizing the utility from the outside good to zero

# Market Shares

- individual is defined as a vector of demographics and product-specific shocks  $(D_i, v_i, \epsilon_{i0t}, \dots, \epsilon_{ijt})$ , this implicitly defines the set of individual attributes that lead to the choice of good  $j$ .
- Formally, the set is

$$A_{jt}(x_t, p_t, \delta_t; \theta_2) = \{(D_i, v_i, \epsilon_{i0t}, \dots, \epsilon_{ijt}) \mid u_{ijt} \geq u_{ilt}, \forall l = 0, 1, \dots, J\}$$

where  $x_t = (x_{1t}, \dots, x_{Jt})'$ ,  $p_t = (p_{1t}, \dots, p_{Jt})'$ , and  $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})'$  are observed characteristics, prices, and mean utilities of all brands, respectively. The set  $A_{jt}$  defines the individuals who choose brand  $j$  in market  $t$ .

## Market Shares (Cont'd)

- Assuming ties occur with zero probability, the market share of the  $j$ -th product is just an integral over the mass of consumers in the region  $A_{jt}$

$$\begin{aligned} s_{jt}(x_t, p_t, \delta_t; \theta_2) &= \int_{A_{jt}} dP^*(D, v, \epsilon) \\ &= \int_{A_{jt}} dP^*(\epsilon|D, v) dP^*(v|D) dP_D^*(D) \quad (8) \\ &= \int_{A_{jt}} dP^*(\epsilon|D, v) dP^*(v|D) dP_D^*(D) \end{aligned}$$

- where  $P^*(\cdot)$  denotes population distribution functions. Given assumptions on the distribution of the (unobserved) individual attributes, we can compute the integral, either analytically or numerically



## Market Shares (Cont'd)

- Therefore, for a given set of parameters, the equation above predicts the market share of each product in each market, as a function of product characteristics, prices, and unknown parameters
- One possible estimation strategy is to choose parameters that minimize the distance between the market shares predicted by the equation and the observed shares
- This estimation strategy does not account for the correlation between prices and the unobserved product characteristics
- The BLP method accounts for this correlation

## Straightforward Approach

- As previously pointed out, a straightforward approach to the estimation is to solve

$$\min_{\theta} \|s(x, p, \delta(x, p, \xi; \theta_1); \theta_2) - S\|,$$

where  $s(\cdot)$  are the market shares given by Equation 8, and  $S$  are the observed market shares

- Unobserved variables include individual-level characteristics,  $(D_i, v_i, \epsilon_i)$  and unobserved product characteristics,  $\xi_j$ 
  - $(D_i, v_i, \epsilon_i)$  were integrated over. The econometric error term will be the unobserved product characteristics,  $\xi_{jt}$
  - prices are potentially correlated with this term, the econometric estimation will have to take account of this

## Straightforward Approach (Cont'd)

- Straightforward approach is usually not taken due to costly minimization
  - all the parameters enter the minimization problem nonlinearly. In some applications the inclusion of brand and time dummy variables results in a large number of parameters and a costly nonlinear minimization problem
- BLP method avoids this problem by transforming the minimization problem so that some (or all) of the parameters enter the objective function linearly

# BLP Method

- Let  $Z = [z_1, \dots, z_M]$  be a set of instruments such that

$$E[Z_m \omega(\theta^*)] = 0,$$

where  $\omega$ , a function of the model parameters, is an error term, and  $\theta^*$  denotes the "true" parameters

- GMM estimate

$$\hat{\theta} = \arg \min_{\theta} \omega(\theta)' Z \Phi^{-1} Z' \omega(\theta), \quad (9)$$

where  $\Phi$  is a consistent estimate of  $E[Z' \omega \omega' Z]$

- The error term is defined as the structural error,  $\xi_{jt}$  as mentioned
- In order to use Equation 9, we need to express the error term as an explicit function of the parameters of the model and the data
  - The key insight is that the error term  $\xi_{jt}$  enters only the mean utility level,  $\delta(\cdot)$

# BLP Method Procedure

- The mean utility level is a linear function of  $\xi_{jt}$ . We solve for each market the implicit system of equations:

$$s(\delta_t; \theta_2) = S_t, \quad t = 1, \dots, T$$

where  $s(\cdot)$  are the market shares given by 8, and  $S$  are the observed market shares.

- First, predict market share

$$s_{jt}(p_t, x_t, \delta_t, P_{ns}; \theta_2) = \frac{1}{ns} \sum_{i=1}^{ns} s_{jti} = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp \left( \delta_{jt} + \sum_{k=1}^K x_{jt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right)}{1 + \sum_{m=1}^J \exp \left( \delta_{mt} + \sum_{k=1}^K x_{mt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id}) \right)}$$

## BLP Method Procedure (Cont'd)

- $(v_i^1, \dots, v_i^K)$  and  $(D_{i1}, \dots, D_{id})$ ,  $i = 1, 2, \dots, ns$ , are draws from  $P_v^*(v)$  and  $P_D^*(D)$ , respectively, while  $x_{jt}^k$ ,  $k = 1, \dots, K$ , are the variables that have random slope coefficients.
- Second, using the computation of the market share, we invert the system of equations by the contraction mapping

$$\delta_t^{h+1} = \delta_t^h + \ln S_t - \ln s(\delta_t^h; p_t, x_t, \delta_t, P_{ns}; \theta_2), \quad h = 0, \dots, H$$

where  $s(\cdot)$  are the predicted market shares,  $H$  is the smallest integer such that  $\|\delta_t^H - \delta_t^{H-1}\|$  is smaller than the tolerance level, and  $\delta_t^H$  is the approximation to  $\delta_t$ .

## BLP Method Procedure (Cont'd)

- Third, define the error term:

$$\omega_{jt} = \delta_{jt}(\theta_2) - (x_{jt}\beta - \alpha p_{jt}) \equiv \xi_{jt}$$

- Note:
  - the observed market shares,  $S$ , enter this equation.
  - The reason for distinguishing between  $\theta_1$  and  $\theta_2$ :  $\theta_1$  enters this term, and the GMM objective, in a linear fashion, while  $\theta_2$  enters nonlinearly.

## Detailed Procedure

- See Algorithm.pdf



# Welfare Change

- Recall utility function

$$u_{ijt} = \alpha_i y_i + \delta_{jt}(x_{jt}, p_{jt}, \xi_{jt}; \theta_1) + \mu_{ijt}(x_{jt}, p_{jt}, v_i, D_i; \theta_2) + \epsilon_{ijt}$$

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

$$\mu_{ijt} = [-p_{jt}, x_{jt}] \cdot (\Pi D_i + \Sigma v_i)$$

## Welfare Change (Cont'd)

### 1 Estimate Parameters Before Policy Shock

- Estimate  $\delta_j^{\text{pre}}, \beta^{\text{pre}}, \alpha^{\text{pre}}$
- Calculate  $\mathbb{E}[\max_j u_{ij}^{\text{pre}}]$ :

$$\mathbb{E}[\max_j u_{ij}^{\text{pre}}] = \ln \left( \sum_j \exp(\delta_j^{\text{pre}} + \mu_{ij}^{\text{pre}}) \right) + C$$

### 2 Estimate Parameters After Policy Shock

- Estimate  $\delta_j^{\text{post}}, \beta^{\text{post}}, \alpha^{\text{post}}$
- Calculate  $\mathbb{E}[\max_j u_{ij}^{\text{post}}]$ :

$$\mathbb{E}[\max_j u_{ij}^{\text{post}}] = \ln \left( \sum_j \exp(\delta_j^{\text{post}} + \mu_{ij}^{\text{post}}) \right) + C$$

### 3 Calculate Welfare Change

- Compute the difference in expected maximum utility:

$$\Delta W_n = \mathbb{E}[\max_j u_{nj}^{\text{post}}] - \mathbb{E}[\max_j u_{nj}^{\text{pre}}]$$

## Python Example

- See Python script

Thanks!

- Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.
- Farronato, C. and Fradkin, A. (2022). The welfare effects of peer entry: the case of airbnb and the accommodation industry. *American Economic Review*, 112(6):1782–1817.
- Judd, K. L. (1998). *Numerical methods in economics*. MIT press.
- Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., and Ng, A. Y. (2011). On optimization methods for deep learning. In *Proceedings of the 28th international conference on international conference on machine learning*, pages 265–272.
- MCFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.
- Train, K. E. (1998). Recreation demand models with taste differences over people. *Land economics*, pages 230–239.