Practical note on specification of discrete choice model

Toshiyuki Yamamoto

Nagoya University, Japan

Contents

- Comparison between binary logit model and binary probit model
- Comparison between multinomial logit model and nested logit model
- Comparison between nested logit model and mixed logit model

Comparison between binary logit model and binary probit model

Random utility models

- Random utility

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

  V_{jn} : deterministic part of utility
  ε_{jn} : stochastic part of utility

- Conventional linear utility function

  $V_{jn} = \beta X_{jn}$

  X_{jn} : vector of explanatory variables
  β : vector of coefficients

Binary choice models

When the choice set contains <u>only two alternatives</u>

- Probability for individual n to choose alternative i

$$
\begin{aligned}
P_{in} &= \text{Prob}(U_{in} > U_{jn}) \\
&= \text{Prob}(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}) \\
&= \text{Prob}(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn})
\end{aligned}
$$

- If ε_{jn} and ε_{in} follow normal distribution, $\varepsilon_{jn} - \varepsilon_{in}$ also follows normal distribution -> Binary probit model
- If ε_{jn} and ε_{in} follow iid Gumbel distribution, $\varepsilon_{jn} - \varepsilon_{in}$ follows <u>logistic</u> distribution -> Binary logit model

Gumbel distribution: G(η, μ)

- Probability density function

$$f\left(\varepsilon\right) = \mu \exp\{-\mu(\varepsilon-\eta)\}\exp[-\exp\{-\mu(\varepsilon-\eta)\}]$$

  – Mode = η, Mean = $\eta + r/\mu$, variance = $\pi^2/6\mu^2$, where r ≈ 0.577 (Euler's constant)

- Cumulative density function

$$F\left(\varepsilon\right) = \exp\left[-\exp\left\{-\mu\left(\varepsilon-\eta\right)\right\}\right]$$

Binary logit model

- If ε_{in} and ε_{jn} follow G(η_{i}, μ) and G(η_{j}, μ) respectively, $\varepsilon_{jn} - \varepsilon_{in} = \varepsilon_n$ follows logistic distribution as below

$$F\left(\varepsilon_n\right) = \frac{1}{1 + \exp\left\{\mu\left(\eta_j - \eta_i - \varepsilon_n\right)\right\}}$$

- Assuming $\eta_{\mathrm{i}} = \eta_{\mathrm{j}} = 0$, probability to choose i is

$$P_{in} = \Pr\left(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}\right) = F\left(V_{in} - V_{jn}\right)$$

$$= \frac{1}{1 + \exp\left\{-\mu\left(V_{in} - V_{jn}\right)\right\}} = \frac{\exp\left(\mu V_{in}\right)}{\exp\left(\mu V_{in}\right) + \exp\left(\mu V_{jn}\right)}$$

Normal distribution: N(m, σ^2)

- Probability density function

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon - m}{\sigma}\right)^2 \right]$$

  – Mode = Mean = m, variance = σ^2

- Cumulative density function

$$F(\varepsilon) = \int_{e=-\infty}^{\varepsilon} f(e) de$$

Binary probit model

- $\varepsilon_{jn} - \varepsilon_{in} = \varepsilon_n$ is assumed to follow N(0, σ^2) where m = 0

$$P_{in} = \Pr\left(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}\right)$$

$$= \int_{\varepsilon_n = -\infty}^{V_{in} - V_{jn}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_n}{\sigma}\right)^2\right] d\varepsilon_n$$

$$= \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right)$$

If V(ε_{in}) = V(ε_{jn}) and COV(ε_{in}, ε_{jn}) = 0 (i.i.d.),
V(ε_{in}) = V(ε_{jn}) = $\sigma^2/2$

Identifiability of parameters

- Binary logit model:

$$P_{in} = \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \exp(\mu V_{jn})} = \frac{\exp(\mu \beta X_{in})}{\exp(\mu \beta X_{in}) + \exp(\mu \beta X_{jn})}$$

- Binary probit model:

$$P_{in} = \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right) = \Phi\left(\frac{\beta X_{in} - \beta X_{jn}}{\sigma}\right) = \Phi\left(\frac{\beta}{\sigma} X_{in} - \frac{\beta}{\sigma} X_{jn}\right)$$

μ and σ are always connected with β
Thus, μ and σ cannot be identified

Standardization

- Binary logit model: $\mu = 1$ ➡ $V(\varepsilon_{in}) = \pi^2/6$

$$P_{in} = \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \exp(\mu V_{jn})} = \frac{\exp(V_{in})}{\exp(V_{in}) + \exp(V_{jn})}$$

- Binary probit model: $\sigma = 1$ ➡ $V(\varepsilon_{in}) = 1/2$

when $V(\varepsilon_{in}) = V(\varepsilon_{jn})$ and $COV(\varepsilon_{in}, \varepsilon_{jn}) = 0$ (i.i.d.)

$$P_{in} = \Phi\left(\frac{V_{in} - V_{jn}}{\sigma}\right) = \Phi(V_{in} - V_{jn})$$

Estimates of $V_{jn} = \beta X_{jn}$ have different sizes

Also applies when comparing multinomial logit and probit models

Comparison between multinomial logit model and nested logit model

Multinomial logit model

$$P_{in} = \frac{\exp(\mu V_{in})}{\sum_{j=1}^{J} \exp(\mu V_{jn})}$$

where ε_{in} follows G(0, μ)

$$= \frac{\exp(\mu \beta X_{in})}{\sum_{j=1}^{J} \exp(\mu \beta X_{jn})}$$

μ Is always connected with β
Thus, μ cannot be identified

$$\rightarrow \frac{\exp(\beta X_{in})}{\sum_{j=1}^{J} \exp(\beta X_{jn})}$$

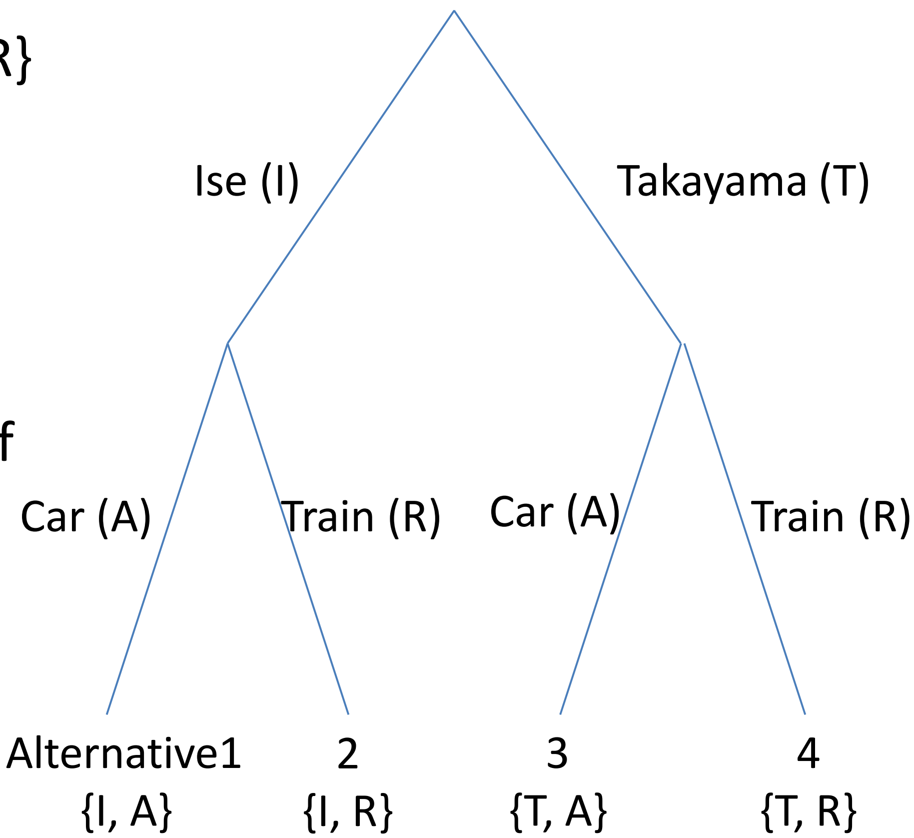
standardized by $\mu = 1$

$V(\varepsilon_{in}) = \pi^2/6$

Nested logit model

- Joint choice of trip destination and mode
- Destination d = {I, T}, mode m = {A, R}
- Utility function:
  $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- V_d: utility specific to destination d
- V_m: utility specific to mode m
- V_{dm}: utility specific to combination of destination d and mode m (such as travel time)
- ε_d: stochastic utility specific to destination d
- ε_{dm}: stochastic utility specific to combination of destination d and mode m

- Tree structure



Identifiability of parameters

$$P(d,m) = \frac{\exp\left\{\mu_{dm}\left(V_m + V_{dm}\right)\right\}}{\sum_{m' \in \{A,R\}} \exp\left\{\mu_{dm}\left(V_{m'} + V_{dm'}\right)\right\}}$$

$$\times \frac{\exp\left\{\mu V_d + \boxed{\frac{\mu}{\mu_{dm}}} \ln \sum_{m \in \{A,R\}} \exp\left\{\mu_{dm}\left(V_m + V_{dm}\right)\right\}\right\}}{\sum_{d' \in \{I,T\}} \exp\left\{\mu V_{d'} + \boxed{\frac{\mu}{\mu_{d'm}}} \ln \sum_{m \in \{A,R\}} \exp\left\{\mu_{d'm}\left(V_m + V_{d'm}\right)\right\}\right\}}$$

- ε_{dm} follows $G(0, \mu_{dm})$ and $\varepsilon_d + \varepsilon_{dm}$ follows $G(0, \mu)$ which means $\mu \le \mu_{dm}$
- One of μ and μ_{dm} can be identified, and the other should be fixed

Two ways of standardization

- $\mu_{dm} = 1$ → $0 \le \mu \le 1$ → $V(\varepsilon_d + \varepsilon_{dm}) \ge \pi^2/6$

$$P(d,m) = \frac{\exp(V_m + V_{dm})}{\sum_{m' \in \{A,R\}} \exp(V_{m'} + V_{dm'})} \times \frac{\exp\left\{\mu V_d + \mu \ln \sum_{m \in \{A,R\}} \exp(V_m + V_{dm})\right\}}{\sum_{d' \in \{I,T\}} \exp\left\{\mu V_{d'} + \mu \ln \sum_{m \in \{A,R\}} \exp(V_m + V_{d'm})\right\}}$$

- $\mu = 1$ → $1 \le \mu_{dm}$ → $V(\varepsilon_d + \varepsilon_{dm}) = \pi^2/6$

$$P(d,m) = \frac{\exp\{\mu_{dm}(V_m + V_{dm})\}}{\sum_{m' \in \{A,R\}} \exp\{\mu_{dm}(V_{m'} + V_{dm'})\}} \times \frac{\exp\left\{V_d + \frac{1}{\mu_{dm}} \ln \sum_{m \in \{A,R\}} \exp\{\mu_{dm}(V_m + V_{dm})\}\right\}}{\sum_{d' \in \{I,T\}} \exp\left\{V_{d'} + \frac{1}{\mu_{d'm}} \ln \sum_{m \in \{A,R\}} \exp\{\mu_{d'm}(V_m + V_{d'm})\}\right\}}$$

$\mu = 1$ is recommended to keep the size of β comparable with multinomial logit model

Comparison between nested logit model and mixed logit model

Stochastic terms of nested logit model and mixed logit model

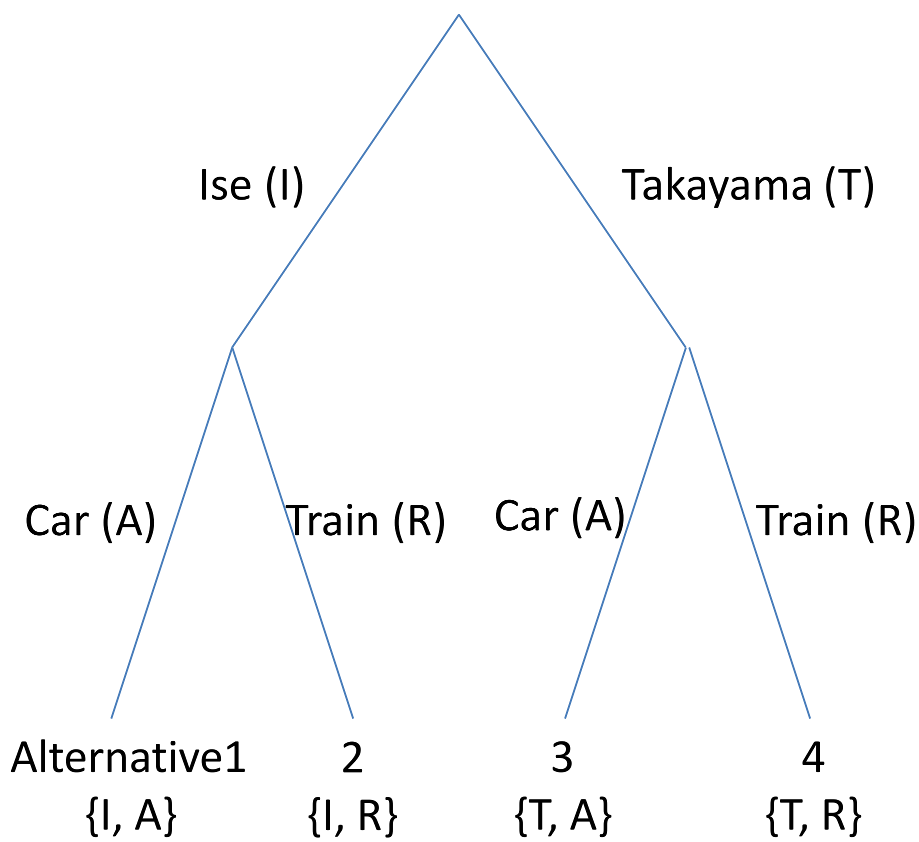
Nested logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- $\varepsilon_d + \varepsilon_{dm}$ follows $G(0, \mu)$

Mixed logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- ε_d follows $N(0, \sigma_d^2)$

- Tree structure



Stochastic terms of nested logit model and mixed logit model

Nested logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- $\varepsilon_d + \varepsilon_{dm}$ follows $G(0, \mu)$

Mixed logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- ε_d follows $N(0, \sigma_d^2)$

- Different probability distributions are mixed
- Distributions other than normal can be used, but normal is often used
- Standardized by $\mu_{dm} = 1$, $V(\varepsilon_d + \varepsilon_{dm}) = \sigma_d^2 + \pi^2/6$
- Size of β becomes different from nested logit model

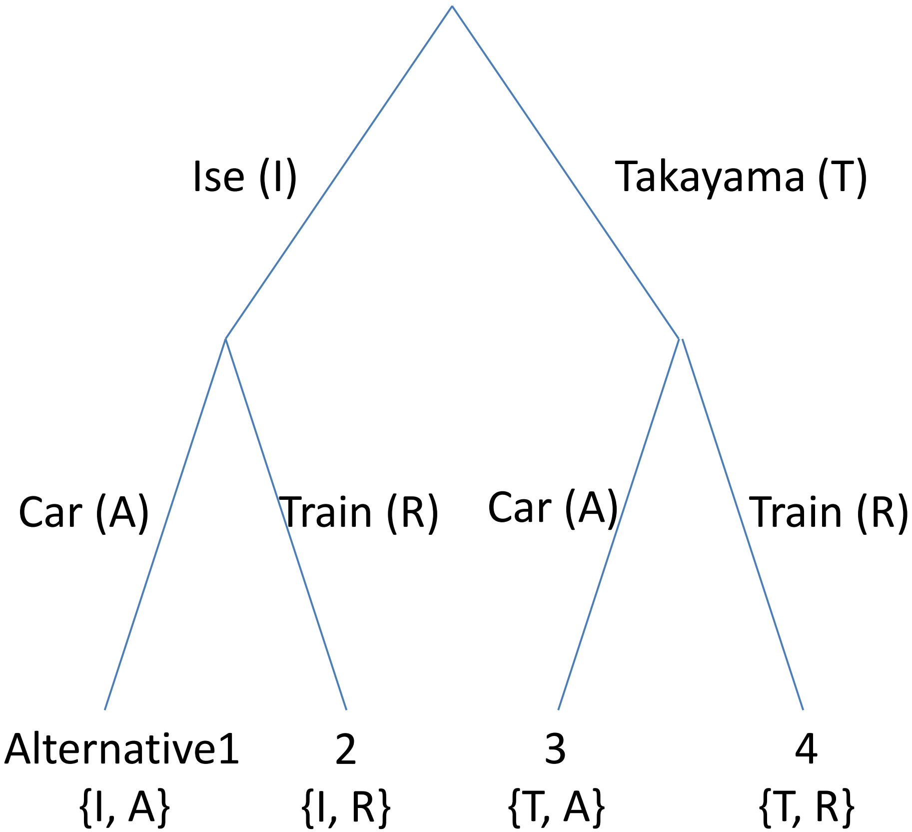
Nested logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- $\varepsilon_d + \varepsilon_{dm}$ follows $G(0, \mu)$

Utility function for each alternative

1. $U_{IA} = V_I + V_A + V_{IA} + \varepsilon_I + \varepsilon_{IA}$
2. $U_{IR} = V_I + V_R + V_{IR} + \varepsilon_I + \varepsilon_{IR}$
3. $U_{TA} = V_T + V_A + V_{TA} + \varepsilon_T + \varepsilon_{TA}$
4. $U_{TR} = V_T + V_R + V_{TR} + \varepsilon_T + \varepsilon_{TR}$

- Tree structure



Nested logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows $G(0, \mu_{dm})$
- $\varepsilon_d + \varepsilon_{dm}$ follows $G(0, \mu)$

Utility function for each alternative

1. $U_{IA} = V_I + V_A + V_{IA} + \varepsilon_I + \varepsilon_{IA}$
2. $U_{IR} = V_I + V_R + V_{IR} + \varepsilon_I + \varepsilon_{IR}$
3. $U_{TA} = V_T + V_A + V_{TA} + \varepsilon_T + \varepsilon_{TA}$
4. $U_{TR} = V_T + V_R + V_{TR} + \varepsilon_T + \varepsilon_{TR}$

- ε_I is common for alt. 1 & 2, so $V(\varepsilon_{IA}) = V(\varepsilon_{IR})$
- ε_T is common for alt. 3 & 4, so $V(\varepsilon_{TA}) = V(\varepsilon_{TR})$
- However, $V(\varepsilon_I)$ and $V(\varepsilon_T)$ can be different
- It means μ_{dm} and $\mu_{d'm}$ can be different

Mixed logit model

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows G(0, 1)

$$P\left(d,m\middle|\varepsilon_I,\varepsilon_T\right)=\frac{\exp\left(V_d+V_m+V_{dm}+\varepsilon_d\right)}{\sum_{d'm'\in\{IA,IR,TA,TR\}}\exp\left(V_{d'}+V_{m'}+V_{d'm'}+\varepsilon_{d'}\right)}$$

- ε_d follows $N(0, \sigma_d^2)$

$$P\left(d,m\right)=\int_{\varepsilon_I=-\infty}^{\infty}\int_{\varepsilon_T=-\infty}^{\infty}P\left(d,m\middle|\varepsilon_I,\varepsilon_T\right)\frac{1}{\sigma_I}\phi\left(\frac{\varepsilon_I}{\sigma_I}\right)\frac{1}{\sigma_T}\phi\left(\frac{\varepsilon_T}{\sigma_T}\right)d\varepsilon_I d\varepsilon_T$$

Numerical integration is needed for 2 dimensions

Identifiability of parameters

- $U_{dm} = V_d + V_m + V_{dm} + \varepsilon_d + \varepsilon_{dm}$
- ε_{dm} follows G(0, 1)
- ε_d follows N(0, σ_d^2)

Utility function for each alternative

1. $U_{IA} = V_I + V_A + V_{IA} + \varepsilon_I + \varepsilon_{IA}$
2. $U_{IR} = V_I + V_R + V_{IR} + \varepsilon_I + \varepsilon_{IR}$
3. $U_{TA} = V_T + V_A + V_{TA} + \varepsilon_T + \varepsilon_{TA}$
4. $U_{TR} = V_T + V_R + V_{TR} + \varepsilon_T + \varepsilon_{TR}$

- Different from nested logit model, σ_I^2 and σ_T^2 cannot be estimated together
- Considering [only difference in utility matters], setting $\varepsilon_I' = \varepsilon_I - \varepsilon_T$ and $\varepsilon_T' = 0$ gives the same β

Then, why can both be estimated in nested logit model?

Reference

- Carrasco, J.A. and Ortuzar, J. de D. (2002) Review and assessment of the nested logit model, Transport Reviews, Vol. 22(2), pp. 197-218.
- Walker, J.L., Ben-Akiva, M. and Bolduc, D. (2007) Identification of parameters in normal error component logit-mixture (NECLM) models, Journal of Applied Econometrics, Vol. 22, pp. 1095-1125.