Probabilistic Choice Models

James J. Heckman
University of Chicago

Econ 312, Spring 2023

- This chapter examines different models commonly used to model probabilistic choice, such as eg the choice of one type of transportation from among many choices available to the consumer.
- Section 1 discusses derivation and limitations of conditional logit models.
- Section 2 discusses probit models and Section 3 discusses the nested logit (generalized extreme value models), which address some of the limitations of the conditional logit models.

The Conditional Logit Model

- In this section we investigate conditional logit models.
- We discuss its derivation from a random utility model with Extreme Value Type I distributed shocks.
- The relevant properties of the Extreme Value Type I distribution are discussed.
- We also derive the conditional logit model from the Luce axioms.
- We discuss some of the limitations of the conditional logit models.

The Extreme Value Type I Distribution

- Suppose ε is independent (not necessarily identical) Extreme Value Type I random variable.
- Then the CDF of ε is:

$$\Pr(\varepsilon < c) = F(c) = \exp\left(-\exp\left(-\left(c + \alpha_i\right)\right)\right)$$

where α_i is a parameter of the Extreme Value Type I CDF.

- Also, by the assumption of independence, we can write:

$$F\left(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\right) = \prod_{i=1}^{n} F\left(\varepsilon_i\right) = \prod_{i=1}^{n} \exp\left(-\exp\left(-\left(\varepsilon_i + \alpha_i\right)\right)\right)$$

- The Extreme Value Type I distribution has two useful features.
- First, the difference between two Extreme Value Type I random variables is a logit.
- Second, Extreme Value Type Is are closed under maximization, since (assuming independence):

$$\begin{aligned}
\Pr\left(\max_i\{\varepsilon_i\} \leq \varepsilon\right) &= \prod_{i=1}^{n} \Pr(\varepsilon_i \leq \varepsilon) \\
&= \prod_{i=1}^{n} \exp\left(-\exp\left(-\left(\varepsilon + \alpha_i\right)\right)\right) \\
&= \exp\left(-\left(\sum_{i=1}^{n} \exp\left(-\left(\varepsilon + \alpha_i\right)\right)\right)\right) \\
&= \exp\left(-\exp\left(-\varepsilon\right)\sum_{i=1}^{n} \exp(-\alpha_i)\right) \quad (1)
\end{aligned}$$

- Consider $\sum_{i=1}^{n} \exp(-\alpha_i)$.
- We can solve for α in the following equation:

$$\sum_{i=1}^{n} \exp(-\alpha_i) = \exp(-\alpha)$$

which implies:

$$-\alpha = \log\left(\sum_{i=1}^{n} \exp(-\alpha_i)\right).$$

- We can then substitute this value of α into equation (1) to get:

$$\begin{aligned} \Pr\left(\max_i\{\varepsilon_i\} \leq \varepsilon\right) &= \exp\left(-\left(\exp\left(-\varepsilon\right)\right)\exp(-\alpha)\right) \\ &= \exp\left(-\exp\left(-\left(\varepsilon+\alpha\right)\right)\right) \end{aligned}$$

which is indeed a Extreme Value Type I random variable.

Random Utility Model

- An individual with characteristics s has a choice set B; with element $x \subseteq B$, B is a feasible set.
- We write:

$$\Pr(x \mid s, B)$$

as the probability that a person of characteristics s chooses x from the feasible set.

- We also suppose that:

$$U(s,x) = v(s,x) + \varepsilon(s,x)$$

where ε is independent Extreme Value Type I.

- From our information on Extreme Value Type Is in section 1, we know that $\varepsilon_i + v_i$, (and thus U_i), has an Extreme Value Type I distribution with parameter $\alpha_i - v_i$, as shown below:

$$\begin{aligned} F_{U_i}(\varepsilon) = \Pr(\varepsilon_i + v_i < \varepsilon) &= \Pr(\varepsilon_i < \varepsilon - v_i) \\ &= \exp(-\exp(-(\varepsilon + \alpha_i - v_i))) \end{aligned}$$

- Let us now suppose that there are two goods and two corresponding utilities.
- Consumers govern their choices by the obvious decision rule: choose good one if $U_1 > U_2$.
- More generally, if there are n goods, then good j will be selected if $U_j \in \text{argmax } \{U_i\}_{i=1}^{n}$.

- Specifically, in our two good case:

$$\Pr(1 \text{ is chosen}) = \Pr(U_1 > U_2) = \Pr(\varepsilon_1 + v_1 > \varepsilon_2 + v_2)$$

- Imposing that ε is independent Extreme Value Type I, we can be much more precise about this probability:

$$\begin{aligned} & \Pr\left(\varepsilon_1 + v_1 > \varepsilon_2 + v_2\right) \qquad (2) \\ = \; & \Pr\left(\varepsilon_1 + v_1 - v_2 > \varepsilon_2\right) \\ = \; & \int_{-\infty}^{\infty} f(\varepsilon_1) \left(\int_{-\infty}^{\varepsilon_1 + v_1 - v_2} f(\varepsilon_2) d\varepsilon_2 \right) d\varepsilon_1 \\ = \; & \int_{-\infty}^{\infty} f(\varepsilon_1) \exp\left(-\exp-\left(\varepsilon_1 + v_1 - v_2 + \alpha_2\right)\right) d\varepsilon_1 \end{aligned}$$

- Observe that $F(\varepsilon_1) = \exp\left(-\exp-(\varepsilon_1 + \alpha_1)\right)$, which implies:

$$\begin{aligned} f(\varepsilon_1) &= \frac{\partial F(\varepsilon_1)}{\partial \varepsilon_1} \\ &= \exp\left(\exp-(\varepsilon_1 + \alpha_1)\right)\left(\exp-(\varepsilon_1 + \alpha_1)\right) \\ &= \exp-(\varepsilon_1 + \alpha_1)\left(\exp\left(-\exp-(\varepsilon_1 + \alpha_1)\right)\right) \end{aligned}$$

- Substituting this into equation (2) gives us:

$$\begin{aligned}\Pr(1 \text{ is chosen}) &= \int_{-\infty}^{\infty} \exp-(\varepsilon_1+\alpha_1)\left(\exp\left(-\exp-(\varepsilon_1+\alpha_1)\right)\right) \\ &\qquad \exp\left(-\exp-(\varepsilon_1+v_1-v_2+\alpha_2)\right) d\varepsilon_1 \\ &= e^{-\alpha_1}\int_{-\infty}^{\infty}\left(e^{-\varepsilon_1}\right) e^{[-\exp(-\varepsilon_1)][\exp(-\alpha_1)-\exp-(v_1-v_2+\alpha_2)]} d\varepsilon_1\end{aligned}$$

$$
\begin{aligned}
&= \exp(-\alpha_1)\left[\frac{1}{\exp(-\alpha_1)+\exp-(v_1-v_2+\alpha_2)}\right] \\
&\qquad\qquad \left[e^{[-\exp(-\varepsilon_1)][\exp(-\alpha_1)-\exp-(v_1-v_2+\alpha_2)]}\right]_{-\infty}^{\infty} \\
&= \frac{\exp(-\alpha_1)}{\exp(-\alpha_1)+\exp-(v_1-v_2+\alpha_2)} \\
&= \frac{\exp(v_1-\alpha_1)}{\exp(v_1-\alpha_1)+\exp(v_2-\alpha_2)}
\end{aligned}
$$

- This result generalizes, because the max over $(n-1)$ choices is still an Extreme Value Type I, so we can make a two stage maximization argument, as follows:

$$
\begin{aligned}
& \Pr\left(\varepsilon_1 + v_1 > \varepsilon_i + v_i, \quad i = 1, 2, \cdots, n\right) \\
= \; & \Pr\left(\varepsilon_1 + v_1 > \max_{i=2,\cdots,n} \left(\varepsilon_i + v_i\right)\right) \\
= \; & \frac{\exp(v_1 - \alpha_1)}{\exp(v_1 - \alpha_1) + \exp(v_2 - \alpha_2) + \cdots + \exp(v_n - \alpha_n)} \\
= \; & \frac{\exp(\tilde{v}_1)}{\sum_{i=1}^{n} \exp\left(\tilde{v}_i\right)}
\end{aligned}
$$

where $\tilde{v}_j = v_j - \alpha_j$.

- This type of model of probabilistic choice is called a conditional or multinomial logit model.
- The difference between "conditional" and "multinomial" is simply that in the "conditional" logit case, the values of the variables (usually choice characteristics) vary across the choices, while the parameters are common across the choices.

- In the "multinomial" logit case, the values of the variables are common across choices for the same person (usually individual characteristics) but the parameters vary across choices.

- For e.g. we have in the linear v_i case, the probability of individual j making choice i from among m choices is:

  Conditional Logit case: $P_{ij} = \frac{\exp(\beta' c_{ij})}{\sum_{k=1}^{m} \exp(\beta' c_{kj})}$, where c_{ij} is the vector of values of characteristics of choice i as perceived by individual j.

  Multinomial Logit case: $P_{ij} = \frac{\exp(\alpha_i' s_j)}{\sum_{k=1}^{m} \exp(\alpha_k' s_j)}$, where s_j is a vector of individual characteristics for individual j.

- Note that we can easily combine the two cases under one model, as described below:

  Generalized case: We can combine the conditional and multinomial logit models by generalizing either one of the two types of models. For eg, we could permit the coefficients in the multinomial logit case to depend on choice characteristics, ie have:

$$\alpha_i = \phi_i + c_{ij}'\theta$$

- Then we get the generalized case, where the probability of choice i by individual j depends on both individual as well as choice characteristics (as well as interaction terms):

$$P_{ij} = \frac{\exp(\alpha_i' s_j)}{\sum_{k=1}^{m} \exp(\alpha_k' s_j)} = \frac{\exp(\phi_i' s_j + \theta' c_{ij} s_j)}{\sum_{k=1}^{m} \exp(\phi_k' s_j + \theta' c_{kj} s_j)}$$

- We could similarly modify the coefficients in the conditional logit case to obtain the generalized version.

Derivation of Logit from the Luce Axioms

- We will now show how the conditional logit can be derived from the random utility model and the Luce Axioms presented below.

Luce Axioms

Axiom 1: Independence of Irrelevant Alternatives(IIA)

Suppose that $x, y \in B$, $s \in S$.

Then,

$$\Pr(x \mid s, \{x, y\}) \Pr(y \mid s, B) = \Pr(y \mid s, \{x, y\}) \Pr(x \mid s, B)$$

or, we have:

$$\frac{\Pr(x \mid s, \{x, y\})}{\Pr(y \mid s, \{x, y\})} = \frac{\Pr(x \mid s, B)}{\Pr(y \mid s, B)}.$$

- The term on the left is the odds ratio; the ratio of probabilities of choosing x to y given characteristics s and $\{x, y\}$.
- This axiom has been named "Independence of Irrelevant Alternatives" for an obvious reason — the odds of our choice are not effected by adding additional alternatives.
- Note that this assumes that the additional choices entering in B affect probability of choosing x in the same manner as they affect the probability of choosing y; implicitly we are assuming that the additional choices have equivalent relationship with choice x and choice y.
- We will see how this assumption is a limitation below.

Axiom 2: Positivity

This axiom states that the probability of choosing any one of the choices is strictly greater than zero:

$$\Pr(y \mid s, B) > 0 \quad \forall\ y \in B$$

Derivation of Logit

- With the Luce assumptions set out in the proceeding section, we can now proceed to our derivation of the logit.
- Define $P_{yx} = \Pr(y \mid s, \{x, y\})$.
- Then by Axiom 1 above, we know:

$$\left(\frac{P_{yx}}{P_{xy}}\right) \Pr(x \mid s, B) = \Pr(y \mid s, B) \tag{3}$$

- Summing over y, we get:

$$\Pr(x \mid s, B) \sum_{y \in B} \left(\frac{P_{yx}}{P_{xy}} \right) = 1$$

$$\Longrightarrow \quad \Pr(x \mid s, B) = \frac{1}{\sum_{y \in B} \left(\frac{P_{yx}}{P_{xy}} \right)} \tag{4}$$

- Again using Axiom 1, for $z \in B$:

$$\left(\frac{P_{yz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right) = \Pr\left(y \mid s, B\right) \tag{5a}$$

$$\left(\frac{P_{xz}}{P_{zx}}\right) \Pr\left(z \mid s, B\right) = \Pr\left(x \mid s, B\right) \tag{5b}$$

- Substituting these in equation (3), we get:

$$\left(\frac{P_{yx}}{P_{xy}}\right) = \frac{\Pr\left(y \mid s, B\right)}{\Pr\left(x \mid s, B\right)} = \frac{\left(\frac{P_{yz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right)}{\left(\frac{P_{xz}}{P_{zy}}\right) \Pr\left(z \mid s, B\right)} = \frac{\frac{P_{yz}}{P_{zy}}}{\frac{P_{xz}}{P_{zx}}} \tag{6}$$

- Now, in terms of the random utility model , define the mean utility of a person with characteristics s choosing x from set $\{x, z\}$ as:

$$v(s, x, z) \equiv \ln \frac{P_{xz}}{P_{zx}}$$

$$\Longrightarrow \frac{P_{xz}}{P_{zx}} = \exp\left(v(s, x, z)\right)$$

- Define a comparable expression for $\frac{P_{yz}}{P_{zy}}$.
- Replacing this into equation (6) produces:

$$\frac{P_{yx}}{P_{xy}} = \frac{\exp\left(v(s, y, z)\right)}{\exp\left(v(s, x, z)\right)}$$

- Then from equation (4), we get:

$$\begin{aligned}
\Pr(x \mid s, B) &= \frac{1}{\sum_{y \in B} \left(\frac{\exp(v(s,y,z))}{\exp(v(s,x,z))} \right)} \\
&= \frac{1}{\left(\frac{1}{\exp(v(s,x,z))} \right) \sum_{y \in B} (\exp(v(s,y,z)))} \\
&= \frac{\exp(v(s,x,z))}{\sum_{y \in B} (\exp(v(s,y,z)))}.
\end{aligned}$$

- Assume additionally, additive separability of $v(s, x, z)$ as follows:

$$v(s, x, z) = v(s, x) - v(s, z)$$

- Note that this is equivalent to assuming irrelevance of the benchmark. From this assumption, we get:

$$\begin{aligned}
\Pr(x \mid s, B) &= \frac{\exp\left(v(s,x) - v(s,z)\right)}{\sum_{y \in B}\left(\exp\left(v(s,y) - v(s,z)\right)\right)} \\
&= \frac{\exp v(s,x)\exp\left(-v(s,z)\right)}{\exp\left(-v(s,z)\right)\left(\sum_{y \in B}\exp\left(v(s,y)\right)\right)} \\
&= \frac{\exp v(s,x)}{\sum_{y \in B}\exp\left(v(s,y)\right)} \qquad (7)
\end{aligned}$$

which gives the multinomial logit.

- McFadden (1974) shows that Luce Axioms and a condition on ε ("Translation Completeness") produce the Extreme Value Type I (which he mistakenly referred to as the Weibull).

Consequences of Independence: Limitations of Logit Models

- We just showed that:

$$P_i = \frac{\exp(v_i)}{\sum_i \exp(v_i)}$$

so that:

$$\frac{P_i}{P_j} = \frac{\frac{\exp(v_i)}{\sum_i \exp(v_i)}}{\frac{\exp(v_j)}{\sum_i \exp(v_i)}} = \frac{\exp(v_i)}{\exp(v_j)} = \exp\left(v_i - v_j\right) \Rightarrow \ln\left(\frac{P_i}{P_j}\right) = v_i - v_j$$

- A common specification for v_i is $v_i = z_i\beta$. Thus:

$$\ln\left(\frac{P_i}{P_j}\right) = (z_i - z_j)\,\beta \Rightarrow \frac{\partial \ln\left(\frac{P_i}{P_j}\right)}{\partial z_j} = -\beta$$

or, changes in characteristics z_j have a common effect on the ratio of log probabilities.

- This allows for estimation of the probabilities of purchasing a new good.
- (One could obtain an estimate of β from the existing goods. This estimate can then be combined with the characteristics, z_{new}, of the new good to estimate the probability of selection, as in equation 7).

- Further, from equation (7):

$$\Pr(2 \mid \{1,2\}) = \frac{e^{v_2}}{e^{v_1} + e^{v_2}}$$

and:

$$\Pr(2 \mid \{1,2,3\}) = \frac{e^{v_2}}{e^{v_1} + e^{v_2} + e^{v_3}} < \Pr(2 \mid \{1,2\})$$

- This leads us to a restrictive property of the conditional logit model – we have assumed independence of the ε_i, when in fact, they may be correlated.

- This is illustrated by McFadden's famous red bus, blue bus problem:
- Suppose we are modelling transportation choice and our alternatives consist of $\{$car, bus, train$\}$.
- If the alternatives are replaced by $\{$car, red bus, blue bus$\}$, then we have violated our assumption of dissimilar alternatives; if $U_2 > U_1$, then the event $U_3 > U_1$ is more likely.

- One can see by the preceding equation that adding more bus colors continually decreases the probability that car travel is chosen.
- We can deal with the problem of similar alternatives by using the nested logit model (Nested Logit) or the random coefficient probit model.

Probit: Random Coefficients

- In this section (as above), we make v_i a simple linear function of the choice characteristics alone, we can easily generalize this to include individual characteristics as well as interactions).
- Then we have, utility from choice i is:

$$U_i = Z_i\beta + \eta_i$$

where: $\eta_i \sim N(0, \sigma_i^2)$, $\eta_i \perp\!\!\!\perp Z_i, \beta, \eta_j, \forall\ i, j$.

- Moreover, β is a random variable, with $\beta \sim (\bar{\beta}, \sum_{\beta})$, so that:

$$U_i = Z_i \bar{\beta} + Z_i \left(\beta - \bar{\beta}\right) + \eta_i$$

- It follows that:

$$U_1 - U_2 \geq 0 \Longleftrightarrow (Z_1 - Z_2)\,\bar{\beta} + (Z_1 - Z_2)\left(\beta - \bar{\beta}\right) + (\eta_1 - \eta_2) \geq 0$$

$$U_1 - U_3 \geq 0 \Longleftrightarrow (Z_1 - Z_3)\,\bar{\beta} + (Z_1 - Z_3)\left(\beta - \bar{\beta}\right) + (\eta_1 - \eta_3) \geq 0.$$

- Further:

$$
\begin{aligned}
\text{Var}\,(U_1 - U_2) &= E\left\{ \begin{bmatrix} (U_1 - U_2) \\ -E\,(U_1 - U_2) \end{bmatrix}' \begin{bmatrix} (U_1 - U_2) \\ -E\,(U_1 - U_2) \end{bmatrix} \right\} \\
&= E\left\{ [(Z_1 - Z_2)\,(\beta - \bar{\beta}) + (\eta_1 - \eta_2)]' \right. \\
&\qquad \left. [(Z_1 - Z_2)\,(\beta - \bar{\beta}) + (\eta_1 - \eta_2)] \right\} \\
&= E\left\{ \begin{array}{c} (Z_1 - Z_2)\,(\beta - \bar{\beta})(\beta - \bar{\beta})'(Z_1 - Z_2)' \\ +(\eta_1 - \eta_2)(\eta_1 - \eta_2)' \end{array} \right\} \\
&= (Z_1 - Z_2) \sum\nolimits_{\beta} (Z_1 - Z_2)' + \sigma_1^2 + \sigma_2^2
\end{aligned}
$$

(since $\sigma_{12} = 0$).

- Similarly:

$$\text{Var}\left(U_1 - U_3\right) = \left(Z_1 - Z_3\right) \sum \beta \left(Z_1 - Z_3\right)' + \sigma_1^2 + \sigma_3^2$$

- Thus:

$$\text{Cov}\left(U_1 - U_2, U_1 - U_3\right) = \left(Z_1 - Z_2\right) \sum \beta \left(Z_1 - Z_3\right)' + \sigma_1^2$$

so:

$$\rho = \text{Corr}\left(U_1 - U_2, U_1 - U_3\right) = \frac{\left(Z_1 - Z_2\right) \sum \beta \left(Z_1 - Z_3\right)' + \sigma_1^2}{\sqrt{\text{Var}\left(U_1 - U_2\right) \text{Var}\left(U_1 - U_3\right)}}$$

- We now seek to derive the probability of choosing good 1 in a three good case:

$$\Pr\left(1 \mid \{1,2,3\}\right) = \Pr\left(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0\right).$$

- From before, we know that:

$$\begin{aligned} U_1 - U_2 &\sim N\left(\left(Z_1 - Z_2\right)\bar{\beta}, \text{Var}\left(U_1 - U_2\right)\right) \\ U_1 - U_3 &\sim N\left(\left(Z_1 - Z_3\right)\bar{\beta}, \text{Var}\left(U_1 - U_3\right)\right). \end{aligned}$$

- Thus:

$$\begin{aligned} & \Pr\left(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0\right) \\ = \quad & \Pr\left[\begin{array}{c} \sqrt{\text{Var}\left(U_1 - U_2\right)} t_1 + \left(Z_1 - Z_2\right)\bar{\beta} \geq 0 \\ \text{and } \sqrt{\text{Var}\left(U_1 - U_3\right)} t_2 + \left(Z_1 - Z_3\right)\bar{\beta} \geq 0 \end{array}\right], \end{aligned}$$

where t_1 and t_2 are standard normal.

- Thus, the above equation reduces to:

$$\Pr\left(t_1 \geq -\frac{(Z_1 - Z_2)\,\bar{\beta}}{\sqrt{\mathrm{Var}\,(U_1 - U_2)}} \text{ and } t_2 \geq -\frac{(Z_1 - Z_3)\,\bar{\beta}}{\sqrt{\mathrm{Var}\,(U_1 - U_3)}}\right)$$

$$= \Pr\left(t_1 \leq \frac{(Z_1 - Z_2)\,\bar{\beta}}{\sqrt{\mathrm{Var}\,(U_1 - U_2)}} \text{ and } t_2 \leq \frac{(Z_1 - Z_3)\,\bar{\beta}}{\sqrt{\mathrm{Var}\,(U_1 - U_3)}}\right)$$

- As t_1 and t_2 may be correlated, we integrate over the joint density to get the probability:

$$\begin{aligned}
&\Pr(\text{choosing } 1) \\
&= \int_{-\infty}^{a} \left(\int_{-\infty}^{b} \left(\frac{1}{2\pi\sqrt{1\text{-}\rho^2}} \, e^{-\frac{1}{2}\left(\frac{t_1^2\text{-}2\rho t_1 t_2 + t_2^2}{1\text{-}\rho^2}\right)} \right) dt_2 \right) dt_1
\end{aligned}$$

where:

$$a = \frac{(Z_1 - Z_2)\,\bar{\beta}}{\sqrt{\text{Var}\,(U_1 - U_2)}}, \text{ and } b = \frac{(Z_1 - Z_3)\,\bar{\beta}}{\sqrt{\text{Var}\,(U_1 - U_3)}}$$

- Now consider adding a third good to the two good case, under two alternative scenarios.

  Case 1: Non-random utility, random coefficients.

  If the third good has identical characteristics as the first, then $Z_2 = Z_3$.
  If there is no stochastic component (no utility innovation), then $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$.
  Therefore, in this case:

$$\begin{aligned} \Pr(1 \text{ chosen}) &= \Pr(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0) \\ &= \Pr(U_1 - U_2 \geq 0) \end{aligned}$$

- Thus, there is no change in the probability of choosing good 1 despite the addition of a third good.
- Again focusing on the two good case, we observe:

$$\begin{aligned}
\Pr(1 \mid \{1,2\}) &= \Pr(U_1 - U_2 \geq 0) \\
&= \Pr\left(t_1 \leq \frac{(Z_1 - Z_2)\,\bar{\beta}}{\sqrt{\mathrm{Var}(U_1 - U_2)}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{[(Z_1 - Z_2)\Sigma_\beta (Z_1 - Z_2)' + \sigma_1^2 + \sigma_2^2]^{1/2}}} \exp\left(-\frac{t_1^2}{2}\right) dt
\end{aligned}$$

which can be evaluated to derive the desired probability.

cients.

- Here we consider a McFadden-Luce type of set up, where one imposes $\sum \beta = 0$.
- Defining $\sigma^* = \sqrt{\sigma_1^2 + \sigma_2^2}$, we observe that the probability of choosing good 1 in the two-good case is:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{\sigma^*}} \left(\exp\left(-\frac{t^2}{2} \right) dt \right)$$

- Adding a third good to the scene with identical characteristics, $(Z_2 = Z_3)$, yields the probability for good 1 being purchased as:

$$\int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{\sigma^*}} \left(\int_{-\infty}^{\frac{(Z_1 - Z_2)\bar{\beta}}{\sigma^*}} \left(\frac{1}{2\pi\sqrt{1-\rho^2}} \exp -\frac{1}{2} \left(\frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{1-\rho^2} \right) \right) \right.$$

- One can show that, upon evaluation of these integrals, the probability derived from addition of the third good is less than the probability in the two good case.
- This leads us to a similar problem as the multinomial/conditional logit—adding alternatives decreases the probability of choice, despite the fact that the alternatives are quite similar.

- Thus, in the probit case we are able to avoid the limitation of the logit models with regard to addition of an identical good, through the covariance structure of the random coefficients.
- As illustrated in case 2, probit models without random coefficients suffer from the same limitation.

- Note that while the richer covariance structure is able to capture the relationship between choices in the probit model, applications involving many choices are practically limited as evaluation of higher-order multivariate normal integrals is difficult (refer discussion in Greene, Section 19.6.2.a).

Nested Logit: Generalized Extreme Value (GEV) Mode

- Consider a function $G(y_1, y_2, \cdots, y_J)$, where G satisfies:
  - i. Non-negativity:
    $G(y_1, y_2, \cdots, y_J) \geq 0 \; \forall \; (y_1, y_2, \cdots, y_J) \geq 0.$
  - ii. Homogeneous of degree 1:

$$G(\alpha y_1, \alpha y_2, \cdots, \alpha y_J) = \alpha G(y_1, y_2, \cdots, y_J).$$

  - iii. Derivative property:

$$\begin{aligned} \frac{\partial^k G}{\partial y_1 \partial y_2 \cdots \partial y_J} &\geq 0 \quad \text{if } k \text{ even} \\ &\leq 0 \quad \text{if } k \text{ odd.} \end{aligned}$$

- If G satisfies these conditions, then we get the following probability:

$$P(y_i \mid \{y_1, Y_2, ..., y_J\}) \equiv P_i = \frac{y_i G_i\left(y_1, y_2, \cdots, y_J\right)}{G\left(y_1, y_2, \cdots, y_J\right)},$$

where P_i is a probability that can be derived from utility maximization.

- We can use the theorem above to derive a special case of the nested logit model.
- Define:

$$
\begin{aligned}
& G\left(\exp(v_1), \exp(v_2), \cdots, \exp(v_J)\right) \\
\equiv \quad & \exp(v_1) + \left[\begin{array}{l} \exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right) \\ + \cdots + \exp\left(\frac{v_J}{1-\sigma}\right) \end{array} \right]^{1-\sigma} \\
= \quad & \exp(v_1) + \left[\left(\exp(v_2)\right)^{\frac{1}{1-\sigma}} + \cdots + \left(\exp\left(v_J\right)\right)^{\frac{1}{1-\sigma}} \right]^{1-\sigma}
\end{aligned}
$$

- Observe that $\sigma = 0$ is the ordinary logit model.
- (With G defined in this way, we are assuming that ε_1 is uncorrelated with all of the other ε_j, while the remaining ε_i may be correlated.
- The parameter σ is a kind of measure of correlation between the remaining ε_i .

- It is this correlation structure that would allow the GEV model to tackle the limitation of the ordinary conditional/multinomial logit models.
- This function obviously meets the conditions for the GEV model.

- For

i. Non-negativity: obvious as $0 < \sigma < 1$

ii. Homogeneity:

$$
\begin{aligned}
&G\left(\alpha \exp(v_1), \alpha \exp(v_2), \cdots, \alpha \exp(v_J)\right) \\
&\quad = \alpha \exp(v_1) + \left[(\alpha \exp(v_2))^{\frac{1}{1-\sigma}} + \cdots + (\alpha \exp(v_J))^{\frac{1}{1-\sigma}}\right]^{1-\sigma} \\
&\quad = \alpha \exp(v_1) + \left[\left(\alpha^{\frac{1}{1-\sigma}}\right)(\exp(v_2))^{\frac{1}{1-\sigma}} + \cdots + \left(\alpha^{\frac{1}{1-\sigma}}\right)(\exp(v_J))^{\frac{1}{1-\sigma}}\right]^{1-\sigma} \\
&\quad = \alpha \exp(v_1) + \alpha\left[\left(\exp\left(\frac{v_2}{1-\sigma}\right)\right) + \cdots + \left(\exp\left(\frac{v_J}{1-\sigma}\right)\right)\right]^{1-\sigma} \\
&\quad = \alpha\left(\exp(v_1) + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \cdots + \exp\left(\frac{v_J}{1-\sigma}\right)\right]^{1-\sigma}\right) \\
&\quad = \alpha G\left(\exp(v_1), \exp(v_2), \cdots, \exp(v_J)\right)
\end{aligned}
$$

iii. By inspection, one can see that this derivative property will hold. (It is obvious when differentiating with respect to $\exp(v_1$.

- For other derivatives, the fact that $0 < \sigma < 1$ gives the needed alternation in sign.
- Note that y_i in the definition of the property is analogous to $\exp(v_i$ here.)

- Thus, we can now proceed to derive our probabilities. First, consider:

$$\Pr(1 \mid \{1,2\}) = \frac{e^{v_1}}{e^{v_1} + \left(e^{\frac{v_2}{1-\sigma}}\right)^{1-\sigma}} = \frac{e^{v_1}}{e^{v_1} + e^{v2}}$$

which is simply our binomial logit model.

- Also note that in the three good case:

$$
\begin{aligned}
G_2 &= (1-\sigma)\left(\left[\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}\right)\frac{1}{1-\sigma}\exp \\
&= \exp\left(\frac{\sigma v_2}{1-\sigma}\right)\left(\left[\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}\right)
\end{aligned}
$$

- Now suppose that we eliminate choice 1 (by letting $v_1 \to -\infty$).
- Then:

$$\Pr(2 \mid \{2,3\}) = \frac{\exp(v_2)\exp\left(\frac{\sigma v_2}{1-\sigma}\right)\left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{-\sigma}}{\left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{1-\sigma}}$$

$$= \frac{\exp\left(\frac{v_2}{1-\sigma}\right)}{\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)}$$

- Observe that:

$$
\begin{aligned}
\Pr\left(1 \mid \{1,2,3\}\right) &= \frac{e^{v_1}}{e^{v_1} + \left(e^{\frac{v_2}{1-\sigma}} + e^{\frac{v3}{1-\sigma}}\right)^{1-\sigma}} \\
&= \frac{e^{v_1}}{e^{v_1} + \left\{e^{\frac{v_2}{1-\sigma}}\left(1 + e^{\frac{v_3 - v_2}{1-\sigma}}\right)\right\}^{1-\sigma}} \\
&= \frac{e^{v_1}}{e^{v_1} + e^{v_2}\left(1 + \left(\frac{e^{v_3}}{e^{v_2}}\right)^{\frac{1}{1-\sigma}}\right)^{1-\sigma}} \qquad (8)
\end{aligned}
$$

- Letting $\sigma \to 1$, and supposing $e^{v_2} > e^{v_3}$, we get:

$$\left(\frac{e^{v_3}}{e^{v_2}}\right) < 1 \Longrightarrow \left(\frac{e^{v_3}}{e^{v_2}}\right)^{\frac{1}{1-\sigma}} \to 0 \text{ as } \sigma \to 1$$

and thus from equation (8), we have:

$$\Pr\left(1 \mid \{1,2,3\}\right) \longrightarrow \frac{e^{v_1}}{e^{v_1} + e^{v_2}} \tag{9}$$

- Conversely, if $e^{v_3} > e^{v_2}$, just reverse the roles of v_2 and v_3 so:

$$\Pr(1 \mid \{1,2,3\}) \longrightarrow \frac{e^{v_1}}{e^{v_1} + e^{v_2}\left(\frac{e^{v_3}}{e^{v_2}}\right)} = \frac{e^{v_1}}{e^{v_1} + e^{v_3}} \tag{10}$$

- Combining equations (9) and (10), we get, as $\sigma \to 1$:

$$\Pr\left(1 \mid \{1,2,3\}\right) \to \frac{e^{v_1}}{e^{v_1} + \max\left\{e^{v_2}, e^{v_3}\right\}} \tag{11}$$

- Equations (9), (10) & (11) imply that in this GEV model, the probability of choice 1 on addition of a choice 3 identical to choice 2, does not necessarily fall, as was the case in the ordinary conditional/multinomial logit case.

- Equations (9) shows that if the added choice 3 is highly correlated to choice 2 ($\sigma \rightarrow 1$) but yields less utility, then the probability in the three choice case reduces to the binomial logit (the probability in the two choice case), with choice 3 dropping out, as one would intuitively expect.

- What about the probability of choice 2 – how does this change when we add an identical choice 3 in this GEV model?
- To answer this, consider:

$$
\begin{aligned}
&\Pr(2 \mid \{1,2,3\}) \\
&= \frac{e^{v_2}\left[(1-\sigma)\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{-\sigma}\right]\frac{1}{1-\sigma}\exp\left(\frac{\sigma v_2}{1-\sigma}\right)}{e^{v_1}+\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{1-\sigma}} \\
&= \frac{\exp\left(\frac{v_2}{1-\sigma}\right)\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{-\sigma}}{e^{v_1}+\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{1-\sigma}} \\
&= \frac{\exp\left(\frac{v_2}{1-\sigma}\right)}{\left(e^{v_1}+\left\{\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right\}^{1-\sigma}\right)\left(\exp\left(\frac{v_2}{1-\sigma}\right)+\exp\left(\frac{v_3}{1-\sigma}\right)\right)^{\sigma}}
\end{aligned}
$$

- When $\sigma = 0$, ie when there is no correlation between choice 2 and choice 3, we have ordinary conditional/multinomial logit.
- Suppose $v_2 > v_3$ and $\sigma \to 1$.
- By appealing to the result derived in equation (11), we get: $P(2 \mid \{1, 2, 3\})$

$$= \left(\frac{\exp\left(\frac{v_2}{1-\sigma}\right)}{\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)} \right) \tag{12}$$

$$\times \left[\frac{\left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{1-\sigma}}{\exp v_1 + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_3}{1-\sigma}\right)\right]^{1-\sigma}} \right]$$

- We know for $v_2 > v_3$:

$$\left(\frac{e^{v_3}}{e^{v_2}}\right) < 1 \Rightarrow \left(\frac{e^{v_3}}{e^{v2}}\right)^{\frac{1}{1-\sigma}} \to 0, \text{ as } \sigma \to 1$$

and thus, from equation (12), we get:

$$\Pr\left(2 \mid \{1,2,3\}\right) \to \frac{\exp(v_2)}{\exp(v_1) + \exp(v_2)}, \text{ as } \sigma \to 1$$

- (One could derive a similar result be assuming that $v_3 > v_2$).
- This equation tells us that in the GEV model, if choices 2 and 3 are very similar, if utility from 2 is greater than that from 3, then choice 3 gets disregarded (same as in Equation 9 earlier), which agrees with our intuition.

- Finally, supposing that $v_2 = v_3$, we get:

$$\begin{aligned} G &= e^{v_1} + \left[\exp\left(\frac{v_2}{1-\sigma}\right) + \exp\left(\frac{v_2}{1-\sigma}\right)\right]^{1-\sigma} \\ &= e^{v_1} + \left[2\exp\left(\frac{v_2}{1-\sigma}\right)\right]^{1-\sigma} \\ &= \exp(v_1) + 2^{1-\sigma}\exp\left(v_2\right). \end{aligned}$$

Thus:

$$\begin{aligned} \Pr\left(2 \mid \{1,2,3\}\right) &= \frac{\exp\left(v_2\right)2^{-\sigma}}{\exp(v_1) + 2^{1-\sigma}\exp\left(v_2\right)} \\ &= \frac{\exp v_2}{2^{\sigma}\exp v_1 + 2\exp v_2} \\ \Longrightarrow \lim_{\sigma\to 1}\Pr\left(2 \mid \{1,2,3\}\right) &\longrightarrow \frac{1}{2}\frac{\exp(v_2)}{\exp(v_1) + \exp(v_2)} \end{aligned}$$

- This final equation tells us if the characteristics are identical in the nested logit model, then the probability, in the three choice case, of choosing one of the two identical choices is equal half the probability of the two choice case, which is again what is intuitively expected.
- Thus the nested logit (GEV) model is able to avoid the key limitation of the conditional/multinomial logit imposed by the IIA assumption.