Econométrie 3

Multiple Choices Modelling

Iragaël Joly

Grenoble-Inp, Génie Industriel - Gael

iragael.joly@grenoble-inp.fr

Nov. 2018

Table of contents

Multiple Choice Model

Choice between multiple alternatives (1)

Choice between multiple alternatives

- Utility is defined for all $i \in \mathcal{C}_n$: $U_{in} = V_{in} + \varepsilon_{in}$
- We need to define
  - Choice set : $\mathcal{C}_i n$
  - Error term : ε_{in}
  - Systematic part : V_{in}

Choice set (1)

Choice set

- How to model "awareness" ?
- What does "long distance" exactly mean ?
- What does "unreachable" exactly mean ?

We assume here deterministic rules

- Car is available if *n* has a driver license and a car is available in the household
- Walking is available if trip length is shorter than 4km.

Error term (1)

Main assumption

ε_{in} are

- extreme value $EV(0, \mu)$,
- independent and
- identically distributed (i.i.d.).

Comments

- Independence : across *i* and *n*.
- Identical distribution : same scale parameter μ across *i* and *n*.
- Scale must be normalized : $\mu = 1$.

Derivation of the logit model (1)

Assumptions

- $C_n = \{1, \ldots, J_n\}$
- $U_{in} = V_{in} + \varepsilon_{in}$
- $\varepsilon_{in} \sim EV(0, \mu)$
- ε_{in} i.i.d.

Choice model

$$P(i|\mathcal{C}_n) = Pr(V_{in} + \varepsilon_{in} \geq max_{j=1,\ldots,J_n} V_{jn} + \varepsilon_{jn})$$

Assume without loss of generality that $i = 1$

$$P(1|\mathcal{C}_n) = P(V_{1n} + \varepsilon_{1n} \geq max_{j=2,\ldots,J_n} V_{jn} + \varepsilon_{jn})$$

Derivation of the logit model (2)

Composite alternative

- Define a composite alternative : "anything but alternative one"
- Associated utility :

$$U^* = max_{j=2,\dots,J_n}(V_{jn} + \varepsilon_{jn})$$

- From a property of the EV distribution

$$U^* \sim EV\left(\frac{1}{\mu} ln \sum_{j=2}^{J_n} exp\{\mu V_{jn}\}, \mu\right)$$

Derivation of the logit model (3)

Composite alternative

From another property of the EV distribution

$$U^* = V^* + \varepsilon^*$$

where

$$V^* = \frac{1}{\mu} ln \sum_{j=2}^{J_n} exp\{\mu V_{jn}\}$$

and

$$\varepsilon^* \sim EV(0, \mu)$$

Derivation of the logit model (4)

Binary choice

$$P(1|\mathcal{C}_n) = P(V_{1n} + \varepsilon_{1n} \geq max_{j=2,\ldots,J_n} V_{jn} + \varepsilon_{jn})$$

$$P(V_{1n} + \varepsilon_{1n} \geq V^* + \varepsilon^*)$$

ε_{1n} and ε^* are both $EV(0,\mu)$.

Binary logit

$$P(1|\mathcal{C}_n) = \frac{exp\{\mu V_{1n}\}}{exp\{\mu V_{1n}\} + exp\{\mu V^*\}}$$

where

$$V^* = \frac{1}{\mu} ln \sum_{j=2}^{J_n} exp\{\mu V_{jn}\}$$

Derivation of the logit model (5)

We have

$$exp\{\mu V^*\} = exp\{ln \sum_{j=2}^{J_n} exp\{\mu V_{jn}\}\}$$

and

$$\begin{aligned} P(1|\mathcal{C}_n) &= \frac{exp\{\mu V_{1n}\}}{exp\{\mu V_{1n}\} + exp\{\mu V^*\}} \\ &= \frac{exp\{\mu V_{1n}\}}{exp\{\mu V_{1n}\} + \sum_{j=2}^{J_n} exp\{\mu V_{jn}\}} \\ &= \frac{exp\{\mu V_{1n}\}}{\sum_{j=1}^{J_n} exp\{\mu V_{jn}\}} \end{aligned}$$

Derivation of the logit model (1)

Assumptions

- The scale parameter μ is not identifiable : $\mu = 1$.
- Warning : not identifiable $\neq$ not existing

Derivation of the logit model (2)

$\mu \to 0$, that is variance goes to infinity

$$lim_{\mu \to 0} P(i|\mathcal{C}_n) = \frac{1}{J_n} \quad \forall i \in \mathcal{C}_n$$

$\mu \to +\infty$, that is variance goes to zero

$$\begin{aligned} lim_{\mu \to +\infty} P(i|\mathcal{C}_n) &= lim_{\mu \to +\infty} \frac{1}{1 + \sum_{j \neq i} exp\{\mu(V_{jn} - V_{in})\}} \\ &= 1 \text{if} V_{in} > max_{j \neq i} V_{jn} \\ &= 0 \text{if} V_{in} < max_{j \neq i} V_{jn} \end{aligned}$$

Systematic part of the utility function

Systematic part of the utility function (1)

$$V_{in} = V(z_{in}, S_n)$$

- z_{in} is a vector of attributes of alternative i for individual n
- S_n is a vector of socio-economic characteristics of n

Functional form : linear utility

$$x_{in} = (z_{in}, S_n)$$

Linear-in-parameters utility functions

$$V_{in} = V(z_{in}, S_n) = V(x_{in}) = \sum_k \beta_k (x_{in})_k$$

Explanatory variables : alternatives attributes (1)

Alt. Attributes : Numerical and continuous

$$(z_{in})_k \in \mathbb{R}, \forall i, n, k$$

Associated with a specific unit

Examples

- Auto in-vehicle time (in min.)
- Auto out-of-pocket cost (in cents)
- Walking time to the bus stop (in min.)

Explanatory variables : alternatives attributes (2)

V_{in} is unitless

- Therefore, β depends on the unit of the associated attribute
- Example : consider two specifications

$$V_{in} = \beta_1 TT_{in} + \dots \qquad \text{or} \qquad V_{in} = \beta_1' TT_{in}' + \dots$$

- If TT_{in} is a number of minutes, the unit of β_1 is 1/min
- If TT_{in}' is a number of hours, the unit of β_1' is 1/hour
- Both models are equivalent, but the estimated value of the coefficient will be different

$$\beta_1' TT_{in}' = \beta_1 TT_{in} \Rightarrow \frac{TT_{in}'}{TT_{in}} = \frac{\beta_1}{\beta_1'} = 60$$

Explanatory variables : alternatives attributes (3)

Generic and alternative specific parameters

Modeling assumption : a minute has/has not the same marginal utility

$$V_{auto} = \beta_1 TT_{auto} + \dots \qquad \text{and} \qquad V_{bus} = \beta_1 TT_{bus} + \dots$$

or

$$V_{auto} = \beta_1 TT_{auto} + \dots \qquad \text{or} \qquad V_{bus} = \beta_2 TT_{bus} + \dots$$

Explanatory variables : socio-eco. characteristics (1)

Socio-eco. Characteristics : Numerical and continuous

- $(S_n)_k \in \mathbb{R}, \forall n, k$
- Associated with a specific unit
- Examples : Annual income ; Age (in years) ; etc

S_n do not depend on alternative i

Explanatory variables : socio-eco. characteristics (2)

Numerical and continuous

They are constant between alternative,

- hence they cannot be associated with the same effect in all utility functions

$$\begin{aligned} V1 &= \beta_1 x_{11} + \beta_2 income + \dots & & V_1' = \beta_1 x_{11} \\ V2 &= \beta_1 x_{21} + \beta_2 income + \dots & \Leftrightarrow \quad & V_2' = \beta_1 x_{21} \\ V3 &= \beta_1 x_{31} + \beta_2 income + \dots & & V_3' = \beta_1 x_{31} \end{aligned}$$

- In general : alternative specific characteristics

$$\begin{aligned} V_1 &= \beta_1 x_{11} + \beta_2 income + \beta_4 age \\ V_2 &= \beta_1 x_{21} + \beta_3 income + \beta_5 age \\ V_3 &= \beta_1 x_{31} \end{aligned}$$

Discrete variables (1)

Discrete variables

- Mainly used to capture qualitative attributes or characteristics
  Level of comfort for the train, Color, etc ; Sex, Education, etc.
- Carrefull model specification :
  - If a qualitative attribute has n levels, we introduce $n-1$ dummies variables (0/1) in the model
  - Introduce a 0/1 attribute for all levels except the base case

Discrete variables (2)

Discrete variables

- For example :

  Base : level 1

$$V_{in} = \cdots + 0 z_{i,lev1} + \beta_{lev2} z_{i,lev2} + \beta_{lev3} z_{i,lev3}$$

  - β_{lev2} : difference of utility between level 2 and reference level (level 1)
  - β_{lev3} : difference of utility between level 3 and reference level (level 1)
  - if levels are ordered, signs are in coherence

- **Base** : level 2

$$V_{in} = \cdots + \beta_{lev1} z_{i,lev1} + 0 z_{i,lev2} + \beta_{lev3} z_{i,lev3}$$

Nonlinear transformations of the variables (1)

Nonlinear transformations of the variables

- Introduction of variable in linear additive utiliy function leads to implicit behavioural assumptions
- Example with travel time
  Behavioral assumption : One more minute of travel is not perceived the same way for short trips as for long trips

Nonlinear transformations of the variables (2)

- Assumption 1 : the marginal impact of travel time is constant

$$V_i = \beta_T time_i + \ldots$$

- Assumption 2 : the marginal impact of travel time decreases with travel time

$$V_i = \beta_T ln(time_i) \ldots$$

- It is still a linear-in-parameters form
- The unit, the value, and the interpretation of β_T are different

Nonlinear transformations of the variables (3)

- Data can be preprocessed to account for nonlinearities

$$V_{in} = V(h(z_{in}, S_n)) = \sum_k \beta_k (h(z_{in}, S_n))_k$$

- It is linear-in-parameter, even with h nonlinear.

Specifications

- Categories with constants (inferior solution)
- Piecewise linear specification (spline)
- Numerous tranforms (Box-Cox transforms)

Categories with constant (1)

Nonlinear transformations of the variables

Same specification as for discrete variables

$$V_i = \beta_{T1} x_{T1} + \beta_{T2} x_{T2} + \beta_{T3} x_{T3} + \beta_{T4} x_{T4} + \dots$$

with

- $x_{T1} = 1$ if $TT_i \in [0; 90[$, 0 otherwise
- $x_{T2} = 1$ if $TT_i \in [90; 180[$, 0 otherwise
- $x_{T3} = 1$ if $TT_i \in [180; 270[$, 0 otherwise
- $x_{T4} = 1$ if $TT_i \in [270; +\infty[$, 0 otherwise
- One β must be normalized to 0.

Categories with constant (2)

Drawbacks

- No sensitivity to travel time within the intervals
- Discontinuous utility function (jumps)
- Need for many small intervals
- Results may vary significantly with the definition of the intervals

Appropriate when

- Categories have been used in the survey (income, age)
- Definition of categories is natural (weekday)

Piecewise linear specification (1)

Features

- Capture the sensitivity within the intervals
- Enforce continuity of the utility function

$$V_i = \beta_{T1} x_{T1} + \beta_{T2} x_{T2} + \beta_{T3} x_{T3} + \beta_{T4} x_{T4} + \ldots$$

where x_{Ti} is t in four classes of range 90 $(0; 90; 180; 270)$

$$x_{T1} = \begin{cases} t & \text{if } t < 90 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T2} = \begin{cases} 0 & \text{if } t < 90 \\ t - 90 & \text{if } 90 \leq t < 180 \\ 90 & \text{otherwise} \end{cases}$$

$$x_{T3} = \begin{cases} 0 & \text{if } t < 180 \\ t - 180 & \text{if } 180 \leq t < 270 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T4} = \begin{cases} 0 & \text{if } t < 270 \\ t - 270 & \text{otherwise} \end{cases}$$

Piecewise linear specification (2)

Coding :

$$x_{T1} = min(t, 90)$$
$$x_{T2} = max(0, min(t - 90, 90))$$
$$x_{T3} = max(0, min(t - 180, 90))$$
$$x_{T4} = max(0, t - 270)$$

t	TT1	TT2	TT3	TT4
40	40	0	0	0
100	90	10	0	0
200	90	90	20	0
300	90	90	90	30

Box-Cox transforms (1)

Box-Cox transforms

Other power transforms are possible :

- Box-Cox, Box-Tukey, Taylor expansion,...

Interactions (1)

Motivation

- All individuals in a population are not alike
- Socio-economic characteristics define segments in the population
- How to capture heterogeneity ?

The $h()$ transform can interact attributes and characteristics : cost/income, distance/time,...

$$V_{in} = V(h(z_{in}, S_n)) = \sum_k \beta_k (h(z_{in}, S_n))_k$$

Segmentation (1)

The population is divided into a finite number of segments

- Each individual belongs to exactly one segment

Specification

$$\beta_{M,m} TT_{M,m} + \beta_{M,s} TT_{M,s} + \beta_{M,p} TT_{M,p}$$

$$+\beta_{F,m} TT_{F,} m + \beta_{F,s} TT_{F,s} + \beta_{F,p} TT_{F,p}$$

$TT_i = TT$ if indiv. belongs to segment i , and 0 otherwise

- For a given individual, exactly one of these terms is non zero.
- The number of segments grows exponentially with the number of variables.

Heteroscedasticity (1)

Heteroscedasticity

- Logit is homoscedastic : ε_{in} i.i.d. across both i and n.
- But for ex. : people have different level of knowledge
- Different sources of data

How can we specify the model in order to use logit ?

Heteroscedasticity (2)

Assumption

- Variance of error terms is different across two different groups :

$$U_{in1} = V_{in1} + \varepsilon_{in1}$$

$$U_{in2} = V_{in2} + \varepsilon_{in2}$$

and $Var(\varepsilon_{in2}) = \alpha^2 Var(\varepsilon_{in1})$

- Solution : include scale parameters

$$\alpha U_{in1} = \alpha V_{in1} + \alpha\varepsilon_{in1} = \alpha V_{in1} + \varepsilon'_{in1}$$

$$U_{in2} = V_{in2} + \varepsilon_{in2} = V_{in2} + \varepsilon'_{in2}$$

where ε'_{in1} and ε'_{in2} i.i.d.

Heteroscedasticity (3)

Assumption

- $\alpha V_{in1} = \sum_j \alpha\beta_j x_{jin1}$ is not linear-in-parameters
- Normalization : a different scale parameter can be estimated for each segment of the population, except one that must be normalized.

Table of contents

Relaxing the independence assumption

The independence from irrelevant alternatives (IIA) Property (1)

The IIA Property

- For any two alternatives i and k, the ratio of the logit probabilities is

$$\frac{P(i|\mathcal{C})}{P(k|\mathcal{C})} = \frac{\frac{exp\{V_{in}\}}{\sum_{j=1}^{J_n} exp\{V_{jn}\}}}{\frac{exp\{V_{kn}\}}{\sum_{j=1}^{J_n} exp\{V_{jn}\}}} = \frac{exp\{V_{in}\}}{exp\{V_{kn}\}} = exp\{V_{in} - V_{kn}\}$$

- these relative odds are independent from ’irrelevant’ alternatives

Blue Bus Red Bus Paradox (1)

Imagine a mode choice for commuting : car or blue bus.

- Two alternatives : car and bus.
- There are red buses and blue buses.
- Car and bus travel times are equal : T.
- Only travel time is considered in the utility function.

Blue Bus Red Bus Paradox (2)

Imagine a mode choice for commuting : car or blue bus.

- For simplicity we assume that the representative utility of the two modes are the same

$$U_{car} = \beta_T T + \varepsilon_{car} \qquad U_{bus} = \beta_T T + \varepsilon_{bus}$$

- Then
$P(\text{car}|\{\text{car, bus}\}) = P(\text{bus}|\{\text{car, bus}\}) = \frac{exp\{\beta_T T\}}{exp\{\beta_T T\}+exp\{\beta_T T\}} = \frac{1}{2}$
- Then $\frac{P(\text{car})}{P(\text{bus})} = 1$

Blue Bus Red Bus Paradox (3)

introduction of a red bus

- Suppose introduction of a red bus (considered to be exactly like the blue bus)

$$U_{car} = \beta_T T + \varepsilon_{car} \qquad U_{bluebus} = \beta_T T + \varepsilon_{bluebus}$$

$$U_{redbus} = \beta_T T + \varepsilon_{redbus}$$

- Then

$$P(\text{car}|\{\text{car, blue bus, red bus}\}) = \frac{1}{3}$$
$$P(\text{blue bus}|\{\text{car, blue bus, red bus}\}) = \frac{1}{3}$$
$$P(\text{red bus}|\{\text{car, blue bus, red bus}\}) = \frac{1}{3}$$

- Then $\frac{P(\text{car})}{P(\text{bus})} \neq 1$

Blue Bus Red Bus Paradox (1)

introduction of a red bus

- With IIA : the logit model predicts :
  $P(\text{red bus}) = P(\text{blue bus}) = P(\text{car}) = 1/3$
- $\frac{P(\text{red bus})}{P(\text{blue bus})} = 1$ and $\frac{P(\text{car})}{P(\text{blue bus})} = 1$
- we would expect :$P(\text{red bus}) = P(\text{blue bus}) = 1/4$ and $P(\text{car}) = 1/2$
- Logit overestimates $P(\text{bus})$ and underestimates $P(\text{car})$

Explaining the paradox

- Shared attributes are captured by the error terms ($\varepsilon_{bluebus}\varepsilon_{redbus}$) : fare, comfort, etc.
- Logit model assumes that $\varepsilon_{bluebus}$ and ε_{redbus} are independent.

Interpreting the paradox (1)

Proportional substitution

- The cross-elasticities of logit probabilities :
  Consider the effect of changing an attribute of alt. j on the probabilities for all the other alternatives.
- Deriving $P(i|\mathcal{C})$:

$$E_{iz_{nj}} = \frac{\partial ln(Pi|\mathcal{C})}{\partial ln(z_{nj})} = -\beta_z z_{nj} P(j|\mathcal{C})$$

- This cross-elasticity is the same for all i.

Interpreting the paradox (2)

- Improvement in one alternative draws proportionaly the other alternatives.
- If one probability drops of x % then all probabilities drop of x %, except for alternative j.
- Behavioral implication of IIA :
  If an alternative is removed, probability is spread equally across the remaining alternatives.

Discrete Choice Model Extensions

Discrete Choice Model Extensions (1)

Discrete Choice Model Extensions

- Heteroscedasticity and other forms of heterogeneity
  - Across individuals
  - Across alternatives
- Panel data (Repeated measures)
  - Random and fixed effects models
  - Building into a multinomial logit model
- The nested logit model
- Latent class model
- Mixed logit, error components and multinomial probit models

Ordinal Logistic Regression

Background

Ordinal Response Y

- represents levels of a standard measurement scale
- constructed by specifying a hierarchy of separate endpoints
- rank-based methods to continuous responses
- do not assume a spacing between levels of Y
- ordinal models use only the rankordering of values of Y
- same β and P-values result from an analysis of Y having levels 0, 1, 2 and levels recoded 0, 1, 20

Proportional Odds Model

***PO* model :**

- for a response variable having levels $j = 1, 2, \ldots, k$:

$$P(Y \geq j|X) = \frac{1}{1 + exp[-(\alpha_j + X\beta)]}$$

- k intercepts (αs)
- For fixed j, the model logistic model for the event $Y \geq j$
- common vector of β connecting probabilities for varying j (parsimonious modeling of the distribution of Y)
- implicit assumption : the β are independent of j, the cutoff level for Y
- the log odds that $Y \geq j$ is linearly related to each X and that there is no interaction between the Xs

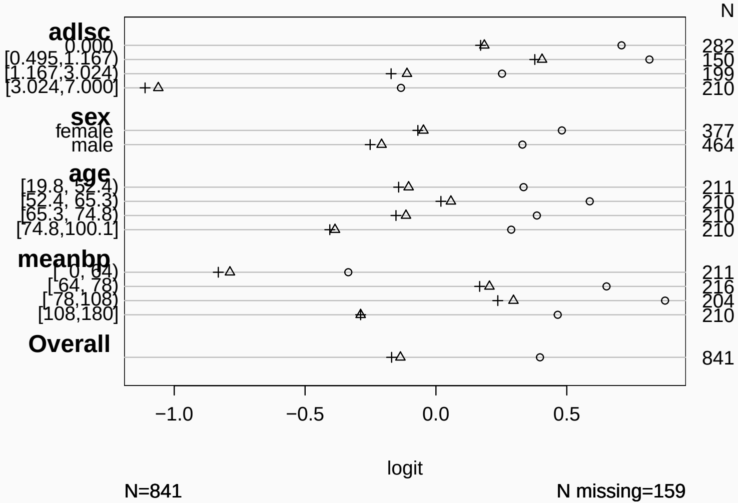
***PO* model :**

- a single odds ratio apply equally to all events $Y \geq j$, $j = 1, 2, \ldots, k$
- If *linearity* and *additivity* hold, the $X_m + 1 : X_m$ odds ratio for $Y \geq j$ is $exp(\beta_m)$, whatever the cutoff j
- BUT, proportional hazards assumption is frequently violated

Assessment of Model Fit (1)

Determining whether the PO assumption is likely

- comparing means of $X|Y$ with and without assuming PO,
- stratifying on each predictor and computing the logits of all proportions of the form $Y \geq j$, $j = 1, 2, \ldots, k$.
  When PO holds, the differences in logits between different values of j should be the same at all levels of X, because the model dictates that $logit(Y \geq j|X) - logit(Y \geq i|X) = \alpha_j - \alpha_i$, for any constant X.


N
adlsc
0.000
[0.495,1.167)
[1.167,3.024)
[3.024,7.000]
sex
female
male
age
[19.8, 52.4)
[52.4, 65.3)
[65.3, 74.8)
[74.8,100.1]
meanbp
[0, 64)
[64, 78)
[78,108)
[108,180]
Overall
282
150
199
210
377
464
211
210
210
210
211
216
204
210
841
−1.0
−0.5
0.0
0.5
logit
N=841
N missing=159

- Checking PO assumption separately for a series of predictors. The circle, triangle, and plus sign correspond to $Y \geq 1, 2, 3$, respectively
- PO is checked by examining the vertical constancy of distances between any two of these three symbols.

Continuation Ratio Model

***CR* model :**

- PO model is based on *cumulative probabilities*
  CR model is based on *conditional probabilities*
- the forward CR model :

$$P(Y = j | Y \geq j, X) = \frac{1}{1 + exp[-(\theta_j + X\gamma)]}$$

$$logit(Y = 0 | Y \geq 0, X) = logit(Y = 0 | X) = \theta_0 + X\gamma$$

$$logit(Y = 1 | Y \geq 1, X) = \theta_1 + X\gamma$$

...

$$logit(Y = k-1 | Y \geq k-1, X) = \theta_{k-1} + X\gamma$$

Exercise (1)

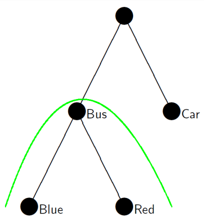
Exercise MROZ - Woolridge

- Reproduce results from the textbook

Table of contents

Nested Logit

What about the choice between bus and car ?



$$U_{car} = \beta T + \varepsilon_{car}$$
$$U_{bus} = V_{bus} + \varepsilon_{bus}$$
$$\text{with } V_{bus} = V_{bus}(V_{bluebus}, V_{redbus})$$
$$\varepsilon_{bus} = ?$$

- Use a logit model at the higher level.
- Define V_{bus} as the expected maximum utility of red bus and blue bus
- Within a branch

Capturing the correlation (2)

  - Identical variances (IIA applies)
  - Covariance (all same) = variance at higher level
- Branches have different variances (scale factors)
- Nested logit probabilities : Generalized Extreme Value
- $P(Alt, Branch) = P(Branch) \times P(Alt|Branch)$

Definition

For a set of alternative $\mathcal{C}$, $U_{k|b}$ defines utility of alt. k in branch b

$$U_{k|b} = \alpha_{k|b} + \beta' x_{k|b}$$

and at branch level

$$U_b = \gamma_b' z_i$$

Capturing the correlation (3)

- Leaf probability : $P(k|b) = \frac{exp\{\alpha_{k|b}+\beta'_{k|b}x_{k|b}\}}{\sum_{m=1}^{K|b} exp\{\alpha_{m|b}+\beta'_{m|b}x_{k|b}\}}$
- Inclusive value for branch b : $IV(b) = ln(\sum_{m=1}^{K|b} exp\{\alpha_{m|b} + \beta'_{m|b}x_{k|b}\})$
- Branch level probability :

$$P(b) = \frac{exp\{\lambda_b(\gamma'_b z_i + IV(b))\}}{\sum_{q=1}^{Q} exp\{\lambda_q(\gamma'_q z_i + IV(q))\}}$$

- MNL results if the inclusive parameters, λ_q are all equal to one.

Nested logit version 2 (1)

The nested logit model is obtained

by assuming that the vector of unobserved utility, ε_n has cumulative distribution :

$$exp\{-\sum_{k=1}^{K}(\sum_{j\in B_k} exp\{\varepsilon_{nj}/\lambda_k\})^{\lambda_k}\}$$

- This distribution is a type of GEV distribution.
- A generalization of the distribution that gives rise to the logit model.
- For logit, each ε_{nj} is independent with a univariate extreme value distribution.

Nested logit version 2 (2)

- For this GEV, the marginal distribution of each ε_{nj} is univariate extreme value.
- However, the ε_{nj} are correlated within nests.
  - For any two alternatives j and m in nest B_k, ε_{nj} is correlated with ε_{nm}.
  - For any two alternatives in different nests, the unobserved portion of utility is still uncorrelated : $Cov(\varepsilon_{nj}, \varepsilon_{nm}) = 0$ for any $j \in B_k$ and $m \in B_l$ with $l \neq k$.
- The parameter λ_k is a measure of the degree of independence in unobserved utility among the alternatives in nest k.
- A higher value of λ_k means greater independence and less correlation.

Nested logit version 2 (3)

- $\lambda_k = 1$ leads to the logit model

Choice probability for alternative $i \in B_k$

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_n} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_n} e^{V_{nj}/\lambda_k}\right)^{\lambda_k}}$$

Mixed Logit

Mixed Logit

Mixed logit model advantages

ML resolves the three limitations of standard logit by allowing :

1. random taste variation
   - some parameters are assumed to be randomly distributed
2. unrestricted substitution patterns
   - relaxes IIA hypothesis of MNL by allowing ε to be correlated,
   - while maintaining the assumption of identical distribution *(Greene, 2008)* .
3. correlation in unobserved factors over time
   - takes into account the repeated nature of the choices made by respondents. (Var-Cov matrix form)

(McFadden and Train, 2000) : any discrete choice model can be approximated by a mixed logit model, including the multinomial logit.

Mixed Logit (1)

Mixed logit and error component models

- Basic form of the MNL model, with alternative specific constants α_{ij} and attributes x_{ij} :

$$P(j) = \frac{exp\{\alpha_{ij} + \beta_i' x_{ij}\}}{\sum_{q=1}^{J_i} exp\{\alpha_{iq} + \beta_i' x_{iq}\}}$$

- The random parameters model emerges as the form of the individual specific parameter vector, β_i, is developed.
- The most familiar, simplest version of the model :

$$\beta_{ik} = \beta_k + \sigma_k \nu_{ik}$$

$$\alpha_{ij} = \alpha_j + \sigma_j \nu_{ij}$$

Mixed Logit (1)

Mixed logit and error component models

- Introducing random taste heterogeneity in the parameters the probability that an alternative j is chosen by an individual i, is the integral of the following probability over all possible values of β :

$$P_i(j) = \int P_i(j|x_{ij}, \beta) f(\beta|\Omega) d\beta$$

  where $f(\beta|\Omega)$ is the conditional density function of the β parameters and Ω is a vector of parameters of the distribution of β.

  - error components (EC) version, $U_{ij} = \alpha' x_{ij} + \theta_i' z_{ij} \varepsilon_{ik}$
  - random coefficients (RC) structure $U_{ij} = \theta_{ij}' x_{ij} + \varepsilon_{ij}$

Mixed Logit (1)

- where β_k is the population mean
- ν_{ik} is the individual specific heterogeneity, with zero mean and s.d. one
- σ_k is s.d. of the distribution of β_{ik} around β_k
- The choice-specific constants, α_{ij}, and the elements of β_i are distributed randomly across individuals with fixed means
- Normal distribution or other may be specified.
  for ex : $\beta_{ik} \sim \mathcal{N}(b, W)$ or $ln(\beta) \sim \mathcal{N}(b, W)$

Discret Choice and Stated Preferences

Many Data about Preference

Data about Preference

Relying on what consumers say they will do or observing what they actually do.

- Revealed Preference (RP) data : *Market data*

  (Samuelson, 1948) demonstrate that under some conditions, system demand equations can be estimated consistently...
- Stated Preference data : mainly hypothetical choice situation surveys.

Many Data about Preference

SP data pro

- Estimate demand for new products, or with new attributes or new features

  the 3.5” disk example
- Explanatory variables have little variability in the marketplace

  RP are of limited use to model behavioral change in response to variables change.
- Explanatory variables are higly collinear in the marketplace

  Attributes of products should become negatively correlated in competitive market

  Technological constraints may leads correlation betweeen attributes
- Observational data cannot satisfy model assumptions and/or contain statistical ’nasties’

  SP data can be designed to eliminate such problem

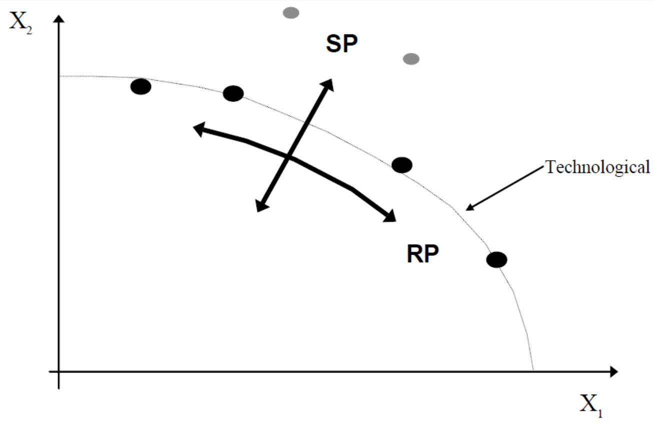


FIGURE 1 – RP and SP studying techno. frontier *(Louviere et al., 2000)*

Many Data about Preference (1)

RP and SP Advantages

RP data

- observation of market (current equilibrium)
- inherent relationships between attributes (techno. constraints)
- only existing alternatives are observable
- embody market and personal constraints on the decision maker
- have high reliability and face validity
- mainly one observation per individual at each observation point.

Many Data about Preference (2)

SP data

- hypothetical or virtual decision contexts
- experimentalist controls relationships between attributes (mapping of utility functions outside the existing technologies)
- can include existing and/or proposed and/or generic (i.e., unbranded or unlabelled) choice alternatives,
- cannot (or hardly) represent changes in market and personal constraints effectively,
- seem to be reliable when respondents understand, are committed to and can respond to tasks,
- (usually) yield multiple observations per respondent at each observation point.

Preference data and RUT (1)

"[...]measurement in the absence of theory is at best uninterpretable, and at worst meaningless." *(Louviere et al., 2000)*

Measurement in choices and preference modelling

- of preferences and choice
- of attributes
- of consumers or decision-making units
- of decision environments.

Preference data and RUT (1)

Measurement of choices and preference

- measures called dominance measures
- one or more objects are indistinguishable in degree of preference from one another (i.e., equal) or are more/less preferred to one another
- Many types of dominance measures are (or can be transformed to be) consistent with RUT

Some examples

- Discrete choice of one option from a set of competing ones. Most preferred option no relative preferences among the non-chosen.

Preference data and RUT (2)

- Yes/No : ‘Yes, I like this option’ and ‘No, I do not like this option’.
- Complete ranking of options from most to least preferred.
- Degrees of preference for each option. rate on a scale or psychometric methods
- Allocation of some of resources, such as money, time, trips, chips...

Discrete Choice (1)

Discrete Choice

- one option among competing (exclusive) options
- options differ in attributes
- 'weakly ordered' data : complete preference ordering cannot be determined
- to know complete ordering or change in response : need of more responses (same and/or other individuals)
- see example

Discrete Choice

auto > bus, train, ferry, carpool

Table 2.1. *Discrete choice of commuting option*

'Brands' for journey to work	Consumer chooses
Take bus	
Take train	
Take ferry	
Drive own auto	✓
Carpool	

FIGURE 2 – *(Louviere et al., 2000)*

Yes / No response (1)

Yes / No response

- binary discrete response
- categorises into two groups
- 'weakly ordered' data to be completed
- example can be obtained by asking to consider the 5 options or by 5 days observation

Yes / No response

auto > bus, train, ferry,

carpool > bus, train, ferry

Table 2.2. *Acceptance or rejection of commuting options*

'Brands' for journey to work	Consumer will consider (y/n)
Take bus	no
Take train	no
Take ferry	no
Drive own auto	yes
Carpool	yes

FIGURE 3 – *(Louviere et al., 2000)*

Complete ranking of options (1)

Complete ranking

- some issues : task difficulty ; response reliability ; realibility and validity for option that would never be chosen or option that are not well known
- best ranked option may be not chosen
- 'strongly ordered' data, but *(Louviere et al., 2000)* advise not to use them
- example : asking or observing

Complete ranking

auto > carpool > ferry > train > bus

Table 2.3. *Complete preference ranking of commuting options*

'Brands' for journey to work	Ranking by likelihood of use
Take bus	5
Take train	4
Take ferry	3
Drive own auto	1
Carpool	2

FIGURE 4 – *(Louviere et al., 2000)*

Degrees of preference (1)

Degrees of preference

- Most popular : category rating scale
- Report preference differences or ratios, or...
- Rating leads to ordinal information
- but cardinal interpretation have strong behavioral assumptions
- example : asking

Degrees of preference

auto > carpool > ferry > train = bus

Table 2.4. *Scale rating of commuting options*

'Brands' for journey to work	Consumer likelihood to use (0–10)
Take bus	4
Take train	4
Take ferry	6
Drive own auto	10
Carpool	7

FIGURE 5 – *(Louviere et al., 2000)*

Ressource allocation (1)

Ressource allocation

- ’stronger order’
- behavioral assumptions relative to the resource, the ’numeraire’

Data provided (1)

Implied choice data provided by each response

- For each choice set faced by each consumer, one must identify the chosen and rejected option
- Analogue to RP data : one *observes* some preference (choice) from a sample of individuals, *measures* (observes) attributes associated with choice alternatives, *measures* characteristics of the individual choosers

Data provided

Table 2.5. *Creating choice sets and coding choices from response data*

Implied choice set	Alternative	Implied choice
Discrete choice		
1	Auto	1
1	Bus	0
1	Train	0
1	Ferry	0
1	Carpool	0
Yes/No		
1	Auto	1
1	Bus	0
1	Train	0
1	Ferry	0
1	Carpool	0
2	Auto	0
2	Bus	0
2	Train	0
2	Ferry	0
2	Carpool	1

FIGURE 6 – *(Louviere et al., 2000)*

Data provided

Complete ranking		
1	Auto	1
1	Bus	0
1	Train	0
1	Ferry	0
1	Carpool	0
2	Bus	0
2	Train	0
2	Ferry	0
2	Carpool	1
3	Bus	0
3	Train	0
3	Ferry	1
4	Bus	0
4	Train	1
Rating		
1	Auto	1
1	Bus	0
1	Train	0
1	Ferry	0

Références

Greene, W. (2008). *Econometric Analysis, 6th*. Prentice-HallOxford : Clarendon Press edition.

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated Choice Methods : Analysis and Application*. Cambridge University Press, New York, NY, USA.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of of Applied Econometrics*, 64 :207–240.

Samuelson, P. (1948). *Foundations of economic analysis*. Cambridge, Mass :McGraw-Hill edition.

Train, K. (2009). *Discrete choice methods with simulation (2nd ed.)*. UK : Cambridge University Press, Cambridge edition.