

# Multinomial Discrete Choice: Nested Logit and GEV

---

Chris Conlon

Fall 2025

Grad IO

# Multinomial Logit: IIA

The multinomial logit is frequently criticized for producing unrealistic substitution patterns

- ▶ Suppose we got rid of a product  $k$  then  $s_{ij}(\mathcal{J} \setminus k) = s_{ij}(\mathcal{J}) \cdot \frac{1}{1 - s_{ik}}$ .
- ▶ Substitution is just proportional to your pre-existing shares  $s_j$
- ▶ No concept of “closeness” of competition!

# Can we do better?

## Multinomial Probit?

- ▶ The probit has  $\varepsilon_i \sim N(0, \Sigma)$ .
- ▶ If  $\Sigma$  is unrestricted, then this can produce relatively flexible substitution patterns.
- ▶ Flexible is relative: still have normal tails, only pairwise correlations, etc.
- ▶ It might be that  $\rho_{12}$  is large if 1, 2 are similar products.
- ▶ Much more flexible than Logit

## Downside

- ▶  $\Sigma$  has potentially  $J^2$  parameters (that is a lot)!
- ▶ Maybe  $J * (J - 1)/2$  under symmetry. (still a lot).
- ▶ Each time we want to compute  $s_j(\theta)$  we have to simulate an integral of dimension  $J$ .
- ▶ I wouldn't do this for  $J \geq 5$ .

# 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.
- ▶ Do not try this at home: it turns out the required distribution is a special case of **GEV** (more on this later) and the resulting model is known as the **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!

## Alternative Interpretation

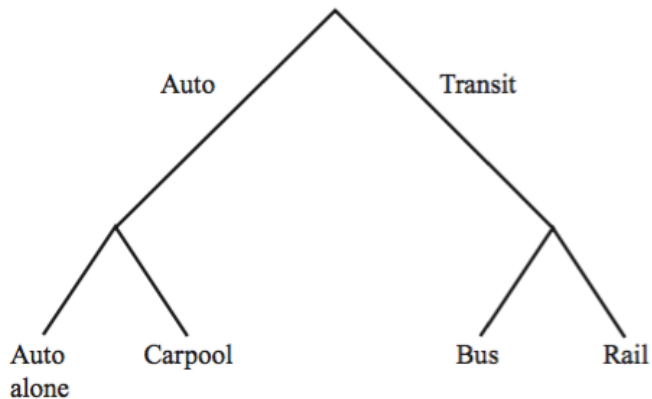


Figure 4.1. Tree diagram for mode choice.

# 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  $\varepsilon_{ij}$ .
- ▶ Now  $\varepsilon_i \sim F(\varepsilon)$  where  $F(\varepsilon) = \exp[-\sum_{g=G}^G \left(\sum_{j \in J_g} \exp[-\varepsilon_{ij}/\lambda_g]\right)^{\lambda_g}]$ . 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:

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

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

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

But for different groups we do not:

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



# Nested Logit

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

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

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. (If you report one in your paper I will reject it).
- ▶  $\lambda$  is often interpreted as a correlation parameter and this is almost true but not exactly!
- ▶ 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  $\varepsilon_i$ .
- ▶ You need to assign products to categories **before you estimate** and you can't make mistakes!

# Parametric Identification

Look at derivatives:

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

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

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

- ▶ 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  $\tilde{\beta} = \beta/(1 - \lambda_g)$

# A Confusing Gotcha

An alternative version of the nested logit is popular in IO (Cardell 1991)  $\sigma \approx 1 - \lambda$ :

$$s_{ij|g} = \frac{e^{V_{ij}/(1-\sigma)}}{D_{ig}}$$

$$s_{ig} = \frac{D_{ig}^{(1-\sigma)}}{\sum_g D_{ig}^{(1-\sigma)}}$$

$$D_{ig} = \sum_{j \in \mathcal{G}} e^{V_{ij}/(1-\sigma)}$$

$$s_{ij} = s_{ij|g} \cdot s_{ig} = \frac{\exp\left(\frac{V_{ij}}{1-\sigma}\right)}{D_g^\sigma \left[ \sum_g D_g^{(1-\sigma)} \right]}$$

Derivatives for nested logit are complicated and worked out at  
<http://www.nathanhmilller.org/nlnotes.pdf>.

# Substitution Patterns

It is helpful to define:  $Z(\sigma, s_g) = [\sigma + (1 - \sigma)s_g] \in (0, 1]$  and note that  $Z(0, s_g) = s_g$  and  $Z(1, s_g) = 1$ . If two products are in the same nest or different nests respectively:

$$-\frac{\frac{\partial s_k}{\partial p_j}}{\frac{\partial s_j}{\partial p_j}} \Big|_{\text{same}} = \frac{s_{k|g}}{Z^{-1}(\sigma, s_g) - s_{j|g}} \equiv D_{jk}^*$$
$$-\frac{\frac{\partial s_k}{\partial p_j}}{\frac{\partial s_j}{\partial p_j}} \Big|_{\text{different}} = \frac{s_k(1 - \sigma)}{1 - s_{j|g} \cdot Z(\sigma, s_{g(j)})} \equiv D_{jk}^{**}$$

These are related by:

$$D_{jk}^{**} = D_{jk}^* \cdot \frac{s_{g(k)} \cdot (1 - \sigma)}{Z(\sigma, s_{g(j)})}$$

There are more potential generalizations though they are less frequently used:

- ▶ You can have multiple levels of nesting: first I select a size car (compact, mid-sized, full-sized) then I select a manufacturer, finally a car.
- ▶ You can have potentially overlapping nests: Yogurt brands are one nest, Yogurt flavors are a second nest. This way strawberry competes with strawberry and/or Dannon substitutes for Dannon.

# McFadden (1978) and GEV

In case you are wondering where these things come from...

$$s_{ij}(\mathcal{J}) = \frac{y_{ij} \cdot \frac{\partial G_i}{\partial y_j}(y_{i1}, \dots, y_{iJ})}{\mu \cdot G(y_{i1}, \dots, y_{iJ})}$$

With conditions on the **generator function**  $G$ :

1.  $G(\cdot)$  is homogenous of degree  $\mu > 0$  so that  $G(\alpha y) = \alpha^\mu G(y)$
2.  $\lim_{y_j \rightarrow +\infty} G(y_1, \dots, y_j, \dots, y_J) = +\infty$ , for each  $j \in \mathcal{J}$
3. the  $k$  th partial derivative with respect to  $k$  distinct  $y_j$  is **non-negative if  $k$  is odd** and **non-positive if  $k$  is even** that is, for any distinct indices  $i_1, \dots, i_k \in \mathcal{J}$ , we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \dots \partial x_{i_k}}(x) \leq 0, \forall x \in \mathbb{R}_+^J$$

The objects are more mathematical than economic...

# McFadden (1978) and GEV

This is much easier with an example:

$$s_{ij}(\mathcal{J}) = \frac{y_{ij} \cdot \frac{\partial G_i}{\partial y_j}(y_{i1}, \dots, y_{iJ})}{\mu \cdot G_i(y_{i1}, \dots, y_{iJ})}$$

- ▶ If  $y_j = e^{V_{ij}}$  and  $G_i = \log \sum_{j \in \mathcal{J}} y_{ij}$  we get the IIA logit.
- ▶ If  $y_j = e^{V_{ij}}$  and  $G_i = \sum_{h=1}^H \left( \sum_{j \in B_h} Y_{ij}^{1/\lambda_h} \right)^{\lambda_h}$  we get the nested logit.
- ▶ ... if  $G_i = \sum_{h=1}^H \left( \sum_{j \in B_h} (\alpha_{jh} Y_{ij})^{1/\lambda_h} \right)^{\lambda_h}$  we get **generalized nested logit** (GNL).
- ▶ ... if  $G = \sum_{k=1}^{J-1} \sum_{l=k+1}^J \left( y_{ik}^{1/\lambda_{kl}} + y_{il}^{1/\lambda_{kl}} \right)^{\lambda_{kl}}$  we get **pairwise combinatorial logit** (PCL).
- ▶ there are a number of other **cross nested logit** variants with slightly different setups (from each other).



What's next?

- ▶ Many of these GEV and variants are found in the engineering literature (particularly traffic problems, civil engineering, and industrial engineering).
- ▶ Economists tend to use either nested logit or mixed logit (next lecture).
- ▶ Part of the issue is that it is hard to understand the restrictions on the  $G$  function and the economic meaning of the patterns produced by some of these models.
- ▶ But they may be more parsimonious and easier to estimate than the alternatives.