Lecture 2: Discrete Choice.

Sam Kortum - Fall 2002

Greene (ch. 19.7.4), McFadden, Small, Goldberg

This Draft: October 26, 2002.

For many applications the IIA restriction imposed by the multinomial logit model is not reasonable. There are 3 main ways of getting around it:

(i) Maintain the extreme value setup but drop the assumption of independent draws across choices. A popular example is Nested Logit, but many other possibilities can be derived from the Generalized Extreme Value distribution of McFadden. Small (1987) lays out the theory clearly. The Goldberg (1995) paper is a nice application to automobile demand.

(ii) Introduce random coefficients into the multinomial logit model. This approach yields simple choice probabilities conditional on a draw of the coefficients, but the aggregate choice probabilities must be calculated numerically. This approach can lead to nice economic interpretation of results, see, Berry (1994) and Berry, Levinsohn, and Pakes (1995).

(iii) Abandon the extreme value distribution and replace it with the joint Normal and a rich covariance structure across choices. This approach is very computationally intensive and I would argue not very revealing about the underlying economics.

In this lecture I focus on (i). In the following lecture I do (ii).

1 Generalized Extreme Value Distribution

An individual chooses one from the set of alternatives $j = 0, 1, ..., J$. He gets utility u_j by choosing j, where

$$u_j = v_j + \varepsilon_j$$

The econometrician observes v_j but not ε_j. (Linking this notation to that from the previous lecture: $v_j = \ln g(x_{ij})$ and ε_j is now additive so we want the Type I extremal distribution.)

Interpreting the $J+1$-vector ε as a realization of a random vector X, we want to calculate

$$P_k = \Pr\left[\arg\max_j \{U_j\} = k\right]$$

We say X has the Generalized Extreme Value distribution (of Type I) if

$$\Pr\left[X_0 \le \varepsilon_0, X_1 \le \varepsilon_1, \dots, X_J \le \varepsilon_J\right] = F(\varepsilon) = e^{-H(e^{-\varepsilon_0}, e^{-\varepsilon_1}, \dots, e^{-\varepsilon_J})}$$

where H is homogeneous of degree one, tends to infinity if any argument tends to infinity, and has n'th cross partial derivatives that are nonnegative for odd n and nonpositive for even n (we will eventually see the need for this last restriction for the case of $n = 1$.)

Can show that the marginal distribution of X_k is a Type 1 extemal distribution:

$$\Pr\left[X_k \le \varepsilon_k\right] = F^k(\varepsilon_k) = e^{-H(0, \dots 0, e^{-\varepsilon_k}, 0 \dots, 0)} = e^{-c_k e^{-\varepsilon_k}}$$

where $c_k = H(0, \dots 0, 1, 0 \dots, 0)$. It is the correlation structure that makes this approach more general.

From the joint distribution of X we get the joint distribution of U

$$G(u) = F(u - v) = e^{-H(e^{v_0} e^{-u_0}, e^{v_1} e^{-u_1}, \dots, e^{v_J} e^{-u_J})}$$

Let $G_k(u)$ denote the partial derivative of this joint distribution with respect to its k'th argument.

The probability that all the U_j's are less than a scalar m is just $G(m, \dots, m) = Q(m) = e^{-H(e^{v_0}, e^{v_1}, \dots, e^{v_J}) e^{-m}}$. Thus

$$E\left[\max_j \{U_j\}\right] = \int_{-\infty}^{\infty} m dQ(m) = \ln H(e^{v_0}, e^{v_1}, \dots, e^{v_J}) + \zeta$$

Intuitively, the choice probability P_k should be the derivative of this expectation with respect to v_k (since an increase in v_k only does you any good when you make choice $j = k$). Thus, we conjecture that $P_k = \frac{H_k(e^{v_0}, e^{v_1}, \dots, e^{v_J}) e^{v_k}}{H(e^{v_0}, e^{v_1}, \dots, e^{v_J})}$.

We can verify this intuition using a more direct attack

$$\begin{aligned}
P_k &= \Pr\left[U_1 \le U_k, \dots, U_{k-1} \le U_k, U_{k+1} \le U_k, \dots, U_J \le U_k\right] \\
&= \int_{-\infty}^{\infty} G_k(u, \dots, u) du \\
&= \int_{-\infty}^{\infty} e^{-H(e^{v_0}, e^{v_1}, \dots, e^{v_J}) e^{-u}} H_k(e^{v_0}, e^{v_1}, \dots, e^{v_J}) e^{v_k} e^{-u} du \\
&= \frac{H_k(e^{v_0}, e^{v_1}, \dots, e^{v_J}) e^{v_k}}{H(e^{v_0}, e^{v_1}, \dots, e^{v_J})}
\end{aligned}$$

So it works.

A Special Case: If $H(y) = \sum_{j=0}^{J} y_j$ then get the multinomial logit choice probabilities $P_k = \frac{e^{v_k}}{\sum_{j=0}^{J} e^{v_j}}$.

2 Nested Logit

Let $\{0, 1, \ldots, J\}$ be subdivided into $0 < R + 1 \leq J + 1$ non-intersecting sets B_r for $r = 0, \ldots, R$. Then, for parameters $0 < \rho_r \leq 1$ let

$$H(y) = \sum_{r=0}^{R} \left[\sum_{j \in B_r} y_j^{1/\rho_r} \right]^{\rho_r}$$

hence, if $k \in B_s$ then $H_k(y) = \left[\sum_{j \in B_s} y_j^{1/\rho_s}\right]^{\rho_s - 1} y_k^{(1/\rho_s)-1}$. The choice probability for k is thus

$$P_k = \frac{\left[\sum_{j \in B_s} e^{v_j/\rho_s}\right]^{\rho_s - 1} e^{v_k/\rho_s}}{\sum_{r=0}^{R} \left[\sum_{j \in B_r} e^{v_j/\rho_r}\right]^{\rho_r}}$$

This specifiction is called Nested Logit. It allows one to avoid the Independence of Irrelevant Alternatives restriction as the following example reveals. Consider again the case of $J = 2$ with Red Apple, Green Apple, and Donut. Suppose $R = 1$ and B_0 contains the two colors of apple. The relative probability of choosing a Red Apple to a Green Apple is simply

$$\frac{P_0}{P_1} = \frac{e^{v_0/\rho_0}}{e^{v_1/\rho_0}}$$

For comparing the choice probabilities within the same group, the characteristics of the other choices (the donut) are irrelevant. However, consider the relative probability of choosing a Red Apple to a Donut,

$$\frac{P_0}{P_2} = \frac{\left[e^{v_0/\rho_0} + e^{v_1/\rho_0}\right]^{\rho_0 - 1} e^{v_0/\rho_0}}{e^{v_2}}$$

Here we get away from IIA as the characteristics of the Green Apple matter (if $\rho_0 < 1$) for the relative probability of choosing the Red Apple relative to the Donut. As the Green Apple becomes more desireable (say its price falls) the probability of choosing the Red Apple falls relative to the probability of choosing the Donut.

The term "Nested" Logit comes from the following decomposition of the choice probabilities:

$$P_k = \frac{\left[\sum_{j\in B_s} e^{v_j/\rho_s}\right]^{\rho_s}}{\sum_{r=0}^{R}\left[\sum_{j\in B_r} e^{v_j/\rho_r}\right]^{\rho_r}} \frac{e^{v_k/\rho_s}}{\sum_{j\in B_s} e^{v_j/\rho_s}}$$

The second fraction is the probability of choosing k from within B_s. It has the form of Multinomial Logit with parameters $v_j' = v_j/\rho_s$. The first fraction is the probability of choosing the subset B_s. It has the form of Multinomial Logit with parameters $v_r'' = \rho_r \ln \sum_{j\in B_0}^{1/\rho_r} e^{v_j/\rho_r}$. Imagine a tree structure for decision making with the major branches representing the r's.

We can probe this decomposition a bit more by noting

$$\Pr\left[\max_{j\in B_r}\{U_j\} \leq m_r\right] = e^{-\left[\sum_{j\in B_r} e^{v_j/\rho_r}\right]^{\rho_r} e^{-m_r}}$$

From our result earlier

$$E\left[\max_{j\in B_r}\{U_j\}\right] = \ln\left[\sum_{j\in B_r} e^{v_j/\rho_r}\right]^{\rho_r} + \zeta = v_r'' + \zeta$$

Thus the parameter v_r'' captures the expected value of the r'th branch. These terms (or v_r''/ρ_r) are sometimes called the "inclusive values".

There are several special cases of Nested Logit that reduce to MNL:

(a) If $\rho_r = 1$ for all r.

(b) If each set B_r contains only one element. Then

$$P_k = \frac{\left[e^{v_k/\rho_k}\right]^{\rho_k - 1} e^{v_k/\rho_k}}{\sum_{r=0}^{R}\left[e^{v_r/\rho_r}\right]^{\rho_r}} = \frac{e^{v_k}}{\sum_{j=0}^{J} e^{v_j}}$$

(c) If $R = 0$, i.e. $B_0 = \{0, 1, \ldots, J\}$. Then

$$P_k = \frac{\left[\sum_{j\in B_0} e^{v_j/\rho_0}\right]^{\rho_0 - 1} e^{v_k/\rho_0}}{\left[\sum_{j\in B_0} e^{v_j/\rho_0}\right]^{\rho_0}} = \frac{e^{v_k/\rho_0}}{\sum_{j=0}^{J} e^{v_j/\rho_0}}$$

where, since ρ_0 is not identified it can be normalized to 1.

3 Multiple Nests

Goldberg

4 Extensions

In deriving Nested Logit from the GEV framework, we were considering only a very special case. Another special case is considered in Small and another in Bresnahan, Stern, and Trajtenberg (RAND, 1997). It would seem that many other useful cases have yet to be explored.