

# UG Empirical Industrial Organization

## Lecture 6: Demand Systems: Discrete Choice Models

Anubhav Jha<sup>1</sup>  
Ashoka University

March 11, 2024

---

<sup>1</sup>This course is based on Victor Aguirregabiria's Empirical IO book. The slides (in my first year of teaching) are extremely close to his slides. These will change in the future iterations of this course.

# Outline on today's lecture

1. Limitations of demand systems from consumer preferences in product space
2. Discrete choice models of consumer demand & consumer preferences in characteristics space
  - 2.1. General Model
  - 2.2. Standard Logit Model
  - 2.3. Nested Logit Model
  - 2.4. Model with heterogeneous coefficients

# Limitations: Demand systems product space

- ▶ Demand models in 'product space' have several practical limitations in IO applications.
  1. Positive consumption of all products.
  2. “Too many” parameters.
  3. Finding instruments for prices.
  4. Problems to predict demand of new varieties.

# [1] Positive consumption of all products

- ▶ The model assumes the marginal conditions of optimality hold for every product.
- ▶ This means that  $q_j > 0$  for every product  $j$ .
- ▶ This condition is very unrealistic when we consider the demand of differentiated products within a product category, e.g., the demand of automobiles.
- ▶ In this context, a consumer buys only one unit of a single variety (e.g., one Toyota Corolla) or of a few varieties (e.g., one Toyota Corolla, and one KIA Sorento minivan).
- ▶ To account for this type of consumer decisions, we need to model the consumer problem as a discrete choice model.

## [2] Too many parameters

$$w_{jt} = \beta_j^{(0)} + \beta_j^{(y)} \ln \left( \frac{y_t}{P_t} \right) + \sum_{k \in E_j} \beta_{jk}^{(p)} \ln(p_{kt}) + \xi_{jt}$$

- ▶ The number of parameters increases quadratically with the number of products  $J$ .
- ▶ It is not possible to estimate demand systems for differentiated products with many varieties.
- ▶ For instance, demand system for car models. With  $J = 100$ , #parameters = 5, 250.
- ▶ Even with the restrictions of multi-stage budgeting and multiple groups and nests, the number of parameters is large.

### [3] Finding instruments for prices

$$w_{jt} = \beta_j^{(0)} + \beta_j^{(y)} \ln \left( \frac{y_t}{P_t} \right) + \sum_{k \in E_g} \beta_{jk}^{(p)} \ln(p_{kt}) + \xi_{jt}$$

- ▶ In the AIDS model, for the equation of a product, we have as many endogenous regressors (prices) as products.
- ▶ If we deal with the endogeneity of prices using Instrumental Variables, we need as many instruments as products,  $J$ .
- ▶ The ideal case is when we have information on production costs for each individual good. However, that information is very rarely available.

## [4] Problems to predict demand of new varieties

$$w_{jt} = \beta_j^{(0)} + \beta_j^{(y)} \ln \left( \frac{y_t}{P_t} \right) + \sum_{k \in E_g} \beta_{jk}^{(p)} \ln(p_{kt}) + \xi_{jt}$$

- ▶ A problem that has received substantial attention in IO is the prediction of the demand of a new product.
- ▶ In a demand system in product space, estimating the demand of a new good requires estimates of the parameters associated with that good. If the new product is  $J + 1$ :

$$\{\beta_{J+1}^{(0)}, \beta_{J+1}^{(y)}, \beta_{j,J+1}^{(p)} : j = 1, 2, \dots, J + 1\}.$$

- ▶ This makes it impossible to make predictions on demand of a product that has not been introduced yet in any market.
- ▶ It also limits the applicability of this model in cases where the new product has been introduced very recently or in very few markets, because we may not have enough data to estimate these parameters.

# Discrete choice demand models: Basic Assumptions

- ▶ The basic assumptions are:
  1. A product, say a laptop computer, can be described in terms of *a bundle of physical characteristics*: e.g., CPU speed, memory, screen size, etc. These characteristics determine a product.
  2. Consumers have *preferences on characteristics of products*, not on the products per se.
  3. A product category (e.g., laptops) has  $J$  different products and *each consumer buys at most one variety of the product per period*. This implies that the  $J$  products are substitutes with each other.



## Model:

## Products

- ▶ We index varieties by  $j \in \{1, 2, \dots, J\}$ .
- ▶ We can distinguish two sets of product characteristics.
- ▶ **Characteristics observable to the researcher:**

$$X_j = (X_{1j}, X_{2j}, \dots, X_{Kj})$$

$X_{kj}$  represents the amount of attribute  $k$  in product  $j$ .

- ▶ Example (Laptops):  $X_{1j}$  = CPU speed;  $X_{2j}$  = RAM memory;  $X_{3j}$  = storage;  $X_{4j}$  = weight;  $X_{5j}$  = screen size;  $X_{6j}$  = dummy 'Intel inside'; etc.
- ▶ **Other characteristics are not observable to the researcher** but they are known and valued by consumers. We use  $\xi_j$  to represent the unobservable (to the researcher) characteristics of product  $j$ .

- ▶ The utility of household  $h$  if they purchase product  $j$  is:

$$U_{hj} = u(c_h) + X_j \beta_h + \xi_j + \varepsilon_{hj}$$

- ▶  $c_h$  is consumption of the outside good and  $u(c_h)$  is the corresponding utility.
- ▶  $X_j \beta_h = \beta_{1h}X_{1j} + \beta_{2h}X_{2j} + \dots + \beta_{Kh}X_{Kj}$  is the utility from the observable characteristics.
- ▶ Coefficient  $\beta_{kh}$  represents the **marginal utility of this product characteristic** (e.g., RAM memory) for consumer  $h$ .
- ▶  $\xi_j$  is utility from unobservable characteristics, the same for all consumers.
- ▶  $\varepsilon_{hj}$  is an idiosyncratic taste of consumer  $h$  for product  $j$  that is unobservable to the researcher.

- ▶ Budget constraint of consumer  $h$ , with income  $y_h$ , if they buy product  $j$ :

$$c_h + p_j \leq y_h$$

To represent no purchase of any of the  $J$  products we use  $j = 0$ , with  $p_0 = 0$ ,  $X_j = 0$ , and  $\xi_j = 0$ .

- ▶ The utility of purchasing product  $j$ , after taking into account the budget constraint ( $c_h = y_h - p_j$ ) becomes:

$$U_{hj} = u(y_h - p_j) + X_j \beta_h + \xi_j + \varepsilon_{hj}$$

## Model:

## Consumer decision problem

- ▶ A consumer chooses the alternative  $j \in \{0, 1, \dots, J\}$  that maximizes her utility  $U_{hj}$ .
- ▶ Consumer  $h$  chooses product  $j$  if and only if:

$$U_{hj} > U_{hi} \quad \text{for any } i \neq j$$

- ▶ Or equivalently, if and only if:

$$u(y_h - p_j) + X_j \beta_h + \xi_j + \varepsilon_{hj} > u(y_h - p_i) + X_i \beta_h + \xi_i + \varepsilon_{hi} \quad \text{for any } i \neq j$$

## Model: Individual consumer (unit) demands

- ▶ Let  $q_{hj}$  represent the quantity of product  $j$  demanded by consumer  $h$ .
- ▶ In this model,  $q_{hj}$  is either 0 or 1:  $q_{hj} \in \{0, 1\}$ .
- ▶ The demand is:

$$q_{hj} = \begin{cases} 1 & \text{if } U_{hj} > U_{hi} \text{ for any } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

with

$$U_{hj} = u(y_h - p_j) + X_j \beta_h + \xi_j + \varepsilon_{hj}.$$

- ▶ These unit demands depend on income, all the product prices, and all the product characteristics  $X_j$  and  $\xi_j$ .

## Model: Aggregate demand / Market shares

- ▶ Let  $s_j$  be the proportion of consumers who purchase product  $j$ .

$$s_j = \frac{\sum_{h=1}^H q_{hj}}{H} = \Pr(q_{hj} = 1)$$

- ▶ We describe  $s_j$  as the **market share of product  $j$** . According to the model:

$$s_j = \Pr(U_{hj} > U_{hi} \text{ for any } i \neq j)$$

$$= \Pr\{u(y_h - p_j) + X_j\beta_h + \xi_j + \varepsilon_{hj} > u(y_h - p_i) + X_i\beta_h + \xi_i + \varepsilon_{hi} \text{ for any } i \neq j\}$$

- ▶ This probability (market share) results from integrating the individual demand  $q_{hj}$  over the distribution of  $(\beta_h, \varepsilon_h, y_h)$ .

# Different types of discrete choice models

- ▶ The particular type of discrete choice model depends on three types of assumptions.
- [1] The specification of the "utility of money"  $u(\cdot)$ ,  
e.g., linear or logarithmic function.
  - [2] The probability distribution of the variables  $\varepsilon_{hj}$ ,  
e.g., Logit or Probit.
  - [3] The assumptions on the coefficients  $\beta_h$ ,  
e.g., constant ( $\beta_h = \beta$  for every  $h$ ) or heterogeneous coefficients.

- ▶ We will study three examples of discrete choice demand models that have received a lot of attention in empirical applications.

## [1] Standard Logit model

Linear  $u(\cdot)$ ;  $\varepsilon_{hj} \sim$  i.i.d. Extreme Value; and  $\beta_h = \beta$ .

## [2] Nested Logit model

Linear  $u(\cdot)$ ;  $\varepsilon_{hj} \sim$  Sum of two i.i.d. Extreme Values; and  $\beta_h = \beta$ .

## [3] Random coefficients Logit model

$u(\cdot)$  may be nonlinear;  $\varepsilon_{hj} \sim$  i.i.d. Extreme Values; and  $\beta_h$  are heterogeneous across individuals according to some distribution.



# Extreme value distribution: Some properties

- ▶ We say that a random variable  $\varepsilon$  has an Extreme Value type 1 distribution (or Gumbel distribution) iff its Cumulative Distribution Function is:

$$F(\varepsilon) = \exp \{ - \exp \{ - \varepsilon \} \}$$

- ▶  $\varepsilon$  has support  $(-\infty, +\infty)$ . The density function is bell shaped but asymmetric.
- ▶ Suppose that the variables  $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_J$  are all Extreme Value and independently distributed. These variables satisfy two properties that are useful for discrete choice models.

## Extreme value distribution: Some properties [2]

- ▶ For any vector of constants  $(\delta_0, \delta_1, \dots, \delta_J)$ , we have that:

$$\Pr(\delta_j + \varepsilon_j > \delta_i + \varepsilon_i \text{ for any } i \neq j) = \frac{\exp\{\delta_j\}}{\sum_{i=0}^J \exp\{\delta_i\}}$$

- ▶ And:

$$\mathbb{E} \left( \max_{j \in \{0, 1, \dots, J\}} [\delta_j + \varepsilon_j] \right) = \ln \left( \sum_{j=0}^J \exp\{\delta_j\} \right)$$

# Standard Logit Demand Model

- ▶ Suppose that  $u(c) = \alpha c$  [Linear];  $\beta_h = \beta$  for every  $h$ ; and all the  $\varepsilon_{hj}$  are *i.i.d.* Extreme Value Type 1. Then,

$$U_{hj} = -\alpha p_j + X_j \beta + \xi_j + \varepsilon_{hj}$$

- ▶ If we define  $\delta_j = -\alpha p_j + X_j \beta + \xi_j$ , then  $\delta_j$  represents the average utility of purchasing product  $j$  for every consumer in the market.
- ▶ The only idiosyncratic component of utility is  $\varepsilon_{hj}$ .
- ▶ The model implies that the market shares are:

$$s_j = \frac{\exp\{\delta_j\}}{1 + \sum_{i=1}^J \exp\{\delta_i\}} = \frac{\exp\{-\alpha p_j + X_j \beta + \xi_j\}}{1 + \sum_{i=1}^J \exp\{-\alpha p_i + X_i \beta + \xi_i\}}$$

- ▶ because  $\delta_0 = -\alpha p_0 + X_0 \beta + \xi_0 = 0$ .

- ▶ When the indirect utility has this form,

$$U_{hj} = -\alpha p_j + \beta_1 X_{1j} + \dots + \beta_K X_{Kj} + \xi_j + \varepsilon_{hj}$$

we have that the parameters  $\beta$  and  $\alpha$  have a very clear economic interpretation.

- ▶  $\alpha$  is the *marginal utility of income* (or marginal utility of money), and it is measured in utils per dollar.
- ▶  $\beta_k$  is the *marginal utility of attribute  $k$* , and it is measured in utils per unit of this attribute.
- ▶ Therefore,

$$\frac{\beta_k}{\alpha} = \frac{\text{utils / unit of } k}{\text{utils / \$}} = \frac{\$}{\text{unit of } k}$$

is the **marginal value of attribute  $k$**  measured in dollars per unit of the attribute.

- In the logit model,  $s_j = \frac{\exp\{\delta_j\}}{1 + \sum_{i=1}^J \exp\{\delta_i\}}$ , implies that (prove it):

$$\frac{ds_j}{d\delta_j} = s_j(1 - s_j) \quad \text{and} \quad \frac{ds_j}{d\delta_i} = -s_j s_i \quad \text{for } i \neq j$$

- Therefore, the own price elasticities are:

$$\frac{ds_j}{dp_j} \cdot \frac{p_j}{s_j} = \frac{ds_j}{d\delta_j} \cdot \frac{d\delta_j}{dp_j} \cdot \frac{p_j}{s_j} = s_j(1 - s_j) \cdot [-\alpha] \cdot \frac{p_j}{s_j} = -\alpha(1 - s_j)p_j$$

- And the cross price elasticities are:

$$\frac{ds_j}{dp_i} \cdot \frac{p_i}{s_j} = \frac{ds_j}{d\delta_i} \cdot \frac{d\delta_i}{dp_i} \cdot \frac{p_i}{s_j} = -s_j s_i \cdot [-\alpha] \cdot \frac{p_i}{s_j} = \alpha s_i p_i$$

- ▶ According to the model, for any product  $j = 1, 2, \dots, J$ :

$$\frac{s_j}{s_0} = \exp \{-\alpha p_j + X_j \beta + \xi_j\}$$

- ▶ such that:

$$\ln(s_j) - \ln(s_0) = -\alpha p_j + X_j \beta + \xi_j$$

- ▶ This is a linear regression model with regressors  $(p_j, X_j)$ , parameters  $(\alpha, \beta)$ , and error term  $\xi_j$ .
- ▶ If the number of products  $J$  is large enough relative to the number of observable product characteristics  $K$ , then we can estimate the parameters of this model using a cross-section of consumers (one market).

## Logit Demand Model:

## Some advantages

- ▶ Let's revisit the limitations of the traditional demand systems and see that the Logit model overcomes these limitations.
- 1 **Positive consumption of all products.** As a discrete choice model, the logit does not have this unrealistic prediction.
- 2 **"Too many" parameters.** The number of parameters  $(\alpha, \beta)$  is  $K + 1$  that does not increase with the number of products  $J$ . Having many products (large  $J$ ) is a blessing for this model (more degrees of freedom, more precise estimates).
- 3 **Finding instruments for prices.** There is only one endogenous regressor in the model. We need only one instrument.
- 4 **Predict demand of new varieties.** Given  $(\alpha, \beta)$ , we can predict the demand of any hypothetical product with characteristics  $(p^*, X^*)$ . [More on this in Problem Set-2].

## Estimation:

## Endogeneity

- ▶ We have the regression model:

$$\ln(s_j) - \ln(s_0) = \beta^{(p)} p_j + X_j \beta + \xi_j$$

with  $\beta^{(p)} = -\alpha$ .

- ▶ **Endogeneity of prices:** Products with better (higher) unobserved quality  $\xi_j$  tend to have higher prices:

$$\text{cov}(p_j, \xi_j) > 0$$

We will see this formally in the topic of models of price competition.

- ▶ This endogeneity problem implies that the OLS estimate of the parameter associated to price is upward biased:

$$\hat{\beta}^{(p)} = \beta^{(p)} + \frac{\text{cov}(p_j, \xi_j)}{\text{var}(p_j)}$$

- ▶ We could even get  $\hat{\beta}^{(p)}$  is positive or not significantly different from zero.



- ▶ Berry, Levinsohn, & Pakes (1996) proposed the following instrument for price in discrete choice models: **BLP instruments**.
- ▶ Suppose that price is the only endogenous regressor such that

$$\mathbb{E}[X_j \xi_j] = 0.$$

- ▶ Firms' price competition imply that the price of a product depends on its product characteristics  $(X_j, \xi_j)$ , but also on the characteristics of the competing products,  $(X_i, \xi_i : i \neq j)$ .
- ▶ Intuitively, if the other products have characteristics very similar to  $j$ , competition is more intense and  $p_j$  will be small. But if product  $j$  has very different characteristics from the other products, firm  $j$  can enjoy some market power from product differentiation and  $p_j$  can be larger.
- ▶ The idea is using  $(X_i : i \neq j)$  as an instrument for  $p_j$ .

## IV Estimation:

## BLP instruments [2]

- ▶ Suppose that we use the vector of instruments

$Z_j = (Z_{1j}, Z_{2j}, \dots, Z_{Kj})$  where for any product characteristic  $k$ ,  $Z_{kj}$

- ▶ We could consider other functions of  $(X_i : i \neq j)$  as instruments, such as the average distance:

$$Z_{kj} = \frac{\sum_{i \neq j} |X_{ki} - X_{kj}|}{J - 1}.$$

- ▶  $Z_j$  is a vector of valid instruments for  $p_j$  in the regression equation:

$$\ln(s_j) - \ln(s_0) = \beta^{(p)} p_j + X_j \beta + \xi_j$$

because:

1. They are not regressors in this regression function (the utility of product  $j$  depends only on characteristics of product  $j$ );
2. They are not correlated with the error term  $\xi_j$  (by assumption);
3. They are correlated with  $p_j$  (testable restriction).

- ▶ The standard Logit model imposes some strong restrictions on price elasticities:

$$\frac{ds_j}{dp_j} \cdot \frac{p_j}{s_j} = -\alpha(1 - s_j)p_j \quad \text{and} \quad \frac{ds_j}{dp_i} \cdot \frac{p_i}{s_j} = \alpha s_i p_i$$

- ▶ An increase in the price of product  $i$  (e.g., a luxury car) implies the same increase in the demand of product  $j$  (another luxury car) as in the demand of product  $b$  (a very low quality car).
- ▶ We can introduce two extensions in the model that relax this restriction:
  - A. Consumer heterogeneous coefficients  $\beta_h$
  - B. Nested Logit model for  $\varepsilon_h$