Topics in Econometrics— Discrete Choice Models
Multinomial Models and Count Data

Yao Thibaut Kpegli

Master II Advanced Economics, ENS Lyon
2022–2023

Context and Objective

Context

1. You have observed a sample $\boldsymbol{y} = \{y_1, ..., y_n\}$ and y_i has J possible values (say 1,2,..J)

$$y_i = \begin{cases} 1 & \text{with probability} & p_{i1} \\ 2 & \text{with probability} & p_{i2} \\ ... & ... & \\ J & \text{with probability} & p_{iJ} \end{cases} \tag{1}$$

2. You have also observed a vector of characteristics $\boldsymbol{x}_i$ $(K, 1)$ associated to the individual i
3. You suspect that $\boldsymbol{x}_i$ determines p_{ij} :

- and you assume this is through a function g_j depending on vectors of parameters $\boldsymbol{\theta}$ $(K,1)$

$$p_{ij} = g_j(\boldsymbol{x}_i, \boldsymbol{\theta}) \qquad \text{with} \qquad \sum_{j=1}^{J} p_{ij} = \sum_{j=1}^{J} g_j(\boldsymbol{x}_i, \boldsymbol{\theta}) = 1$$

Context and Objectives

Objective

Provide an estimate of $\boldsymbol{\theta}$

Example

Examples of multiple-choice context:

- Opinions: strongly opposed / opposed / neutral / support (ranked choices),
- Occupational field: lawyer / farmer / engineer / doctor / ...,
- Occupational field: lawyer / farmer / engineer / doctor / ...,
- Alternative shopping areas,
- Transportation types
- ...

General modeling

General modeling

1. From $\boldsymbol{y}$, we can extract J binary variables $\boldsymbol{y}^1, ..., \boldsymbol{y}^J$ as follows

$$y_i^j = \begin{cases} 1 & \text{if} \quad y_i = j \\ 0 & \text{otherwise} \end{cases} \qquad j = 1, 2, ..., J \tag{2}$$

2. The log-likelihood function associated to $(\boldsymbol{y}, \boldsymbol{x})$ is

$$log\, \mathcal{L}(\boldsymbol{\theta}, ...; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_i^j log\Big(g_j(\boldsymbol{x}_i, \boldsymbol{\theta})\Big) \tag{3}$$

The ordered case

Context

1. The J alternatives feature a natural ordering
   - example: income groups (low, lower-middle, upper-middle, and high-income countries)

Latent variable approach

1. A continuous but unobservable variable y_i^*:

$$y_i^* = \boldsymbol{\theta}' \boldsymbol{x}_i + \epsilon_i$$

   - ϵ_i a continuous random variable (e.g. normal or logistic) with F the c.d.f of ϵ_i

2. Assume that $-\infty = c_0 < c_1 < c_2 < ... < c_J = +\infty$, such that $\forall j \in \{1, 2, ..., J\}$

$$\mathbb{P}(y_i = j) = \mathbb{P}(c_{j-1} < y_i^* \leq c_j) = F(c_j - \boldsymbol{\theta}' \boldsymbol{x}_i) - F(c_{j-1} - \boldsymbol{\theta}' \boldsymbol{x}_i)$$

The ordered case

Estimation

1. **Ordered Logit**: $F(.)$ is the c.d.f of logistic distribution
2. **Ordered Probit**: $F(.)$ is the c.d.f of standard normal distribution
3. If one component of $\boldsymbol{x}_i$ is 1 (intercept), then set $c_1 = 0$ for identifiability (like in binary logit/probit cases)
4. The log-likelihood equation (3) becomes

$$log\,\mathcal{L}(\boldsymbol{\theta}, c_2, ..., c_{J-1}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_i^j log\Big(g_j(\boldsymbol{x}_i, \boldsymbol{\theta})\Big) \tag{4}$$

   with $g_j(\boldsymbol{x}_i, \boldsymbol{\theta}) = F(c_j - \boldsymbol{\theta}'\boldsymbol{x}_i) - F(c_{j-1} - \boldsymbol{\theta}'\boldsymbol{x}_i)$
5. Apply MLE to estimate $\boldsymbol{\theta}$ as well as the cutoffs $, c_2, ..., c_{J-1}$

Marginal Effects

Marginal Effects

1. The marginal effects:

$$\frac{\partial \mathbb{P}(y_i = j)}{\partial \boldsymbol{x}_i} = \underbrace{\Big(f(c_j - \boldsymbol{\theta}'\boldsymbol{x}_i) - f(c_{j-1} - \boldsymbol{\theta}'\boldsymbol{x}_i)\Big)}_{A} \boldsymbol{\theta} \qquad (5)$$

   - therefore the signs of the marginal effects are not necessarily those of the components of $\boldsymbol{\theta}$. (The sign of A is a priori unknown.)

2. The sign of of the components of $\boldsymbol{\theta}$ allows to know whether or not the latent variable y^* increases with the the corresponding regressor.

Application 1

FIGURE: Data

variable name	type	format	label	variable label
warm	byte	%10.0g	SD2SA	**Mom can have warm relations with child**
yr89	byte	%10.0g	yrlbl	**Survey year: 1=1989 0=1977**
male	byte	%10.0g	sexlbl	**Gender: 1=male 0=female**
white	byte	%10.0g	race2lbl	**Race: 1=white 0=not white**
age	byte	%10.0g		**Age in years**
ed	byte	%10.0g		**Years of education**
prst	byte	%10.0g		**Occupational prestige**

warm: 1 (cold relations) to 4 (warm relations)

Application 1

Ordered logit

Table: Estimation results : ologit

Variable	Coefficient	(Std. Err.)
yr89	0.524**	(0.080)
male	-0.733**	(0.078)
white	-0.391**	(0.118)
age	-0.022**	(0.002)
ed	0.067**	(0.016)
prst	0.006†	(0.003)
c_1	-2.465**	(0.239)
c_2	-0.631**	(0.233)
c_3	1.262**	(0.234)

Significance levels : † : 10% $*$: 5% $**$: 1%

Application 1 (Marginal effect)

	MEM		AME	
yr89				
1._predict	-0.0518***	(0.00814)	-0.0557***	(0.00879)
2._predict	-0.0773***	(0.0122)	-0.0608***	(0.00918)
3._predict	0.0582***	(0.00962)	0.0435***	(0.00691)
4._predict	0.0708***	(0.0109)	0.0730***	(0.0112)
male				
1._predict	0.0725***	(0.00813)	0.0780***	(0.00881)
2._predict	0.108***	(0.0125)	0.0851***	(0.00892)
3._predict	-0.0815***	(0.01000)	-0.0609***	(0.00678)
4._predict	-0.0992***	(0.0109)	-0.102***	(0.0112)
white				
1._predict	0.0387**	(0.0118)	0.0416**	(0.0127)
2._predict	0.0577**	(0.0176)	0.0454***	(0.0137)
3._predict	-0.0435**	(0.0134)	-0.0325**	(0.00997)
4._predict	-0.0529***	(0.0160)	-0.0545***	(0.0165)
age				
1._predict	0.00214***	(0.000255)	0.00230***	(0.000275)
2._predict	0.00319***	(0.000390)	0.00252***	(0.000282)
3._predict	-0.00241***	(0.000313)	-0.00180***	(0.000214)
4._predict	-0.00293***	(0.000340)	-0.00302***	(0.000350)
ed				
1._predict	-0.00664***	(0.00160)	-0.00714***	(0.00171)
2._predict	-0.00990***	(0.00240)	-0.00780***	(0.00185)
3._predict	0.00746***	(0.00184)	0.00558***	(0.00133)
4._predict	0.00908***	(0.00217)	0.00936***	(0.00224)
prst				
1._predict	-0.000600	(0.000326)	-0.000646	(0.000351)
2._predict	-0.000895	(0.000487)	-0.000705	(0.000382)
3._predict	0.000675	(0.000368)	0.000505	(0.000275)
4._predict	0.000821	(0.000445)	0.000846	(0.000458)

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Multinomial logit (MNL)

1. **Alterantive-invariant $\boldsymbol{x}$**: the explanatory variables vary only with i, but not with j
2. The probabilities are given by:

$$\mathbb{P}(y_i = j) = \frac{\exp(\boldsymbol{\theta}_j' \boldsymbol{x}_i)}{\sum_{k=1}^{J} \exp(\boldsymbol{\theta}_k' \boldsymbol{x}_i)} \quad , \; j = 1, ..., J \tag{6}$$

   with $\boldsymbol{\theta}_1 = \mathbf{0}$ (normalization)
3. The log-likelihood equation (3) becomes

$$log\, \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_i^j log\Big(\mathbb{P}(y_i = j)\Big) \tag{7}$$

4. Apply MLE to estimate $\boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_J$

Multinomial logit (MNL)

interpretation of Coefficients

$$\frac{\mathbb{P}(y_i = j)}{\mathbb{P}(y_i = 1)} = \exp(\boldsymbol{\theta}_j' \boldsymbol{x}_i)$$

The sign of coefficient indicates how the probability of alternative j varies in comparison to the baseline ($y = 1$).

Marginal Effects

1. The marginal effects:

$$\frac{\partial \mathbb{P}(y_i = j)}{\partial \boldsymbol{x}_i} = \mathbb{P}(y_i = j)\left(\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}_i}\right) \tag{8}$$

with $\overline{\boldsymbol{\theta}_i} = \sum_{l=1}^{J} \mathbb{P}(y_i = l)\boldsymbol{\theta}_l$

- the sign of the marginal effect is the one of A which is priori unknown.)

Conditional logit (CL)

1. **Alternative-varying $\boldsymbol{x}$**: the explanatory variables vary with j

$$\mathbb{P}(y_i = j) = \frac{\exp(\boldsymbol{\theta}'\boldsymbol{x}_i^j)}{\sum_{k=1}^{J}\exp(\boldsymbol{\theta}'\boldsymbol{x}_i^k)} \ , \ j = 1, ..., J \tag{9}$$

2. The log-likelihood equation (3) becomes

$$log\,\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_i^j log\Big(\mathbb{P}(y_i = j)\Big) \tag{10}$$

3. Apply MLE to estimate $\boldsymbol{\theta}$

Conditional logit (CL)

Interpretation of Coefficients and Marginal Effects

1. The marginal effects:

$$\frac{\partial \mathbb{P}(y_i = j)}{\partial \boldsymbol{x}_i^j} = \mathbb{P}(y_i = j)\Big(1 - \mathbb{P}(y_i = j)\Big)\boldsymbol{\theta} \tag{11}$$

$$\frac{\partial \mathbb{P}(y_i = j)}{\partial \boldsymbol{x}_i^k} = -\mathbb{P}(y_i = j)\mathbb{P}(y_i = k)\boldsymbol{\theta} \tag{12}$$

- The sign of own-effect is the one of the corresponding component of $\boldsymbol{\theta}$
- The sign of cross-effect is the opposite of the one of the corresponding component of $\boldsymbol{\theta}$

Mixture of MNL and CL

1. Mixture of MNL and CL

$$\mathbb{P}(y_i = j) = \frac{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i^j + \boldsymbol{\delta}_j'\boldsymbol{x}_i)}{\sum_{k=1}^{J}\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i^k + \boldsymbol{\delta}_k'\boldsymbol{x}_i)} \;,\; j = 1, ..., J \qquad (13)$$

   - $\boldsymbol{z}$: Alternative-varying variables
   - $\boldsymbol{x}$: Alternative-invariant variables

2. The log-likelihood equation (3) becomes

$$log\,\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_i^j\, log\Big(\mathbb{P}(y_i = j)\Big) \qquad (14)$$

3. Apply MLE to estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$
4. Interpret coefficients and marginal effects as in MNL and CL

Justification with random utility

Definition (Gumbel distribution)

A random variable X follows the Gumbel distribution ($\mathcal{W}$) if its c.d.f (F) and p.d.f (f) are given by

$$F(x) = \exp\Big(- \exp(-x)\Big) \quad \text{and} \quad f(x) = \exp\Big(- x - \exp(-x)\Big)$$

Definition (random utility)

The random utility $U_{i,j}$ associated to the alternative $j \in \{1, 2, ..., J\}$ of the unit i is:

$$U_{i,j} = V_{i,j} + \epsilon_{i,j}$$

with $V_{i,j}$ a deterministic component of the utility and $\epsilon_{i,j}$ a random variable (response error)

Justification with random utility

Proposition (Utility Model justification)

In the context of the random utility, if $\epsilon_{i,j}$ are independent and follow the Gumbel distribution ($\mathcal{W}$), then

$$\mathbb{P}(y_i = j) = \frac{\exp(V_j)}{\sum_{k=1}^{J} \exp(V_k)} \quad , \ j = 1, \ldots, J \tag{15}$$

Proof

$$\begin{aligned} \mathbb{P}(y_i = j) &= \mathbb{P}\Big(U_{i,j} > U_{i,k}; \forall k \neq j\Big) \\ &= \mathbb{P}\Big(\epsilon_{i,k} < V_{i,j} - V_{i,k} + \epsilon_{i,j}; \forall k \neq j\Big) \end{aligned}$$

After computing

$$\mathbb{P}(y_i = j) = \int \prod_{k \neq j} F\Big(V_{i,j} - V_{i,k} + \epsilon\Big) f(\epsilon) d\epsilon$$

Justification with random utility

After computing, we have

$$\prod_{k\neq j} F\Big(V_{i,j} - V_{i,j} + \epsilon\Big) f(\epsilon) \quad = \quad \exp\Big(- \epsilon - \exp(-\epsilon + \lambda_{i,j})\Big)$$

with $\lambda_{i,j} = -\log \left(\frac{\exp(V_{i,j})}{\sum_{k=1}^{J} \exp(V_{i,k})} \right)$

$$\begin{aligned}
\mathbb{P}(y_i = j) &= \int \exp\Big(- \epsilon - \exp(-\epsilon + \lambda_{i,j})\Big) d\epsilon \\
&= \int \exp\Big(- t - \lambda_{i,j} - \exp(-t)\Big) dt \\
&= \exp\Big(- \lambda_{i,j}\Big) \int f(t) dt = \exp(-\lambda_{i,j}) \\
\mathbb{P}(y_i = j) &= \frac{\exp(V_{i,j})}{\displaystyle\sum_{k=1}^{J} \exp(V_{i,k})}
\end{aligned}$$

Justification with random utility

1. MNL : $V_{i,j} = \boldsymbol{\theta}_j' \boldsymbol{x}_i$
2. CL : $V_{i,j} = \boldsymbol{\theta}' \boldsymbol{x}_i^j$
3. MNL & CL : $V_{i,j} = \boldsymbol{\gamma}' \boldsymbol{z}_i^j + \boldsymbol{\delta}_j' \boldsymbol{x}_i$

Independence of Irrelevant Alternatives (IIA)

IIA

1. For any individual, the ratio of probabilities of choosing two alternatives is independent of the availability or attributes of any other alternatives:

$$\frac{\mathbb{P}(y_i = j)}{\mathbb{P}(y_i = k)} = \exp(V_{i,j} - V_{i,k})$$

$$\mathbb{P}(y_i = j | y_i \in \{j, k\}) = \frac{\exp(V_{i,j})}{\exp(V_{i,j}) + \exp(V_{i,k})}$$

2. IIA is driven by Independence of errors

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color
2. IIA: the conditional probability of commute by car given commute by car or red bus, i.e. $\mathbb{P}(y = 1|y \in \{1, 2\})$ is independent on whether there is a blue bus option or not

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color
2. IIA: the conditional probability of commute by car given commute by car or red bus, i.e. $\mathbb{P}(y = 1|y \in \{1, 2\})$ is independent on whether there is a blue bus option or not
3. But, in "fact" ...
   - how would the probability of red bus commuting changes when introducing a blue bus ? Halve ?

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color
2. IIA: the conditional probability of commute by car given commute by car or red bus, i.e. $\mathbb{P}(y = 1|y \in \{1, 2\})$ is independent on whether there is a blue bus option or not
3. But, in "fact" ...
   - how would the probability of red bus commuting changes when introducing a blue bus ? Halve ?
   - how would the probability of car commuting change when introducing a blue bus ? no effect ?

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color
2. IIA: the conditional probability of commute by car given commute by car or red bus, i.e. $\mathbb{P}(y = 1 | y \in \{1, 2\})$ is independent on whether there is a blue bus option or not
3. But, in "fact" ...
   - how would the probability of red bus commuting changes when introducing a blue bus ? Halve ?
   - how would the probability of car commuting change when introducing a blue bus ? no effect ?
   - how would $\mathbb{P}(y = 1 | y \in \{1, 2\})$ change ? Increases ?

Independence of Irrelevant Alternatives (IIA)

Example: the red bus vs blue bus problem

1. Transportation: car (y=1), red bus (y=2), blue bus (y=3)
   - The only difference between red and blue bus is the color
2. IIA: the conditional probability of commute by car given commute by car or red bus, i.e. $\mathbb{P}(y = 1|y \in \{1, 2\})$ is independent on whether there is a blue bus option or not
3. But, in "fact" ...
   - how would the probability of red bus commuting changes when introducing a blue bus ? Halve ?
   - how would the probability of car commuting change when introducing a blue bus ? no effect ?
   - how would $\mathbb{P}(y = 1|y \in \{1, 2\})$ change ? Increases ?
4. While IIA is driven by independence of errors, we expect that ϵ_2 and ϵ_3 are highly correlated.

Nested Logit

Intuition: the red bus vs blue bus problem

1. First, the individual chooses between two modes of transportation: "car" vs "bus"
2. Second, if the individual chooses the mode "bus"; then he needs to chose the color of the bus

Context: nesting structure

The decision three is as follows:

1. The individual chooses a **limb** j among the available J limbs
2. If the individual chooses the limb j, then he will choose one of the K_j **branches** available in the limb j

Nested Logit

1. Specification of the deterministic part of utility associated to limb j and branch k:

$$U_{j,k} = \boldsymbol{z}_j'\boldsymbol{\alpha} + \boldsymbol{x}_{j,k}'\boldsymbol{\beta}_j + \epsilon_{j,k} \quad , \quad j = 1, 2, ..., J \quad \text{and} \quad k = 1, 2, ..., K_j$$

   - $\boldsymbol{z}$ vary only across limbs.
   - $\boldsymbol{x}$ vary across limbs and branches.

2. The joint probability of being in limb j and choosing branch k is the product of the probability of being in limb j times the conditional probability of choosing branch k given that limb j have been chosen.

Nested Logit

- The probability of choosing limb j and branch k:

$$p_{j,k} = \underbrace{\frac{exp(\boldsymbol{z}_j'\boldsymbol{\alpha} + \tau_j I_j)}{\sum_{m=1}^{J} exp(\boldsymbol{z}_m'\boldsymbol{\alpha} + \tau_m I_m)}}_{p_j} \times \underbrace{\frac{exp(\boldsymbol{x}_{j,k}'\boldsymbol{\beta}_j/\tau_j)}{\sum_{l=1}^{K_j} exp(\boldsymbol{x}_{j,l}'\boldsymbol{\beta}_j/\tau_j)}}_{p_{k|j}} \tag{16}$$

  - I_j is the **inclusive value** or the **log-sum**

$$I_j = \ln\left(\sum_{l=1}^{K_j} exp(\boldsymbol{x}_{j,l}'\beta_j/\tau_j)\right)$$

  - dissimilarity parameters: $\tau_j = \sqrt{1 - Corr(\epsilon_{j,k}, \epsilon_{j,l})}$, which is inversely related to the correlation within each limb
  - Apply MLE to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_j$ and τ_j

Nested Logit

1. IIA holds within the limbs

$$\frac{p_{j,k}}{p_{j,t}} = \frac{exp(\boldsymbol{x}'_{j,k}\boldsymbol{\beta}_j/\tau_j)}{exp(\boldsymbol{x}'_{j,t}\boldsymbol{\beta}_j/\tau_j)}$$

2. But, IIA does not hold between the limbs as $\frac{p_{j,k}}{p_{j',t'}}$ depends on all alternatives in nest j and j'.

Application 2

Data description (Cameron and Trivedi)

1. **mode**: 4 mutually exclusive categories (fishing mode)
   - beach (y=1)
   - pier (y=2)
   - private boat (y=3)
   - charter boat (y=4)
2. alternative-varying regressors :
   - price (pbeach, ppier, pprivate and charter)
   - catch rate (qbeach, qpier, qprivate and qcharter)
3. alternative-invarying regressor: income

Application 2

MNL (alternative-invariante regressor)

Table: Estimation results : mlogit

Variable	Coefficient	(Std. Err.)
Equation 1 : beach		
o.income	0.000	(0.000)
o._cons	0.000	(0.000)
Equation 2 : pier		
income	-0.143**	(0.053)
Intercept	0.814**	(0.229)
Equation 3 : private		
income	0.092*	(0.041)
Intercept	0.739**	(0.197)
Equation 4 : charter		
income	-0.032	(0.042)
Intercept	1.341**	(0.195)

Significance levels : † : 10% $*$: 5% $**$: 1%

MNL (marginal effects)

	MEM		AME	
income				
1._predict	0.0000750	(0.00393)	0.000165	(0.00376)
2._predict	-0.0207***	(0.00487)	-0.0208***	(0.00514)
3._predict	0.0326***	(0.00569)	0.0318***	(0.00526)
4._predict	-0.0120*	(0.00608)	-0.0112	(0.00594)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Application 2

CL (alternative-varying regressors)

Table: Estimation results : asclogit

Variable	Coefficient	(Std. Err.)
p	-0.020**	(0.001)
q	0.953**	(0.089)

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

CL (MARGINAL EFFECTS)

- Beach

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	-.002641	.000089	-29.53	0.000	-.002816	-.002465	103.42
charter	.001013	.000034	29.88	0.000	.000947	.001079	84.379
pier	.000439	.00001	43.50	0.000	.00042	.000459	103.42
private	.001188	.000067	17.67	0.000	.001057	.00132	55.257
q							
beach	.122915	.010319	11.91	0.000	.10269	.14314	.24101
charter	-.04715	.005116	-9.22	0.000	-.057177	-.037122	.62937
pier	-.020454	.001623	-12.60	0.000	-.023635	-.017272	.16222
private	-.055312	.00395	-14.00	0.000	-.063054	-.04757	.17121

- Pier

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	.000439	.00001	43.50	0.000	.00042	.000459	103.42
charter	.00094	.000027	35.04	0.000	.000887	.000992	84.379
pier	-.002481	.000083	-29.93	0.000	-.002644	-.002319	103.42
private	.001102	.000064	17.31	0.000	.000978	.001227	55.257
q							
beach	-.020454	.001623	-12.60	0.000	-.023635	-.017272	.24101
charter	-.043739	.004456	-9.82	0.000	-.052472	-.035005	.62937
pier	.115503	.009035	12.78	0.000	.097795	.133211	.16222
private	-.051311	.003303	-15.53	0.000	-.057784	-.044837	.17121

CL (MARGINAL EFFECTS)

- Private

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	.001188	.000067	17.67	0.000	.001057	.00132	103.42
charter	.002541	.000219	11.61	0.000	.002112	.00297	84.379
pier	.001102	.000064	17.31	0.000	.000978	.001227	103.42
private	-.004832	.000336	-14.39	0.000	-.00549	-.004174	55.257
q							
beach	-.055312	.00395	-14.00	0.000	-.063054	-.04757	.24101
charter	-.118281	.013546	-8.73	0.000	-.14483	-.091732	.62937
pier	-.051311	.003303	-15.53	0.000	-.057784	-.044837	.16222
private	.224904	.020701	10.86	0.000	.18433	.265477	.17121

- Charter

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	.001013	.000034	29.88	0.000	.000947	.001079	103.42
charter	-.004494	.000276	-16.26	0.000	-.005036	-.003952	84.379
pier	.00094	.000027	35.04	0.000	.000887	.000992	103.42
private	.002541	.000219	11.61	0.000	.002112	.00297	55.257
q							
beach	-.04715	.005116	-9.22	0.000	-.057177	-.037122	.24101
charter	.209169	.022337	9.36	0.000	.16539	.252949	.62937
pier	-.043739	.004456	-9.82	0.000	-.052472	-.035005	.16222
private	-.118281	.013546	-8.73	0.000	-.14483	-.091732	.17121

Application 2

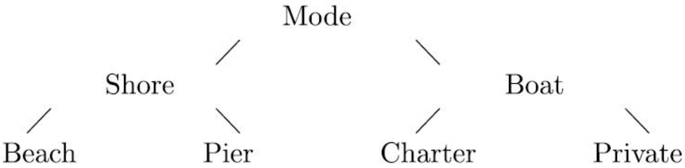
Mixture of MNL and CL

Table: Estimation results : asclogit

Variable	Coefficient	(Std. Err.)
Equation 1 : fishmode		
p	-0.025**	(0.002)
q	0.358**	(0.110)
Equation 2 : charter		
income	-0.033	(0.050)
Intercept	1.694**	(0.224)
Equation 3 : pier		
income	-0.128*	(0.051)
Intercept	0.778**	(0.220)
Equation 4 : private		
income	$0.089^{\dagger}$	(0.050)
Intercept	0.527*	(0.223)

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Application 2 (Nested Logit)



Application 2 (Nested Logit)

Table: Estimation results : nlogit

Variable	Coefficient	(Std. Err.)
	Equation 1 : fishmode	
p	-0.027**	(0.002)
q	1.347**	(0.283)
	Equation 2 : charter	
income	-5.469	(12.490)
Intercept	48.310	(96.493)
	Equation 3 : pier	
income	-6.426	(12.884)
Intercept	39.889	(81.529)
	Equation 4 : private	
income	-1.289	(2.035)
Intercept	28.221	(56.683)
	Equation 5 : shore_tau	
Intercept	55.932	(118.953)
	Equation 6 : boat_tau	
Intercept	32.626	(86.884)

Significance levels : $\dagger$: 10% $\quad *$: 5% $\quad **$: 1%

Even though the model is mathematically correct, with probabilities between 0 and 1 that add up to 1, the fitted model is not consistent with the random utility framework (the dissimilarity parameters "tau" are much greater than 1).

Context and objective

Context

1. You have a sample $\boldsymbol{y} = \{y_1, ..., y_n\}$
2. y_i is in $\mathbb{N}$, often with a large proportion of zeros. The data is usually skewed to the right.
3. You have also observed a vector of characteristics $\boldsymbol{x}_i$ $(K, 1)$ associated to the individual i

Objective

Propose and estimate parametric models accounting for the distribution of y_i conditional on explanatory variables x_i

Examples

Examples

- number of doctor visits
- number of customers
- number of hospital stays
- number of borrowers' defaults
- number of recreational trips
- number of accidents
- ...

Poisson model

Poisson model

1. Natural model: $Y \rightsquigarrow \mathcal{P}(\lambda)$, i.e.

$$\mathbb{P}(Y = k) = \frac{\lambda^k \exp(-\lambda)}{k!}$$

   implying $E(Y) = Var(Y) = \lambda$ (equidispersion property)

2. Standard assumption:

$$Y_i \rightsquigarrow \mathcal{P}(\lambda_i) \quad \text{with} \quad \lambda_i = \exp(\boldsymbol{\theta}' \boldsymbol{x}_i)$$

Poisson model

Estimation

1. Log-likelihood of the sample:

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \Big(y_i \boldsymbol{\theta}' \boldsymbol{x}_i - \exp(\boldsymbol{\theta}' \boldsymbol{x}_i) - \ln(y_i!) \Big)$$

2. F.O.C of of MLE :

$$\sum_{i=1}^{n} \Big(y_i - \exp(\boldsymbol{\theta}' \boldsymbol{x}_i) \Big) \boldsymbol{x}_i = \mathbf{0}$$

Poisson model

Marginal effect and coefficient

1. The sign of marginal effects are the ones of the components of $\boldsymbol{\theta}$

$$\frac{\partial \mathbb{E}(y_i|\boldsymbol{x}_i)}{\partial x_{i,j}} = \theta_j \exp(\boldsymbol{\theta}'\boldsymbol{x}_i)$$

- depends on the considered individual, then AME and MEM

Application (Cameron and Trivedi)

Figure: description of data

variable name	storage type	display format	value label	variable label
docvis	float	%9.0g		**# doctor visits**
private	byte	%8.0g		**=1 if has private supplementary insurance**
medicaid	byte	%8.0g		**=1 if has Medicaid public insurance**
age	byte	%8.0g		**Age**
age2	float	%9.0g		**Age-squared**
educyr	byte	%8.0g		**Years of education**
actlim	byte	%8.0g		**=1 if activity limitation**
totchr	byte	%8.0g		**# chronic conditions**

Application 3

Table: result: poisson model

	Coef.		MEM		AME	
main						
private	0.142***	(0.0143)	0.891***	(0.0896)	0.970***	(0.0980)
medicaid	0.0970***	(0.0189)	0.608***	(0.119)	0.662***	(0.129)
age	0.294***	(0.0260)	1.841***	(0.162)	2.004***	(0.178)
age2	-0.00193***	(0.000172)	-0.0121***	(0.00108)	-0.0132***	(0.00118)
educyr	0.0296***	(0.00188)	0.185***	(0.0117)	0.202***	(0.0129)
actlim	0.186***	(0.0146)	1.168***	(0.0911)	1.272***	(0.0997)
totchr	0.248***	(0.00464)	1.557***	(0.0280)	1.695***	(0.0334)
_cons	-10.18***	(0.972)				
N	3677		3677		3677	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Limit

Limit

1. The problem of Poisson model is that the distribution of y_i conditional on $\boldsymbol{x}_i$ depends on a single parameter λ_i, so cannot account for the fact that data often :
   - suggest $\mathbb{E}(y_i) < V(y_i)$ (overdispersion)
   - contain important proportion of zero
2. Generalizations: negative binomial model, hurdle Models, zero-inflated models, etc...

Good Luck!