CIVE707 – Theory of Transport Demand Modelling

Disaggregate Choice: Mixed Logit

Jason Hawkins

Outline:

- Mixed Logit/Logit Kernel
- Heteroskedastic Mixed Logit
- Error-Component Mixed Logit
- Random Parameter Mixed Logit
- Identification
- Issues with Estimation
- Other Mixed Logit Models

Closed Form Discrete Choice Model

- Choice probability of a Multinomial Choice Context:

$$\Pr(j) = \Pr\left(\varepsilon_k \le \left(V_j - V_k + \varepsilon_j\right)\right)$$

- Probability distribution of random utility (ε) needs to be fully specified to get the unconditional probability:

$$\Pr(j) = \int_{\varepsilon_j=-\infty}^{+\infty} \int_{\varepsilon_k=-\infty}^{\left(V_j - V_k + \varepsilon_j\right)} f(\varepsilon_j, \varepsilon_k)\, d\varepsilon_j\, d\varepsilon_k$$

  - Specification of univariate distribution, $f(\varepsilon_k)$, and the joint distribution, $f(\varepsilon_j,\varepsilon_k)$, are necessary

- Possible distributions for closed form formulation:
  - Type I Extreme Value distribution
  - Generalized Extreme Value

Closed Form Discrete Choice Model

- Advantages:
  - Standard probability equation
  - Estimation does not need simulation

- Disadvantage:
  - IIA assumption persists: Even within Nested/GEV structure
    - Various nesting structure allows overcoming group-specific IIA, but within a nest IIA exists
    - a priori specification of nesting is required
  - Closed form models are (mostly) homoscedastic

Mixing Distribution: Discrete Choice

- Utility functions of closed-form logit model:

$$
\begin{aligned}
U_1 &= V_1 + \varepsilon_1 \\
U_2 &= V_2 + \varepsilon_2 \\
U_3 &= V_3 + \varepsilon_3 \\
&\dots \\
U_J &= V_J + \varepsilon_J
\end{aligned}
$$

- J alternatives with J, ε that are of IID Type I EV random variables with scale μ

- Mix additional random errors (as opposed to considering the main random errors are multivariate normal)

$$
\begin{aligned}
U_1 &= V_1 + \xi_1 + \varepsilon_1 \\
U_2 &= V_2 + \xi_2 + \varepsilon_2 \\
U_3 &= V_3 + \xi_3 + \varepsilon_3 \\
&\dots \\
U_J &= V_J + \xi_4 + \varepsilon_J
\end{aligned}
$$

- Additional random variables, ξ_j have multivariate distributions

Mixed Logit Model

- Unconditional probability of closed form choice model:

$$\Pr(j|C_i) = \frac{\exp(\mu V_j)}{\sum_{k\in C_i} \exp(\mu V_k)}$$

- Resulting conditional choice models after mixing distribution:

$$\Pr(j|C_i) = \frac{\exp\left(\mu\left(V_j + \xi_j\right)\right)}{\sum_{k\in C_i} \exp \mu(V_k + \xi_k)}$$

- Unconditional choice models after mixing distribution:

$$\Pr(j|C_i) = \int \frac{\exp\left(\mu\left(V_j + \xi_j\right)\right)}{\sum_{k\in C_i} \exp \mu(V_k + \xi_k)} f(\xi) d\xi$$

Mixed Logit Model

- Mixed Logit model:

$$\Pr(j|C_i) = \int \frac{\exp\left(\mu\left(V_j + \xi_j\right)\right)}{\sum_{k \in C_i} \exp \mu(V_k + \xi_k)} f(\xi) d\xi$$

- Mixed Logit model (using Monte-Carlo):

$$\Pr(j|C_i) = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left(\mu\left(V_j + \xi_{j-r}\right)\right)}{\sum_{k \in C_i} \exp \mu(V_k + \xi_{k-r})}$$

R numbers of random draws from the multivariate distribution of ξ_j

- Different assumptions on multivariate distribution assumptions of ξ_j results in different mixed logit model

- As the core structure of the model remains a Logit model, Mixed logit is also called Logit Kernel Model

Heteroskedastic Mixed Logit

- Considering a multivariate normal random error with zero off-diagonals

$$U_j = V_j + \xi_j + \varepsilon_j \qquad j = 1,2,3,\ldots.J \in C_i$$

$$= \beta_{0j} + \sum(\beta x)_j + \xi_j + \varepsilon_j$$

IID Type I EV error

$$= (\beta_{0j} + \xi_j) + \sum(\beta x)_j + \varepsilon_j$$

Alternative Specific Constant: ASC

A multivariate normal error with **full** variance-covariance

- Example: for *J*=3

$$\xi = MVN\left(\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}\right)$$

Heteroskedastic Mixed Logit

- Mixing error specification:

$$Covariance(\xi) = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} . \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

- Simulating mixing error:

$$\xi = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

3 univariate (un-correlated) standard normal draws

Un-known parameters to estimate

$$U_1 = V_1 + \xi_1 + \varepsilon_1$$
$$U_2 = V_2 + \xi_2 + \varepsilon_2$$
$$U_3 = V_3 + \xi_3 + \varepsilon_3$$

$$\Pr(j) = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left(\mu(V_j + \sigma_j \eta_{j-r})\right)}{\sum_{k=1}^{J} \exp(\mu(V_k + \sigma_k \eta_{k-r}))}$$

Heteroskedastic Mixed Logit

- Presenting in the form of a factor loading: $U_j = V_j + F\xi + \varepsilon_j$
  - Example: for J=3

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \xi = \begin{bmatrix} \xi_1 & 0 & 0 \\ 0 & \xi_2 & 0 \\ 0 & 0 & \xi_3 \end{bmatrix} = T\eta \quad T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

  - Example: for J=4

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \xi = \begin{bmatrix} \xi_1 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 \\ 0 & 0 & 0 & \xi_4 \end{bmatrix} = T\eta \quad T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}$$

Error-Component Mixed Logit

- Considering a multivariate normal random error with non-zero off-diagonals

$$U_j = V_j + \xi_j + \varepsilon_j \qquad j = 1,2,3,\ldots.J \in C_i$$

$$= \beta_{0j} + \sum(\beta x)_j + \xi_j + \varepsilon_j$$

IID Type I EV error

$$= (\beta_{0j} + \xi_j) + \sum(\beta x)_j + \varepsilon_j$$

ASC

A multivariate normal error with **full** variance-covariance

- Example: for *J*=3

$$\xi \text{ follows a } MVN\left(\begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}\right)$$

Error-Component Mixed Logit

- Mixing error specification: $Variance(\xi)$

$$= \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} L_1 & 0 & 0 \\ L_{12} & L_2 & 0 \\ L_{13} & L_{23} & L_3 \end{bmatrix} . \begin{bmatrix} L_1 & L_{12} & L_{13} \\ 0 & L_2 & L_{23} \\ 0 & 0 & L_3 \end{bmatrix}$$

Cholesky factorization of the square matrix

- Simulating mixing error:

$$\xi = \begin{bmatrix} L_1 & 0 & 0 \\ L_{12} & L_2 & 0 \\ L_{13} & L_{23} & L_3 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

3 univariate (un-correlated) standard normal draws

Lower-triangular Cholesky factor as Unknown parameters to estimate

Error-Component Mixed Logit

- Mixed utility function:

$$U_1 = V_1 + L_1\eta_1 + \varepsilon_1$$
$$U_2 = V_2 + L_{12}\eta_1 + L_2\eta_2 + \varepsilon_2$$
$$U_3 = V_3 + L_{13}\eta_1 + L_{23}\eta_2 + L_3\eta_3 + \varepsilon_3$$

$$\Pr(1) = \frac{1}{R}\sum_{r=1}^{R}\frac{e^{(\mu(V_1+L_1\eta_{1-r}))}}{e^{(\mu(V_1+L_1\eta_{1-r}))} + e^{(\mu(V_2+L_{12}\eta_{1-r}+L_2\eta_{2-r}))} + e^{\left(\mu(V_3+L_{13}\eta_{1-r}+L_{23}\eta_{2-r}+L_3\eta_{3-r})\right)}}$$

$$\Pr(2) = \frac{1}{R}\sum_{r=1}^{R}\frac{e^{(\mu(V_2+L_{12}\eta_{1-r}+L_2\eta_{2-r}))}}{e^{(\mu(V_1+L_1\eta_{1-r}))} + e^{(\mu(V_2+L_{12}\eta_{1-r}+L_2\eta_{2-r}))} + e^{\left(\mu(V_3+L_{13}\eta_{1-r}+L_{23}\eta_{2-r}+L_3\eta_{3-r})\right)}}$$

$$\Pr(3) = \frac{1}{R}\sum_{r=1}^{R}\frac{e^{\left(\mu(V_3+L_{13}\eta_{1-r}+L_{23}\eta_{2-r}+L_3\eta_{3-r})\right)}}{e^{(\mu(V_1+L_1\eta_{1-r}))} + e^{(\mu(V_2+L_{12}\eta_{1-r}+L_2\eta_{2-r}))} + e^{\left(\mu(V_3+L_{13}\eta_{1-r}+L_{23}\eta_{2-r}+L_3\eta_{3-r})\right)}}$$

Error-Component Mixed Logit

- In the form of a factor loading: $U_j = V_j + F\xi + \varepsilon_j = V_j + F(T\eta) + \varepsilon_j$
  - Example: for J=3

$$F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad T = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{12} & L_{22} & 0 \\ L_{13} & L_{23} & L_{33} \end{bmatrix} \qquad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

  - for J=4

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad T = \begin{bmatrix} L_{11} & 0 & 0 & 0 \\ L_{12} & L_{22} & 0 & 0 \\ L_{13} & L_{23} & L_{33} & 0 \\ L_{14} & L_{24} & L_{24} & L_{44} \end{bmatrix} \qquad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}$$

- Heteroskedastic Mixed Logit is a special case of Error-Component Mixed Logit with all off-diagonal elements of T matrix forced to be zero

Error-Component Mixed Logit

- Full Error-Component model refers to a fully cross-nested mixed logit model. For example:

$$U_j = V_j + F\xi + \varepsilon_j = V_j + F(T\eta) + \varepsilon_j$$

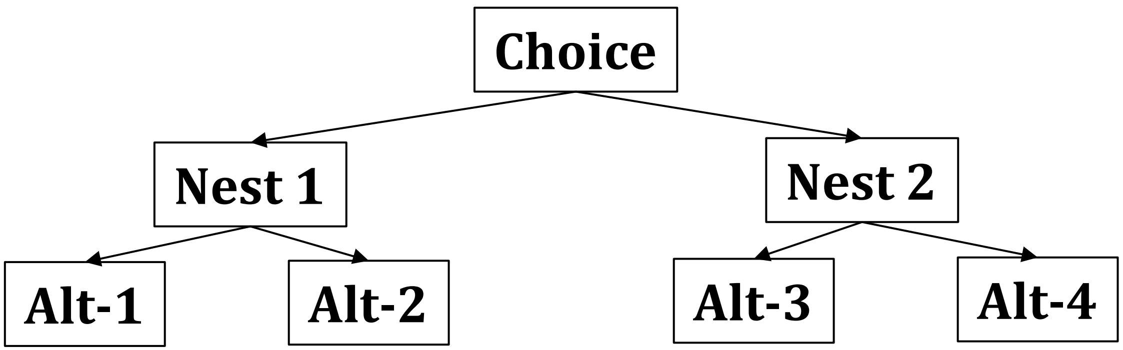
$$F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad T = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{12} & L_{22} & 0 \\ L_{13} & L_{23} & L_{33} \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

  - All 3 utility functions are correlated with each other
  - A full-blown cross-nested choice model

- Obviously, such full-scale correlation is never identified:
  - Primarily only utility differences matter
  - Numerical conditions: order, rank, positive definiteness

Simplified Error-Component Mixed Logit: Capturing Nesting

- Use targeted mixing to achieve specific nesting structure



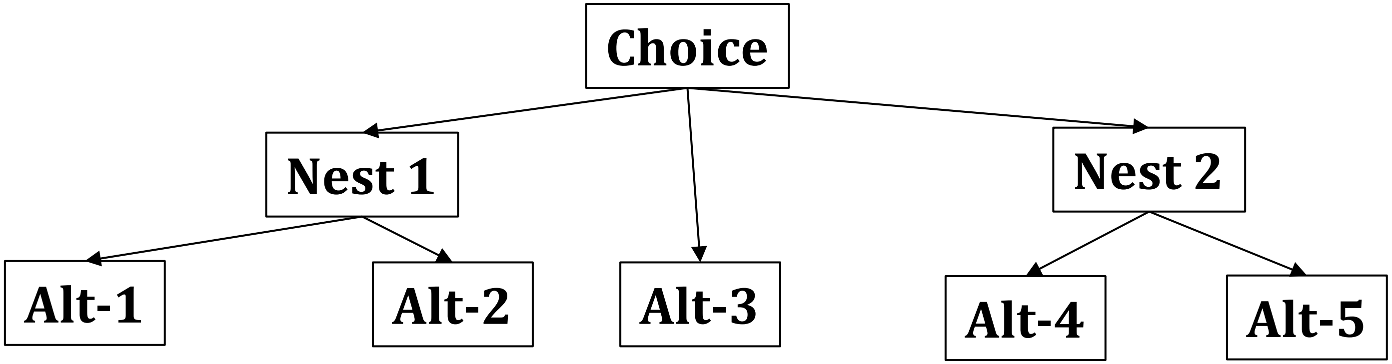
$$U_1 = V_1 + \xi_1 + \varepsilon_1$$
$$U_2 = V_2 + \xi_1 + \varepsilon_2$$
$$U_3 = V_3 \qquad + \xi_2 + \varepsilon_3$$
$$U_4 = V_4 \qquad + \xi_2 + \varepsilon_4$$

$$U_j = V_j + F(T\eta) + \varepsilon_j$$

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

Simplified Error-Component Mixed Logit: Capturing Nesting

- Use targeted mixing to achieve specific nesting structure



$$
\begin{aligned}
U_1 &= V_1 + \xi_1 + \varepsilon_1 \\
U_2 &= V_2 + \xi_1 + \varepsilon_2 \\
U_3 &= V_3 + \xi_2 + \varepsilon_3 \\
U_4 &= V_4 + \xi_3 + \varepsilon_4 \\
U_5 &= V_5 + \xi_3 + \varepsilon_5
\end{aligned}
$$

$$U_j = V_j + F(T\eta) + \varepsilon_j$$

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

Random Parameter Logit

- Considering a multivariate normal random distribution of the coefficient of a variable x (instead of ASC)

$$U_j = V_j + \varepsilon_j \qquad j = 1,2,3,\ldots.J \in C_i$$

$$= \beta_{0j} + \sum(\beta_j + \xi_j)\, x_j + \varepsilon_j$$

IID Type I EV error (ε_j)

A multivariate normal error (ξ_j) with **full, partial** or **Diagonal** variance-covariance

- For a full variance-covariance error, it captures random heterogeneity, heteroskedasticity and competitions
- For full or partial variance-covariance, but with unit diagonal element, it captures heterogeneity and competitions
- For only diagonal variance-covariance, it captures random heterogeneity and heteroskedasticity

Random Parameter Logit

$$U_j = V_j + \varepsilon_j \qquad j = 1,2,3,\dots.J \in C_i$$

$$= \beta_{0j} + \sum(\beta_j + \xi_j)\, x_j + \varepsilon_j$$

$$= \beta_{0j} + \sum(\beta_j + F\xi_j)\, x_j + \varepsilon_j$$

Following factor loading approach of specification

$$= \beta_{0j} + \sum(\beta_j + F(T\eta))\, x_j + \varepsilon_j$$

- Classical random coefficient Mixed Logit model

$$F = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \qquad T = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_J \end{pmatrix} \qquad \eta = \begin{pmatrix} \eta_1 \\ \cdot \\ \cdot \\ \eta_J \end{pmatrix}$$

- Frequently, only 1 or 2 alternatives have random coefficients and so the rest of the diagonal elements are forced to zero

Identification

Of the variance-covariance matrix (Ω) of random utility functions alternatives in the choice set, the following conditions are assessed:

- **Order condition** ('the number of rows time the number of columns'): defines the maximum number of parameters that can be estimated based on the number of alternatives in the choice set.
- **Rank condition** (max number of linearly independent columns of a matrix'): defines the actual number of parameters that can be estimated based on the number of independent equations available for estimation
- **Positive definiteness** ('non-zero determinant of a matrix'): Restrictions for maintaining the same covariance structure before and after normalization for identification restrictions
- **Empirical Identification**: None of the above can ensure an estimation unless empirical data supports the model structure. So, empirically, more restrictions may be necessary

Order Condition

Maximum number of parameter that can be identified based on the number of alternatives in the choice set: The number of distinct cells in the symmetric covariance matrix of random utility difference(Ω_Δ)

- As per order condition, maximum of $\left(S = \frac{J(J-1)}{2} - 1\right)$ alternative-specific parameters of the mixing covariance matrix (Ω) can be estimated
- $\left(S = \frac{J(J-1)}{2} - 1\right)$ is the total number of distinct cells in the covariance matrix of the utility differences (Ω_Δ) minus 1
    - ❑ 1 is deducted to set the scale parameter of the IID Gumbel of logit error.

- For 2 alts, *J*=2, s = 0: No alt-specific covariance term is identified
- For 3 alts, *J*=3, s = 2: up to 2 alt-specific covariance terms are identified
- For 4 alts, *J*=4, s = 5: up to 5 alt-specific covariance terms are identified
- For 5 alts, *J*=5, s = 9: up to 9 alt-specific covariance terms are identified

Rank Condition

- Rank of the Jacobian of the covariance matrix of utility differences (Ω_Δ) need to be derived:
    - Jacobian: the first derivative of vectorized covariance matrix of utility differences with respect to all unknown parameters of random errors
        - Rank: the maximum of the number of linearly independent rows and columns of the matrix
        - Rank can be automatically calculated using Gaussian elimination (by reducing the matrix into simple row echelon form) method
- Total number of parameters that can be identified is equal to the rank of the Jacobian minus 1
- Rank condition is more restrictive than the order condition

Positive Definiteness

- Following the Order and Rank conditions, the positive definiteness condition is necessary to normalization of un-identified parameters.
- There could be infinite possible solutions that can generate a particular hypothesized covariance structure
- So, normalization is necessary to establish the existence of a unique solution without changing the underlying preference structure
- Requires investigating the normalization necessary to the unique solution of the system of linear equation derived from the covariance matrix of utility differences

Identification: Order Condition

- **Heteroskedastic Mixed Logit: 2 alts**

Mixing errors: $F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$

Unknown Parameters(3): σ_1, σ_2 & μ

→ S = 0, so no variance are identified

- **Heteroskedastic Mixed Logit: 3 alts**

Mixing errors: $F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$

Unknown Parameters(4): σ_1, σ_2, σ_3& μ

  - S = 2, So, up to 2 variances are identified

Identification: Order Condition

- **Heteroskedastic Mixed Logit: 4 alts**

Mixing errors:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix}$$

Unknown Parameters (5): $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \& \mu$

- S = 5, so, **potentially** all variances are identified

- **Heteroskedastic Mixed Logit: 5 alts**

Unknown Parameters(6): $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5 \& \mu$

$\rightarrow$ S = 9, so, **potentially** all variances are identified

Covariance Matrix of Utility Difference

- Considering J alternatives with J^{th} as the reference: Ω_{Δ_j}

g= Constant of IID Gumbel error of logit

I: (J×J) identity matrix

$$\Omega = Covariance(U_j) = (FTT^T F^T) + \left(\frac{g}{\mu^2} I_J\right)$$

(J×1) utility functions

F: (J×M) Factor loading

T: (M×M) *Cholesky factors*

μ :The scale of IID Gumbel error of logit

$$\Omega = \quad (\Sigma) \quad + \quad (\Gamma)$$

Covariance Matrix of Utility Functions

- Considering J alternatives with J^{th} as the reference: Ω_{Δ_j}

g= Constant of IID Gumbel error of logit

I: (J-1)×(J-1) identity matrix

$$\Omega_{\Delta_j} = Covariance(\Delta_j U) = (\Delta_j F T T^T F^T \Delta_j{}^T) + \left(\Delta_j \frac{g}{\mu^2} I_J \Delta_j{}^T\right)$$

Δ_j: The Linear operator that transforms J utilities into a (J-1) utility differences taking J^{th} as the reference

F:Factor loading

T:Cholesky factor of the variance-covariance of mixed error term

μ :The scale of IID Gumbel error of logit

$$\Omega_{\Delta_j} = \quad (\Sigma) \quad + \quad (\Gamma)$$

Covariance Matrix of Utility Difference

The case of Heteroskedastic Mixed Logit Model

$$\Omega_\Delta = \left(\Delta_j F T T^T F^T \Delta_j{}^T\right) + \left(\Delta_j \frac{g}{\mu^2} I_J \Delta_j{}^T\right)$$

- Δ_J is a $(J\text{-}1)\times(J-1)$ identity matrix with a column vector of (-1) inserted as an additional J^{th} columns

For $J = 2$: $\Delta_J = [1 \quad -1]$

$$\boldsymbol{\Omega_\Delta} = \left(\Delta_j F T T^T F^T \Delta_j{}^T\right) + \left(\Delta_j \frac{g}{\mu^2} I_J \Delta_j{}^T\right)$$

$$\Omega_\Delta = [\sigma_{11} + \sigma_{22} + 2g/\mu^2]$$

Covariance Matrix of Utility Difference

The case of Heteroskedastic Mixed Logit Model

$$\Omega_\Delta = \left(\Delta_j F T T^T F^T \Delta_j{}^T\right) + \left(\Delta_j \frac{g}{\mu^2} I_J \Delta_j{}^T\right)$$

- Δ_J is a $(J\text{-}1)\times(J-1)$ identity matrix with a column vector of (-1) inserted as an additional J^{th} columns

$\mathbf{For}\ \boldsymbol{J = 3}$:

$$\Delta_J = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 & 0 \\ \sigma_{33} + g/\mu^2 & \sigma_{22} + \sigma_{33} + 2g/\mu^2 \end{bmatrix}$$

Covariance Matrix of Utility Difference

The case of Heteroskedastic Mixed Logit Model

$$\Omega_\Delta = \left(\Delta_j F T T^T F^T \Delta_j{}^T\right) + \left(\Delta_j \frac{g}{\mu^2} I_J \Delta_j{}^T\right)$$

- Δ_J is a $(J\text{-}1)\times(J-1)$ identity matrix with a column vector of (-1) inserted as an additional J^{th} columns

For $J = 4$:

$$\Delta_J = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{44} + 2g/\mu^2 & 0 & 0 \\ \sigma_{44} + g/\mu^2 & \sigma_{22} + \sigma_{44} + 2g/\mu^2 & 0 \\ \sigma_{44} + g/\mu^2 & \sigma_{44} + g/\mu^2 & \sigma_{33} + \sigma_{44} + 2g/\mu^2 \end{bmatrix}$$

Identification: Rank Condition

- Rank Condition: The total number of parameters that can be estimated is equal to the rank (r) of the Jacobian (with respect to the parameters of random errors) of the vectorized, Ω_{Δ} minus 1: (r-1) number of parameters are identified

- Rank is the number of linearly independent equations

The case of Heteroskedastic Mixed Logit Model

For $J = 2$:

- Rank condition is not necessary to check as the order condition proves that no alternative-specific covariance terms are identified

Identification: Rank Condition

The case of Heteroskedastic Mixed Logit Model

For $J = 3$:

$$\text{Vectorized}(\Omega_\Delta) = \text{Vec}(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 \\ \sigma_{22} + \sigma_{33} + 2g/\mu^2 \\ \sigma_{33} + g/\mu^2 \end{bmatrix}$$

- Rank is the rank (r) of the Jacobian (with respect to the unknown parameters of random errors) of the vectorized, Ω_Δ

$$Jacobian = \begin{bmatrix} \partial Vec(\Omega_\Delta)/\partial\sigma_{11} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{33} \\ \partial Vec(\Omega_\Delta)/\partial(g/\mu^2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$\rightarrow$: Rank =3. So, only (3-1)=2 can be estimated and rest should be normalized

Identification: Rank Condition

The case of Heteroskedastic Mixed Logit Model

For $J = 4$:

$$\text{Vectorized}(\Omega_\Delta) = \text{Vec}(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{44} + 2g/\mu^2 \\ \sigma_{22} + \sigma_{44} + 2g/\mu^2 \\ \sigma_{33} + \sigma_{44} + 2g/\mu^2 \\ \sigma_{44} + g/\mu^2 \end{bmatrix}$$

- Rank is the rank (r) of the Jacobian (with respect to the unknown parameters of random errors) of the vectorized, Ω_Δ

$$Jacobian = \begin{bmatrix} \partial Vec(\Omega_\Delta)/\partial\sigma_{11} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{22} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{33} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{44} \\ \partial Vec(\Omega_\Delta)/\partial(g/\mu^2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- : Rank =4. So, only (4-1)=3 can be estimated and rest should be normalized

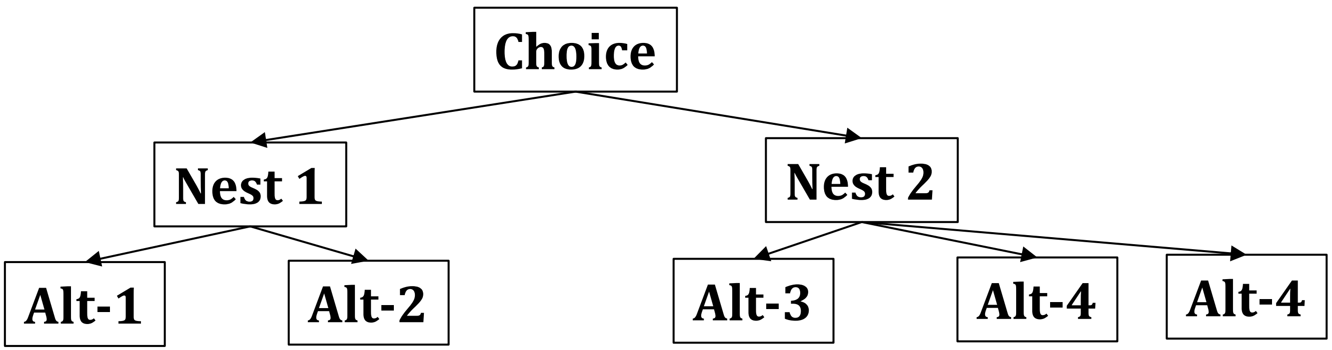
Identification: Positive Definiteness

The case of Heteroskedastic Mixed Logit Model

- Preferred normalization requires that the heteroskedastic term of the minimum variance alternative is restricted to zero:
  - A priori knowledge on lowest variance alternative does not exists
  - One has to try either normalizing the heteroskedastic term of different alternative and identify the best one that gives the best goodness of fit
  - Or, try estimating unidentified model to have clear idea of lowest variance alternative

Identification: Order Condition

- **Error-Component Mixed Logit: Nesting structure**



$$F = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

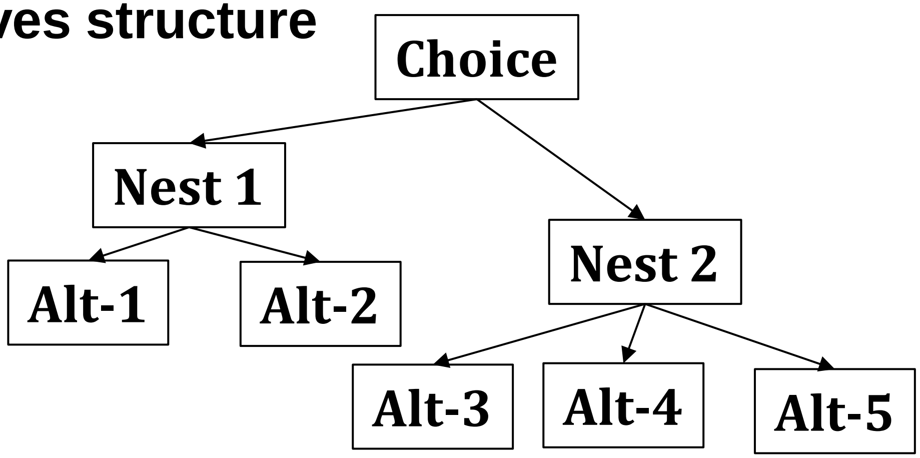
$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \frac{g}{\mu^2} & & & & \\ \sigma_{11} & \sigma_{11} + \frac{g}{\mu^2} & & & \\ 0 & 0 & \sigma_{22} + \frac{g}{\mu^2} & & \\ 0 & 0 & \sigma_{22} & \sigma_{22} + \frac{g}{\mu^2} & \\ 0 & 0 & \sigma_{22} & \sigma_{22} & \sigma_{22} + \frac{g}{\mu^2} \end{bmatrix}$$

Unknown Parameters: σ_1, σ_2 & μ

- S = 9, so, potentially all are identified

Covariance Matrix of Utility Difference

- **The case of 2 nest of 5 alternatives structure**



$$F = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

- Covariance of utility differences need to be carefully formed

  Considering the 5th as the reference:

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{22} + 2g/\mu^2 & & & \\ \sigma_{11} + \sigma_{22} + g/\mu^2 & \sigma_{11} + \sigma_{22} + 2g/\mu^2 & & \\ g/\mu^2 & g/\mu^2 & 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix}$$

Identification: Rank Condition

The case of 2 Nests and 5 Alternatives

$$\text{Vectorized}(\Omega_\Delta) = \text{Vec}(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{22} + 2g/\mu^2 \\ \sigma_{11} + \sigma_{22} + g/\mu^2 \\ g/\mu^2 \\ 2g/\mu^2 \end{bmatrix}$$

- Rank is the rank (r) of the Jacobian (with respect to the unknown parameters of random errors) of the vectorized, Ω_Δ

$$Jacobian = \begin{bmatrix} \partial Vec(\Omega_\Delta)/\partial\sigma_{11} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{22} \\ \partial Vec(\Omega_\Delta)/\partial(g/\mu^2) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

- : Rank =2. So, only (2-1) =1 can be estimated and rest should be normalized

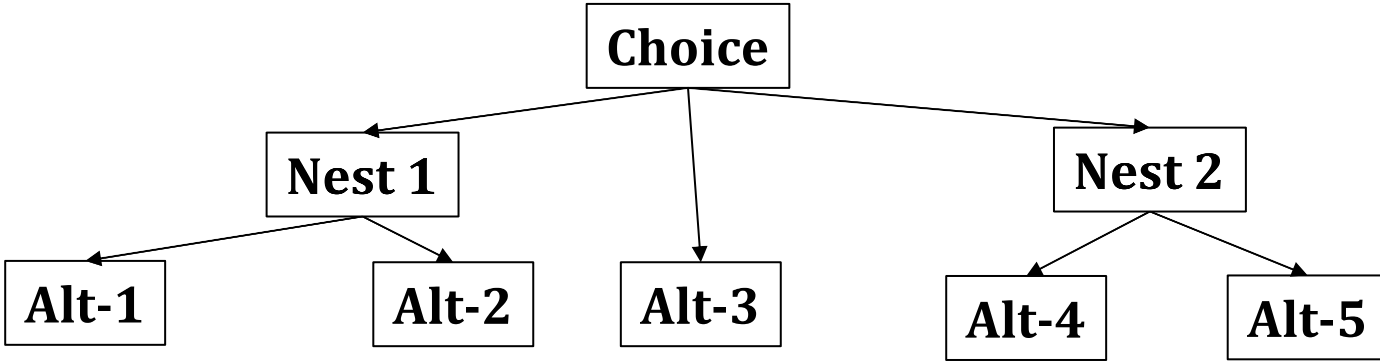
Identification: Positive Definiteness

The case of 2 nest Error-Component Model

- Three possible ways of normalization that satisfy positive definiteness condition:
  - Normalize σ_{11}= 0, estimate σ_{22}
  - Or, Normalize σ_{22}= 0, estimate σ_{11}
  - Normalize σ_{11}= σ_{22}, estimate σ
- This is valid for any two nest structure irrespective of number of alternatives in either nest

Identification: Order Condition

- **Error-Component Mixed Logit: Nesting structure**



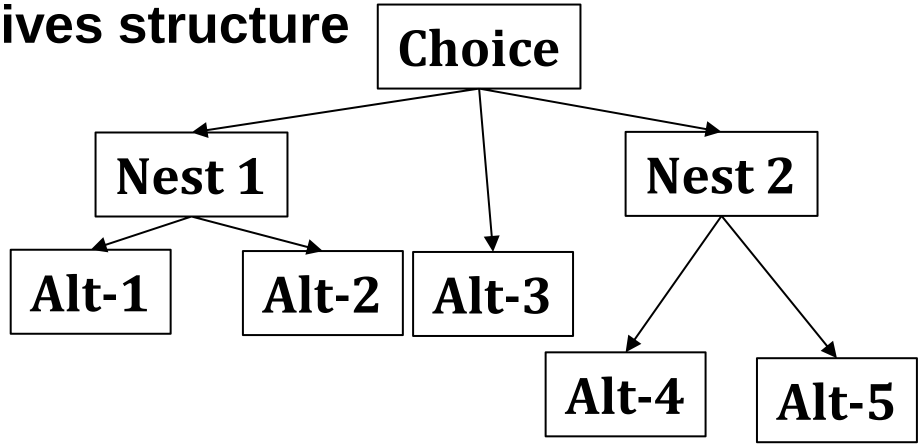
$$F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

Unknown Parameters: 3 $Cholesky\ factors + \ \mu$

- S = 9, so, potentially all are identified

Covariance Matrix of Utility Difference

- **The case of 3 nests of 5 alternatives structure**



$$F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad T = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

- Covaraince of Utility differences need to be carefully formed

  Considering the 5th as the reference:

$$\Omega_\Delta = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 & & & \\ \sigma_{11} + \sigma_{33} + g/\mu^2 & \sigma_{11} + \sigma_{33} + 2g/\mu^2 & & \\ \sigma_{33} + g/\mu^2 & \sigma_{33} + g/\mu^2 & \sigma_{22} + \sigma_{33} + 2g/\mu^2 & \\ g/\mu^2 & g/\mu^2 & g/\mu^2 & 2g/\mu^2 \end{bmatrix}$$

Identification: Rank Condition

The case of 3 Nests and 5 Alternatives

$$\text{Vectorized}(\Omega_\Delta) = \text{Vec}(\Omega_\Delta) = \begin{bmatrix} \sigma_{11} + \sigma_{33} + 2g/\mu^2 \\ \sigma_{11} + \sigma_{33} + g/\mu^2 \\ \sigma_{33} + g/\mu^2 \\ \sigma_{22} + \sigma_{33} + 2g/\mu^2 \\ g/\mu^2 \\ 2g/\mu^2 \end{bmatrix}$$

- Rank is the rank (r) of the Jacobian (with respect to the unknown parameters of random errors) of the vectorized, Ω_Δ

$$Jacobian = \begin{bmatrix} \partial Vec(\Omega_\Delta)/\partial\sigma_{11} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{22} \\ \partial Vec(\Omega_\Delta)/\partial\sigma_{33} \\ \partial Vec(\Omega_\Delta)/\partial(g/\mu^2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

- : Rank =4. So, only (4-1)=3 can be estimated. Potentially, only μ needs to be normalized

Identification: Positive Definiteness

The case of 2 nest Error-Component Model

- Investigating the normalization necessary to the unique solution of the system of linear equation derived from the covariance matrix of utility differences:

  - All three heteroskedastic terms (σ_{11}, σ_{22}, σ_{33}) are identified and can be estimated
  - This can be extended to any 3 nest case, where nests can have only 1 alternative as long as only one nest has at least 3 alternatives

Random Parameter Mixed Logit

- Random parameter mixed logit that does not capture competition between different alternatives, the identification restrictions related to Heteroskedastic model are valid
- Random parameter mixed logit can also use error-nesting approach to capture parameter distribution as well as competition.
    - Identification issues of error-nesting models are valid here
- In some cases, random parameters may need to have non-normal distribution:
    - Lognormal distribution, Triangular distribution or Truncated normal distribution for travel time/cost

Issues with Estimation

- Maximum Simulated Likelihood (MSL) estimation requires simulating random variables:
  - Monte Carlo simulation of any random variable (of any distribution type) requires generating uniform random variables
  - MSL requires that the generated uniform random variables are uncorrelated
  - Pseudo-random number generators induce serial correlations
    - Alternative methods to overcome serial correlations: Halton sequence, Scrambled Halton sequence, Latin hypercube sampling etc.

Other Logit Kernel Models

- Latent variable mixed logit model
- Random scale mixed logit model
- Latent class mixed logit model

Mixing Distribution: Others

- Closed-form model: Continuous, Ordered, Count

$$y_j = \sum \beta x + \varepsilon_j$$

- ε_j follows any legitimate marginal distribution

- Mix additional random errors

$$y_j = \sum \beta x + \xi_j + \varepsilon_j$$

$$y_j = \sum (\beta + \xi_j) x + \varepsilon_j$$

- Additional random variables, ξ_j can have multivariate distribution

- Such error mixing can allow
  - Random parameter (Random heterogeneity)
  - Correlated data points (e.g. panel data, time series, autocorrelation, etc.)

Further Topics

Bayesian Estimation of Mixed Logit model: Using Hierarchical Bayes Approach

- All random parameters
- Mix of random and fixed parameters