Lecture Note

TRANSPORT PLANNING AND MODELING

--- MODAL SPLIT MODEL 02 ---

Dr.Eng. Muhammad Zudhy Irawan, S.T., M.T.
Master Programme in Transportation System and Engineering, Gadjah Mada University

DISAGGREGATE CASE

- Generally known as discrete choice
- The calculation method is as follows:

$$\Pr\left(V_{nm} > V_{nb}\right) = \frac{\exp\left(V_{nm}\right)}{\exp\left(V_{nm}\right) + \sum_{b \neq m} \exp\left(V_{nb}\right)}$$

1. Disaggregate demand models (DM) are based on theories of individual behaviour and do not constitute physical analogies of any kind. Therefore, as an attempt is made to explain individual behaviour, an important potential advantage over aggregate models is that it is more likely that DM models are stable (or transferable) in time and space.
2. DM models are estimated using individual data and this has the following implications:
   - DM models may be more efficient than aggregate models in terms of information usage; fewer data points are required as each individual choice is used as an observation. In aggregate modelling one observation is the average of (sometimes) hundreds of individual observations.
   - As individual data are used, all the inherent variability in the information can be utilised.
   - DM models may be applied, in principle, at any aggregation level; however, although this appears obvious, the aggregation processes are not trivial
   - DM models are less likely to suffer from biases due to correlation between aggregate units. A serious problem when using aggregate information is that individual behaviour may be hidden by unidentified characteristics associated with the zones; this is known as *ecological correlation*.

The example in Figure 7.1 shows that if a trip generation model was estimated using zonal data, we would obtain that the number of trips decreases with income; however, the opposite would be shown to hold if the data were considered at a household level.
This phenomenon, which is of course exaggerated in the figure, might occur for example if the land-use characteristics of zone B are conducive to more trips on foot.

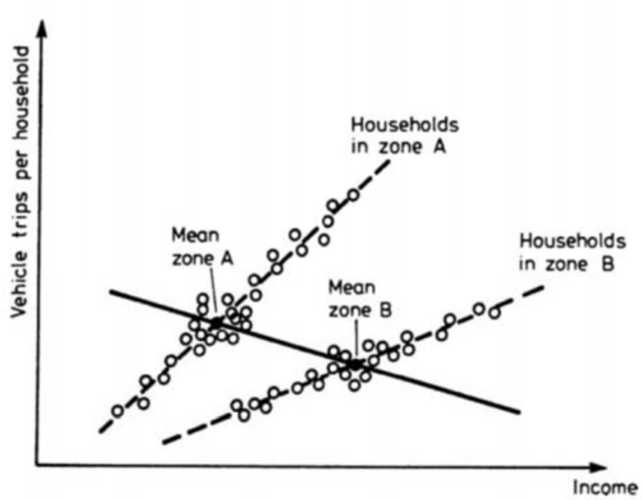


Figure 7.1 Example of ecological fallacy

3. Disaggregate models are probabilistic; furthermore, as they yield the probability of choosing each alternative and do not indicate which one is selected, use must be made of basic probability concepts such as
   - The expected number of people using a certain travel option equals the sum over each individual of the probabilities of choosing that alternative:

$$N_i = \sum_n P_{in}$$

   - An *independent* set of decisions may be modelled separately considering each one as a conditional choice; then the resulting probabilities can be multiplied to yield joint probabilities for the set, such as in:

$$P(f, d, m, r) = P(f)\, P(d/f)\, P(m/d, f)\, P(r/m, d, f)$$

with f = frequency; d = destination; m = mode; r = route.

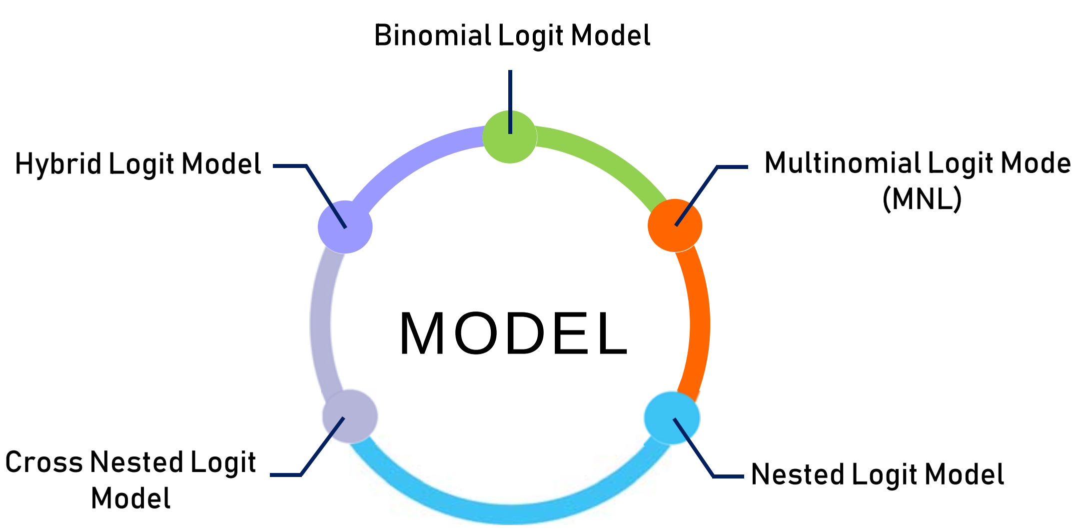
4. The explanatory variables included in the model can have explicitly estimated coefficients. In principle, the utility function allows any number and specification of the explanatory variables, as opposed to the case of the generalised cost function in aggregate models which is generally limited and has several fixed parameters. This has implications such as the following:
   - DM models allow a more flexible representation of the policy variables considered relevant for the study.
   - The coefficients of the explanatory variables have a direct marginal utility interpretation (i.e. they reflect the relative importance of each attribute).

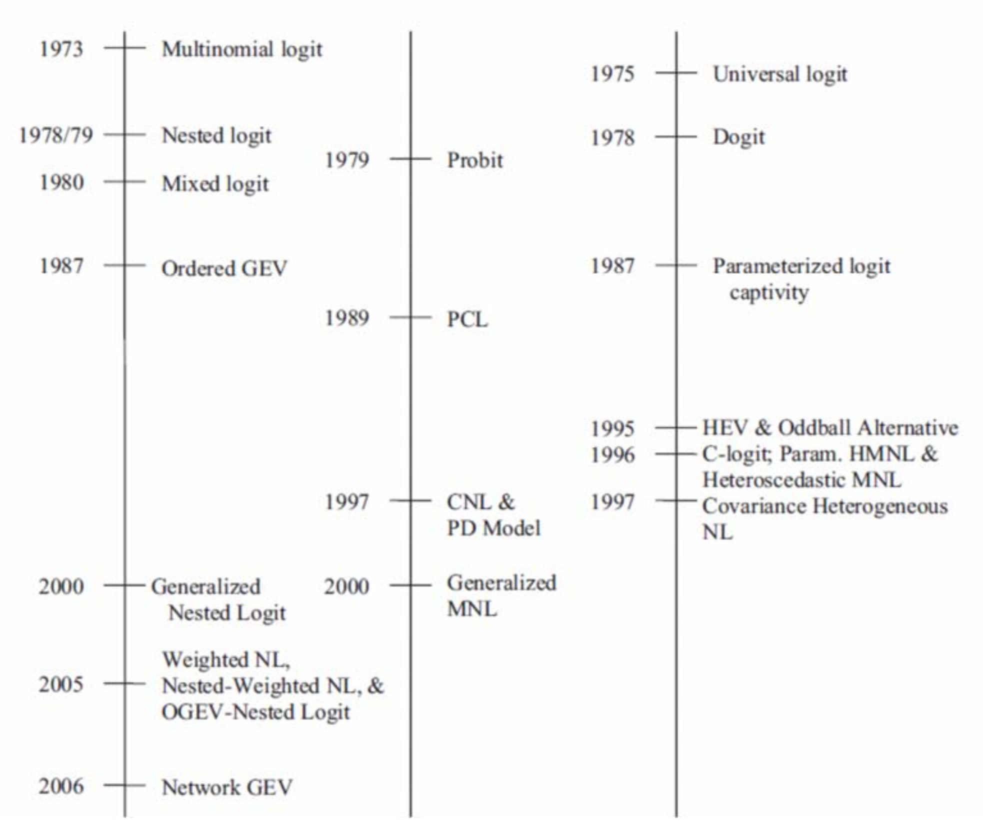
Exercise

- Estimate the travel mode chosen by each traveler.

Traveler	U1 (bus)	U2 (rail)	U3 (airplane)	Received Utility : Max (U1, U2, U3)	Alternative Chosen
1	3.5	2.25	1.57		
2	2.5	3.25	2.75		
3	3.5	2.25	4.0		
4	3.5	3.25	3		
5	2.5	3.25	3.5		
6	2.5	3.25	3.5		
7	3.5	2.25	2.5		

Types of DM





Exercise : Binary Logit Model

Discouraged because only 6 percent of the workers at the new office park at the edge of Jakarta use the bus way from a certain white-collar neighborhood, the Trans Jakarta asks the student of MSTT GMU to conduct a survey of persons who are commuting to the new development. We find that 2 factors affect commuter mode choice the most: *out-of-pocket cost (OPC) and total travel time (TTT)*. The IMPC computation for a MNL model result in a utility function $V_m = a_0 - (0.47 * OPC_m) - (0.22 * TTT_m)$, where $a_{0,auto}$ = 0.73, OPC_{bus} = \$0.75, TTT_{auto} = 10.5 minutes, and TTT_{bus} = 18 minutes. All other value is zero.

a. Does the MNL model developed by the student of MSTT GMU replicate the actual bus mode share

b. According to the MNL model, what would the bus mode share be if the Trans Jakarta reduce the bus fare to zero

Exercise : Binary Logit Model

Three strategies for off peak service were under active consideration by the Trans Jakarta.

A. Increase fares from 75 cents to $1.00, in hopes of increasing revenues

B. Decrease service frequency from four times per hour to twice per hour, to reduce operating costs

C. Increase service to six times per hour, in hopes of attracting more rider ship and more revenue.

Which of these alternative strategies would help Trans Jakarta financial situation the most.

Using data from the last mode choice survey done by us. Trans Jakarta has developed the MNL mode choice model shown here:

$U_b = a_0 - (0.41*OPC_b) + (0.24*FREQ_b) - (0.68*TTT_b)$

$U_m = a_0 - (0.47*OPC_m) - (1.22*TTT_m)$

It should be noted that $a_{0,auto}$ = 0.73, TTT_{auto} = 10.5 minutes, and TTT_{bus} = 18 minutes. VOC = $40/bus/route. Number of traveler = 1000 per hour

Estimation of Models from Random Samples

- To estimate the coefficient (θ), the maximum likelihood method is used.
- Let us assume a sample of Q individuals for which we observe their choice (0 or 1) and the values of x_{jkq} for each available alternative, such that for example:
  - individual 1 selects alternative 2
  - individual 2 selects alternative 3
  - individual 3 selects alternative 2
  - individual 4 selects alternative 1, etc.
- As the observations are independent the likelihood function is given by the product of the model probabilities that each individual chooses the option they actually selected:

  $L(\theta) = P_{21}P_{32}P_{23}P_{14} \ldots$

- Defining the following dummy variable: $$g_{jq} = \begin{cases} 1 & \text{if } A_j \text{ was chosen by } q \\ 0 & \text{otherwise} \end{cases}$$

- the above expression may be written more generally as:

$$L(\theta) = \prod_{q=1}^{Q} \prod_{A_j \in \mathbf{A}(q)} (P_{jq})^{g_{jq}}$$

- To maximise this function we proceed as usual, differentiating L(θ) partially with respect to the parameters θ and equating the derivative to 0. As in other cases we normally maximise l(θ), the natural logarithm of L(θ), which is more manageable and yields the same optima θ^*.

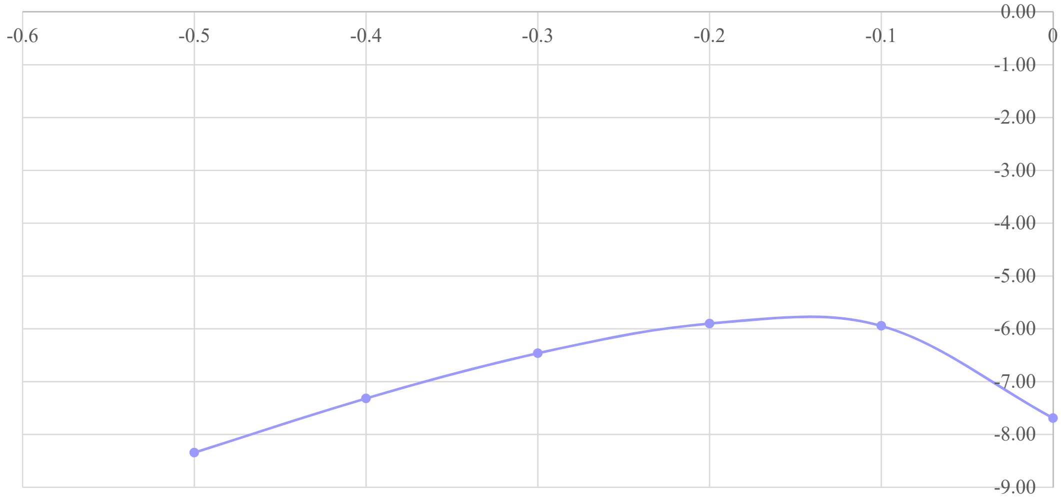
- Therefore, the function we seek to maximise is:

$$l(\theta) = \log L(\theta) = \sum_{q=1}^{Q} \sum_{A_j \in \mathbf{A}(q)} g_{jq} \log P_{jq}$$

Exercise : Estimating Coefficient

- A regional transportation agency wishes to calibrate a utility function that can be used with the logit model to predict modal choice between bus, auto, and rail.
- Survey data were obtained by interviewing seven people identified as persons A through G who reported the travel time for three modes they considered (car, bus, and rail) and the mode that they used.
- The results of the survey are shown in the following table. The agency has proposed to select a utility function of the form U = b (time).
- Use the method of maximum likelihood estimation to calibrate this utility function for the parameter, b.
- Sample Interview Survey Data:

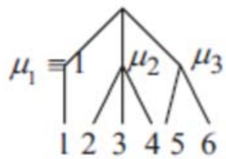
Respondent	*Auto Time (min)*	*Bus Time (min)*	*Rail Time (min)*	*Mode Used*
A	10	13	15	Auto
B	12	9	8	Auto
C	35	32	20	Rail
D	45	15	44	Bus
E	60	58	64	Bus
F	70	65	60	Auto
G	25	20	15	Rail



Multinomial Logit Model

$$\frac{e^{V_i}}{\sum_J e^{V_j}}$$

Nested Logit Model



$$\frac{e^{\frac{V_i}{\mu_m}}}{\sum_{j\in A_m} e^{\frac{V_j}{\mu_m}}} \times \frac{e^{V_m+\mu_m\Gamma_m}}{\sum_{l=1}^{M} e^{V_l+\mu_l\Gamma_l}},$$

$$\Gamma_m = \ln\left(\sum_{j\in A_m} e^{\frac{V_j}{\mu_m}}\right), 0<\mu_m\le 1$$

MULTINOMIAL LOGIT

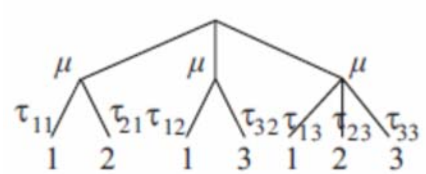
	Utility	Exp (U)	Prob
Choice 1	-2	0.1353	33.15%
Choice 2	-1.5	0.2231	54.65%
Choice 3	-3	0.0498	12.20%
	Total	0.4083	100%

NESTED LOGIT MODEL

	Lamda	0.85	
	Lamda / U	Exp (L / U)	Prob
Choice 1	-2.3529412	0.0951	30.96%
Choice 2.1	-1.7647059	0.1712	55.75%
Choice 2.2			13.29%
		0.2663	100%

Lamda antara 0 - 1 karena jika 1 nilainya akan sama dengan MNL

Cross Nested Logit Model



$$\sum_m \left[\frac{\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu}}}{\sum_{j\in A_m}\left(\tau_{jm}e^{V_j}\right)^{\frac{1}{\mu}}} \times \frac{\left(\sum_{j\in A_m}\left(\tau_{jm}e^{V_j}\right)^{\frac{1}{\mu}}\right)^{\mu}}{\sum_l \left(\sum_{j\in A_l}\left(\tau_{jl}e^{V_j}\right)^{\frac{1}{\mu}}\right)^{\mu}} \right],$$

$$0<\mu\le 1, \sum_m \tau_{jm}=1$$

To visualize the calculations of GNL probabilities, assume an individual's decision of whether to take economy train, premium train, economy air, or premium air is expressed as a function of time and cost (as well as an intercept term):

$$V_i^n = \alpha_i + \beta_1\left(Time_i^n\right) + \beta_2\left(Cost_i^n\right)$$

Suppressing the index representing individual n for notational convenience, assume the utility function for the four alternatives (faced with specific time and costs) is given as:

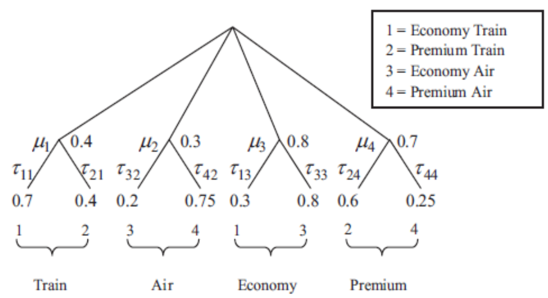
$$V_1 = 1 - 0.075(5\text{ hrs}) - 0.0015(\$300) = 0.175$$

$$V_2 = 0.5 - 0.075(5\text{ hrs}) - 0.0015(\$400) = -0.475$$

$$V_3 = 2.5 - 0.075(3\text{ hrs}) - 0.0015(\$350) = 1.75$$

$$V_4 = 0 - 0.075(3\text{ hrs}) - 0.0015(\$750) = -1.275$$

$$P_i = \sum_m \left[\frac{\overbrace{\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}}}^{\text{A TERM}}}{\underbrace{\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}}_{\text{B TERM}}} \times \frac{\overbrace{\left(\sum_{j \in A_m} \left(\tau_{jm} e^{V_j}\right)^{\frac{1}{\mu_m}}\right)^{\mu_m}}^{\text{C TERM}}}{\underbrace{\sum_l \left(\sum_{j \in A_l} \left(\tau_{jl} e^{V_j}\right)^{\frac{1}{\mu_l}}\right)^{\mu_l}}_{\text{D TERM}}} \right]$$



where:

A TERM is computed for each alternative in nest m,
B TERM is the sum of all A TERMS in nest m,
C TERM is the B TERM raised to the logsum coefficient for nest m, or μ_m,
D TERM is the sum, overall all nests, of C TERMS.

A TERM for alternative one in nest one = $\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}} = \left(0.7e^{0.175}\right)^{\frac{1}{0.4}} = 0.6350$

A TERM for alternative two in nest one = $\left(\tau_{im} e^{V_i}\right)^{\frac{1}{\mu_m}} = \left(0.4e^{-0.475}\right)^{\frac{1}{0.4}} = 0.0309$

The B TERM associated with nest one is simply the sum of these two A TERMS, or 0.6350 + 0.0309 = 0.6658. The C TERM associated with nest one is also straightforward, i.e., the B TERM raised to the logsum coefficient for nest one, or:

C TERM for nest one = $(0.6658)^{0.4} = 0.8499$

The process is repeated for each of the nests, after which the D TERM is calculated as the sum of all of the C TERMS, or 0.8499 + 1.1521 + 4.7540 + 0.3967 = 7.153. With the intermediate calculations of Table 4.2 completed, the probability of alternative one is given as the sum of probabilities calculated for nests one and three (the nests that alternative one belongs to), or:

$$P_1 = \frac{A_1}{B_1} \times \frac{C_1}{D} + \frac{A_3}{B_3} \times \frac{C_3}{D}$$

$$= \frac{0.6350}{0.6658} \times \frac{0.8499}{7.1526} + \frac{0.2763}{7.0198} \times \frac{4.7540}{7.1526} = 0.1395.$$

Similar calculations apply for alternatives two, three, and four.

	A TERM				B TERM	C TERM	D TERM
	ALT 1	ALT 2	ALT 3	ALT 4			
Nest 1	0.6350	0.0309	0	0	0.6658	0.8499	7.153
Nest 2	0	0	1.5977	1.6031	1.6031	1.1521	7.153
Nest 3	0.2763	0	6.7434	0	7.0198	4.7540	7.153
Nest 4	0	0.2446	0	0.0223	0.2669	0.3967	7.153
Probability	0.1395	0.0563	0.7990	0.0052			

THANK YOU