1 Multinomial Logit

Remember the binomial RUM model. Now extended to choice between more than two alternatives. Probability for choosing alternative i is

$$P(i) = P(\widetilde{U}_i = \max_{j=1,\dots,n} \widetilde{U}_j),$$

where $\widetilde{U}_i = V_i + \varepsilon_i$.

It can be shown that if ε_i is iid and has a type II extreme value (or Gumbel) distribution, i.e.,$f(\varepsilon_i) = e^{-\varepsilon_i} e^{-e^{-\varepsilon_i}}$ and $F(\varepsilon_i) = e^{-e^{-\varepsilon_i}}$, where the variance of the distribution is set to $\pi^2/6$, then the choice probability of alternative i

$$P_i = P(V_i + \varepsilon_i > V_j + \varepsilon_j, \forall j \neq i) = P(\varepsilon_j < \varepsilon_i + V_i - V_j, \forall j \neq i).$$

can be written on closed form as

$$P_i = \frac{e^{V_i}}{\sum_j e^{V_j}}.$$

This is the multinomial logit model. This result was first shown by Holman and Marley. Mc Fadden susequently showed uniqueness, i.e., that the type II extreme value distribution is the *only* one giving a closed form solution for the choice probability. Proof is given in Maddala (1983) or Anderson, de Palme and Thisse.

V_i can be made as a function of observables:

$$V_i = x_i'\beta.$$

Example: Herriges and Kling (1999): Choice of Fishing Mode:

Four different fishing modes:

Beach

Pier

Private boat

Charter

Three different Characteristics: Price (varies between choices), Catch probability (varies between choices) and Income (choice invariant).

Conditional Logit Model (using the alternative specific covariates):

$$p_{ij} = \frac{\exp\left(\beta_P P_{ij} + \beta_C C_{ij}\right)}{\sum_{k=1}^{4} \exp\left(\beta_P P_{ik} + \beta_C C_{ik}\right)}.$$

Results is given as marginal effects or elasticities, since they suffer fromthe non-identification of the variance problem.

Marginal effects:

$$\frac{\partial p_i}{\partial x_{ik}} = p_{ij}\left(1 - p_{ij}\right)\boldsymbol{\beta}_k.$$

Elasticities:

$$E_{ix_i} = \frac{\partial p_{ij}}{\partial x_{ik}} \frac{x_{ik}}{p_{ij}} = x_{ik}\left(1 - p_{ij}\right)\boldsymbol{\beta}_k.$$

Averages can be calculated in the sample.

Multinomial Logit Model (Individual specific covariates):

$$p_{ij} = \Pr\left[y_i = j\right] = \frac{\exp\left(\alpha_j + \beta_{1j} I_i\right)}{\sum_{k=1}^{4} \exp\left(\alpha_k + \beta_{1k} I_i\right)}$$

Relative to a base alternative. Normalization: set $\alpha_1 = 0$ and $\beta_{I1} = 0$:

Marginal effects:

$$\frac{\partial p_{ij}}{\partial x_{ik}} = p_{ij}\left(\beta_j - \overline{\beta}_i\right),$$

where $\overline{\beta}_i = \sum p_{il}\beta_l$.

Mixed Model (Both Individual and choice specific covariates):

$$\begin{aligned} p_{ij} &= \Pr\left[y_i = j\right] = \frac{\exp\left(\beta_P P_{ij} + \beta_C C_{ij} + \alpha_j + \beta_{1j} I_i\right)}{\sum_{k=1}^{4} \exp\left(\beta_P P_{ik} + \beta_C C_{ik} + \alpha_k + \beta_{1k} I_i\right)} = \\ &= \frac{\exp(\beta_P P_{ij} + \beta_C C_{ij}) + \sum_{l=1}^{4}(\alpha_j d_{ijl} + \beta_{1j} d_{ijl} I_{ijl})}{\sum_{k=1}^{4} \exp \beta_P P_{ik} + \beta_C C_{ik} + \sum_{l=1}^{4}(\alpha_l d_{ijl} + \beta_{Il} d_{ijl} I_{ijl})}, \end{aligned}$$

where $d_{ijl} = 1$ if $j = l$ and zero otherwise.

Conditional logit

Example:

Boskin (1974)

Choice and individual specific independent variables. Additional sub-index j for individual:

$$V_i = x'_{ij}\beta + w'_i\alpha$$

Purpose: Examine what affects different groups in their occupational choice.

$$p_{ij} = f\left(E_{i1}, ..., E_{in}, u_{i1}, ..., u_{in}, T_{i1}/w_i, ..., T_{in}/w_i\right)$$

$i-$ individual.

$j-$ occupation.

$E-$ lifetime earnings

$u-$ cost of unemployment (different unemployment probabilities).

T/w_i- training cost relative to wealth (i indicates that Boskin uses personal wealth).

He could (should!) have done interaction!

All coefficients have the expected signs: $E(+)$

$u(-)$

$T/w_i(-)$

- White men are least sensitive to training costs. Boskin interpret that as that they have better access to credit markets compared to the other groups.
- Black females are most sensitive to cost of unemployment.
- White males are most sensitive to life-time earnings.

Multinomial logit:

Individual specific independent variables only.

Example:

Schmidt and Strauss (1975)

What determines occupational choice. Does discrimination occur?

Five different occupations

Professional (5)

White collar (4)

Craft (3)

Blue collar (2)

Menial (1)

Relative risk measure same as for the binomial logit model:

$$\frac{\Pr[y_i = j]}{\Pr[y_i = 1]} = e^{\mathbf{x}_i'\beta_j},$$

$$\begin{aligned}\log(p_2/p_1) &= \beta_{11} + \beta_{12} Educ_t + \beta_{13} Exp_t + \beta_{14} Race_t + \beta_{15} Sex_t \\ \log(p_3/p_1) &= \beta_{21} + \beta_{22} Educ_t + \beta_{23} Exp_t + \beta_{24} Race_t + \beta_{25} Sex_t\end{aligned}$$

Normalized by "Menial"

From this it is possible to obtain log-odds ratios for different choices:

$\log(p_3/p_2) = (\beta_{31} - \beta_{21}) + (\beta_{32} - \beta_{22}) Educ_t$......

Results:

Education increases the probability of obtaining a higher ranked occupations. Except for menial versus blue collars. This unexpected result is explained by the fact that blue collar workers often have informal education.

Gender: Women are more likely to be white collar and menial. Being women makes it more likely being in any occupation relative to occupations lower on the list:

White collar

Menial

Professional

Blue collar

Craft

Race: Being black makes it more likely being in any occupation relative to any occupation lower on the list:

Menial

Blue collar

Crafts, Professional

White collar

Change over time. The analysis is done for 1960 and 1970. Decreasing discrimination over time.

2 Estimation

Multinomial density for one observation:

$$f(y) = p_1^{y_1} \times p_2^{y_2} \ldots\ldots \times p_m^{y_m} = \prod_{j=1}^{m} p_j^{y_j}$$

Gives the log likelihood

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij},$$

where $p_{ij} = F_j(\mathbf{x}_i, \boldsymbol{\beta})$.

This gives the first order conditions:

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i=1}^{N} \sum_{j=1}^{m} \frac{y_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \beta_k} = 0.$$

From the second order condition it is possible to obtain the variance-covariance matrix for the β vector.

3 Goodness-of-fit

Pseudo-R^2 measure defined as

$$Pseudo - R^2 = 1 - \ln L_{fit} / \ln L_0,$$

where $\ln L_{fit}$ is the log likelihood of the fitted model and $\ln L_0$ is the log likelihood of the model with a constant only.

Can be rewritten as

$$Pseudo - R^2 = \frac{\ln L_{fit} - \ln L_0}{\ln L_{\max} - \ln L_0},$$

since $\ln L_{\max} = 0$.

4 Welfare Analysis

What is the monetary value of changing one or more choice characteristics. The welfare cost can be much smaller that the "mechanical" cost, since the individuals can change their behavior as a result of the change.

Indirect utility function

$$V_j = V\left(I - p_j, \mathbf{x}_j\right) + \varepsilon_j,$$

where I is Income, p_j is the price of alternative j and $\mathbf{x}_j$ is the characteristics vector of alternative j.

Suppose we change the characteristics vector from $\mathbf{x}_j'$ to $\mathbf{x}_j''$. CV can be defined as

$$\max_{j=1,\ldots,m} U\left(I - p_j, \mathbf{x}_j', \varepsilon_j\right) = \max_{j=1,\ldots,m} U\left(I - CV - p_j, \mathbf{x}_j'', \varepsilon_j\right).$$

Simplest possible example. Two alternatives (1 and 2) $U_j = I + x_j + \varepsilon_j$ and one characteristic (x_j) that changes from x_j' to x_j''.

Four possible alternatives.

Alternative 1 is chosen before and after. $CV = x_1'' - x_1'$, since $U_1'' = I - CV + x_1'' + \varepsilon_1 = I + x_1' + \varepsilon_1 = U_1'$.

Alternative 2 is chosen before and after. $CV = x_2'' - x_2'$, since $U_2'' = I - CV + x_2'' + \varepsilon_2 = I + x_2' + \varepsilon_2 = U_2'$.

Switching from alternative 1 to alternative 2. $U_1'' = I - CV + x_1' + \varepsilon_1 = I + x_2'' + \varepsilon_2 = U_2'$. Implies that $CV = x_2'' - x_1' + \varepsilon_2 - \varepsilon_1$.

Switching from alternative 2 to alternative 1. $U_2'' = I - CV + x_2' + \varepsilon_2 = I + x_1'' + \varepsilon_1 = U_1'$. Implies that $CV = x_1'' - x_2' + \varepsilon_1 - \varepsilon_2$.

The unobservables can be eliminated by computing $E(CV)$. For some models there are no analytical solution $E(CV)$. Has to be solved numerically.

5 IIA

The (classical) Debreu (1960) Red bus-Blue bus example.

Choice between Red bus (X_1), Blue bus (X_2) and Car (X_3). The choice maker is indifferent between the Red and the Blue bus. We have

$$P(1 \mid X_1, X_2) = P(1 \mid X_1, X_3) = P(2 \mid X_2, X_3) = \frac{1}{2}$$

and if all three choices are available:

$$P(1 \mid X_1, X_2, X_3) = P(2 \mid X_1, X_2, X_3) = \frac{1}{4}.$$

This means that the odds ratio between alternative 1 and 3 is 1:1 if alternative 2 is not present, but 1:2 if alternative 2 is present. An obvious inconsistency and depends on the fact that alternatives 1 and 2 are correlated choices.

Test for IIA

Two types of tests.

1. Hausman and McFadden (1984). The idea is that if choices are really independent, the odds ratios will not change if the model is estimated on a sub-set of alternatives. This means that the coefficients should not change. Hausman type test:

$$\left(\widehat{\beta}_s - \widehat{\beta}_f\right)' \left[\widehat{V}_s - \widehat{V}_f\right]^{-1} \left(\widehat{\beta}_s - \widehat{\beta}_f\right) \sim \chi^2(K).$$

2. Include the characteristics of an "irrelevant" variable into the choice probabilities of the included ones. Should be insignificant if it is independent of the alternative tested for.

6 Nested Logit

Choice alternatives can be grouped into nests.

Examples: Transportation modes and colors of different vehicles; Location studies; Consumer brands.

1. For any two alternatives in the same nest, IIA holds.
2. For any alternatives in different nests, IIA does not need to hold.

Extension of the extreme value distribution to Generalized Extreme Value (GEV) distribution:

$$F\left(\varepsilon_1, \varepsilon_m, ...\varepsilon_m\right) = \exp\left[-G\left(e^{-\varepsilon_1}, e^{-\varepsilon_2}, ...e^{-\varepsilon_m}\right)\right],$$

where $G(Y_1, Y_2.......Y_m)$ is a nonnegative function (homogenous of degree one) in $(Y_1, Y_2.......Y_m) \geqslant 0$.

A special cas is

$$G(Y_1, Y_2.......Y_m) = \sum Y_i,$$

yields the multinimial logit model.

Suppose we have K non overlapping nests $B_1....B_K$ and choices indexed by i. We then have the following cumulative distribution for the vactor of te unobserved utility $(\varepsilon_1, \varepsilon_2, ...\varepsilon_I)$:

$$\exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\varepsilon_i/\lambda_k}\right)^{\lambda_k}\right).$$

ε_i can be correlated within any nest, but are independent for any two alternatives in different nests.

λ_k measures the the degree of independence between any two alternatives within the same nest k. $\lambda_k = 1$ indicates complete independence within nest k. $1 - \lambda_k$ measures the correlation. This specification gives the following choice probabilities:

$$P_i = \frac{e^{V_i/\lambda_k} \left(\sum_{i \in B_k} e^{V_i/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K} \left(\sum_{i \in B_\ell} e^{V_i/\lambda_\ell}\right)^{\lambda_\ell}}.$$

Consider two alternatives i and m in different nests, B_k and B_ℓ, respectively. Then we get the following odds ratios:

$$\frac{P_i}{P_m} = \frac{e^{V_i/\lambda_k} \left(\sum_{j \in B_k} e^{V_j/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_m/\lambda_\ell} \left(\sum_{j \in B_\ell} e^{V_j/\lambda_\ell}\right)^{\lambda_\ell - 1}}.$$

If $k = \ell$, then the parentheses cancel out

$$\frac{P_i}{P_m} = \frac{e^{V_i/\lambda_k}}{e^{V_m/\lambda_\ell}}.$$

Decomposition into two logits

The observable part of the utility function can be decomposed into one part depending on the nest (W_k) and one part on the choice alternatives within the nests, i.e.,

$$V_i = W_k + Y_i.$$

The probability of choosing alternative i can be written as

$$P_i = P_{i|B_k} P_{B_k}.$$

These marginal and conditional allows us to write the model in two separate logits:

$$
\begin{aligned}
P_{B_k} &= \frac{e^{W_k+\lambda_k I_k}}{\sum_{\ell=1}^{K} e^{W_\ell+\lambda_\ell I_\ell}} \\
P_{i|B_k} &= \frac{e^{Y_k/\lambda_k}}{\sum_{i\in B_k} e^{Y_i/\lambda_k}},
\end{aligned}
$$

where

$$I_k = \ln \sum_{i\in B_k} e^{Y_i/\lambda_k}.$$

I_k links the information from the lower model to the upper model. It is the log of the denominator in the lower model. $\lambda_k I_k$ is the utility from the choice situation faced by the individual. The inclusive value of nest B_k.If there is a high degree of correlation between the alternatives in the nest, λ_k and the inclusive value is small.

7 Multinomial Probit Models

Extension of the binomial probit model. Suppose we have the choice between three, rather than two alternatives. Then we have

$$\begin{aligned} U_1 &= V_1 + \varepsilon_1 \\ U_2 &= V_2 + \varepsilon_2 \\ U_3 &= V_3 + \varepsilon_3 \end{aligned}$$

Where the unobserved components are allowed to be correlated, i.e.,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

The probability that the first alternative is chosen is the joint probability

$$P(V_1+\varepsilon_1 > V_2+\varepsilon_2, V_1+\varepsilon_1 > V_3+\varepsilon_3) = P(\varepsilon_2-\varepsilon_1 > V_1-V_2, \varepsilon_3-\varepsilon_1 > V_1-V_3).$$

Define

$$\begin{aligned} \eta_{21} &= \varepsilon_2 - \varepsilon_1 \\ \eta_{31} &= \varepsilon_3 - \varepsilon_1 \\ V_{12} &= V_1 - V_2 \\ V_{13} &= V_1 - V_3 \end{aligned}$$

Bivariate normal distribution for η_{21} and η_{31}

$$\Omega = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} \\ \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix}$$

This gives us the probability that alternative 1 is chosen as

$$P_1 = \int_{-\infty}^{V_{12}} \int_{-\infty}^{V_{13}} f(\eta_{21}, \eta_{31}) d\eta_{21} d\eta_{31}.$$

8 Mixed Logit

Another strategy is to enter the heterogeneity through the coefficients.

$$U_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_i + \varepsilon_{ij},$$

where ε_{ij} is iid extreme value, but permits the $\boldsymbol{\beta}_i$'s to be random.

Can be rewritten as

$$\begin{aligned} U_{ij} &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \nu_{ij}, \\ \nu_{ij} &= \mathbf{x}'_{ij}\mathbf{u}_i + \varepsilon_{ij}, \end{aligned}$$

where $cov\left[\nu_{ij}, \nu_{ik}\right] = \mathbf{x}'_{ij} \sum_{\beta} \mathbf{x}_{ij}$.This means that the introduction of randomness in the parameters allows for correlation between choice alternatives.

The mixed logit model is defined as

$$P_{ij} = \int L_{ij}(\beta) f\left(\beta\right) d\beta,$$

where $L_i(\beta)$ is the logit probability evaluated at the parameter vector β, such that

$$L_{ij}(\beta) = \frac{e^{V_{ij}(\beta)}}{\sum_{k=1}^{K} e^{V_{ik}(\beta)}}$$

and $f(\beta)$ is the density function for the distribution of β.For a linear observed part of the utility function we have

$$P_{ij} = \int \left(\frac{e^{\beta' x_{ii}}}{\sum_{k=1}^{K} e^{\beta' x_{ik}}} \right) f\left(\beta\right) d\beta.$$

Mixed logit is a mixture of logit functions evaluated at different $\beta's$. That is why it is called the mixed logit.

Standard logit is a special case when $f(\beta) = 1$.

Suppose that there are M different β vectors: $\beta_1, \beta_2 .... \beta_M$ with the corresponding probabilities $s_1, s_2 .... s_M$.Then we have a *latent class model*. The different categories can also correspond to different observable groups in the society. We then have

$$P_{ij} = \sum_{m=1}^{M} s_m \left(\frac{e^{b'_m x_{ij}}}{\sum_j e^{b'_m x_{ij}}} \right).$$

It is also possible to use a continuous distribution for $f(\beta)$, such as the normal, log normal, gamma or uniform. In this case we estimate the average β vector along with the covariance for the distribution of this vector.