

# Models of Multinomial Qualitative Response

## Multinomial Logit Models

October 22, 2015

## Dependent Variable as a Multinomial Outcome

Suppose we observe an economic choice that is a binary signal from amongst  $M$  discrete alternatives. The focus on the course is to estimate parameters to test economic theories and/or predict the impact of exogenous change due to policy.

Numerous examples:

- Regulator's choice of where to site a hazardous waste incinerator from among Virginia's 134 counties and independent cities.
- Student's choice of which of  $M$  colleges to attend (where  $M > 2$ ).
- Individual's choice from among  $M$  different private health care plans (where  $M > 2$ ).
- REI's choice of where to site their next retail store from among 366 Metropolitan statistical areas in the US.

## Terminology and Definitions

Terminology: The choice set,  $C$ , is the set comprised of the  $M$  feasible alternatives that could be chosen. For binary discrete outcomes, the choice set is  $C = \text{Yes}, \text{No} = 1, 0$ , while for multinomial outcomes, it is

$C = \text{Red}, \text{Blue}, \dots, \text{Orange} = 1, 2, \dots, M$ .

For each choice alternative  $m$ , define a  $1 \times K_M$  vector of *alternative specific* independent variables,

$$\mathbf{x}_m = [x_{m1} \quad x_{m2} \quad \dots \quad x_{mK}]_{1 \times K_M} \quad (1)$$

For each individual  $i$ , define a  $1 \times K_I$  vector of *individual specific* independent variables,

$$\mathbf{z}_i = [x_{i1} \quad x_{i2} \quad \dots \quad x_{iK}]_{1 \times K_I} \quad (2)$$

Additionally, we could also define  $\mathbf{y}_{im}$  as a vector of individual *and* alternative specific data, but for brevity we omit this type of information without ruling it out.

# The Random Utility Model (RUM) for Multinomial Outcomes

In the Multinomial RUM, an individual  $i$  looks at the indirect utility (an index of well-being) of each of the  $M$  alternatives and chooses the best one:

$$d_{im} = 1 \quad \text{if} \quad V(\mathbf{x}_m, \mathbf{z}_i, \epsilon_m | \beta) > V(\mathbf{x}_j, \mathbf{z}_i, \epsilon_j | \beta) \quad \forall j \in C \\ = 0 \quad \text{otherwise}$$

For individual  $i$ ,  $\sum_{m=1}^M d_{im} = 1$ , since only one alternative can be chosen.

## The Random Utility Model (RUM): An example

Suppose we observe student  $i$ , having characteristics  $\mathbf{z}_i$  considering  $M$  potential universities, each having characteristics  $\mathbf{x}_m$ . Simplify by assuming  $M=3$  (3 choice alternatives) and a linear in parameters form for  $V$ . So, we observe the individual choosing alternative 3 ( $d_{i3} = 1$ ) iff

$$\mathbf{x}_3\beta_x + \mathbf{z}_i\beta_z + \epsilon_{i3} > \mathbf{x}_2\beta_x + \mathbf{z}_i\beta_z + \epsilon_{i2} \text{ and}$$

$$\mathbf{x}_3\beta_x + \mathbf{z}_i\beta_z + \epsilon_{i3} > \mathbf{x}_1\beta_x + \mathbf{z}_i\beta_z + \epsilon_{i1}$$

Note: variables that do not vary over the choice alternatives drop out of the difference, given our linear function. So, we can't identify  $\beta_z$ .

## Econometrics Step 1

As researchers, we can observe  $\mathbf{z}_i$  and/or  $\mathbf{x}_m$  (for all choice alternatives), but we can't observe the  $\epsilon$ 's. Nor is there a straightforward way to construct estimated errors as in OLS, since the multinomial qualitative choice signals if the indirect utility is higher or not, not the degree to which it is higher. But we can tackle this problem in a maximum likelihood framework. In a RUM context, write the probability that individual  $i$  choose '3' as

$$Prob(3) =$$

$$Prob((\mathbf{x}_3 - \mathbf{x}_2)\beta_z > \epsilon_{i2} - \epsilon_{i3}, (\mathbf{x}_3 - \mathbf{x}_1)\beta_z > \epsilon_{i1} - \epsilon_{i3})$$

## Econometrics Step 2

In an analogous manner to what we did before with binary probit and logit models and letting  $f(\epsilon_m)$  be the pdf for unobservables,

$$\begin{aligned} \text{Prob}(3|\beta, \epsilon, \mathbf{x}, \mathbf{z}_i) = & \int_{-\infty}^{\infty} \int_{-\infty}^{(\mathbf{x}_3 - \mathbf{x}_2)\beta_x + \epsilon_{i3}} \int_{-\infty}^{(\mathbf{x}_3 - \mathbf{x}_1)\beta_x + \epsilon_{i3}} f(\epsilon_{i3})f(\epsilon_{i2})f(\epsilon_{i1}) \dots \\ & d\epsilon_{i1}d\epsilon_{i2}d\epsilon_{i3} \end{aligned}$$

In words: Given the alternative and individual characteristics and a guess for  $\beta$ , find the likelihood that a draw of  $\epsilon_{i3}$ ,  $\epsilon_{i2}$  and  $\epsilon_{i1}$  are consistent with the observed choice.

## Econometrics Step 3: Assume a distribution for the errors

Continuing to assume that '3' was chosen:

Probit :

$$Prob(3) = \int_{-\infty}^{\infty} \int_{-\infty}^{(\mathbf{x}_3 - \mathbf{x}_2)\beta_x + \epsilon_{i3}} \int_{-\infty}^{(\mathbf{x}_3 - \mathbf{x}_1)\beta_x + \epsilon_{i3}} \phi(\epsilon_{i3})\phi(\epsilon_{i2})\phi(\epsilon_{i1})d\epsilon_{i1}d\epsilon_{i2}d\epsilon_{i3}$$

Logit :  $Prob(3) = \frac{e^{x_3\beta}}{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta}}$

With that, we can construct the log-likelihood over the entire sample:

$$\ln(L(\beta|\mathbf{d}, \mathbf{x}_i\beta)) = \sum_{i=1}^N \ln \left( \prod_{m=1}^M Prob(m)^{d_{im}} \right) \quad (3)$$

## Interpreting Parameters

Here, the marginal effects can be a bit complicated (all of these are for the multinomial logit model). For example, how changing an attribute at one alternative, changes the probability of another alternative:

$$\frac{\partial \ln[P(m)]}{\partial x_{jk}} = -P(j) \times P(m) \times \beta_k \quad (4)$$

Or, how changing an attribute at one alternative, changes the probability of choosing that alternative:

$$\frac{\partial \ln[P(m)]}{\partial x_{mk}} = P(m) \times \beta_k \quad (5)$$

And finally,

$$\frac{\partial \ln[P(m)/P(j)]}{\partial [x_{mk} - x_{jk}]} = \beta_k \quad (6)$$

## There is no such thing as a free lunch.

The logit probability is quite easy to work with compared to the multiple integrals required for the probit model. But, there is some baggage that comes with it. Inherent in the multinomial logit model is the IIA property- Independence of Irrelevant Alternatives. Consider the ratio of any two probabilities, such as  $\frac{P(3)}{P(1)}$ :

$$\frac{\frac{e^{x_3\beta}}{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta}}}{\frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta} + e^{x_3\beta}}} = \frac{e^{x_3\beta}}{e^{x_1\beta}} \quad (7)$$

So what?? The  $\beta$ 's estimated in the model *must* adhere to this potentially restrictive condition and by extension, preferences and economic inference may be *biased*.

## IIA Example

Suppose there are 7 health plans an individual might buy. Plans 1-5 are HMO plans while Plans 6 and 7 are PPO. Many people prefer PPO plans because it offers flexibility in choosing out of network physicians. Suppose we estimate a multinomial logit model and obtain the following predicted probabilities for individual  $i$ :

Plan	$P(m)$	$P(m)/P(PPO_2)$
$HMO_1$	.13	.65
$HMO_2$	.09	.45
$HMO_3$	.14	.7
$HMO_4$	.09	.45
$HMO_5$	.15	.75
$PPO_1$	.2	1
$PPO_2$	.2	1
Total	1	

## IIA Example, continued

Now suppose that  $PPO_1$  is no longer available. How does the Multinomial Logit and the IIA property reappportion the 20% likelihood of choosing  $PPO_1$  amongst the remaining 6 alternatives?

Plan	7 Options Avail.		6 Options Avail.	
	$P(m)$	$P(m)/P(PPO_2)$	$\tilde{P}(m)$	$\tilde{P}(m)/\tilde{P}(PPO_2)$
$HMO_1$	.13	.65	.16	.65
$HMO_2$	.09	.45	.11	.45
$HMO_3$	.14	.70	.18	.70
$HMO_4$	.09	.45	.11	.45
$HMO_5$	.15	.75	.19	.75
$PPO_1$	.20	1.00	N/A	N/A
$PPO_2$	.20	1.00	.25	1.00
Total	1		1	

## Mechanics of IIA

Assume 3 alternatives, with initial probabilities  $P(1)$ ,  $P(2)$ , and  $P(3)$ . Suppose that alternative 2 is eliminated, denote the *new* probabilities as  $\tilde{P}(1)$  and  $\tilde{P}(3)$ . Given the multinomial logit model, it must be the case that following the elimination of alternative 2, the following 2 conditions must hold:

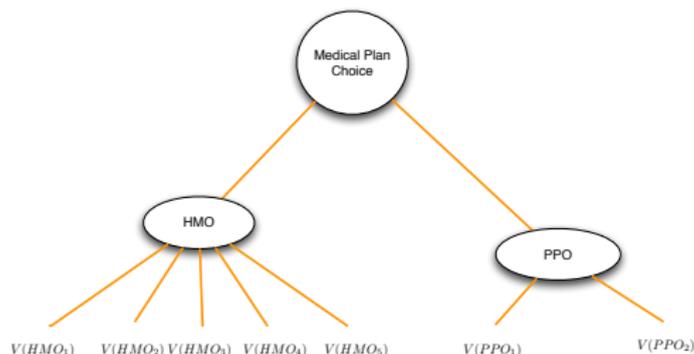
Adding Up:  $\tilde{P}(1) + \tilde{P}(3) = 1$

IIA:  $\frac{\tilde{P}(1)}{\tilde{P}(3)} = \frac{P(1)}{P(3)}$ , which implies  $\tilde{P}(1) = \frac{\tilde{P}(3)P(1)}{P(3)}$

Using these, it can be shown that

- $\tilde{P}(3) = \frac{P(3)}{P(1)+P(3)}$
- $\tilde{P}(1) = \frac{P(1)}{P(1)+P(3)}$

# The Nested Logit Model



$$V(PPO_m) = \mathbf{x}_m\beta + \mathbf{z}_{PPO}\gamma + \epsilon_m^{PPO}$$

$$V(HMO_m) = \mathbf{x}_m\beta + \mathbf{z}_{HMO}\gamma + \epsilon_m^{HMO}$$

Basic Idea: Relax IIA for alternatives in different nests, while keeping IIA within nests.

## Choice Probabilities

The nested logit model writes the probability of choosing the second PPO alternative, for example, as

$$P(PPO_2) = P(PPO) \times P(PPO_2|PPO) \quad (8)$$

where

$$P(PPO) = \frac{e^{TPPO(z_{PPO}\gamma + IV(PPO))}}{e^{TPPO(z_{PPO}\gamma + IV(PPO))} + e^{THMO(z_{HMO}\gamma + IV(HMO))}} \quad (9)$$

$$P(PPO_2|PPO) = \frac{e^{XPPO,2\beta}}{e^{XPPO,1\beta} + e^{XPPO,2\beta}} \quad (10)$$

and  $IV(B) = \ln \left[ \sum_{m \in B} e^{XB,m} \right]$

## What is the role of $\tau$

$\tau_{PPO}$  and  $\tau_{HMO}$  dictate the degree to which it is easy to substitute from alternatives within one branch to alternatives in another branch. It can be shown that the IIA property

- ① Is imposed for alternatives *within* a branch
- ② Is relaxed (does not hold) for alternatives across branches

If  $\tau_{PPO} = \tau_{HMO} = 1$ , then the nested logit collapses to the multinomial logit model. Therefore, the suitability of the IIA property can be tested in a maximum likelihood framework:

$$H_0: \tau_{PPO} = \tau_{HMO} = 1$$

$$H_1: \tau_{PPO} \neq \tau_{HMO} \neq 1$$

Using a likelihood ratio test  $\chi^2$ , in this case having 2 degrees of freedom since we have two restrictions.

## Practical Issues and Extensions

- For large M, the clogit model is almost universally used (more than 10 choice alternatives)
- Stata has two multinomial logit commands: mlogit and clogit. Mlogit is consistent with the varying parameters model and clogit with the RUM model.
- The R command mlogit handles both types of models
- Some other extensions (also relevant for the binary Probit and Logit models):
  - 1 Random Parameters (or mixed models) that relax IIA
  - 2 Heteroskedastic error models