Open-form
discrete choice models
3. Genealogy of DCMs (again)

- **Multinomial logit (MNL)** (Luce, 1959)
  - Generalization
  - **Nested logit (NL)** (Ben-Akiva, 1973)
    - Generalization
    - **Generalized extreme value (GEV)** (McFadden, 1978)
      - Special case
      - **Paired combinational logit (PCL)** (Chu, 1981)
      - **Cross-nested logit (CNL)** (Vovsha, 1997)
        - Generalization
        - Generalized nested logit (GNL), recursive nested logit extreme value model (RNEV), network-GEV (Wen & Koppelman, 2001; Daly, 2001; Bierlaire, 2002)

- **Multinomial Probit (MNP)** (Thurstone, 1927)
  - Normal to Gumbel

- **Heteroscedastic/mixed distributions**
  - Error component logit (ECL); Mixed logit (MXL); Kernel logit (KL);
    - Heteroscedastic logit (HL)
      - (Boyd and Mellman, 1980; Cardell and Dunbar, 1980; McFadden, 1989; Bhat, 1995; See Train (2009) for details)
  - Gumbel to Weibull
  - Weibull to GEV (not MEV)
  - q-generalized logit
    - (Nakayama, 2013)

- **Multinomial weibit (MNW)** (Castillo, et al., 2008)
  - Generalization

- **Generalized G function** (Mattsson et al., 2014)

- **Variance stabilization** (Li, 2011)

- Derived from McFadden’s G function or “choice probability generating functions” (Fosgerau et al., 2013)
  - Derived from the generalized G function
Non-GEV (Open-form) Probit model
3. Difference of GEV and Non-GEV

**GEV model (Closed-form)**

**Multinomial Logit (MNL)**

\[
P(i) = \frac{\exp(\mu V_i)}{\sum_{j \in C} \exp(\mu V_j)}
\]

- Luce (1959), McFadden (1974)
- Not consider correlation of choice alternatives’ (IIA)
- Easy and fast estimation
- High operability
  (easy evaluation for new additional choice alternative ⇒ benefit of IIA)

**Non-GEV model (Open-form)**

**Multinomial Probit (MNP)**

\[
P(i) = \int_{\varepsilon_1 = -\infty}^{\infty} \cdots \int_{\varepsilon_i = -\infty}^{\infty} \cdots \int_{\varepsilon_j = -\infty}^{\infty} \phi(\varepsilon) d\varepsilon_j \cdots d\varepsilon_1
\]

\[
\phi(\varepsilon) = \frac{1}{(\sqrt{2\pi})^{J-1} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \varepsilon \Sigma^{-1} \varepsilon'\right)
\]

- Thurstone (1927)
- Consider correlation of choice alternates’ based on Variance-Covariance matrix
- Hard and slow estimation
  (need calculation of multi-dimensional interrelation depend on N of alternatives')

Non-GEV model has high power of expression, however parameter estimation cost is high.
3. Structured Covariance MNP (1)


- This model applied to practical demand forecasting in real Tokyo network, and it used for decision making of railway policy toward 2015. (2000)
Tokyo Metropolitan has highly dense railway network!
⇒ route overlapping problem

Railway line: about 130
Station: about 1800
Passengers: 40 million/day
Cong. rate: max over 200%
In the overlap network that has correlation between routes, Logit model is susceptible to error by IIA property.

Overlap = correlation

Probit is better?

- Difficult to setting covariance matrix for each OD pair ⇒ structured covariance by divide into two error
- Difficult to parameter estimation. (multi-dimensional Integral) ⇒ reduce computational time using simulation methods

$$U_i = V_i + \varepsilon_i$$

Error of depend on route length

$$\varepsilon_i = \varepsilon_i^{Length} + \varepsilon_i^{Route}$$

Error of route specific
3. Structured Covariance MNP (4)

Variance-Covariance structure in Error term

\[ \varepsilon_r = \varepsilon_r^1 + \varepsilon_r^0 \]
\[ \Sigma = \Sigma^1 + \Sigma^0 \]

**Error of depend on route length**
Variance of route utility increases in proportion to the route length.

\[ \text{Var}(\varepsilon_r^1) = L_r \sigma^2 \]

Covariance between routes increases in proportion to the length of route overlap.

\[ \text{Cov}(\varepsilon_r^1, \varepsilon_q^1) = L_{rq} \sigma^2 \]

**Error of route specific**
- independent of each route (cov=0)

\[ \text{Cov}(\varepsilon_r^0, \varepsilon_q^0) = \sigma_0^2, \quad q = r \]
\[ = 0, \quad q \neq r \]

\[ \Sigma = \sigma^2 \begin{pmatrix} L_1 & L_{12} & \cdots & L_{1R} \\ L_{12} & L_2 & \cdots & L_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ L_{1R} & L_{2R} & \cdots & L_R \end{pmatrix} + \sigma_0^2 I \]

Simplify use cov. ratio

\[ \Sigma = \sigma_0^2 \begin{pmatrix} \eta L_1 + 1 & \eta L_{12} & \cdots & \eta L_{1R} \\ \eta L_{12} & \eta L_2 + 1 & \cdots & \eta L_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \eta L_{1R} & \eta L_{2R} & \cdots & \eta L_R + 1 \end{pmatrix} \]

\[ \eta = \frac{\sigma^2}{\sigma_0^2} \]

Estimate only cov. ratio!

- \( L_r \) : length of route \( r \)
- \( L_{rq} \) : overlap length between route \( r \) and \( q \)
- \( \sigma^2 \) : variance of unit length
3. Structured Covariance MNP (5)

Apply to the SCMNL for The 18th master plan for urban railway network in TMA (2000)

Ex : Oomiya to Kanda station

Estimation results

<table>
<thead>
<tr>
<th>parameter</th>
<th>parameter</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-vehicle time</td>
<td>−0.0943</td>
<td>−8.09</td>
</tr>
<tr>
<td>access/egress time</td>
<td>−0.127</td>
<td>−11.7</td>
</tr>
<tr>
<td>transfer time</td>
<td>−0.112</td>
<td>−10.7</td>
</tr>
<tr>
<td>cost</td>
<td>−0.002</td>
<td>−3.98</td>
</tr>
<tr>
<td>congestion index</td>
<td>−0.00869</td>
<td>−3.34</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.436</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Prediction results

<table>
<thead>
<tr>
<th>Route</th>
<th>Obs</th>
<th>MNL</th>
<th>SCMNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utsunomiya + Yamanote</td>
<td>33%</td>
<td>48%</td>
<td>28%</td>
</tr>
<tr>
<td>Utsunomiya + Keihin-Tohoku</td>
<td>15%</td>
<td>24%</td>
<td>52%</td>
</tr>
<tr>
<td>Keihin-Tohoku</td>
<td>53%</td>
<td>47%</td>
<td>20%</td>
</tr>
</tbody>
</table>

To achieve a high prediction accuracy by the relaxation of route overlap (Obs ±10% in all route)
Non-GEV (Open-form) Mixed Logit
Mixed Loigt  (Train 2000)

High flexible structure using **two error** term.

Utility function

\[ U_i = V_i + \eta_i + \nu_i \]

\( \nu \) dist.: assume any G function
  - IID Gamble (Logit Kernel) \( \Rightarrow \) MNL
  - any G function (GEV Kernel) \( \Rightarrow \) NL, PCL, CNL, GNL…

\( \eta \) dist.: basically assume “**Normal dist.**”
  In the case of normal distribution takes a non-realistic value, it can assume a variety of probability distribution (triangular distribution, cutting normal distribution, lognormal distribution, Rayleigh distribution, etc.).

• Error Component: approximate to any GEV model
• Random Coefficient: Consider the heterogeneity
Approximation of Nested Logit (NL)

Describe the nest (covariance) using structured $\eta$.

**Ex: model choice**

- **Car**
  \[ U_{\text{car}} = \beta X_{\text{car}} + \nu_{\text{car}} \]

- **Bus**
  \[ U_{\text{bus}} = \beta X_{\text{bus}} + \sigma_{\text{transit}} \eta_{\text{transit}} + \nu_{\text{bus}} \]

- **Rail**
  \[ U_{\text{rail}} = \beta X_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}} + \nu_{\text{rail}} \]

**Choice prob. (open-form)**

\[
P_{\text{rail}} = \int_{\eta_{\text{transit}}} \frac{e^{V_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}}}}{e^{V_{\text{car}}} + e^{V_{\text{bus}} + \sigma_{\text{transit}} \eta_{\text{transit}}} + e^{V_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}}}} f(\eta_{\text{transit}}) d\eta_{\text{transit}}
\]

**Choice prob. (Simulated)**

\[
P_{\text{rail}} = \frac{1}{N} \sum_{N} \frac{e^{V_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}}^N}}{e^{V_{\text{car}}} + e^{V_{\text{bus}} + \sigma_{\text{transit}} \eta_{\text{transit}}^N} + e^{V_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}}^N}}
\]

$\eta_{\text{transit}} \approx N(0,1)$
Approximation of Nested Logit (NL)

Note that variance-covariance matrix is inconsistent with normal NL

Normal NL

\[
\begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & \sigma^2_{\text{transit}} \\
0 & \sigma^2_{\text{transit}} & \sigma^2
\end{pmatrix}
\]

Diagonal elements (variance of Bus and Rail) is bigger than \(\sigma^2_{\text{transit}}\)

Approximated NL based on MXL

\[
\begin{pmatrix}
0 & 0 & 0 \\
0 & \sigma^2 & \sigma^2 \\
0 & \sigma^2 & \sigma^2
\end{pmatrix} + \begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & 0 \\
0 & 0 & \sigma^2
\end{pmatrix} = \begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 + \sigma^2 & 0 \\
0 & 0 & \sigma^2 + \sigma^2
\end{pmatrix}
\]
Error Component: CNL

Approximation of Cross Nested Logit (CNL)

Describe the nest (covariance) using structured $\eta$. 

Road nest
- Car
- Bus
- Rail

Transit nest

$U_{\text{car}} = \beta X_{\text{car}} + \sigma_{\text{road}} \eta_{\text{road}} + \nu_{\text{car}}$

$U_{\text{bus}} = \beta X_{\text{bus}} + \sigma_{\text{road}} \eta_{\text{road}} + \nu_{\text{bus}}$

$U_{\text{rail}} = \beta X_{\text{rail}} + \sigma_{\text{transit}} \eta_{\text{transit}} + \nu_{\text{rail}}$

\[ \eta = \begin{pmatrix} \sigma^2_{\text{road}} & \sigma^2_{\text{transit}} & 0 \\ \sigma^2_{\text{road}} & \sigma^2_{\text{transit}} & 0 \\ 0 & 0 & \sigma^2_{\text{transit}} \end{pmatrix} \]

\[ \nu = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \]

$\eta_{\text{transit}}, \eta_{\text{road}} \approx N(0,1)$
Error Component: SCL

Approximation of Spatial Correlation Logit

Describe the spatial correlation using structured $\eta$.

$$U_{\text{zorn}1} = \beta X_{\text{zorn}1} + \sigma \eta_1 + \sigma \eta_2 + v_{\text{zorn}1}$$
$$U_{\text{zorn}2} = \beta X_{\text{zorn}2} + \sigma \eta_1 + \sigma \eta_2 + \sigma \eta_3 + v_{\text{zorn}2}$$
$$U_{\text{zorn}3} = \beta X_{\text{zorn}3} + \sigma \eta_2 + \sigma \eta_3 + \sigma \eta_4 + v_{\text{zorn}3}$$
$$U_{\text{zorn}4} = \beta X_{\text{zorn}4} + \sigma \eta_3 + \sigma \eta_4 + v_{\text{zorn}4}$$

$$\eta \sim N(0,1)$$
Approximation of heteroscedastic Logit

Assume the different error variance in each alternatives' parameters to zero at least.

\[
\begin{align*}
U_{\text{car}} &= \beta X_{\text{car}} + \sigma_{\text{car}} \eta_{\text{car}} + \nu_{\text{car}} \\
U_{\text{bus}} &= \beta X_{\text{bus}} + \sigma_{\text{bus}} \eta_{\text{bus}} + \nu_{\text{bus}} \\
U_{\text{rail}} &= \beta X_{\text{rail}} + \sigma_{\text{rail}} \eta_{\text{rail}} + \nu_{\text{rail}} \\
\eta_{\text{car}}, \eta_{\text{bus}}, \eta_{\text{rail}} &\sim N(0,1)
\end{align*}
\]

Assume heteroscedastic in error

\[
\begin{pmatrix}
\sigma_{\text{car}}^2 + \sigma^2 & 0 & 0 \\
0 & \sigma_{\text{bus}}^2 + \sigma^2 & 0 \\
0 & 0 & \sigma_{\text{rail}}^2 + \sigma^2
\end{pmatrix}
\]

\(\sigma_{\text{bus}}^2 = 1\)

- Car: Low travel time reliability
  \(\Rightarrow\) Error variance is large

- Rail: High travel time reliability
  \(\Rightarrow\) Error variance is small

※consider only heteroscedastic (IID assumption is not relaxed)
Taste heterogeneity of decision maker
Parameters defined homogeneously in population. However, decision maker $n$ has different taste (= heterogeneity)

$$U_{car,n} = \beta T_{car,n} + \epsilon_{car,n} \quad \rightarrow \quad U_{car,n} = \beta_n T_{car,n} + \epsilon_{car,n}$$

Segmentation (observable heterogeneity)

- Constant by gender: male’s constant: $\alpha_0 + \alpha_1$
  $$U_{car,n} = \left[ (\alpha_0) + \alpha_1 \times male_n \right] + \beta_1 T_{car,n} + \epsilon_{car,n}$$

Female’s constant: $\alpha_0$

- Parameter by gender: male’s parameter: $\beta_1$
  $$U_{car,n} = \alpha_0 + \left( \beta_1 \right) \times male_n \times T_{car,n} + \left( \beta_2 \right) \times (1 - male_n) \times T_{car,n}$$

Female’s parameter: $\beta_2$

※ $1 - male_n = female_n$
Random Coefficient (2)

Parameter distribution (unobservable heterogeneity)

Assume the heterogeneity of parameter
⇒ In the case of parameter following Normal dist., we estimate the dist.’s hyper-parameter (mean and variance).

\[
U_{car,n} = \beta_n T_{car,n} + \nu_{car,n}
\]

\[
\beta_n \approx N(\bar{\beta}, \sigma^2)
\]

\[
U_{car,n} = \bar{\beta} T_{car,n} + \sigma \eta_n T_{car,n}
\]

\[
U_{bus,n} = \bar{\beta} T_{bus,n} + \sigma \eta_n T_{bus,n}
\]

\[
U_{rail,n} = \bar{\beta} T_{rail,n} + \sigma \eta_n T_{rail,n}
\]

\[
\eta_n \approx N(0,1) \quad \bar{\beta}, \sigma : \text{unknown parameter}
\]

Hyper-parameter can describe using observable variables

\[
\bar{\beta}_n = \gamma_0 + \gamma_1 income_n \quad \beta \text{ depend on observable income variable}
\]
Summary of open-form models

Strengths

- Describe correlation between alternatives’ by EC
  - MNP: all alternatives’ (relax and reduce by structuring)
  - MXL: depend on approximated model

- Describe heterogeneity by RC
  - Segmentation, parameter distribution...

Limitations

- High calculation cost in parameter estimation
  - Open-form model has high dimensional integration.
  - Recently, proposed high speed estimation methods
    Ex: Bayesian estimation (MCMC) ⇒ see Train’s book
    MACML: analytical integration by Bhat et al.(2011)
References

References