

## § 4.4 离散被解释变量数据计量经济学 模型（二）——多元选择模型

### Models with Discrete Dependent Variables—Multiple Choice Model

- 一、多元离散选择模型的经济背景
- 二、一般多元离散选择Logit模型
- 三、嵌套多元离散选择模型
- 四、排序多元离散选择模型

# 一、多元离散选择模型的经济背景

# 1、经济生活中的多元选择问题

- 一般的多元选择问题
- 排序选择问题
  - 将选择对象按照某个准则排队，由决策者从中选择。
  - 决策者对同一个选择对象的偏好程度。
- 嵌套选择问题

## 2、社会生活中的多元选择问题

- 一般的多元选择问题
- 排序选择问题
- 嵌套选择问题

## 二、一般多元离散选择Logit模型

## 说明

- 在多元离散选择模型中，因为Probit模型需要对多元正态分布的整体进行评价，所以它的应用受到限制。
- 逻辑分布更适合于效用最大化时的分布选择，所以应用最多的多元离散选择模型是Logit模型。
- Logit模型的似然函数能够快速可靠地收敛，当方案或者决策个体数量较大时，计算比较简便。

- Logit模型计算的简便性是有条件的。即选择方案是不相关的，具体包括：
  - $U_{ij} = V_{ij} + \varepsilon_{ij}$ ,  $\varepsilon_{ij}$ 是独立的;
  - $P_i/P_k$ 与其它选择方案的属性无关，与选择方案的个数无关;
  - $P_i$ 关于其它选择方案属性的弹性是不变的，与i无关。
- 上述条件只有在选择方案的差异相同的情况下才能得到满足。
- 在相关文献中有数学证明。

## 1. 一般多元选择Logit模型的思路

- 如果决策者*i*在（**J+1**）项可供选择方案中选择了第**j**项，那么其效用模型为：

$$U_{ij} = X_{ij}B + \varepsilon_{ij}$$

选择j的概率

$$P(U_{ij} > U_{ik}) \quad k = 0, 1, 2, \dots, J \quad k \neq j$$

$$P(y_i = j) = \frac{e^{X_{ij}B}}{\sum_{j=0}^J e^{X_{ij}B}}$$

如果（**J+1**）个随机误差项互不相关，并且服从I类极值分布

$$F(\varepsilon_{ij}) = e^{-\varepsilon_{ij}}$$



- 效用模型的解释变量中包括所有影响选择的因素，既包括决策者所具有的属性，也包括备选方案所具有的属性。
- 备选方案所具有的属性是随着方案的变化而变化的。
- 决策者所具有的属性中一部分是随着方案的变化而变化的，而一部分是不随着方案的变化而变化的。
- 用 $\mathbf{Z}_{ij}$ 表示随着方案的变化而变化的那部分解释变量， $\mathbf{W}_i$ 表示不随着方案的变化而变化的那部分解释变量。

$$P(y_i = j) = \frac{e^{Z_{ij}\Gamma + W_i A}}{\sum_{j=0}^J e^{Z_{ij}\Gamma + W_i A}} = \frac{e^{Z_{ij}\Gamma} e^{W_i A}}{\sum_{j=0}^J e^{Z_{ij}\Gamma} e^{W_i A}}$$

$$P(y_i = j) = \frac{e^{Z_{ij}\Gamma} e^{W_i A}}{e^{W_i A} \sum_{j=0}^J e^{Z_{ij}\Gamma}} = \frac{e^{Z_{ij}\Gamma}}{\sum_{j=0}^J e^{Z_{ij}\Gamma}}$$

- 实用的一般多元**Logit**选择模型又分**3**种情况。
- 一是研究选择某种方案的概率与决策者的特征变量之间的关系；
- 二是研究选择某种方案的概率与决策者的特征变量以及方案的特征变量之间的关系；
- 三是考虑到不同方案之间的相关性的情况。

**Multinomial Logit Model**

多项式**Logit**模型

名义**Logit**模型

**Conditional Logit Model**

条件**Logit**模型

**Nested Logit模型**

嵌套模型

## 2. 多元名义Logit离散选择模型及其参数估计

$$P(y_i = j) = \frac{e^{X_i B_j}}{\sum_{j=0}^J e^{X_i B_j}}$$

**X**中未包含备选方案所具有的属性变量，而参数向量**B**对不同的选择方案（即不同的方程）是不同的。

$$P(y_i = j) = \frac{e^{X_i B_j}}{1 + \sum_{k=1}^J e^{X_i B_k}}$$

$$P(y_i = 0) = \frac{1}{1 + \sum_{k=1}^J e^{X_i B_k}}$$

令  $B_0 = 0$ ,  $j = 1, 2, \dots, J$

多元名义 Logit 离散选择模型的参数估计并不复杂。对于第  $i$  个决策者，如果选择了第  $j$  个备选方案，令  $d_{ij}=1$ ；如果没有选择第  $j$  个备选方案，令  $d_{ij}=0$ 。

同时，对于第  $i$  个决策者，在  $(J+1)$  个备选方案中，只能选择其中之一，即只能存在 1 个  $d_{ij}=1$ 。于是，可以写出  $y_{ij}(i=1,2,\cdots,n, j=0,1,2,\cdots,J)$  的联合概率函数，由联合概率函数导出似然函数，进而得到对数似然函数为：

$$\ln L = \sum_{i=1}^n \sum_{j=0}^J d_{ij} \ln P(y_i = j)$$

$$\frac{\partial \ln L}{\partial B_j} = \sum_i (d_{ij} - P_{ij}) X_i \quad j = 1, 2, \dots, J$$

$$\frac{\partial^2 \ln L}{\partial B_j \partial B'_l} = - \sum_{i=1}^n P_{ij} (I(j=l) - P_{il}) X_i X'_i$$

$$I(j=l) = \begin{cases} 1 & \text{如果 } j=l \\ 0 & \text{如果 } j \neq l \end{cases}$$

由对数似然函数最大化的一阶条件，利用**Newton**迭代方法可以迅速地得到方程组的解，得到模型的参数估计量。

## 另一种估计方法

- 可以计算得到相对于基准方案的对数概率比为：

$$\ln\left(\frac{P_{ij}}{P_{i0}}\right) = X_i B_j$$

$$\ln\left(\frac{P_{ij}}{P_{ik}}\right) = X_i (B_j - B_k)$$

如果对每个决策者进行重复观测，可以得到被解释变量的观测值。

如果对每个决策者只进行一次观测，如何得到被解释变量的观测值？

- 两点注意：

假设了原模型中（**J+1**）个随机误差项互不相关。  
对估计结果的解释不同。

## 例题

- 农村异地转移劳动力的迁移目标研究。
- **被解释变量**：迁移目标，即小城镇、县级市、地级市、省级城市和超大城市，依次取值**1、2、3、4、5**。
- **解释变量**：**个人特征和目前所在地属性**。连续变量包括受教育程度、家庭规模、家庭内其他劳动力人数、家庭负担、原有收入、现有收入，目前所在地属性中的所在地农村人口、国内生产总值、城乡居民储蓄余额、粮食产量、中学生在校人数、小学生在校人数等。**离散变量**包括性别、婚姻状况、收入稳定与否，目前所在地所属级别与家乡所在地所属级别等。
- 虽然作为被解释变量的城市规模本身是有序的，但是对于农村劳动力来说，选择进入哪一个级别的城市，本身是无序的，因此对于城市化迁移目标构造**多元名义logit离散选择模型**。



- 调查样本，有效样本**303**份。
- 用**SAS**统计软件进行估计与分析。
- 首先将定义的全部变量放进模型中进行估计，并通过比较各个变量的**P**值来考虑具体剔除哪些变量以及对哪些变量考虑将其交互影响的效应放进模型中去。
- 小城镇、县级市、地级市、省级城市和超大城市依次取值**1、2、3、4、5**。

迁移目标	频数	百分比	累计频数	累计百分比
小城镇	15	4.95	15	4.95
县级市	62	20.46	77	25.41
地级市	84	27.72	161	53.14
省级城市	72	23.76	233	76.90
超大城市	70	23.10	303	100.00

- 由于得到了频数，可以采用“对数概率模型”进行估计。

- 最终模型的估计结果（部分）

变量	模型序号	系数估计	标准差	P 值
常数项	1	1.2137	1.4518	0.4032
	2**	2.7685	1.0998	0.0118
	3**	2.3962	0.9351	0.0104
	4***	3.6742	1.0665	0.0006
教育程度	1**	-0.2475	0.1050	0.0184
	2***	-0.2800	0.0727	0.0001
	3*	-0.1136	0.0660	0.0852
	4	-0.0856	0.0696	0.2184
家庭情况	1	-0.1299	0.1084	0.2310
	2*	-0.0943	0.0552	0.0877
	3	-0.0337	0.0452	0.4556
	4***	-0.1578	0.0586	0.0071

\*代表的是90%的显著性水平，  
 \*\*代表的是95%的显著性水平，  
 \*\*\*代表的是99%的显著性水平。

- 将模型的结果整理出来，并对每个解释变量进行分析。
- 例如：教育程度、家庭情况及现有收入对迁移目标的影响：

因变量	教育程度	家庭情况	现有收入
log(1/5)	-0.2475**	-0.1299	-0.00144*
log(2/5)	-0.2800***	-0.0943*	-0.00032**
log(3/5)	-0.1136*	-0.0337	-0.00028**
log(4/5)	-0.0856	-0.1578***	-0.00030**

- **从教育程度来看**，所有系数都是负值，教育程度越高的农村劳动力越愿意进入规模较大的城市；从显著性水平来看，相对于超大城市来说，县级市被选择的可能性最小，其次是小城镇，然后是地级城市，而教育程度相似的农村劳动力在省级城市与超大城市之间的选择没有明显的差异。
- **从家庭情况来看**，所有系数都是负值，也就是说家庭情况越好的农村劳动力越愿意进入规模较大的城市；从显著性水平来看，相对于超大城市来说，省级城市最不容易被选中，其次是县级市，而小城镇与地级市之间没有明显区别。
- **从现有收入来看**，所有系数都是负值，也就是说目前收入越高的农村劳动力越愿意进入规模较大的城市；再从显著性水平来看，所有系数都是显著的，这说明相对于任何级别的城市而言，农村劳动力都更倾向于超大城市。

### 3. 多元条件Logit离散选择模型及其参数估计

- 选择某种方案的概率不仅与决策者的特征变量有关，而且也与方案的特征变量有关，模型为：

$$P(y_i = j) = \frac{e^{X_{ij}B_j}}{\sum_{j=1}^J e^{X_{ij}B_j}}$$

区别在于  
X的下标

$$\ln L = \sum_{i=1}^n \sum_{j=1}^J d_{ij} \ln P(y_i = j)$$

$$\frac{\partial \ln L}{\partial \mathbf{B}} = \sum_{i=1}^n \sum_{j=1}^J d_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)$$

$$\bar{\mathbf{X}}_i = \sum_{j=1}^J P_{ij} \mathbf{X}_{ij}$$

$$\frac{\partial^2 \ln L}{\partial \mathbf{B} \partial \mathbf{B}'} = - \sum_{i=1}^n \sum_{j=1}^J P_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'$$

- 由对数似然函数最大化的一阶条件，利用**Newton**迭代方法可以迅速地得到方程组的解，得到模型的参数估计量。

### 三、嵌套多元离散选择模型



## 1、问题的提出

- $(J+1)$ 个不同的选择方案之间具有相关性，而且必须考虑这种相关性，表现为模型随机误差项相关。
- 可行的思路是将  $(J+1)$  个选择方案分为  $L$  组，在每组内部的选择方案之间不具有相关性，而组间则具有相关性。
- 就是将条件**Logit**模型中隐含的齐次方差性条件放松，允许方差在组间可以不同，但在组内仍然是同方差的。
- 这样的模型被称为**Nested Logit**模型。

# 1、Nested Logit模型

$$P(j, l) = P_{jl} = \frac{e^{X_{j|l}B + Z_l\Gamma}}{\sum_{l=1}^L \sum_{j=1}^{J_l} e^{X_{j|l}B + Z_l\Gamma}}$$

表示对选择第l组产生影响的变量

表示在第l组内对选择第j种方案产生影响的变量

$$P_{jl} = P_{j|l} P_l = \frac{\left( \frac{e^{X_{j|l}B}}{\sum_{j=1}^{J_l} e^{X_{j|l}B}} \right) \left( \frac{e^{Z_l\Gamma}}{\sum_{l=1}^L e^{Z_l\Gamma}} \right)}{\left( \frac{\sum_{j=1}^{J_l} e^{X_{j|l}B}}{\sum_{l=1}^L \sum_{j=1}^{J_l} e^{X_{j|l}B + Z_l\Gamma}} \right) \left( \frac{\sum_{l=1}^L e^{Z_l\Gamma}}{\sum_{l=1}^L \sum_{j=1}^{J_l} e^{X_{j|l}B + Z_l\Gamma}} \right)}$$

- 定义第 $l$ 组的“内值” (**Inclusive Value**)

$$I_l = \ln \sum_{j=1}^{J_l} e^{X_{j|l}B}$$

$$P_{j|l} = \frac{e^{X_{j|l}B}}{\sum_{j=1}^{J_l} e^{X_{j|l}B}}$$

$$P_l = \frac{e^{Z_l\Gamma + I_l T_l}}{\sum_{l=1}^L e^{Z_l\Gamma + I_l T_l}}$$

### 3、估计方法

- **两阶段最大似然法**，是一种有限信息估计方法。其具体步骤是：
  - 在组内，作为一个简单的条件**Logit**模型，估计参数；
  - 计算每组的“内值”；
  - 将每组看成是一种选择方案，再进行简单的条件**Logit**模型的估计，得到参数 $\Gamma$ 和 $T$ 的估计量。此时用到的贡献变量是 $Z_i$ 和 $I_i$ 。

- 完全信息最大似然法。将对数似然函数写为：

$$\ln L = \sum_{i=1}^n \ln (P_{j|l} \times P_l)_i$$

比两阶段最大似然法更有效

## 四、排序多元离散选择模型

Multivariate Choice Model  
for Ordered Data

## 1、问题的提出

- 作为被解释变量的 ( $J+1$ ) 个选择结果本身是排序的,  $J$  优于 ( $J-1$ ),  $2$  优于  $1$ ,  $1$  优于  $0$ 。
- 决策者选择不同的方案所得到的效用也是排序的。
- 一般多元离散选择模型中的效用关系不再适用。

## 2、效用关系

- 选择不同方案的效用关系:

$y = 0$	如果 $y^* \leq 0$
$= 1$	如果 $0 < y^* \leq u_1$
$= 2$	如果 $u_1 < y^* \leq u_2$
$\vdots$	
$= J$	如果 $u_{J-1} \leq y^*$



### 3、模型

$$y^* = XB + \mu$$

$$P(y = 0) = \Phi(-XB)$$

$$P(y = 1) = \Phi(u_1 - XB) - \Phi(-XB)$$

$$P(y = 2) = \Phi(u_2 - XB) - \Phi(u_1 - XB)$$

$\vdots$

$$P(y = J) = 1 - \Phi(u_{J-1} - XB)$$

假定  $\mu$  服从正态分布，并且标准化为服从期望为0、方差为1的正态分布。那么可以得到选择各个方案的概率

$\Phi$  为正态分布的概率函数

为了保证所有的概率都是正的，必须有：

$$0 < u_1 < u_2 < \cdots < u_{J-1}$$

## 4、估计

- 可以看作二元**Probit**模型的推广；
- 采用最大似然法估计。