Notes on *Discrete Choice Methods with Simulation* (Train, 2009)

Chapter 4. Generalized extreme value (GEV)

- The standard logit model exhibits the property of IIA, which can be seen either as a restriction imposed by the model or as the natural outcome of a well-specified model that captures all sources of correlation over alternatives into representative utility.
- However, the researcher is often unable to capture all sources of correlation explicitly, so that the unobserved portions of utility are correlated and IIA does not hold. Therefore, a more general model than standard logit is needed. The most widely used member of the GEV family is called *nested logit*. GEV models have the advantage that the choice probabilities usually take a closed form, so that they can be estimated without resorting to simulation.

4.1. Nested logit

- A nested logit model is appropriate when the set of alternatives faced by a decision maker can be partitioned into subsets, called *nests*, in such a way that the following properties hold:
    1. For any two alternatives that are in the *same* nest, the ratio of probabilities is independent of the attributes or existence of all other alternatives. That is, IIA holds within each nest.
    2. For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests. IIA does not hold in general for alternatives in different nests.
- The decision-making process can be regarded as a two-step (or multi-step) process: first, choosing a subset (e.g., private or public transit), and then selecting an alternatives within the subset (e.g., bus or rail if public transit is chosen).

4.1.1. Choice probabilities

- Let the set of alternatives j be partitioned into K nonoverlapping subsets denoted $B_1, B_2, \cdots, B_K$ and called nests. The utility that person n obtained from alternative j in nest B_k is denoted, as earlier, as $U_{nj} = V_{nj} + \varepsilon_{nj}$. The nested logit model is obtained by assuming that the vector of unobserved utility, $\varepsilon_n = [\varepsilon_{n1}, \cdots, \varepsilon_{nJ}]'$ has the following cumulative distribution:

$$F(\varepsilon_n) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} \mathrm{e}^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right),$$

  where the parameter λ_k is a measure of the degree of independence in unobserved utility among the alternatives in nest k. When $\lambda_k = 1$ for all k, the nested logit model reduces to the standard logit model. Therefore, by testing the constraint $\lambda_k = 1$ for any k, we can test whether the standard logit model is a reasonable specification against the more general nested logit.
- The choice probability of alternative i for consumer n is

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

Consider alternatives $i \in B_k$ and $m \in B_l$. The ratio of probabilities is

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

If $k = l$, then

$$\frac{P_{ni}}{P_{nm}} = \frac{\mathrm{e}^{V_{ni}/\lambda_k}}{\mathrm{e}^{V_{nm}/\lambda_k}},$$

which is independent of all other alternatives. If $k \neq l$, the ratio depends on the attributes of all alternatives in the nests that contain i and m but not on the attributes of alternatives in other nests. Therefore, a form of IIA holds: the ratio is independent from *irrelevant* nests (IIN).

- The value of λ_k must be positive. If λ_k for any k is between zero and one, the model is consistent with utility maximization for all possible values of the explanatory variables. For λ_k greater than one, the model is consistent with utility-maximizing behavior for some range of the explanatory variables but not for all values.
- In practice, λ_k can be different for different decision makers. To depict this heterogeneity, each λ_k can be specified to be a parametric function of observed demographics or other variables, as long as the function maintains a positive value (e.g., $\lambda_k = \mathrm{e}^{\alpha z_n}$).

4.1.2. Decomposition into two logits

- The observe component of utility can be decomposed into two parts: (1) a part labeled W that is constant for all alternatives within a nest, and (2) a part labeled Y that varies over alternatives within a nest:

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj}.$$

- With this decomposition of utility, the nested logit probability can be written as the product of two standard logit probabilities:

$$P_{ni} = P_{ni|B_k} P_{nB_k},$$

where $P_{ni|B_k}$ is the conditional probability of choosing alternative i given that an alternative in nest B_k is chosen, and P_{nB_k} is the marginal probability of choosing an alternative in nest B_k. These two probabilities is exactly

$$P_{nB_k} = \frac{\mathrm{e}^{W_{nk} + \lambda_k I_{nk}}}{\sum_{l=1}^{K} \mathrm{e}^{W_{nl} + \lambda_l I_{nl}}}$$

and

$$P_{ni|B_k} = \frac{\mathrm{e}^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} \mathrm{e}^{Y_{nj}/\lambda_k}},$$

where $I_{nk} = \ln \sum_{j \in B_k} \mathrm{e}^{Y_{nj}/\lambda_k}$.

4.1.3. Estimation

- The parameters of a nested model can be estimated by standard maximum likelihood techniques. Substituting the choice probabilities into the log-likelihood function gives an explicit function of the parameters of this model. The values of the parameters that maximize this function are, under fairly general conditions, consistent and efficient (Brownstone and Small, 1989).

- Numerical maximization is sometimes difficult, since the log-likelihood function is not globally concave and even in concave areas is not close to a quadratic. The researcher may need to help the routines by trying different algorithms and/or starting values.

4.2. Three-/higher-level nested logit

- In some situations, three- or higher-level nested logit models are appropriate. Three-level models are obtained by partitioning the set of alternatives into nests and then partitioning each nest into subnests.
- The key to derive the choice probability is to decompose it into three/multiple components of conditional probabilities and marginal probabilities (like that in 4.1.2), which take the form of logits and is easier to calculate.

4.3. Overlapping nests

- Sometimes, an alternative could be in several different nests since it shares attributes with alternatives in different nests. In other words, alternatives may show an overlapping-nest structure.
- To capture this structure, we assume that each pair of alternatives form a nest. A parameter λ_{ij} indicates the degree of independence between alternatives i and j, which is analogous to the λ_k in a nested logit model. This is called *paired combinatorial logit* (PCL) model. The choice probabilities for the PCL model are

$$P_{ni} = \frac{\sum_{j \neq i} \mathrm{e}^{V_{ni}/\lambda_{ij}} \left(\mathrm{e}^{V_{ni}/\lambda_{ij}} + \mathrm{e}^{V_{nj}/\lambda_{ij}}\right)^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left(\mathrm{e}^{V_{nk}/\lambda_{kl}} + \mathrm{e}^{V_{nl}/\lambda_{kl}}\right)^{\lambda_{kl}}},$$

  which reduce the probabilities of the standard logit model when all λ_{ij} is 1.
- The researcher can test the hypothesis that $\lambda_{ij} = 1$ for some or all of the pairs, using the likelihood ratio test, to see whether a overlapping-nest structure exists.
- Further, we can propose a *generalized nested logit* (GNL) model. Suppose all nests are labeled $B_1, \cdots, B_K$. Each alternative can be a member of more than one nest. Besides, an alternative can be in a nest to varying degree. An allocation parameter α_{jk} reflects the extent to which alternative j is a member of nest k. We set $\sum_k \alpha_{jk} = 1$ for any j, therefore, α_{jk} reflects the portion of the alternative that is allocated to each nest.
- The probability that person n choose alternative i is

$$P_{ni} = \frac{\sum_k \left(\alpha_{ik}\mathrm{e}^{V_{ni}}\right)^{1/\lambda_k} \left(\sum_{j \in B_k} \left(\alpha_{jk}\mathrm{e}^{V_{nj}}\right)^{1/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} \left(\alpha_{jl}\mathrm{e}^{V_{nj}}\right)^{1/\lambda_l}\right)^{\lambda_l}}.$$

4.4. The GEV family

- Here summarizes the process that McFadden (1978) developed to generate GEV models. For notational simplicity, omit the subscript n denoting the decision maker and denote $\exp(V_j)$ as Y_j. Consider a generating function $G = G(Y_1, \cdots, Y_J)$ that depends on Y_j for all j. Let G_i be the derivative of G with respect to Y_i: $G_i \equiv \partial G / \partial Y_i$. If this function meets certain conditions, then a discrete choice model can be based upon it and the choice probability is

$$P_i = \frac{Y_i G_i}{G}.$$

The conditions that function G must meet are the following:

1. $G \geq 0$ for all positive values of any Y_j.
2. G is homogeneous of degree one. That is, if each Y_j is raised by some proportion ρ, G rises by proportion ρ also: $G(\rho Y_1, \cdots, \rho Y_J) = \rho G(Y_1, \cdots, Y_J)$.
3. $G \to \infty$ as $Y_j \to \infty$ for any j.
4. The cross partial derivatives G change signs in a particular way. That is, $G_i \geq 0$ for all i, $G_{ij} = \partial G_i / \partial Y_j \leq 0$ for all $j \neq i$, $G_{ijk} = \partial G_{ij} / \partial Y_k \geq 0$ for any distinct i, j, and k, and so on for higher order cross partials.

- Any model that can be derived in this way is a GEV model. We can show how logit, nested logit, and PCL model are obtained under appropriate specifications of G.
  - Logit:

$$G_L = \sum_{j=1}^{J} Y_j.$$

  - Nested logit:

$$G_{NL} = \sum_{k=1}^{K} \left(\sum_{j \in B_k} Y_j^{1/\lambda_k} \right)^{\lambda_k}.$$

  - Paired combinatorial logit:

$$G_{PCL} = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left(Y_k^{1/\lambda_{kl}} + Y_l^{1/\lambda_{kl}} \right)^{\lambda_{kl}}.$$

  - Generalized nest logit:

$$G_{GNL} = \sum_{k=1}^{K} \left(\sum_{j \in B_k} (\alpha_{jk} Y_j)^{1/\lambda_k} \right)^{\lambda_k}.$$

- Note that G is closely related to the assumed joint distribution for the unobserved terms:

$$F(\varepsilon_n) = \exp\left(-G\left(\mathrm{e}^{-\varepsilon_{n1}}, \cdots, \mathrm{e}^{-\varepsilon_{nJ}}\right)\right).$$

  This implies that to develop a new GEV model, we can start by specifying a generating function G that satisfies all the conditions mentioned above, and then derive the corresponding distribution for the unobserved terms.
- The conditions required for a generating function G, especially the last one, lack clear economic intuition. However, it is easy to verify whether a function exhibits these properties. The lack of intuition behind the properties is a blessing and a curse. The disadvantage is that the researcher has little guidance on how to specify a G that provides a model that meets the needs of his research. The advantage is that the purely mathematical approach allows the researcher to generate models that he might not have developed while relying only on his economic intuition.