The log likelihood for nested logit can be written as:

$$\ell = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

The log likelihood for nested logit can be written as:

$$\ell = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

Decomposing into the nesting structure:

$$\ell = \sum_{i=1}^{N} \left[d_{iC} \log(P_{iC}) + \sum_{j \in J_B} d_{ij} \log(P_{iB} P_{ij|B}) \right]$$

The log likelihood for nested logit can be written as:

$$\ell = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

Decomposing into the nesting structure:

$$\ell = \sum_{i=1}^{N} \left[d_{iC} \log(P_{iC}) + \sum_{j \in J_B} d_{ij} \log(P_{iB} P_{ij|B}) \right]$$

$$= \sum_{i=1}^{N} \left[d_{iC} \log(P_{iC}) + \sum_{j \in J_B} d_{ij} \left\{ \log(P_{iB}) + \log(P_{ij|B}) \right\} \right]$$

Expanding the likelihood function from the previous slide:

$$\ell = \sum_{i=1}^{N} \left[d_{iC} \log(P_{iC}) + (d_{iBB} + d_{iRB}) \log(P_{iB}) + \sum_{j \in J_B} d_{ij} \log(P_{ij|B}) \right]$$

Expanding the likelihood function from the previous slide:

$$\ell = \sum_{i=1}^{N} \left[d_{iC} \log(P_{iC}) + (d_{iBB} + d_{iRB}) \log(P_{iB}) + \sum_{j \in J_B} d_{ij} \log(P_{ij|B}) \right]$$

Key insight: The likelihood decomposes into separate components that can be estimated sequentially

For sequential estimation, decompose utility as:

$$u_{iRB} = u_{iB} + u_{iRB|B}$$

For sequential estimation, decompose utility as:

$$u_{iRB} = u_{iB} + u_{iRB|B}$$

With normalizations:

$$\begin{aligned} u_{iC} &= 0 \\ u_{iBB|B} &= 0 \end{aligned}$$

For sequential estimation, decompose utility as:

$$u_{iRB} = u_{iB} + u_{iRB|B}$$

With normalizations:

$$\begin{aligned} u_{iC} &= 0 \\ u_{iBB|B} &= 0 \end{aligned}$$

We estimate parameters $(\beta_B, \beta_{RB}, \gamma, \lambda)$ where γ corresponds to alternative-specific variables (Z's)

The normalizations imply:

$$
\begin{aligned}
u_{iC} &= 0 \\
u_{iBB} &= \beta_B X_i + \gamma(Z_{BB} - Z_C) \\
u_{iRB} &= (\beta_B + \beta_{RB}) X_i + \gamma(Z_{RB} - Z_C)
\end{aligned}
$$

The normalizations imply:

$$
\begin{aligned}
u_{iC} &= 0 \\
u_{iBB} &= \beta_B X_i + \gamma(Z_{BB} - Z_C) \\
u_{iRB} &= (\beta_B + \beta_{RB})X_i + \gamma(Z_{RB} - Z_C)
\end{aligned}
$$

Stage 1: Estimate β_{RB} and γ using only observations that chose bus (N_B):

$$
\ell_1 = \sum_{i=1}^{N_B} \left\{ d_{iRB} \left[\frac{u_{iRB|B}}{\lambda} \right] - \log\left(1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right) \right\}
$$

Define the inclusive value I_{iB}:

$$I_{iB} = \log\left(1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right)$$

Define the inclusive value I_{iB}:

$$I_{iB} = \log\left(1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right)$$

This allows us to rewrite the nest probability term as:

$$\begin{aligned}\left[\exp\left(\frac{u_{iBB}}{\lambda}\right) + \exp\left(\frac{u_{iRB}}{\lambda}\right)\right]^{\lambda} &= \exp\left(\frac{u_{iBB}}{\lambda}\right)\left[1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right]^{\lambda} \\ &= \exp\left(\frac{u_{iBB}}{\lambda} + \lambda I_{iB}\right)\end{aligned}$$

Define the inclusive value I_{iB}:

$$I_{iB} = \log\left(1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right)$$

This allows us to rewrite the nest probability term as:

$$\begin{aligned}\left[\exp\left(\frac{u_{iBB}}{\lambda}\right) + \exp\left(\frac{u_{iRB}}{\lambda}\right)\right]^{\lambda} &= \exp\left(\frac{u_{iBB}}{\lambda}\right)\left[1 + \exp\left(\frac{u_{iRB|B}}{\lambda}\right)\right]^{\lambda} \\ &= \exp\left(\frac{u_{iBB}}{\lambda} + \lambda I_{iB}\right)\end{aligned}$$

Note: λI_{iB} represents the expected utility associated with the bus nest

Stage 2: Estimate β_B and λ using the inclusive value from Stage 1:

$$\ell_2 = \sum_i \left\{ d_{iB} \left[\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right] - \log \left(1 + \exp \left(\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right) \right) \right\}$$

Stage 2: Estimate β_B and λ using the inclusive value from Stage 1:

$$\ell_2 = \sum_i \left\{ d_{iB} \left[\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right] - \log \left(1 + \exp \left(\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right) \right) \right\}$$

This sequential estimation works because the log likelihood is additively separable:

$$\log(P_{iB}(\beta_B, \beta_{RB}, \gamma, \lambda)) + \log(P_{iRB|B}(\beta_{RB}, \gamma))$$

Stage 2: Estimate β_B and λ using the inclusive value from Stage 1:

$$\ell_2 = \sum_i \left\{ d_{iB} \left[\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right] - \log \left(1 + \exp \left(\frac{u_{iBB}}{\lambda} + \lambda I_{iB} \right) \right) \right\}$$

This sequential estimation works because the log likelihood is additively separable:

$$\log(P_{iB}(\beta_B, \beta_{RB}, \gamma, \lambda)) + \log(P_{iRB|B}(\beta_{RB}, \gamma))$$

Parameters β_{RB} and γ are identified using only within-nest variation

Then we treat them as given when estimating between-nest parameters β_B and λ

Instead of sequential estimation, can estimate all parameters simultaneously (FIML):

$$\ell(\beta_B, \beta_{RB}, \gamma, \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

Instead of sequential estimation, can estimate all parameters simultaneously (FIML):

$$\ell(\beta_B, \beta_{RB}, \gamma, \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

Where the choice probabilities are:

$$P_{ij} = \Pr(\text{nest}) \times \Pr(j|\text{nest})$$

Instead of sequential estimation, can estimate all parameters simultaneously (FIML):

$$\ell(\beta_B, \beta_{RB}, \gamma, \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log(P_{ij})$$

Where the choice probabilities are:

$$P_{ij} = \mathrm{Pr}(\mathrm{nest}) \times \mathrm{Pr}(j|\mathrm{nest})$$

e.g.

$$P_{i,\mathrm{RB}} = \underbrace{\frac{\left[\exp(u_{i,\mathrm{RB}}/\lambda) + \exp(u_{i,\mathrm{BB}}/\lambda)\right]^{\lambda}}{\left[\exp(u_{i,\mathrm{RB}}/\lambda) + \exp(u_{i,\mathrm{BB}}/\lambda)\right]^{\lambda} + \exp(u_{i,\mathrm{C}})}}_{\mathrm{Pr(Bus)}} \times \underbrace{\frac{\exp(u_{i,\mathrm{RB}}/\lambda)}{\exp(u_{i,\mathrm{RB}}/\lambda) + \exp(u_{i,\mathrm{BB}}/\lambda)}}_{\mathrm{Pr(RB|Bus)}}$$

Advantages of FIML:

- Asymptotically efficient (achieves Cramér-Rao lower bound)
- Provides correct standard errors in one step
- Allows for joint hypothesis testing across stages

Advantages of FIML:

- Asymptotically efficient (achieves Cramér-Rao lower bound)
- Provides correct standard errors in one step
- Allows for joint hypothesis testing across stages

Disadvantages of FIML:

- Computationally more demanding
- Less stable convergence (especially for λ near boundary)
- Requires good starting values

Further considerations:

- Log-likelihood function is not globally concave

Further considerations:

- Log-likelihood function is not globally concave
- Common parameters across submodels estimated separately in sequential approach

Further considerations:

- Log-likelihood function is not globally concave
- Common parameters across submodels estimated separately in sequential approach
- FIML automatically enforces equality constraints

Further considerations:

- Log-likelihood function is not globally concave
- Common parameters across submodels estimated separately in sequential approach
- FIML automatically enforces equality constraints
- Must satisfy $0 < \lambda_k \leq 1$ for model to be consistent with utility maximization