

# ECON-L1300 - Empirical Industrial Organization, PhD

## I: Static models

### Lecture 2

Tanja Saxell, Ari Hyytinen, Otto Toivanen & Iivo Vehviläinen

# About today's lecture

- Today's lecture is on discrete choice models. We discuss
  - ① Discrete choice / ARUM models: binary decisions.
  - ② Extension to many options.
  - ③ The Logit model.
  - ④ Independence of Irrelevant Alternatives.
  - ⑤ Welfare and consumer surplus.
  - ⑥ GEV models and Nested Logit.
  - ⑦ Link from individual level to market level data (Logit, Nested Logit).

# Why discrete choice?

- Many decisions are discrete:

# Why discrete choice?

- Many decisions are discrete:
- To export or not.

# Why discrete choice?

- Many decisions are discrete:
- To export or not.
- To go to college or not.

# Why discrete choice?

- Many decisions are discrete:
- To export or not.
- To go to college or not.
- To choose a product or not.

# Why discrete choice?

- Many decisions are discrete:
- To export or not.
- To go to college or not.
- To choose a product or not.
- Discrete choice the simplest (?) to model theoretically.

# Why discrete choice?

- Many decisions are discrete:
- To export or not.
- To go to college or not.
- To choose a product or not.
- Discrete choice the simplest (?) to model theoretically.
- Above examples all binary.

# 1. Additive Random Utility Models

- We start with simplest models of (discrete) choice: 2 options / binary.

# 1. Additive Random Utility Models

- We start with simplest models of (discrete) choice: 2 options / binary.
- Alternatives are labelled 0 and 1.

# 1. Additive Random Utility Models

- We start with simplest models of (discrete) choice: 2 options / binary.
- Alternatives are labelled 0 and 1.
- How does a utility-maximizing DM choose?

# 1. Additive Random Utility Models

- We start with simplest models of (discrete) choice: 2 options / binary.
- Alternatives are labelled 0 and 1.
- How does a utility-maximizing DM choose?
- Choose  $i$  iff  $U_i \geq U_j$ ,  $i \neq j$ ,  $i, j \in \{0, 1\}$ .

# 1. Additive Random Utility Models

- How to operationalize the relation?

# 1. Additive Random Utility Models

- How to operationalize the relation?
- Slightly differently: Why call the model a random utility model?

# 1. Additive Random Utility Models

- How to operationalize the relation?
- Slightly differently: Why call the model a random utility model?
- There is nothing random about the decision, from the view point of the DM.

# 1. Additive Random Utility Models

- How to operationalize the relation?
- Slightly differently: Why call the model a random utility model?
- There is nothing random about the decision, from the view point of the DM.
- Model called RUM because decision **appears random for an outsider observer** (call her the econometrician).

# 1. Additive Random Utility Model

- Lets specify

$$U_i = V_i + \epsilon_i \quad (1)$$

- By definition,

# 1. Additive Random Utility Model

- Lets specify

$$U_i = V_i + \epsilon_i \quad (1)$$

- By definition,
- $V_i$  is observable to the econometrician (measurable / estimable).

# 1. Additive Random Utility Model

- Lets specify

$$U_i = V_i + \epsilon_i \quad (1)$$

- By definition,
- $V_i$  is observable to the econometrician (measurable / estimable).
- $V_i$  is the **deterministic component** of utility.

# 1. Additive Random Utility Model

- Lets specify

$$U_i = V_i + \epsilon_i \quad (1)$$

- By definition,
- $V_i$  is observable to the econometrician (measurable / estimable).
- $V_i$  is the **deterministic component** of utility.
- $\epsilon_i$  is not observable to the econometrician.

# 1. Additive Random Utility Model

- Lets specify

$$U_i = V_i + \epsilon_i \quad (1)$$

- By definition,
- $V_i$  is observable to the econometrician (measurable / estimable).
- $V_i$  is the **deterministic component** of utility.
- $\epsilon_i$  is not observable to the econometrician.
- $\epsilon_i$  is the **random component** of utility.

## Decision rule

$$\Pr(y = 1) = 1 - F(V_1 - V_0) \quad (2)$$

$$\epsilon_0 - \epsilon_1 \sim F$$

## Decision rule

$$\Pr(y = 1) = 1 - F(V_1 - V_0) \quad (2)$$

$$\epsilon_0 - \epsilon_1 \sim F$$

- Notice benefit from assuming additivity of deterministic and random utility components.

## Decision rule

$$\Pr(y = 1) = 1 - F(V_1 - V_0) \quad (2)$$

$$\epsilon_0 - \epsilon_1 \sim F$$

- Notice benefit from assuming additivity of deterministic and random utility components.
- We need a scale normalization - why?

## More structure

- Let's assume

$$V_i = \mathbf{X}\beta_i \quad (3)$$

## More structure

- Let's assume

$$V_i = \mathbf{X}\beta_i \quad (3)$$

- Then

$$\Pr(y = 1|X) = 1 - F(\mathbf{X}\beta) \quad (4)$$

## More structure

- Let's assume

$$V_i = \mathbf{X}\beta_i \quad (3)$$

- Then

$$\Pr(y = 1|X) = 1 - F(\mathbf{X}\beta) \quad (4)$$

- where

$$\beta = \beta_1 - \beta_0$$

## More structure: Index function interpretation

- Let's assume

$$y^* = \mathbf{X}\beta + u \quad (5)$$

## More structure: Index function interpretation

- Let's assume

$$y^* = \mathbf{X}\beta + u \quad (5)$$

- Then

$$y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0 \end{cases} \quad (6)$$

## Index function c'ed

- Plugging (5) into (6) and solving yields

$$\Pr(y = 1 | X) = P = 1 - F(\mathbf{X}\beta) \quad (7)$$

## Index function c'ed

- Plugging (5) into (6) and solving yields

$$\Pr(y = 1 | X) = P = 1 - F(\mathbf{X}\beta) \quad (7)$$

- Eqn (7) is called a (single-) index model.

## Index function c'ed

- Plugging (5) into (6) and solving yields

$$\Pr(y = 1 | X) = P = 1 - F(\mathbf{X}\beta) \quad (7)$$

- Eqn (7) is called a (single-) index model.
- NOTE: in a single-index model, the LHS variable need not be discrete, but often is.

## Index function c'ed

- Plugging (5) into (6) and solving yields

$$\Pr(y = 1 | X) = P = 1 - F(\mathbf{X}\beta) \quad (7)$$

- Eqn (7) is called a (single-) index model.
- NOTE: in a single-index model, the LHS variable need not be discrete, but often is.
- NOTE #2: Note how we identify difference in parameters.

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .
- $J$  is the choice set. It should have the following properties:
  - ① Alternatives must be mutually exclusive.
  - ② The choice set must be exhaustive.
  - ③ The number of choices must be finite.

We are interested in the setting where consumer  $i$ ,  $i = 1, \dots, I$  chooses among the  $J$  alternatives.

This is actually a very flexible approach.

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .
- Often  $j = 0, \dots, J$ , with  $j = 0 =$  not choosing any of the “inside” goods.

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .
- Often  $j = 0, \dots, J$ , with  $j = 0 =$  not choosing any of the “inside” goods.
- Utility from choosing option  $j$  denoted  $U_{ij} = V_{ij} + \epsilon_{ij}$ .

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .
- Often  $j = 0, \dots, J$ , with  $j = 0 =$  not choosing any of the “inside” goods.
- Utility from choosing option  $j$  denoted  $U_{ij} = V_{ij} + \epsilon_{ij}$ .
- Choose option  $j$  iff  $U_{ij} \geq U_{ik}, \forall k, j \neq k$ .

## 2. Multinomial Additive Random Utility Models

- Now  $J$  options, labelled  $j = 1, \dots, J$ .
- Often  $j = 0, \dots, J$ , with  $j = 0 =$  not choosing any of the “inside” goods.
- Utility from choosing option  $j$  denoted  $U_{ij} = V_{ij} + \epsilon_{ij}$ .
- Choose option  $j$  iff  $U_{ij} \geq U_{ik}, \forall k, j \neq k$ .
- In settings interesting for us, a normalization needed (e.g.  $U_{i0} = 0$ ).

## 2. Multinomial Additive Random Utility Models

- So what is then the probability of  $i$  choosing option  $j$ ?

$$\begin{aligned}\Pr[y_{ij} = 1] &= \Pr[U_{ij} \geq U_{ik}, \forall k, j \neq k] & (8) \\ &= \Pr[U_{ik} - U_{ij} \leq 0, \forall k, j \neq k] \\ &= \Pr[\epsilon_{ik} - \epsilon_{ij} \leq V_{ij} - V_{ik}, \forall k, j \neq k] \\ &= \Pr[\tilde{\epsilon}_{ikj} \leq \tilde{V}_{ikj}, \forall k, j \neq k] \\ &= \int I(\epsilon_{ik} - \epsilon_{ij} \leq V_{ij} - V_{ik}, \forall k, j \neq k) f(\epsilon) d\epsilon\end{aligned}$$

- So now several error terms ( $J - 1$ ). Think of  $\epsilon$  as a  $J$  vector.
- Restriction that needs to be satisfied:  $\sum_j \Pr[y_{ij} = 1] = 1$ .
- Covariance restrictions necessary, and a scale normalization.

## How to further specify?

- Need to decide on the specification of  $V_j$ .
- Need to specify the distribution of error terms (incl covariation).
- Interpretation of  $F(\epsilon)$ :

## How to further specify?

- Need to decide on the specification of  $V_j$ .
  - Need to specify the distribution of error terms (incl covariation).
  - Interpretation of  $F(\epsilon)$ :
- 1 Among DMs with same  $V_s$ , the probability of choosing  $j$ .

## How to further specify?

- Need to decide on the specification of  $V_j$ .
- Need to specify the distribution of error terms (incl covariation).
- Interpretation of  $F(\epsilon)$ :
  - 1 Among DMs with same  $V$ s, the probability of choosing  $j$ .
  - 2 The researcher's subjective probability that unobserved utility takes a particular value.

## How to further specify?

- Need to decide on the specification of  $V_j$ .
- Need to specify the distribution of error terms (incl covariation).
- Interpretation of  $F(\epsilon)$ :
  - 1 Among DMs with same  $V_s$ , the probability of choosing  $j$ .
  - 2 The researcher's subjective probability that unobserved utility takes a particular value.
  - 3 e.g. aspects of bounded rationality that lead to a particular choice, conditional on observables.

## How to further specify?

- For your model to be ARUM, the following needs to hold (Börsch-Supan 1987, Williams 1977, Daly and Zachary 1979, McFadden 1981): A set of choice probabilities  $p_j(V)$  is compatible with maximizing an ARUM if
  - ①  $p_j(V) \geq 0$ ,  $\sum p_j(V) = 1$ , and  $p_j(V) = p_j(V + \alpha)$  for  $\forall \alpha \in R$ . (well-behaved probabilities);
  - ②  $\partial p_j(V) / \partial V_k = \partial p_k(V) / \partial V_j$  (integrability of  $p_j$  / Slutsky condition); and
  - ③  $\partial^{m-1} p_j(V) / \partial V_1 \dots [\partial V_k] \dots \partial V_m \geq 0$  (proper density function).
- Why bother? Allows for a welfare analysis.

## 3. Logit

- As noticed, multinomial choice can lead to higher order integrals.
- Question is, how to effectively deal with these.
- E.g., even with symmetry assumptions, a normal distribution leads to a large number of covariance - parameters.

## 2. Logit

- Let us assume  $\epsilon$  is distributed Type I extreme value (see Train, 2002).
- This gives the following decision rule:

$$\Pr[y_i = j] = \frac{\exp V_{ij}}{\sum_k \exp V_{ik}} \quad (9)$$

## 2. Logit

- Specify  $V_{ij} = \mathbf{x}'_{ij}\beta_j$ .
- In words, some of the regressors vary over choices and others over DMs.
- Also here a normalization of parameters needed to ensure that probabilities sum to 1.

## 2. Discussion of Logit

- What is the interpretation of the assumption of errors being independent?

## 2. Discussion of Logit

- What is the interpretation of the assumption of errors being independent?
- Those things affecting choice that the researcher does not observe for option  $j$  provide no information on the things that affect choice  $k$  but the researcher does not observe.

## 2. Discussion of Logit

- Notice that as  $V_j$  gets “very” large (small) keeping  $V_k$  constant,  $\Pr[y = j]$  gets close to 1 (0).
- The estimated probability of any given choice  $j$  is never exactly 1 or 0.
- Given that the pdf is sigmoid, largest impact of an explanatory variable in the middle region.

### 3. Independence of irrelevant alternatives

- Recall that (omit  $i$ )

$$\Pr[y = j] = \frac{\exp V_j}{\sum_k \exp V_k}$$

### 3. Independence of irrelevant alternatives

- Recall that (omit  $i$ )

$$\Pr[y = j] = \frac{\exp V_j}{\sum_k \exp V_k}$$

- What is the ratio of probabilities  $\Pr[y = j]$  and  $\Pr[y = m]$ ?

$$\frac{\Pr[y = j]}{\Pr[y = m]} = \frac{\frac{\exp V_j}{\sum_k \exp V_k}}{\frac{\exp V_m}{\sum_k \exp V_k}} = \frac{\exp V_j}{\exp V_m} \quad (10)$$

### 3. Independence of irrelevant alternatives

- Recall that (omit  $i$ )

$$\Pr[y = j] = \frac{\exp V_j}{\sum_k \exp V_k}$$

- What is the ratio of probabilities  $\Pr[y = j]$  and  $\Pr[y = m]$ ?

$$\frac{\Pr[y = j]}{\Pr[y = m]} = \frac{\frac{\exp V_j}{\sum_k \exp V_k}}{\frac{\exp V_m}{\sum_k \exp V_k}} = \frac{\exp V_j}{\exp V_m} \quad (10)$$

- The “red bus - blue bus” problem.

### 3. Independence of irrelevant alternatives

- Whatever else is on offer does not matter in the choice between  $j$  and  $m$ .
- Implication: when the probability of choosing a given alternative changes, all other choice probabilities change in proportion.
- Reason: IIA forces the ratios of choice probabilities to stay constant.
- Benefit of IIA: Need not observe all choices!

### 3. Independence of irrelevant alternatives

What is actually identified here?

- For some (Luce (1959)) IIA was an attractive property for axiomatizing choice (A feature or a bug?)
- In fact the logit was derived in search for a statistical model that satisfied various axioms.

### 3. Independence of irrelevant alternatives

As another idea: suppose we add a constant  $C$  to each  $\beta_j$ .

$$P_{ij} = \frac{\exp[\mathbf{x}_i(\beta_j + C)]}{\sum_k \exp[\mathbf{x}_i(\beta_k + C)]} = \frac{\exp[\mathbf{x}_i C] \exp[\mathbf{x}_i \beta_j]}{\exp[\mathbf{x}_i C] \sum_k \exp[\mathbf{x}_i \beta_k]}$$

This has no effect. That means we need to fix a normalization  $C$ .

The most convenient is generally that  $C = -\beta_K$ .

- We normalize one of the choices to provide a utility of zero. This is the outside good.

### 3. Independence of irrelevant alternatives

For the linear  $V_{ij}$  case we have  $\frac{\partial V_{ij}}{\partial z_{ij}} = \beta_z$ .

$$\frac{\partial P_{ij}}{\partial z_{ij}} = P_{ij}(1 - P_{ij}) \frac{\partial V_{ij}}{\partial z_{ij}}$$

And elasticity: 
$$\frac{\partial \log P_{ij}}{\partial \log z_{ij}} = P_{ij}(1 - P_{ij}) \frac{\partial V_{ij}}{\partial z_{ij}} \frac{z_{ij}}{P_{ij}} = (1 - P_{ij}) z_{ij} \frac{\partial V_{ij}}{\partial z_{ij}}$$

### 3. Independence of irrelevant alternatives

An important output from a demand system are elasticities

- The above implies that  $\eta_{jj} = \frac{\partial P_{ij}}{\partial p_j} \frac{p_j}{P_{ij}} = \beta_p \cdot p_j \cdot (1 - P_{ij})$ .
- The price elasticity is increasing in own price (recall  $\beta_p < 0$ ). (Why is this a bad idea?)
- Also mechanical relationship between elasticity and **choice probability** so that popular products necessarily have higher markups (holding prices fixed).

## Own and Cross Elasticity

With cross effects: 
$$\frac{\partial P_{ij}}{\partial z_{ik}} = -P_{ij}P_{ik} \frac{\partial V_{ik}}{\partial z_{ik}}$$

and elasticity : 
$$\frac{\partial \log P_{ij}}{\partial \log z_{ik}} = -P_{ik}z_{ik} \frac{\partial V_{ik}}{\partial z_{ik}}$$

- Notice how the cross-partial is a function of choice probabilities.
- This means that two goods with the same choice probability have the same cross (price) derivative with all other goods. Think of a bad but cheap product and an expensive but high quality product that have the same choice probabilities.
- Notice how choice probabilities also dictate the cross price elasticities.
- The cross-price elasticity with good  $k$  is the same for all  $j$ .

## 4. Welfare analysis

- RUM permits welfare analysis. The deterministic component of utility is the indirect utility function

$$V_{ij} = V(I_i - p_j, x_j, z_i) \quad (11)$$

- where  $I_i$  =income,  $p_j$  =price of good  $j$ ,  $x_j$  = characteristics of good  $j$  and  $z_i$  characteristics of  $i$ . (let us suppress estimated parameters for the time being).

## 4. Welfare analysis

- The utility of alternative  $j$  is given by

$$U_{ij} = V(l_i - p_j, x_j, z_i) + \epsilon_{ij} \quad (12)$$

- Suppose we change characteristic  $x_j$  to  $x'_j$ . Compensating variation (CV) is defined by

$$\max_j V(l_i - p_j, x_j, z_i) + \epsilon_{ij} = \max_j V(l_i + CV_i - p_j, x'_j, z_i) + \epsilon_{ij} \quad (13)$$

- Notice how  $CV_i$  depends on what the choice was before the change, as well as what it is after the change.

## Consumer surplus

- RUM + logit permits a (relatively) straight forward calculation of consumer surplus, therefore popular in IO.
- Consumer surplus = (increase in) utility, in monetary terms, of consuming her best alternative (compared to the outside alternative):

$$CS_i = \frac{1}{\alpha_i} \max_j U_{ij} \quad (14)$$

where  $\alpha_i$  = marginal utility of income for individual  $i$ .

- The researcher can measure  $V_{ij}$ , not  $U_{ij}$ , and can hence calculate (at best) the expected utility

$$\mathbb{E}[CS_i] = \frac{1}{\alpha_i} \mathbb{E}[\max_j U_{ij}] = \frac{1}{\alpha_i} \mathbb{E}[\max_j V_{ij} + \epsilon_{ij}] \quad (15)$$

where the expectation is over all the possible values of  $\epsilon_{ij}$ .

## Consumer surplus

- It has been shown (Williams 1977, Small and Rosen 1981) that if utility is linear in income and  $\epsilon_{ij}$  are extreme value, then we can write

$$\mathbb{E}[CS_i] = \frac{1}{\alpha_j} \ln(\sum \exp V_{ij}) + C \quad (16)$$

- Notice how this can be used to e.g. analyze changes in the choice set (e.g. introduction of a new good, withdrawal of an alternative).
- How to get  $\alpha_j$ ? Imagine having price as a variable; then its coefficient is  $-\alpha_j$ .
- The term  $\ln(\sum \exp V_{ij})$  is called the **inclusive value**.
- Notice how we can calculate CV also in this framework, i.e., CV due to change in either choice set or in characteristic(s) of good(s).

## 6. Nested Logit / GEV models

- McFadden (1978) proposed a powerful way to enrich logit models and to “break free” of the IIA.
- GEV = Generalized extreme value.
- Denote  $Y_j = \exp(V_j)$ .
- Consider some function  $G(Y_1, \dots, Y_J)$ .
- Denote  $G_j = \partial G / \partial Y_j$ .

# GEV models

- Then, if certain conditions are met (see e.g. Train, 2002, pp. 97),

$$p_j = \frac{Y_j G_j}{G} \quad (17)$$

- ...is a choice probability from a well-defined utility maximization problem.
- Any model that satisfies equation (17) is a GEV model.

## Relaxing IIA

Let's make  $\varepsilon_{ij}$  more flexible than IID. Hopefully still have our integrals work out.

$$u_{ij} = V_{ij} + \varepsilon_{ij}$$

- One approach is to allow for a block structure on  $\varepsilon_{ij}$  (and consequently on the elasticities).
- We assign products into groups  $g$  and add a group specific error term

$$u_{ij} = V_{ij} + \eta_{ig} + \varepsilon_{ij}$$

- The trick putting a distribution on  $\eta_{ig} + \varepsilon_{ij}$  so that the integrals still work out.

# Nested Logit

A traditional (and simple) relaxation of the IIA property is the Nested Logit. This model is often presented as two sequential decisions.

- First consumers choose a category (following an IIA logit).
- Within a category consumers make a second decision (following the IIA logit).
- This leads to a situation where while choices within the same nest follow the IIA property (do not depend on attributes of other alternatives) choices among different nests do not!

# Nested Logit

Utility looks basically the same as before:

$$U_{ij} = V_{ij} + \underbrace{[\eta_{ig} + \widetilde{\varepsilon}_{ij}]}_{\varepsilon_{ij}(\lambda_g)}$$

- We add a new term that depends on the group  $g$  but not the product  $j$  and think about it as varying unobservably over individuals  $i$  just like  $\varepsilon_{ij}$ .

# Nested Logit

- Now  $\varepsilon_i \sim F(\varepsilon)$  where  $F(\varepsilon) = \exp[-\sum_{g=G}^G (\sum_{j \in J_g} \exp[-\varepsilon_{ij}/\lambda_g])]^{\lambda_g}$ . This is no longer Type I EV but a special kind of GEV.
- The key is the addition of the  $\lambda_g$  parameters which govern (roughly) the within group correlation.
- This distribution is a bit cooked up to get a closed form result, but for  $\lambda_g \in [0, 1]$  for all  $g$  it is consistent with random utility maximization.

# Nested Logit

The nested logit choice probabilities are:

$$P_{ij} = \frac{e^{V_{ij}/\lambda_g} \left( \sum_{k \in J_g} e^{V_{ik}/\lambda_g} \right)^{\lambda_g - 1}}{\sum_{h=1}^G \left( \sum_{k \in J_h} e^{V_{ik}/\lambda_h} \right)^{\lambda_h}}$$

**Within the same group  $g$**  we have IIA and proportional substitution

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{V_{ij}/\lambda_g}}{e^{V_{ik}/\lambda_g}}$$

But for **different groups** we do not:

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{V_{ij}/\lambda_g} \left( \sum_{k \in J_g} e^{V_{ik}/\lambda_g} \right)^{\lambda_g - 1}}{e^{V_{ik}/\lambda_h} \left( \sum_{k \in J_h} e^{V_{ik}/\lambda_h} \right)^{\lambda_h - 1}}$$

## Nested Logit

We can take the probabilities and re-write them slightly with the substitution that  $\log \left( \sum_{k \in J_g} e^{V_{ik}} \right) \equiv IV_{ig} = E_\varepsilon [\max_{j \in G} u_{ij}]$ :

$$\begin{aligned} P_{ij} &= \frac{e^{V_{ij}/\lambda_g}}{\left( \sum_{k \in J_g} e^{V_{ik}/\lambda_g} \right)} \cdot \frac{\left( \sum_{k \in J_g} e^{V_{ik}/\lambda_g} \right)^{\lambda_g}}{\sum_{h=1}^G \left( \sum_{k \in J_h} e^{V_{ik}/\lambda_h} \right)^{\lambda_h}} \\ &= \underbrace{\frac{e^{V_{ij}/\lambda_g}}{\left( \sum_{k \in J_g} e^{V_{ik}/\lambda_g} \right)}}_{P_{ij|g}} \cdot \underbrace{\frac{e^{\lambda_g IV_{ig}}}{\sum_{h=1}^G e^{\lambda_h IV_{ih}}}}_{P_{ig}} \end{aligned}$$

This is the decomposition into two logits that leads to the “sequential logit” story.

# Nested Logit: Notes

- $\lambda_g = 1$  is the simple logit case (IIA)
- $\lambda_g \rightarrow 0$  implies that all consumers stay within the nest.
- $\lambda < 0$  or  $\lambda > 1$  can happen and usually means something is wrong. These models are not generally consistent with RUM.
- $\lambda$  is often interpreted as a correlation parameter and this is almost true but not exactly!

# Nested Logit: Notes

- Because the nested logit can be written as the within group share  $s_{ij|g}$  and the share of the group  $s_{ig}$  we often explain this model as **sequential choice**. It could just be a **block structure** on  $\varepsilon_j$ .
- You need to assign products to categories **before you estimate** and you can't make mistakes!

# Parametric identification

Look at derivatives:

$$\frac{\partial P_{ij|g}}{\partial X_j} = \beta_x \cdot P_{ij|g} \cdot (1 - P_{ij|g})$$

$$\frac{\partial P_{ig}}{\partial X} = (1 - \lambda_g) \cdot \beta_x \cdot P_{ig}(1 - P_{ig})$$

$$\frac{\partial P_{ig}}{\partial J} = \frac{1 - \lambda_g}{J} \cdot P_{ig} \cdot (1 - P_{ig})$$

- We get  $\beta$  by changing  $x_j$  within group
- We get nesting parameter  $\lambda$  by varying  $X$
- We don't have any parameters left to explain changing number of products  $J$ .
- Estimation happens via MLE. This can be tricky because the model is non-convex. It helps to substitute  $\tilde{\beta} = \beta/(1 - \lambda_g)$

## Linking individual level data and market level data

- We have worked until now with the assumption that we observe the choice of individual  $i$ .
- What is the expected market share of good  $j$ ,  $s_j$ ?

$$\mathbb{E}[s_j] = \sum_i P_{ij}$$

- Thus, a natural mapping from choice probabilities to market shares!
- Clearly, the unit of observation changes. To work with market level data need data from many markets.

## Basic idea with price endogeneity

$$P_j = \frac{\exp[x_j\beta]}{1 + \sum_k \exp[x_k\beta]}$$

- We know prices are set with demand in mind and this can create an endogeneity problem.
- How do we deal with it?
- We would like to instrument in this world but what is the error term exactly?
- An obvious choice might be  $\eta_j = (P_j(\theta) - \tilde{P}_j)$
- Can we find things that are orthogonal to the error between observed and predicted market shares?
- Do we have the usual IV conditions (exogeneity, relevance, monotonicity, etc.)

## Basic idea with price endogeneity

- Let's add an unobservable quality term  $\xi_{jt}$  to our model

$$u_{ijt} = x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ij}$$
$$P_{jt} = \frac{\exp[x_{jt}\beta + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]}$$

- The idea is that  $\xi_{jt}$  is observed to the firm when prices are set, but not to us the econometricians.

## Basic idea with price endogeneity

- We call  $\xi_{jt}$  a vertical component, because all consumers agree on its value.
- This allows for products  $j$  to be better than some other product in a way that is not fully explained by differences in  $x_j$  and  $x_k$ .
- Basically there is something about a BMW that makes it better than a Peugeot in a way that is not fully captured by its mileage, weight, horsepower, etc. that leads to it having higher sales and/or higher prices.

## Basic idea with price endogeneity

- How to deal with endogeneity of price with  $\xi_j$ ?
- With individual level data, we can in principle include product fixed effects.
- Further identification achieved by using instruments.
- With individual level data, there are two methods of choice:
  - ① control function approach.
  - ② two-step approach.

# Basic idea of how to deal with price endogeneity #1

- The control function works in two steps (Petrin and Train, 2010):
  - ① Regress the endogenous variable  $p_{jt}$  on all controls + instrument. Generate estimated error  $\tilde{p}_{jt} = p_{jt} - \hat{p}_{jt}$ , where  $\hat{p}_{jt}$  is the predicted price for good  $j$  in market  $t$ .
  - ② Include the estimated error  $\tilde{p}_{jt}$  into the main estimation equation
- Notice: need to correct the standard errors. Recommendation is to bootstrap.

## Basic idea of how to deal with price endogeneity #2

- Two step approach (Gandhi and Nevo, 2021):
  - ① Estimate  $\delta_j$ .
  - ② Project  $\hat{\delta}_j$  on product characteristics, including price.
  - ③ 2<sup>nd</sup> step is linear (IV).
- Notice: need to correct the standard errors. Recommendation is to bootstrap.

## Market level data / Inversion: IIA Logit

- How to move from individual level choice probabilities to market shares?
- Need a way to transform the choice equation into a market share equation.
- Tricks:
  - 1 Utilize the Logit structure and take logs.
  - 2 Use the outside good as the "benchmark".

## Market level data / Inversion: IIA Logit

- Take logs

$$\ln s_{0t} = -\log \left( 1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}] \right)$$

$$\ln s_{jt} = [x_{jt}\beta - \alpha p_{jt} + \xi_{jt}] - \log \left( 1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}] \right)$$

- Then take the difference:

$$\ln s_{jt} - \ln s_{0t} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

## Inversion: IIA Logit

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{\text{data!}} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}}$$

- The LHS is data! The RHS is now a linear IV problem!
- $\alpha$  is the price coefficient on the endogenous variable.
- We know how to solve this. We need instruments that shift  $p_{jt}$  but are orthogonal to  $\xi_{jt}$ .
- Economic theory tells us how: cost shifters, markup shifters.
- Markups in IIA logit are pretty boring since they only depend on your shares and  $\alpha$ .
- If number of products varies across markets, that works. Otherwise you want cost shifters in cross section or time series.

# Caveats

- We do need a technical condition. This only works if the market size  $N \rightarrow \infty$ .
- That is our data/shares we must believe we are observing without any sampling error.
- This is not necessary for the multinomial MLE where shares have some natural sampling variation.
- In our IV/GMM approach we cannot have this sampling error. (Why?).

## Inversion: Nested Logit (Berry 1994 / Cardell 1991)

This takes a bit more algebra but not much

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{data!} = x_{jt}\beta - \alpha p_{jt} - \sigma \underbrace{\log(s_{j|gt})}_{data!} + \xi_{jt}$$

## Inversion: Nested Logit (Berry 1994 / Cardell 1991)

- Same as logit plus an extra term  $\log(s_{j|g})$  the **within group share**.
- We now have a second endogenous parameter.
- If you don't see it – realize we are regressing  $Y$  on a function of  $Y$ . This should always make you nervous.
- If you forget to instrument for  $\sigma$  you will get  $\sigma \rightarrow 1$  because of **attenuation bias**.
- A good instrument for  $\sigma$  is the number of products within the nest.