Multinomial Logit Model

I. Choice probabilities

II. Characteristics of the logit model

III. Uses of the logit model (may skip for now)

IV. Estimation

Type I Extreme Value Distribution

$$f(x) = \frac{1}{\beta}\exp\left[-\frac{x-\mu}{\beta}\right]\exp\left\{-\exp\left[-\frac{x-\mu}{\beta}\right]\right\}$$

$$\text{mean} = \mu + 0.5772\beta \quad \text{variance} = \frac{\beta^2\pi^2}{6}$$

Standardized Extreme Value: $\mu=0$; $\beta=1$

$$f(x) = \exp[-x]\exp\{-\exp[-x]\}; \quad F(x) = \exp\{-\exp[-x]\}$$

$$\text{mean} = 0.5772 \quad \text{variance} = \frac{\pi^2}{6}$$

Logit Model

Multinomial/conditional logit model arises from the following assumptions:

$$U_j = V_j + \varepsilon_j, \quad j = 1,\ldots,J, \quad \varepsilon_j \sim iid\, EV(\mu = 0, \beta = 1)$$

Note:

$$\varepsilon = \varepsilon_j - \varepsilon_k \sim \text{logistic}$$

$$F(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)}$$

- Difference between two extreme values is very similar to a standard normal
- The iid assumption is what gives logit its distinguishing characteristics.

I. Choice Probabilities

Probability that alternative *j* is chosen:

$$\begin{aligned} P_j &= \Pr(V_j + \varepsilon_j > V_k + \varepsilon_k \forall k \neq j) \\ &= \Pr(\varepsilon_k < \varepsilon_j + V_j - V_k \forall k \neq j) \end{aligned}$$

Suppose ε_j is given:

$$P_j \mid \varepsilon_j = \prod_{k \neq j} \exp\left\{-\exp\left[-(\varepsilon_j + V_j - V_k)\right]\right\}$$

But ε_j is not given…unconditional probability:

$$P_j = \int \left(\prod_{k \neq j} \exp\left\{-\exp\left[-(\varepsilon_j + V_j - V_k)\right]\right\} \right) f(\varepsilon_j) d\varepsilon_j$$

After some vicious algebra…(see Train pp. 78-79):

$$P_j = \frac{\exp(V_j)}{\sum_{k=1}^{J} \exp(V_k)}, \quad j = 1, \ldots, J.$$

Note:

•An easily expressed, closed form solution for the probability.

•Probabilities lie between zero and one and sum to one.

•Probability function is ‘smooth’ in parameters…facilitates estimation.

•Likelihood based on these probabilities globally concave.

Example: Retire or continue to work?

Consider modeling the binary choice of a person deciding to retire or remain in the workforce *(i* subscript ignored):

$$
\begin{aligned}
U_w &= \beta_0 + \beta I_w + \varepsilon_w \\
U_r &= \beta I_r + \varepsilon_r \\
\varepsilon_r, \varepsilon_w &\sim iid\ EV
\end{aligned}
$$

I_w and I_r are income from work and retirement, respectively.

$$P_r = \frac{\exp(\beta I_r)}{\exp(\beta I_r) + \exp(\beta_0 + \beta I_w)}$$

Note: expect $\beta > 0$ (i.e. positive marginal utility of money)

Example: Retire, continue to work fulltime, or work part time?

We now have more than two alternatives:

$$U_w = \beta_0 + \beta I_w + \varepsilon_w$$
$$U_p = \beta_1 + \beta I_p + \varepsilon_p$$
$$U_r = \beta I_r + \varepsilon_r$$
$$\varepsilon_r, \varepsilon_p, \varepsilon_w \sim iid\ EV$$

Probability of retirement:

$$P_r = \frac{\exp(\beta I_r)}{\exp(\beta I_r) + \exp(\beta_0 + \beta I_w) + \exp(\beta_1 + \beta I_p)}$$

II. Characteristics of the Logit Model

Independence of the error terms is the defining characteristic. What does this imply?

•If representative utility V_j is fully specified then error term should just be 'white noise'…nothing is left out to induce correlation in the unknown components of utility.
•Independent errors are in this sense the ideal

Usually we are not so fortunate…what are ramifications of the restrictive error structure?

- •Restricts extent of taste variation that can be characterized
- •Restricts extent of substitution that can be modeled

Taste Variation

Importance of attributes of alternatives probably varies over individuals.

Example: choice of a location for recreation fishing trip

$$U_j = \beta p_j + \delta c_j + \varepsilon_j$$

p_j is travel distance to site j

c_j is the expected fish catch rate at site j

Might expect travel distance and catch rate to matter differentially to people

1. Systematic (observable) variation

$$\beta = \gamma_1 + \gamma_2 R; \quad \delta = \theta_1 + \theta_2 K$$

R=1 if person is retired; K=1 if person has children

$$U_j = \gamma_1 p_j + \gamma_2 p_j R + \theta_1 c_j + \theta_2 c_j K + \varepsilon_j$$

2. Random (unobservable) variation

$\beta = \gamma + \mu; \quad \delta = \theta + \eta; \quad \mu$ and η are random variables

$$\begin{aligned} U_j &= \gamma p_j + \mu p_j + \theta c_j + \eta c_j + \varepsilon_j \\ &= \gamma p_j + \theta c_j + \tilde{\varepsilon}_j; \quad \tilde{\varepsilon}_j = \varepsilon_j + \mu p_j + \eta c_j \end{aligned}$$

note: $\mathrm{COV}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) \neq 0$

Note:

- Logit models can accommodate systematic taste variation.
- Logit models *can not* accommodate unobserved (random) taste variation
- Logit is probably a misspecification when tastes can be expected to vary in the target population and V_j is sparsely parameterized.

Substitution Patterns

Discrete choice models are used to understand substitution patterns between alternatives

- Marketing applications: if a new product is introduced will it substitute for a firm's or its competitors existing products?
- Health applications: to what extent are generic drugs replacements for brand-name drugs?
- Environmental applications: damages from a beach closure due to an oil spill depend on the extent to which other beach sites are close substitutes.

How good is the multinomial logit model at characterizing substitution among alternatives?

Independence of Irrelevant Alternatives (IIA)

The mathematical form of the logit model implies the *ratio of choice probabilities* for any two alternatives j and k is

$$\frac{P_j}{P_k} = \frac{\exp(V_j) \Big/ \sum_s \exp(V_s)}{\exp(V_k) \Big/ \sum_s \exp(V_s)} = \frac{\exp(V_j)}{\exp(V_k)} = \exp(V_j - V_k)$$

The relative odds of choosing j over k is independent of the number and characteristics of all remaining alternatives.

- Imposes a specific structure on the relationship between alternatives

Red Bus/Blue Bus Example

Consider choice of transit to work between a car (C) and a blue bus (B). Suppose initially we predict:

$$P_B = P_C = 0.5 \rightarrow P_B/P_C = 1$$

Suppose a third alternative – a red bus (R) – is introduced and $P_B = P_R$. We would expect:

$$P_B = P_R = 0.25;\ P_C = 0.5$$

But IIA forces the following:

- $P_B/P_C = 1$ (ratio does not change with new alternative)
- $P_B/P_R = 1$ (by construction)
- $P_B = P_R = P_C = 0.33$

Proportional Substitution

Define elasticity of probability j with respect to attribute s in alternative k:

$$\eta_{jk_s} = \frac{\partial P_j(\cdot)}{\partial x_{k_s}} \frac{x_{k_s}}{P_j(\cdot)}$$

Measures the incremental change in the predicted probability person chooses alternative j from a small change in an attribute of alternative k

- Provides a measure of own ($j=k$) and cross ($j \neq k$) attribute responsiveness.

For the linear specification $V_j = \sum_s \beta_s x_{js}$:

$$\eta_{jk_s} = \begin{cases} \left[1 - P_j(\cdot)\right]\beta_s x_{js} & j = k \\ -P_k(\cdot)\beta_s x_{ks} & j \neq k \end{cases}$$

Note:

- This implies $\eta_{jk_s} = \eta_{mk_s} \quad j \neq m$
- Proportional substitution – an improvement in one alternative draws proportionately from all the other alternatives.
- Is this realistic?

Train's Electric Car Example

Consider vehicle ownership with three alternatives:

- Large gas car, small gas car, small electric car

Consider a subsidy program to encourage purchases of electric cars with goal of reducing fuel consumption

- Logit model will predict the same percentage drop in the probability of large and small (gas) car purchases when electric cars get cheaper
- More likely outcome: subsidizing small electric cars will draw more from small gas cars
- Small gas and small electric cars are closer substitutes
- Ramifications of predictions from (misspecified) logit model?

III. Uses of the Logit Model

In spite of its restrictive properties the logit model is the most widely applied discrete choice model. Uses include:

- Prediction of choice probabilities as functions of observed characteristics
- Calculation of elasticity measures
- Welfare analysis – assessing changes in well-being that arise from exogenous changes in attributes of the choice alternatives or the set of available attributes

  Examples:

    Benefits of new products; damages from lost recreation sites, value of improvements in product attributes

Welfare Analysis

Suppose V_j includes the term $M - p_j$ where

- M = person's income available for choice
- p_j = monetary price of alternative j

Utility is now given by: $U_j = V(M - p_j, x_j) + \varepsilon_j \quad j = 1, \ldots, J$

Note:

- Specification is now consistent with budget-constrained util-max problem – Random Utility Maximization (RUM) model
- U_j is the *conditional indirect utility* from choosing (and paying for) alternative j
- (unconditional) indirect utility function is

$$U = \max_{j \in J} \left\{ U_j(M, p_j, x_j, \varepsilon_j) \right\} = \max_{j \in J} \left\{ V(M - p_j, x_j) + \varepsilon_j \right\}$$

Compensating Variation

From the indirect utility function we can define compensating variation (*CV*) for changes in exogenous variables:

$$U(M, \mathbf{p}^0, \mathbf{X}^0, \varepsilon) = U(M - CV, \mathbf{p}^1, \mathbf{X}^1, \varepsilon)$$

$$\Leftrightarrow$$

$$\max_{j \in J^0} \left\{ U_j(M, p_j^0, x_j^0, \varepsilon_j) \right\} = \max_{j \in J^1} \left\{ U_j(M - CV, p_j^1, x_j^1, \varepsilon_j) \right\}$$

Recall:

- *CV* is amount of money we need to take from a person following a change to restore the original utility level.

Note:

$$CV = CV(M, \mathbf{p}^0, \mathbf{p}^1, \mathbf{X}^0, \mathbf{X}^1, \varepsilon)$$

Consumer Surplus

Suppose the conditional indirect utility function is linear in income:

$$U_j = (M - p_j)\beta + \gamma' x_j + \varepsilon_j, \quad j = 1, \ldots, J$$

Recall: $\Delta\text{consumer surplus} \equiv \dfrac{\Delta\text{utility}}{\text{marginal utility of money}}$

Note: β is the (constant) marginal utility of income

$$CS(\mathbf{p}^0, \mathbf{p}^1, \mathbf{X}^0, \mathbf{X}^1, \varepsilon) = \frac{1}{\beta}\left[U^1(\cdot) - U^0(\cdot)\right] = CV$$

Estimating Consumer Surplus

We do not observe ε for the individual…hence we do not observe the maximum utility:

$$U = \max_{j \in J} \left\{ V(M - p_j, x_j) + \varepsilon_j \right\}$$

From the perspective of the analyst the best we can do is calculate the expectation of the maximum utility:

$$E\left[U\right] = E\left[\max_{j \in J} \left\{ V(M - p_j, x_j) + \varepsilon_j \right\} \right]$$

Expected consumer surplus for a person is then:

$$E(CS) = \frac{1}{\beta}\left[E(U^1) - E(U^0) \right]$$

Expected Maximum Utility in Logit Model

If V_j is linear in income and ε_j ~ iid extreme value Small and Rosen (1981) show:

$$E(U) = \ln\left[\sum_{j=1}^{J} \exp(V_j)\right] + C$$

where C is an unknown constant. So…

$$E(CS) = \frac{1}{\beta}\left\{\ln\left[\sum_{j=1}^{J^1} \exp(V_j^1)\right] - \ln\left[\sum_{j=1}^{J^0} \exp(V_j^0)\right]\right\}$$

A Few Notes

Linear-in-income logit RUM models:

- Easy to estimate
- Consistent with the notion of budget-constrained utility maximization
- Straightforward to calculate individual-level expectation of consumer surplus for exogenous changes
- Fairly restrictive characterization of substitution patterns – particularly for sparsely parameterized V_j

Non-linear in income logit RUM models:

- Still easy to estimate
- No closed form for expected utility, compensating variation.