Global Initiative of Academic Networks (GIAN)

UNDERSTANDING THE INCREASING RELEVANCE OF CHOICE MODELS FOR ADVANCING TRANSPORTATION MODELLING IN SMART CITIES

MAY 22 -26, NAGPUR, INDIA

Module 4

Naveen Eluru, *University of Central Florida*

COURSE MODULES

Introduction	• Introduction to Smart City Technologies, their impact on Transportation
Stated Preference Module	• Background on Data Collection Approaches • Stated Preference Design and application
Traditional Discrete Choice Models	• Binary logit, multinomial logit, ordered logit, and count models
Advanced Discrete Choice Models	• Nested logit, mixed logit, maximum simulated likelihood estimation, regret minimization, discrete continuous models
Transportation Planning	• Current state of the art and recent advances

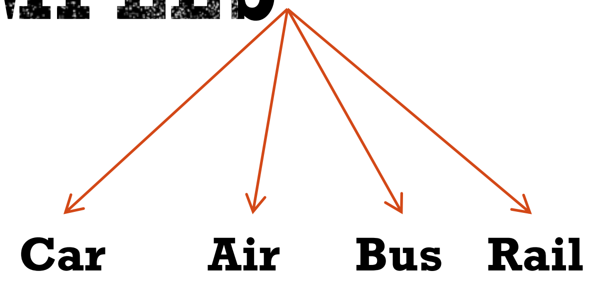
IN THIS MODULE

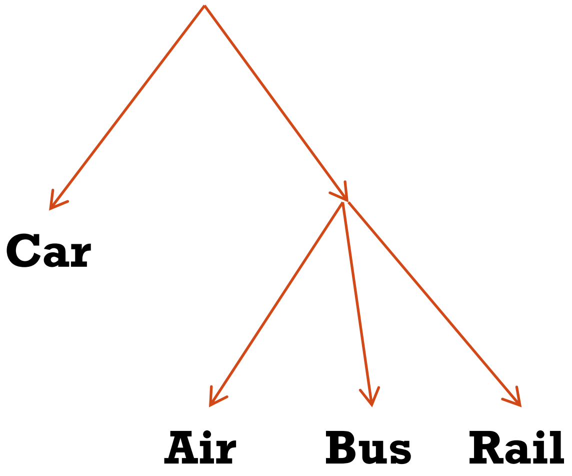
I will build on the basic choice modeling approaches for data analysis and introduce Nested Logit models, GEV models, Mixed Logit model, latent class models, discrete-continuous models and multiple discrete extreme value models

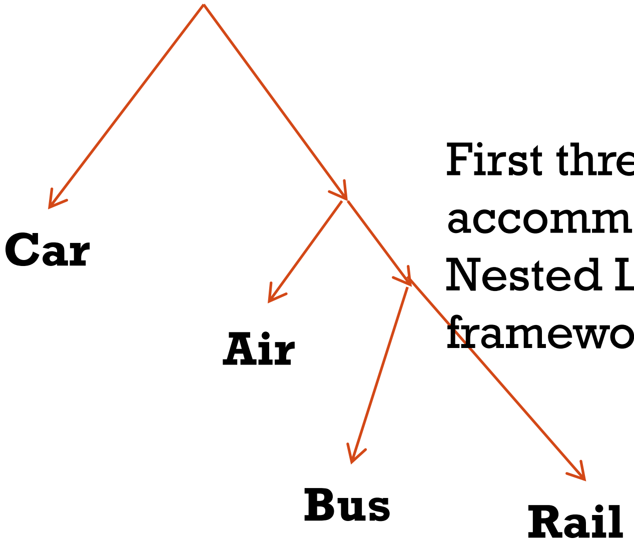
NESTED LOGIT MODEL

- MNL - $P_{in} = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$
- Independent errors
  - Consider mode choice model, with 3 alternatives car, bus and metro. A person whose personality prefers transit modes will assign a higher value to both bus and metro. Neglecting this might have implications for what we are trying to do
    - $V_{in} + \varepsilon_{in}$
- There is a "stickiness" associated to a set of alternatives i.e. the behavior of the alternatives in the "set" is different from the alternatives not in the set
- Within a set however, the behavior is again similar to that in MNL

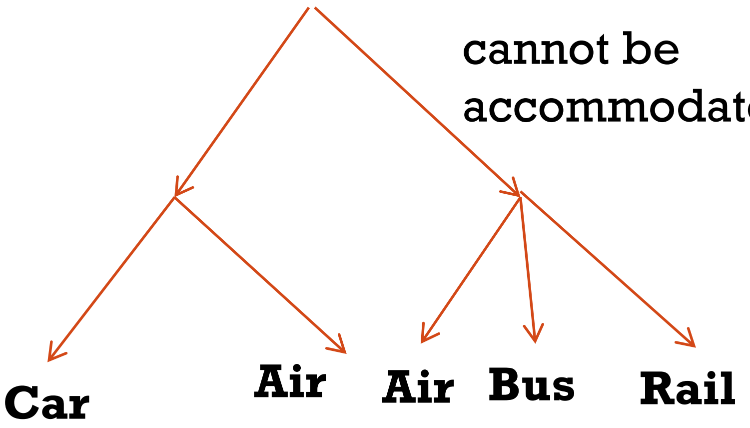
EXAMPLES







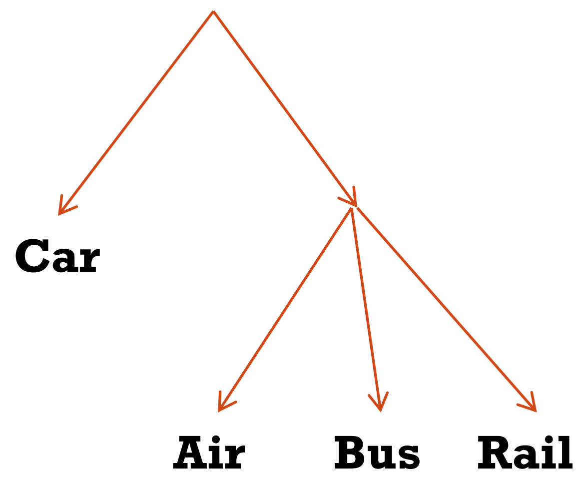
First three can be accommodated in the Nested Logit framework



cannot be accommodated

FORMULATION

- Consider the 4 alternatives: Car, Air, Bus Rail



- $U_C = V_C + \varepsilon_{Car}$
- $U_A = V_A + \varepsilon_A + \varepsilon_{cc}$
- $U_B = V_B + \varepsilon_B + \varepsilon_{cc}$
- $U_R = V_R + \varepsilon_R + \varepsilon_{cc}$
  - where ε_{cc} represents the common error term for the common carriers
- Overall error is still identical i.e. $\varepsilon_{Air} = \varepsilon_A + \varepsilon_{cc}$, $\varepsilon_{Bus} = \varepsilon_B + \varepsilon_{cc}$, $\varepsilon_{Rail} = \varepsilon_R + \varepsilon_{cc}$
- ε_{Car}, ε_{Air}, ε_{Bus}, and ε_{Rail} are distributed G(0,1)
- Now lets say ε_A, ε_B, and ε_R are Gumbel G(0,θ)
  - $(0 < \theta \leq 1)$

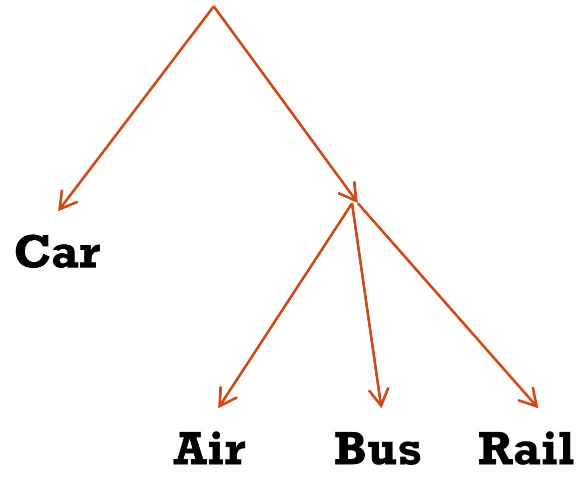
FORMULATION

- Assuming each pair of the error terms in ε_{Air}, ε_{Bus}, and ε_{Rail} are independent we can compute Var(ε_{cc}) as
  - $\frac{\Pi^2}{6} - \frac{\Pi^2\theta^2}{6}$
- Correlation (U_A,U_B) = Correlation (U_A,U_R) = Correlation (U_B,U_R)
- Correlation(a,b) = $\frac{\text{covariance(a,b)}}{[\text{var(a)} * \text{var(b)}]^{1/2}}$

FORMULATION

- In our case, covariance(U_A,U_B) = Var(ε_{cc}) ;
- Var(U_A) = Var(U_B)= Var(U_R) = $\frac{\Pi^2}{6}$
- So Correlation (U_A,U_B) = Correlation (U_A,U_R) = Correlation (U_B,U_R)
  - $= \{ \frac{\Pi^2}{6} - \frac{\Pi^2\theta^2}{6} \} / \frac{\Pi^2}{6} = (1- \theta^2)$
- Correlation is $(1- \theta^2)$
- Hence when we test our hypothesis (which is to see if correlation exists), we do not test if θ is different from 0, but if θ is different from 1
  - Null hypothesis is θ=1

FORMULATION



- Now how do we generate the probability expressions
- Now lets consider the nest
- $P_{air} | cc = \frac{\exp(\frac{V_A}{\theta})}{\exp\left(\frac{V_A}{\theta}\right)+\exp\left(\frac{V_B}{\theta}\right)+\exp(\frac{V_R}{\theta})}$
- Now when we need to generate the probability for the car or cc we somehow need to compute a net utility for the cc
- Now the choice between car and cc is determined as $U_{car} > \mathrm{Max}(U_A, U_B, U_R)$
- For a gumbel distribution $G(V_1,\theta), G(V_2,\theta), G(V_3,\theta)$
- $\mathrm{Max}\ (V1, V2, V3) = G[\theta \ln(\exp\left(\frac{V_1}{\theta}\right) + \exp\left(\frac{V_2}{\theta}\right) + \exp(\frac{V_3}{\theta})), \theta)$

FORMULATION

- In our case
- Max (U1, U2, U3)
- $= \theta\{\ln(\exp\left(\frac{V_A}{\theta}\right) + \exp\left(\frac{V_B}{\theta}\right) + \exp(\frac{V_R}{\theta}))\} + \varepsilon^*$
- This is effectively the composite nest utility
- $\Gamma = \{\ln(\exp\left(\frac{V_A}{\theta}\right) + \exp\left(\frac{V_B}{\theta}\right) + \exp(\frac{V_R}{\theta}))\}$
- P_{Car} = Prob [U_c > Max (U_A, U_B, U_R)]
- = Prob [$V_c + \varepsilon_{Car}$ > Max (U_A, U_B, U_R)]
  G(0,1)
- = Prob [$V_c + \varepsilon_{Car} > \theta\Gamma + \varepsilon^* + \varepsilon_{cc}$]

FORMULATION

- $P_{Car} = \frac{\exp(V_C)}{\exp(V_C)+\exp(\theta\Gamma)}$
- $P_{cc} = \frac{\exp(\theta\Gamma)}{\exp(V_C)+\exp(\theta\Gamma)}$
  - θ – log–sum parameter
  - Γ – log-sum variable
- $P_{air}|cc = \frac{\exp(\frac{V_A}{\theta})}{\exp\left(\frac{V_A}{\theta}\right)+\exp\left(\frac{V_B}{\theta}\right)+\exp(\frac{V_R}{\theta})}$

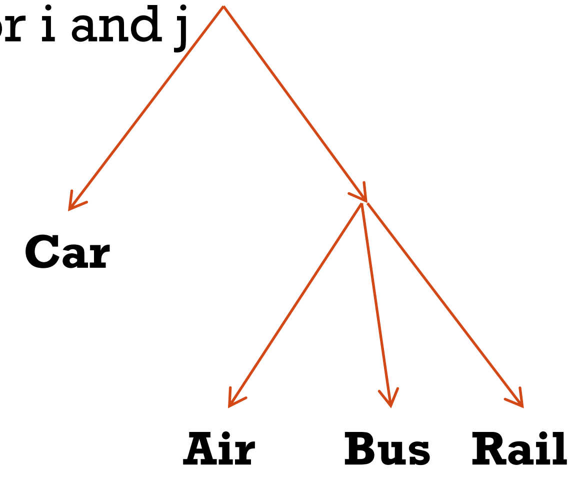
- $P_{air} = P_{air}|cc * P_{cc}$
- Similar to P_{Bus}, P_{Rail}
- To get MNL from NL set $\theta = 1$
- Test it now

FORMULATION

- $P_{air} = P_{air|cc} * P_{cc}$
- $\theta = 1$
- $= \dfrac{\exp(\frac{V_A}{\theta})}{\exp\left(\frac{V_A}{\theta}\right)+\exp\left(\frac{V_B}{\theta}\right)+\exp(\frac{V_R}{\theta})} * \dfrac{\exp(\theta\Gamma)}{\exp(V_C)+\exp(\theta\Gamma)}$
- $= \dfrac{\exp(V_A)}{\exp(V_A)+\exp(V_B)+\exp(V_R)} * \dfrac{\exp(\{\ln(\exp(V_A)+\exp(V_B)+\exp(VR)\})}{\exp(V_C)+\exp\left(\{\ln(\exp\left(\frac{V_A}{\theta}\right)+\exp\left(\frac{V_B}{\theta}\right)+\exp(\frac{V_R}{\theta}))\}\right)}$
- $= \dfrac{\exp(VA)}{\exp(V_C)+\exp(V_A)+\exp(V_B)+\exp(VR)}$ -> MNL

IIA PROPERTY

- **MNL** - Consider the ratio of alternative probabilities for i and j.
- $P_i / P_j = \frac{\exp(Vi)}{\sum_{\forall j}\exp(Vj)} \Big/ \frac{\exp(Vj)}{\sum_{\forall j}\exp(Vj)} = \frac{\exp(Vi)}{\exp(Vj)} = \exp(V_i - V_j)$
- **NL** - Consider the ratio of alternative probabilities for i and j
- $P_A/P_R = [P_{air}|cc * P_{cc}] / [P_{rail}|cc*P_{cc}]$
- $= [P_{air}|cc] / [P_{rail}|cc]$
  - Simplifies exactly like the MNL
- $P_A/P_C = [P_{air}|cc * P_{cc}] / [P_{car}]$
  - Does not simplify
- Alternatives within the nest still act as if they are part of the MNL structure
- Alternatives outside the next exhibit different substitution patterns

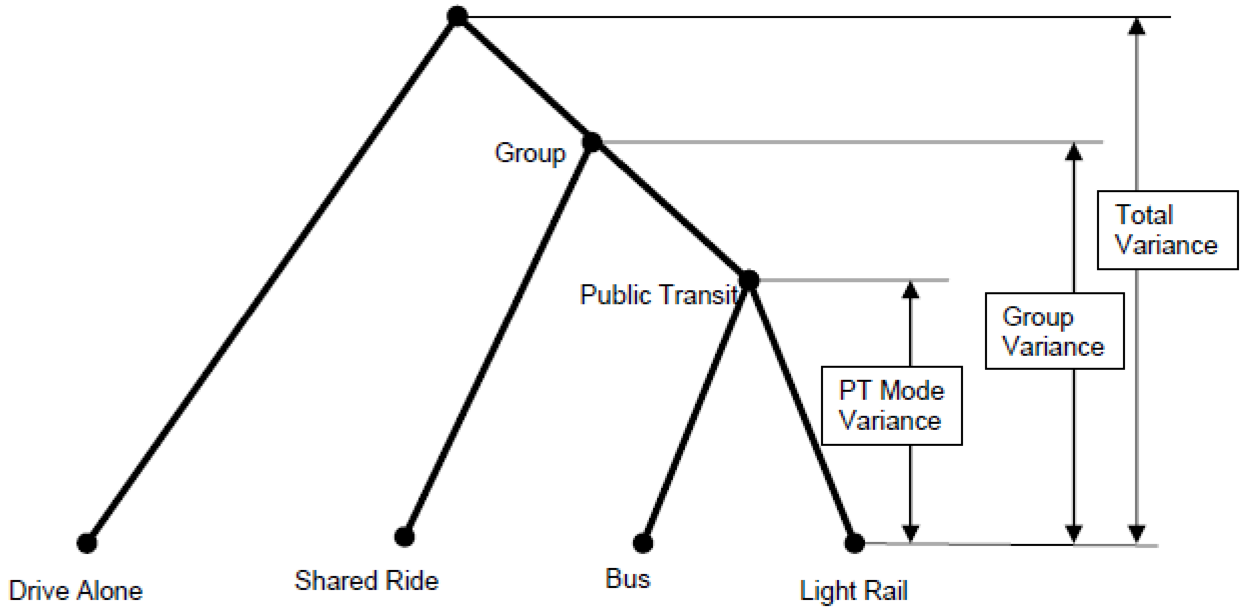


ELASTICITY

- Self-Elasticity
  - For Non-nested alternatives (same as MNL)
    - $\beta_k[1\text{-}P_{\text{in}}]\ \mathrm{x}_{ik}$
  - For Nested alternatives
    - $\beta_k\{[1\text{-}P_{\text{in}}] + \frac{1-\theta}{\theta}(1\text{-}\ P_{\text{in}}|\mathrm{N})\}\ \mathrm{x}_{ik}$

- Cross-Elasticity
  - Effect on Non-nested alts for change in Non-nested alts
    - $-\beta_k[P_{\text{in}}]\ \mathrm{x}_{ik}$
  - Effect on Non-nested alts for change in Nested alts
    - $-\beta_k[P_{\text{in}}]\ \mathrm{x}_{ik}$
  - Effect on Nested alts for change in Non-nested alts
    - $-\beta_k[P_{\text{in}}]\ \mathrm{x}_{ik}$
  - Effect on Nested alts for change in Nested alts
    - $-\beta_k\{[P_{\text{in}}] + \frac{1-\theta}{\theta}(P_{\text{in}}|\mathrm{N})\}\ \mathrm{x}_{ik}$

ANOTHER EXAMPLE

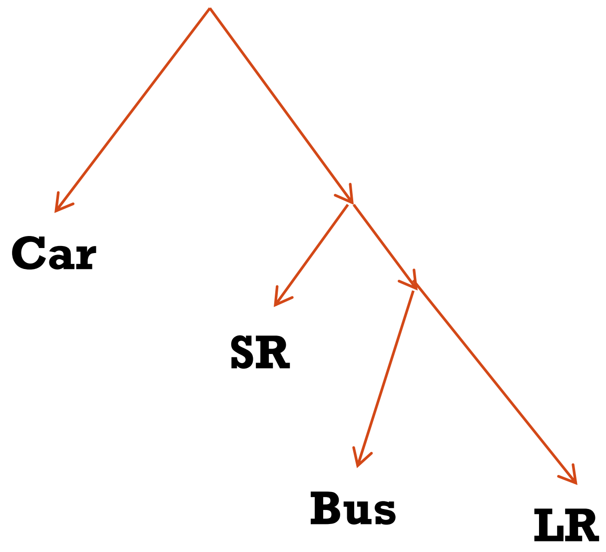


$$
\begin{aligned}
U_{DA} &= V_{DA} + \varepsilon_{DA} \\
U_{SR} &= V_{GRP} + V_{SR} + \varepsilon_{GRP} + \varepsilon_{SR} \\
U_{Bus} &= V_{GRP} + V_{PT} + V_{Bus} + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{Bus} \\
U_{LTR} &= V_{GRP} + V_{PT} + V_{LTR} + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{LTR}
\end{aligned}
$$

→ G(0,1)

ANOTHER EXAMPLE

- ε_{SR} - $G(0,\theta)$
- $\varepsilon_{B,}\ \varepsilon_{LR}\ G(0,\eta)$
- $0< \eta< \theta <1$



- Same approach as the previous case to generate the probabilities
- Read Section 8.3 of Bhat and Koppelman 2006 for exact probability expressions
  - Koppelman, F.S. and Bhat, C., 2006. A self instructing course in mode choice modeling: multinomial and nested logit models

REMARKS

- Multinomial logit model needs to be estimated first
- Account for systematic effects
- Then attempt to incorporate correlation
- Estimate NL model and if $\theta = 1$ it indicates MNL is good enough
- In case $\theta > 1$ then the formulation is not consistent with utility framework

18

GEV CLASS OF MODELS

GEV CLASS OF MODELS

- Generalized Extreme Value class of models
- GEV models constitute a class of models that accommodate different substitution patterns
- The common aspect of these models is the fact that the error terms of the utility equations are extreme value distributed
- There are a whole set of GEV models:
  - MNL, NL, Paired combinatorial logit (PCL), Cross-nested logit (CNL), generalized MNL (GenMNL), ordered GEV model, fuzzy nested logit (FNL)
- Before we start the discussion of specific model structures there are some pre-defined properties that need to be adhered to generate different GEV models

GEV MODEL PROPERTIES

- GEV family of models adhere to a set of rules
- If certain conditions are met then we can create a GEV model based on that with the probability expression as follows
- Consider a function G that depends on Y_j for all j where $G = G(Y_1, Y_2, \ldots Y_J)$ and $Y_j = \exp(V_j)$
- G_j represent the derivative of G w.r.t Y_j ($G_j = \frac{\partial G}{\partial Y_j}$)
- $P_j = \frac{Y_j G_j}{G}$

GEV MODEL PROPERTIES

- Conditions to be met
  - $G \geq 0$ for all positive values of Yj
  - $G(\rho Y_1, \ldots, \rho Y_J) = \rho G(Y_1, \ldots, Y_J)$
  - $G \rightarrow \infty$ as $Yj \rightarrow \infty$ for any j
  - Odd order partial derivatives are non-negative & even order partial derivatives are non-positive (Gi ≥ 0, Gij ≤ 0 and so on)
- There is little intuition behind this approach. However, it can be shown to work very well. Let us illustrate this

GEV CLASS OF MODELS

- The only thing we can say is if you choose a G that satisfies the 4 conditions, you can generate the probability function from that.
- Let G = $\sum_{l=1}^{J} Y_l$
- Now check the four conditions
  - Is G positive for all values of Y_j ---- YES
  - Is it a linear function --- YES
  - G $\rightarrow\infty$ as Yj $\rightarrow\infty$ --- YES
  - Partial derivatives: first derivative is 1 and all subsequent derivatives are 0, hence they satisfy the condition of Gi ≥ 0, Gij ≤ 0 --- YES
- Now what is the probability?
  - $P_j = \frac{Y_j G_j}{G} = \frac{Y_j}{\sum_{l=1}^{J} Y_j} = \frac{Y_j}{\sum_{l=1}^{J} Y_l} = \frac{\exp(Vj)}{\sum_{l=1}^{J} \exp(Vl)}$
- So we have generated the MNL model using the GEV family

GEV CLASS OF MODELS

- Now let us try to generate the NL
- G = $\sum_{p=1}^{K}\left(\sum_{l=1}^{J} Y_l^{1/\lambda_p}\right)^{\lambda_p}$
- You can check all the four conditions
  - G ≥ 0 for all positive values of Yj - Valid
  - $G(\rho Y_1, \ldots, \rho Y_J) = \rho G(Y_1, \ldots, Y_J)$ – adding ρ will let it come out of the function
  - $G \rightarrow \infty$ as $Yj \rightarrow \infty$ for any j - Valid
  - Odd order partial derivatives are non-negative & even order partial derivatives are non-positive – Will work

GEV CLASS OF MODELS

- $G_j = Y_j^{(1/\lambda_j - 1)} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_j - 1}$
- $P_j = \dfrac{Y_j Y_j^{1/\lambda_j - 1} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_j - 1}}{\sum_{p=1}^{K} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_p}} = \dfrac{Y_j^{1/\lambda_j} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_j - 1}}{\sum_{p=1}^{K} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_p}}$

$= \left[\dfrac{Y_j^{1/\lambda_j}}{\left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_p}} \right] * \left[\dfrac{\left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_p}}{\sum_{p=1}^{K} \left(\sum_{l=1}^{J} Y_l^{1/\lambda_p} \right)^{\lambda_p}} \right]$; replace $Y_j = \exp(V_j)$

First component is the within nest choice and second component the nest probability

OTHER GEV MODELS

- Paired Combinatorial Logit
- $G = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left(Y_k^{1/\lambda_{kl}} + Y_l^{1/\lambda_{kl}}\right)^{\lambda_{kl}}$
- $P_i = \sum_{j \neq 1} \left(\frac{(\alpha e^{V_i})^{1/\theta_{ij}}}{(\alpha e^{V_i})^{1/\theta_{ij}} + (\alpha e^{V_j})^{1/\theta_{ij}}} \times \frac{[(\alpha e^{V_i})^{1/\theta_{ij}} + (\alpha e^{V_j})^{1/\theta_{ij}}]^{\theta_{ij}}}{\sum_{k=1}^{J-1} \sum_{m=k+1}^{J} [(\alpha e^{V_k})^{1/\theta_{km}} + (\alpha e^{V_m})^{1/\theta_{km}}]^{\theta_{km}}} \right),$

- Generalized Nested Logit
- $G = \sum_{k=1}^{K} \left(\sum_{j \epsilon Bk} (\alpha_{jk} Y_j)^{1/\lambda_k} \right)^{\lambda_k}$
- $P_i = \sum_{m} \left(\frac{(\alpha_{im} e^{V_i})^{1/\theta_m}}{\sum_{j \in N_m} (\alpha_{jm} e^{V_j})^{1/\theta_m}} \times \frac{\left(\sum_{j \in N_m} (\alpha_{jm} e^{V_j})^{1/\theta_m} \right)^{\theta_m}}{\sum_{m'} \left(\sum_{j \in N_{m'}} (\alpha_{jm'} e^{V_j})^{1/\theta_{m'}} \right)^{\theta_{m'}}} \right),$

ARE GEV MODELS ENOUGH?

- GEV models are advantageous when we expect correlation across alternatives
- However it is possible that individuals exhibit distinct choice behavior of their own for every attribute i.e. Taste variation
- For instance, individuals are negatively sensitive to travel time – we have seen this
- Now in our models we assume that the sensitivity to travel time is same for every body in the population
- Is this truly valid?
- Not necessarily

ARE GEV MODELS ENOUGH?

- How can we account for such taste variations
  - Segmentation approach we discussed – allows for people to be in different segments and hence have distinct taste variation profiles for each segment
  - However, even in this approach, we restrict ourselves to 3-4 possible segments in the data
- Ideally, as data modellers we are interested in identifying the true parameter for every individual in the dataset.
- This is quite tricky and infeasible (unless we collect multiple records for each person)
- So, now we will examine the approach that allows us to capture such to possibly generate individual level parameters

MIXED MULTINOMIAL LOGIT MODEL

- Intuition
  - In the MNL model we estimate a single parameter to determine the influence of an exogenous variable on the choice process
    - For example, we claim that the influence of income is the same for the entire population. However, based on whether the respondent is lavish or conservative with money the influence varies. But, we cannot accommodate for such taste variations in the MNL
- In a MMNL model we allow the coefficients to vary across different individuals
- We accommodate for correlation across the error terms for different alternatives (relaxing the independence assumption)
- We incorporate different error variances (relaxing the identical assumption)

MIXED MULTINOMIAL LOGIT MODEL

- The MMNL model involves the integration of the MNL formulation over the unobserved parameters

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta)\, f(\beta \mid \theta)\, d(\beta),$$

where

$$L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qi}}}$$

- The MMNL model can be formulated from two unique but equivalent formulations:
  - Error components
  - Random coefficients

ERROR COMPONENTS

- Consider utility of person *q* for alternative *i*

$$U_{qi} = \gamma' y_{qi} + \zeta_{qi}$$
$$= \gamma' y_{qi} + \mu' z_{qi} + \varepsilon_{qi}$$

Random component

Systematic component

Induces heteroscedasticity and correlation across error based on how we specify Z_{qi}

Example of Z_{qi}

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

RANDOM COMPONENTS

- Consider utility of person *q* for alternative *i*

$$U_{qi} = \beta'_q x_{qi} + \varepsilon_{qi}$$

Random coefficients

$$\beta_{qk} \sim N(\mu_k, \sigma_k)$$

$$U_{qi} = \alpha_{qi} + \sum_{k=1}^{K} \beta_{qk} x_{qik} + \varepsilon_{qi}$$

$$\beta_{qk} = \mu_k + \sigma_k s_{qk}$$

Fixed coefficients

$$V_{qi} = \alpha_{qi} + \sum_{k} \mu_k x_{qik}$$

$$P_{qi} = \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi} + \sum_k \sigma_k s_{qk} x_{qik}}}{\sum_j e^{V_{qj} + \sum_k \sigma_k s_{qk} x_{qjk}}} d\Phi(s_{q1}) d\Phi(s_{q2}) \ldots d\Phi(s_{qK}) \right\}$$

MIXED MULTINOMIAL LOGIT MODEL

- Estimation

$$P_{qi} = \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi}+\sum_k \sigma_k s_{qk} x_{qik}}}{\sum_j e^{V_{qj}+\sum_k \sigma_k s_{qk} x_{qjk}}} d\Phi(s_{q1}) d\Phi(s_{q2}) \ldots d\Phi(s_{qK}) \right\}$$

- The probabilities are approximated through simulation
- For any given value of σ (1,2,..K), draw a S_q (1,2,..K) and compute P_{qi}. Repeat this multiple times and average the P_{qi}.

$$\tilde{P}_{qi} = \frac{1}{R} \sum_{1}^{R} P_{qi}$$

MIXED MULTINOMIAL LOGIT MODEL

- Log-likelihood function

$$SL = \sum_{1}^{Q} \sum_{i=1}^{I} d_{qi} \tilde{P}_{qi}$$

d_{qi} is 1 if q chose i; else 0

- Now that we have the LL function, we undertake Maximum Likelihood to get our estimates!

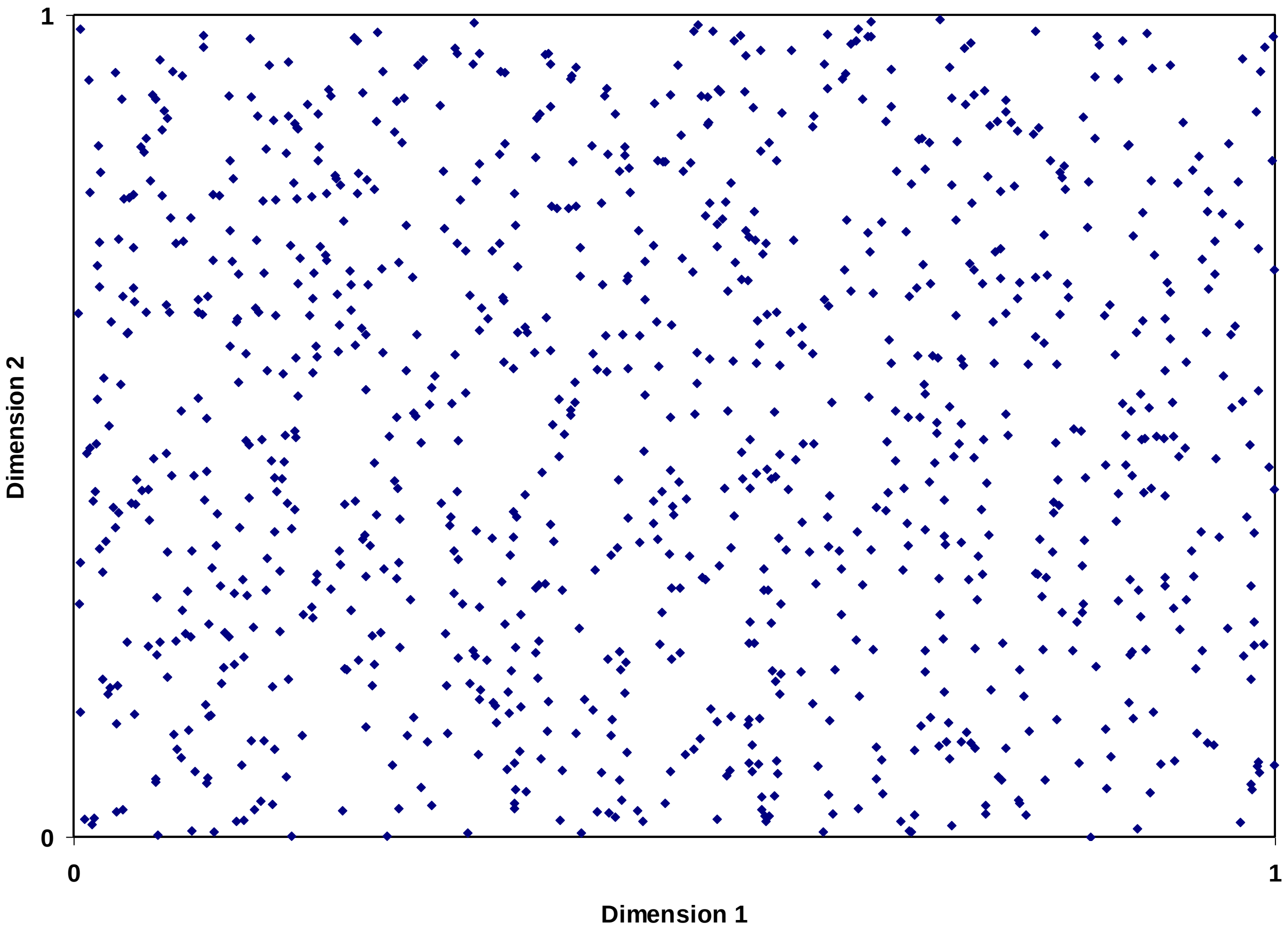
SIMULATION

- The approximation of integrals is undertaken using simulation techniques
- They entail the evaluation of the integrand at a number of draws taken from the domain of integration and computing the average of the resulting integrand values across the different draws.
- The focus of simulation techniques is on generating N sets of S univariate draws for each individual, where N is the number of draws and S is the dimensionality of integration
- Commonly used simulation methods
  - Pseudo Monte-Carlo
  - Quasi Monte-Carlo

PSEUDO MONTE-CARLO (PMC) METHODS

- Computes the average of the integrand over a sequence of "random" points over the domain of integration
- Pseudo-random sequences used in implementations
- Slow asymptotic convergence
  - to increase the accuracy by 1 decimal we need to increase the no. of draws 100 fold!
- Applicable for a wide class of integrands
- Integration error can be easily determined
- These approaches are applicable to any number of dimensions

1000 PSEUDO MONTE CARLO DRAWS



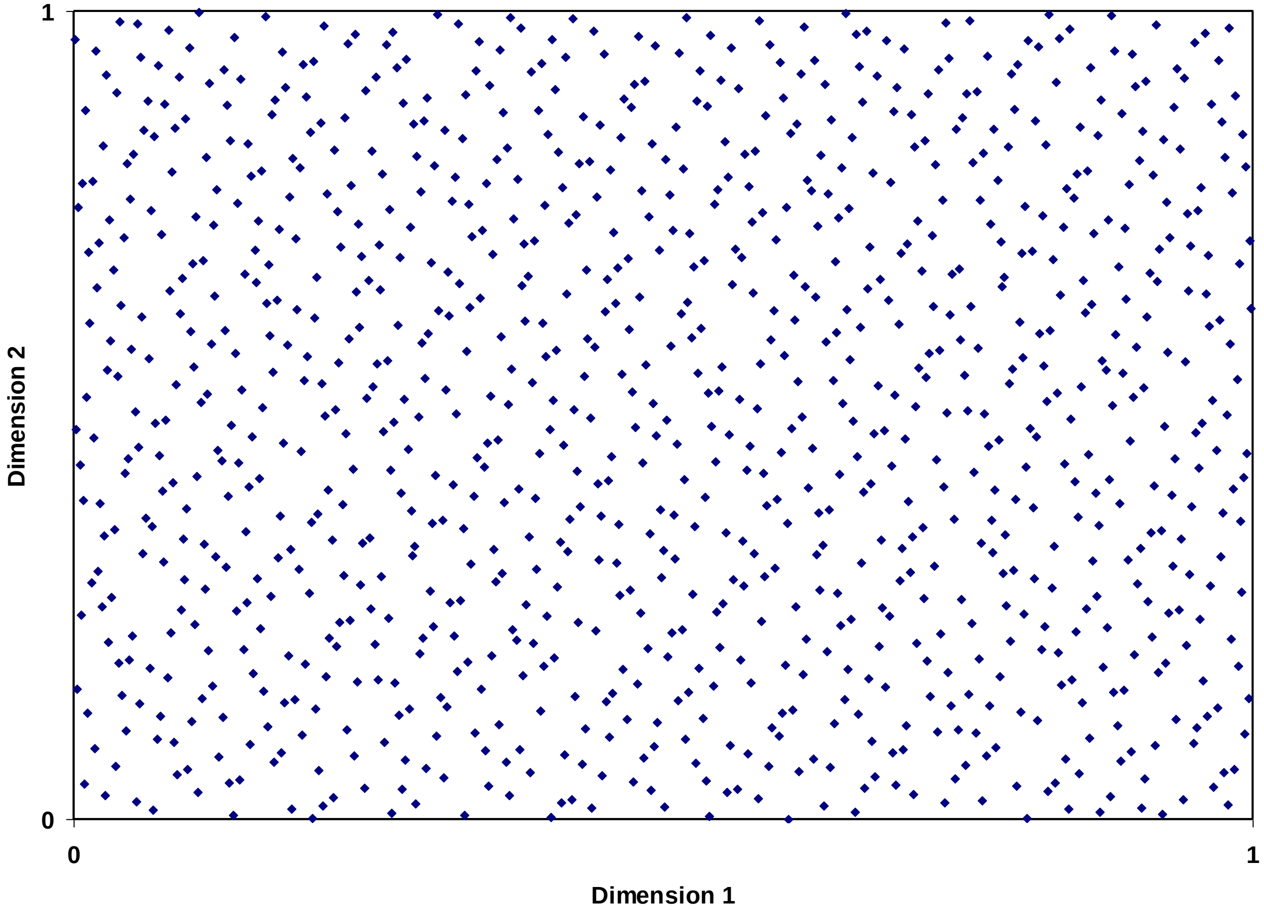
QUASI-MONTE CARLO METHOD

- The approach still entails evaluating different realizations of the integrand and averaging them
- However, as opposed to using pseudo-random draws we use "cleverly" crafted sequences
- Computes the average of the integrand over a non-random, more uniformly distributed, sequence of points over the domain of integration
- These approaches has convergence rates of N^{-1}
- QMC sequences include Halton, Faure, and Sobol

QUASI MONTE-CARLO (QMC) METHODS

- Instead of randomly selecting the draws, in QMC we select draws based on the gaps left in the previous draws.
- Thus we enable better coverage (it does not matter if they are random or not… as long as we cover the full domain well)
- Faster convergence than PMC methods
- Substantially fewer number of draws required
- Integration error cannot be easily determined
- Scrambling improves performance of standard Halton sequences

1000 QUASI MONTE CARLO DRAWS



QUASI-MONTE CARLO METHOD

- Bhat 2001 compared Halton and PMC in their ability to accurately and reliably recover model parameters in a mixed logit model
- Halton sequence outperformed the PMC sequence by a substantial margin
- He found that 125 Halton draws produced more accurate parameters than 2000 PMC draws in estimation

"A phenomenal development in the estimation of complex choice models"… David Hensher

CASE STUDY: INTERCITY MODE CHOICE MODEL

CASE STUDY

- Intercity mode choice between Montreal and Toronto
- Alternatives Car, Train and Air
- Bhat 1998

MIXED MULTINOMIAL LOGIT MODEL

- Small difference in the model

$$P_{qi}|(\nu_{q1},\ldots,\nu_{qk}) = \frac{\exp\left(\alpha_i + \delta_i' z_q + \sum_{k=1}^{K}[\exp(\gamma_k + \beta_k' w_{qk} + \nu_{qk})]x_{qik}\right)}{\sum_{j=1}^{I}\exp\left(\alpha_j + \delta_j' z_q + \sum_{k=1}^{K}[\exp(\gamma_k + \beta_k' w_{qk} + \nu_{qk})]x_{qjk}\right)}$$

$$P_{qi} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left\{\frac{\exp\left(\alpha_i + \delta_i' z_q + \sum_{k=1}^{K}[\exp(\gamma_k + \beta_k' w_{qk} + \nu_{qk})]x_{qik}\right)}{\sum_{j=1}^{I}\exp\left(\alpha_j + \delta_j' z_q + \sum_{k=1}^{K}[\exp(\gamma_k + \beta_k' w_{qk} + \nu_{qk})]x_{qjk}\right)}\right\} \cdot f(\nu_{q1})f(\nu_{q2})\ldots f(\nu_{qK})d\nu_{q1}d\nu_{q2}\ldots d\nu_{qK}$$

MIXED MULTINOMIAL LOGIT MODEL

- Variables considered
  - Socio-demographics
  - Trip information (purpose, party size, origin and destination cities)
  - LOS variables (frequency of service, total cost, in-vehicle time and out-of-vehicle time

Table 2. Intercity mode choice estimation results

Parameter affecting	Parameter on	MNL		FCL		RCL	
		Parm.	*t*-stat.	Parm.	*t*-stat.	Parm.	*t*-stat.
Intrinsic mode preferences	Mode constants						
	Train	−1.172	−2.57	−0.996	−2.77	−0.341	−0.75
	Air	−0.887	−1.30	1.256	1.98	1.466	1.88
	Income ($\times x10^4$)						
	Train	0.011	0.22	–	–	–	–
	Air	0.390	8.21	–	–	–	–
	Traveling alone						
	Train	0.366	1.97	–	–	–	–
	Air	0.436	2.72	–	–	–	–
	Female						
	Train	1.281	6.78	2.313	3.68	2.489	3.26
	Air	0.919	4.70	1.444	1.83	1.417	1.39
	Large city indicator						
	Train	1.105	5.11	1.058	4.90	0.994	4.41
	Air	0.672	3.40	0.699	3.50	0.857	2.88
Response to level-of-service variables	Freq. of service						
	Constant*	−2.435	–	−2.459	–	−2.163	–
	Traveling alone	–	–	−0.154	−1.42	−0.035	−0.34
	Female	–	–	0.395	2.89	0.395	2.47
	Std deviation	–	–	–	–	0.260	4.14
	Travel cost						
	Constant*	−3.165	–	−2.909	–	−2.727	–
	Income ($\times 10^4$)	–	–	−0.077	−3.22	−0.059	−2.45
	Std deviation	–	–	–	–	0.028	0.22
	In-vehicle time						
	Constant*	−4.610	–	−5.174	–	−5.089	–
	Income ($\times 10^4$)	–	–	0.064	1.90	0.115	2.89
	Traveling alone	–	–	0.359	2.63	0.228	1.72
	Std. deviation	–	–	–	–	0.546	6.51
	Out-of-vehicle time						
	Constant*	−3.453	–	−3.543	–	−3.201	–
	Female	–	–	0.428	2.16	0.373	2.02
	Std deviation	–	–	–	–	0.005	0.07

MIXED MULTINOMIAL LOGIT MODEL

Table 3. Comparison of response coefficients and implied money values of time

Level-of-service/money value of time	Average response coefficient values and average implied values of time across individuals				
	MNL	FCL	RCL		
			Mode	Median	Mean
Level-of-service variable					
Frequency of service (dep./day)	0.0876	0.0835	0.1138	0.1218	0.1260
Travel cost (Canadian \$)	−0.0422	−0.0399	−0.0525	−0.0526	−0.0526
In-vehicle time (min)	−0.0100	−0.0105	−0.0105	−0.0140	−0.0162
Out-of-vehicle time (min)	−0.0316	−0.0317	−0.0440	−0.0440	−0.0440
Money value of time					
In-vehicle time (\$ h)	14.15	16.34	12.14	16.38	19.03
Out-of-vehicle time (\$ h)	44.96	48.29	50.62	50.66	50.68

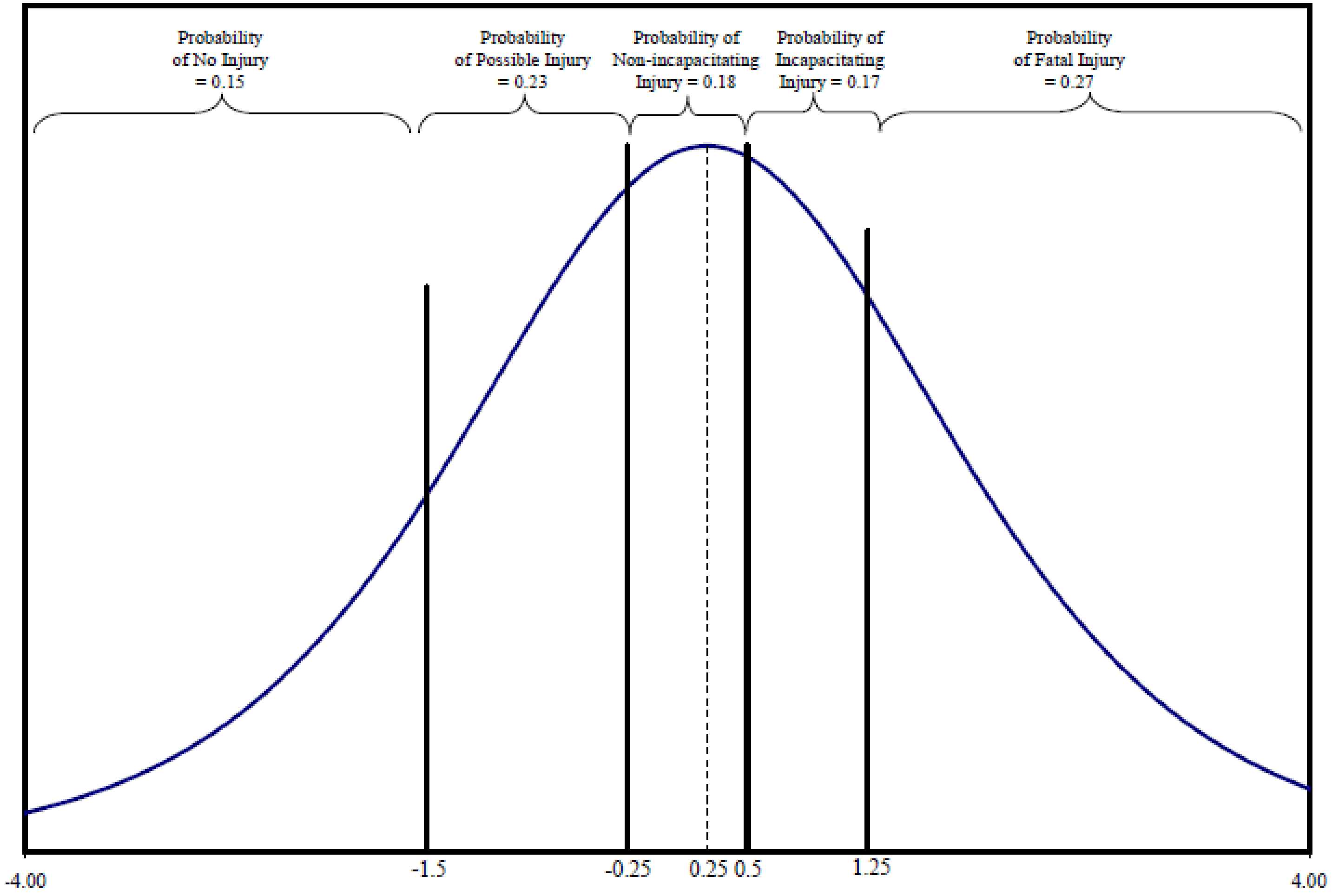
ADVANCED OR MODELS

ADVANCED OR MODELS

- As we discussed earlier, the OR models do not allow for alternative specific effects of various exogenous variables
- Lets consider the following example
- Non-motorist injury severity due to traffic collisions is reported as a five level ordinal variable
  - No injury
  - Possible Injury
  - Non-incapacitating injury
  - Incapacitating injury
  - Fatal injury
- Now we estimated a model and found that
  - A motorist being intoxicated has a coefficient of 0.25 (+ive so increases probability of fatal injury)
  - Coefficient for being hit head-on versus sideways is 0.25
  - Thresholds ψ_i = (-1.5, -0.25, 0.5, 1.25)
  - Let us assume these are the only variables affecting injury severity

ADVANCED OR MODELS

- Now consider two crashes
  - involving a drunk motorist and sideways crash
  - Involving a sober motorist and head-on crash
- Based on our OL model
  - Latent propensity for both crashes is 0.25
- So the probability will be (for standard logistic)
  - No injury (0.15)
  - Possible injury (0.23)
  - Non-incapacitating injury (0.18)
  - Incapacitating injury (0.17)
  - Fatal injury (0.27).



Source: *Eluru, N., C.R. Bhat, and D.A. Hensher (2008), "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes ," Accident Analysis and Prevention, Vol. 40, No. 3, pp. 1033-1054*

ORDERED LOGIT MODELS

- Standard ordered response model

$$y_q^* = \beta' x_q + \varepsilon_q$$

$$\text{where } y_q = k, \ \text{ if } \ \psi_{k-1} < y_q^* < \psi_k, \forall k = 1, 2..K$$

  - K represents the alternatives
  - y_q corresponds to the latent propensity for DM q
  - x_q is an (L x 1)-column vector of attributes (excluding a constant) associated with the DM q
  - β is a corresponding (L x 1)-column vector of variable effects
  - ψ_k corresponds to thresholds (ψ_0 = -∞ and ψ_K = + ∞)
  - ε_q represents the idiosyncratic error term distributed as a logistic

MIXED ORDERED LOGIT MODELS

- In the MGORL model we allow the thresholds to vary across DMs based on the variables
- $y_q^* = (\boldsymbol{\beta} + \boldsymbol{\alpha}_q)\boldsymbol{X}_q + \varepsilon_q$
  - $\tau_{q,k} = \tau_{q,k-1} + exp[(\boldsymbol{\delta_j} + \boldsymbol{\gamma}_{\boldsymbol{q},k})\,\boldsymbol{Z}_{q,k}]$
- $Pr\left(y_q = k \middle| \boldsymbol{\alpha}_q, \boldsymbol{\gamma}_{qk}\right) = \Lambda[(\boldsymbol{\delta_k} + \boldsymbol{\gamma}_{\boldsymbol{q},k})\,\boldsymbol{Z}_{\boldsymbol{q},k} - (\boldsymbol{\beta} + \boldsymbol{\alpha}_q)\boldsymbol{X}_{\boldsymbol{q}}] - \Lambda[(\boldsymbol{\delta}_{\boldsymbol{k}-1} + \boldsymbol{\gamma}_{q,k-1})\,\boldsymbol{Z}_{\boldsymbol{q},k} - (\boldsymbol{\beta} + \boldsymbol{\alpha}_{\boldsymbol{q}})\boldsymbol{X}_{\boldsymbol{q}}]$
- $P_{qk} = \int_{\boldsymbol{\alpha}_{\boldsymbol{q}},\boldsymbol{\gamma}_{\boldsymbol{q}k}} [Pr\left(y_q = k \middle| \boldsymbol{\alpha}_{\boldsymbol{q}}, \boldsymbol{\gamma}_{\boldsymbol{q}\boldsymbol{k}}\right)] * \boldsymbol{dF}\left(\boldsymbol{\alpha}_{\boldsymbol{q}}, \boldsymbol{\gamma}_{\boldsymbol{q}k}\right)\boldsymbol{d}(\boldsymbol{\alpha}_{\boldsymbol{q}}, \boldsymbol{\gamma}_{\boldsymbol{q}\boldsymbol{k}})$
- Simulation approach is same as the MMNL

EXAMINING PEDESTRIAN AND BICYCLIST INJURY SEVERITY LEVEL IN TRAFFIC CRASHES – A MIXED GENERALIZED ORDERED RESPONSE MODEL

SOURCE

- Eluru, N., C.R. Bhat, and D.A. Hensher (2008), "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes", Accident Analysis and Prevention, Vol. 40, No.3, pp. 1033-1054
- *Listed in the Top 50 papers published in Accident Analysis Prevention* - Zou, X., Vu, H.L. and Huang, H., 2020. Fifty years of Accident Analysis & Prevention: a bibliometric and scientometric overview. Accident Analysis & Prevention, 144, p.105568.

MOTIVATION

- Increased personal vehicle dependency in the US leads to
  - Increasing traffic congestion
  - Air quality problems
- Metropolitan organizations encourage non-motorized travel
  - Walking and bicycling for short distance trips
- Safety of non-motorists (pedestrians and bicyclists) in the US
  - Worse record in the US compared to other developed countries
  - Controlling for exposure in terms of miles traveled, US pedestrians are 3 times likely to get killed compared to German pedestrians, and over 6 times more likely compared to Dutch pedestrians (the corresponding numbers for cyclists are 2 and 3)

MOTIVATION

- In terms of absolute numbers, in 2005
  - 4881pedestrian and 784 bicyclist fatalities
  - 110,000 non-motorists are injured
- To put these numbers into perspective
  - A non-motorist is killed every 93 minutes and one is injured every 5 minutes in traffic accidents in the US
- High risk of non-motorists has attracted a lot of attention in the past decade
- Researchers examined the crashes involving non-motorists to:
  - Improve motorized vehicle and roadway design,
  - Enhance control strategies at conflict locations
  - Design good bicycle and pedestrian facilities
  - Formulate driver and non-motorized user education programs

MOTIVATION

- The host of factors that could potentially influence non-motorist injury severity include
  - Pedestrian/bicyclist characteristics (such as age, gender, helmet use, alcohol consumption)
  - Motorized vehicle driver characteristics (such as state of soberness and age)
  - Motorized vehicle attributes (such as vehicle type and speed)
  - Roadway characteristics (such as speed limit and whether the highway is divided or not)
  - Environmental factors (such as time of day, day of week, and weather conditions)
  - Crash characteristics (such as the direction of impact and motorist/non-motorist maneuver type at impact).

EARLIER RESEARCH

- A vastly researched area in the recent decade
- Research classified into two categories
  - Descriptive analyses at an aggregate level
    - A common association across the entire sample is arrived at through frequency analysis or cross-tabulation
  - Multivariate analyses at individual level of accidents
    - A host of factors influencing non-motorist injury severity are examined
- Remarks on earlier studies
  - The more recent studies have employed multivariate analyses
  - In cases where a binary dependent variable is employed (fatal vs non-fatal) logistic regression methods are predominant

EARLIER RESEARCH

- Remarks on earlier studies
  - In cases with ordered injury categories (such as property damage only, no visible injury but pain, non-incapacitating injury, incapacitating injury, and fatal injury) an ordered response model is employed
  - Studies have examined pedestrian or bicyclist injury severity separately
    - It is important from a policy perspective to compare the similarities and differences in the factors, and the magnitude of the impact of factors, affecting injury severity between the two non-motorist user groups
  - Earlier studies have very often, failed to recognize the need to consider motorist vehicle characteristics in the analysis

EARLIER RESEARCH

- Important findings
  - Pedestrians
    - Male, intoxicated, very young and elderly are prone to severe injuries
    - Alcohol-intoxicated driver, non-sedan and high speed vehicles cause severe injuries
  - Bicyclists
    - Similar to pedestrians
    - Accidents at high speed limit, low traffic volume and curved/non-flat roadway locations
    - Conditions of darkness with no lighting, in inclement weather (fog, rain and snow) and accidents in the morning peak period lead to severe injuries

EARLIER RESEARCH

- Current research in perspective
  - Employ a generalized version of the ordered logit model
  - Undertake the analysis for pedestrians and bicyclists
  - Consider the factors from all the six categories identified earlier
  - Allow for the presence of unobserved attributes to influence injury severity
    - For instance, the slower reaction time of being intoxicated may be exacerbated by the use of a walkman. But accident reports may not record or may miss information on walkman use and so walkman use may be unobserved
  - To summarize, develop a generalized model with a comprehensive set of variable to examine injury severity determinants

NOTATION

- Standard ordered response model

$$y_q^* = \beta' x_q + \varepsilon_q$$

$$\text{where } y_q = k, \ \text{ if } \ \psi_{k-1} < y_q^* < \psi_k, \forall\, k = 1, 2..K$$

K represents the number of injury categories

y_q corresponds to the latent injury risk propensity for non-motorist q in the crash she or he was involved in

x_q is an (L x 1)-column vector of attributes (excluding a constant) associated with the non-motorist, driver, vehicle, roadway, environment, and crash characteristics of the crash involving individual q

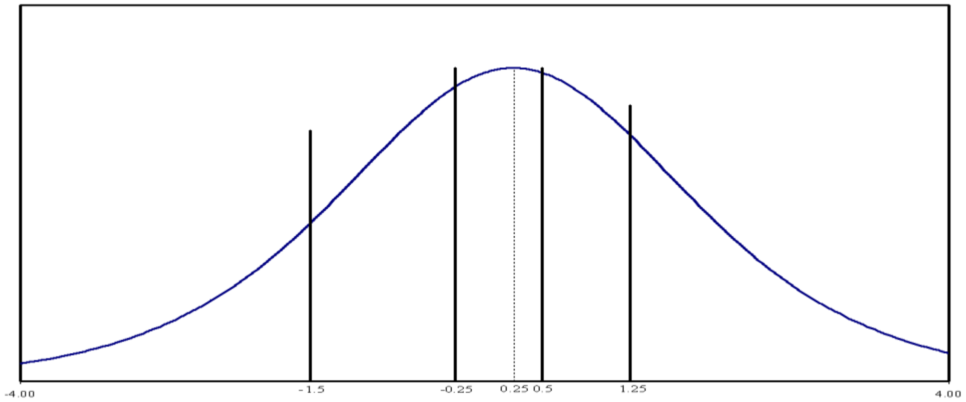
β is a corresponding (L x 1)-column vector of variable effects

ψ_k corresponds to thresholds (ψ_0 = -∞ and ψ_K = + ∞)

ε_q represents the idiosyncratic error term distributed as a logistic

EXAMPLE OF AN ORDERED RESPONSE LOGIT MODEL (ORL)

- For K = 5 injury categories, $\psi_0 = -\infty$ $\psi_1 = -1.5$, $\psi_2 = -0.25$, $\psi_3 = 0.5$, $\psi_4 = 1.25$ and $\psi_K = +\infty$
- Propensity ($\beta' x$) = 0.25
- Probability of injury severity in a particular category is the area under the curve between the corresponding thresholds



- Potential limitations of ORL model
  - The thresholds remain constant across individual accidents

MIXED GENERALIZED ORDERED RESPONSE LOGIT (MGORL) MODEL

- In the MGORL model we allow the thresholds to vary across individual accidents based on the variables

$$y_q^* = \beta_q^{'} x_q + \varepsilon_q, \; y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}$$

Next, we adopt a specific parametric form for the thresholds to guarantee the ordering conditions $(-\infty < \psi_{q,1} < \psi_{q,2} < \ldots < \psi_{q,K-1} < \infty)$ for each crash q.

To do so, we write:

$$\psi_{q,k} = \psi_{q,k-1} + \exp(\alpha_{qk} + \gamma'_{qk} z_{qk}),$$

DATA SOURCE

- 2004 General Estimates System (GES)
  - National Highway Traffic Safety Administration's National Center for Statistics and Analysis
  - Data compiled from a sample of police-reported accidents
  - The injury severity is collected on a five point ordinal scale: (1) No injury, (2) Possible injury, (3) Non-incapacitating injury, (4) Incapacitating injury, and (5) Fatal injury
  - Categories 1 and 2 are collapsed into a single category
- Sample preparation
  - Accidents involving pedestrians and bicyclists
  - Accidents involving a single vehicle and single non-motorist are chosen

DATA SOURCE

- Sample characteristics
- Distribution of non-motorist injury severity by non-motorist type

Injury severity category	Pedestrian	Bicyclist	All Non-motorists
No injury	135 (7.8%)	89 (7.3%)	224 (7.6%)
Non-incapacitating injury	951 (55.3%)	863 (70.6%)	1814 (61.6%)
Incapacitating injury	541 (31.4%)	250 (20.4%)	791 (26.9%)
Fatal injury	94 (5.5%)	21 (1.7%)	115 (3.9%)
Total	1223 (100.0%)	1721 (100.0%)	2944 (100.0%)

DATA SOURCE

- Distribution of Non-Motorist Injury Severity by Non-Motorist Alcohol Intoxication

Injury severity category	Non-motorist was alcohol intoxicated?		All Non-motorists
	No	Yes	
No injury	217 (8.0%)	7 (2.8%)	224 (7.6%)
Non-incapacitating injury	1688 (62.6%)	126 (51.2%)	1814 (61.6%)
Incapacitating injury	699 (25.9%)	92 (37.4%)	791 (26.9%)
Fatal injury	94 (3.5%)	21 (8.5%)	115 (3.9%)
Total	2698 (100.0%)	246 (100.0%)	2944 (100.0%)

EMPIRICAL ANALYSIS

- Results based on the estimation of the MGORL model for variables from all the six categories of variables identified earlier

- Non-motorist characteristics
  - Age is an important consideration. Non-motorists >60 years are prone to severe (even fatal) injuries
  - Gender effect is marginal
  - Alcohol intoxication increases likelihood of injury
  - Pedestrians are more likely to be severely injured

EMPIRICAL ANALYSIS

- Motorist characteristics
  - Alcohol intoxication leads to higher loading of severe injuries

- Motorized vehicle attributes
  - Non-sedan vehicle increases potential injury to non-motorist

- Roadway characteristics
  - Crashes on roads with high speed limits result in severe crashes
  - Signalized intersection reduce the severity of a crash for non-motorist

EMPIRICAL ANALYSIS

- Environment factors
  - Crashes occurring between 6PM – 12AM result in more severe injuries
  - Interestingly, presence of snow reduces the probability of fatality

- Crash characteristics
  - Direction of crash impacts the injury severity
    - Frontal impacts result in more severe crashes

RESULTS

Variables	Latent Propensity	Threshold between Non-incapacitating and Incapacitating injury	Threshold between Incapacitating and Fatal injury
Constant	1.846 (12.94)	1.305 (36.26)	1.645 (11.49)
Non-motorist Characteristics			
Pedestrian (Bicyclist is the base)	---	-0.103 (-2.67)	---
Male	0.159 (1.85)	---	---
Age Variables (age ≤ 60 years is base)			
> 60 years	0.667 (5.26)	---	-0.536 (-4.61)
Under the influence of alcohol	0.455 (3.47)	---	---
Motorized Vehicle Driver Characteristics			
Under the influence of alcohol	0.837 (2.14)	0.271 (2.87)	-0.250 (-1.53)
Motorized Vehicle Attributes			
Sports utility vehicle	0.364 (3.15)	---	---
Pick-up truck	---	-0.070 (-2.18)	-0.197 (-1.98)
Van	---	---	-0.237 (-1.70)

RESULTS

Variables	Latent Propensity	Threshold between Non-incapacitating and Incapacitating injury	Threshold between Incapac and Fatal injury
Roadway Design Characteristics			
Speed Limit			
25-50mph	0.218 (1.97)	---	-0.225 (-2.01)
>50 mph	0.605 (3.06)	---	-0.679 (-3.93)
Speed limit > 25mph * pedestrian	---	-0.117 (-2.61)	---
Accident Location			
Signalized Intersection	-0.300 (-3.32)	---	0.387 (3.43)
Environmental Factors			
6pm - 12am	0.297 (3.43)	---	-0.352 (-3.82)
12am - 6am	---	-0.304 (-4.66)	-0.365 (-2.59)
Snow	---	---	0.538 (1.60)
Crash Characteristics			
Direction of Impact (sideways impact is the base)			
Frontal Impact	0.447 (3.20)	0.072 (1.64)	-0.226 (-2.38)
Other directions of impact	-0.734 (-2.91)	---	-0.603 (-2.23)

VALIDATION EXERCISE

- Comparing the proposed (MGORL) model vs standard (ORL) model

Injury Categories/ Measures of fit	Pedestrians			Bicyclists		
	Actual shares	ORL predictions	MGORL predictions	Actual shares	ORL predictions	MGORL predictions
No injury	7.84	6.04	7.44	7.28	9.89	7.93
Non-incapacitating injury	55.26	57.70	55.55	70.56	65.90	70.07
Incapacitating injury	31.44	31.38	31.73	20.44	21.59	20.28
Fatal injury	5.46	4.94	5.29	1.72	2.62	1.72
Number of observations	1721	1721	1721	1223	1223	1223
Root mean square error (RMSE)	---	1.54	0.30	---	2.77	0.42
Mean absolute percentage error (MAPE)	---	9.28	2.46	---	25.14	2.62

IMPLICATIONS FROM RESEARCH

- Education and training
  - The results reinforce the need to educate non-motorists and motorists of the risks of driving under influence. It is necessary to underscore that alcohol combined with night driving is deadly
  - Encouraging non-motorists to wear "reflector" gear to improve visibility
- Traffic regulation and control
  - Signs need to be posted to communicate to non-motorists information regarding heavy traffic on roadways
  - Restricting speed limits to < 25 mph on roadways with heavy pedestrian and bicycle traffic
  - Good street lighting and illumination, and additional traffic signal installation might alleviate non-motorist injury severity
- Planning and design of pedestrian/bicyclist facilities
  - On roadways with high speed limits bicycle facility need to be separated from roadway. Further bicycle facilities need to be chosen based on roadway speed limits, vehicular mix and presence of lighting

CONCLUSIONS

- The current study addresses the safety of non-motorists
- An advanced econometric framework to address the ordinal category of the reported injury severity is developed. The proposed model generalizes the standard ORL model
- The MGORL model developed is employed on 2004 General Estimates System (GES) database
- The standard ORL model employed produces inconsistent estimates
- It is very interesting to note that the general pattern and relative magnitude of elasticity effects of injury severity determinants are similar for pedestrians and bicyclists

CONCLUSIONS

- Pedestrians are more likely to be injured in the event of a crash
- The most important variables influencing the injury severity are:
  - Non-motorist age
  - Speed limit of the roadway
  - Location of the crash (if a signalized intersection or not)
  - Time of day (evening time being more riskier)
- Important implications for education and training, traffic regulation and control, and planning of pedestrian/bicycle facilities

PANEL DATA

PANEL DATA

- What is Panel Data
  - If we have multiple observations for a single decision maker in the data – the data is considered panel data
  - Also referred to as repeated observations, longitudinal data
- What impacts would this have on our modeling assumptions?
  - First major impact, across multiple observations the error terms are unlikely to be independent
  - To address this we consider mixed models that allow us to consider common unobserved factors
  - Another difference is how the likelihood is to be considered
  - In a simple discrete choice, we try to match the chosen alternative as well as possible; in a panel dataset for the same DM we have multiple observations – so we need to figure out how to match this

ERROR TERMS IN PANEL MODELS

- Consider a choice scenario with DM q, k alternatives and t repetitions
- $U_{qkt} = (x_{qkt}) * (\beta_k) + \varepsilon_{qk} + \varepsilon_{qkt}$
- So to account for this common error term we can adopt the error components approach we talked about
- Important to recognize that the same draw should be repeated for all observations for the individual
  - So the halton draws are created an individual level and copied across multiple observations to obtain panel specific error terms
- It is not necessary that unobserved effects exist only at the DM level – we can have unobserved effects at the repetition level as well – in that case we have some parameters estimated at the DM level and the other at the repetition level; so halton draws have to be drawn separately for these cases

LL IN PANEL MODELS

- If we have one DM repetition the LL is computed as
- LL = Ln(Pchosen)
- Now in a panel case we have multiple instances – in this case our objective is not to match each individual instance but the match the string of instances i.e. for a DM with 3 instances the idea would be to match Pchosen1, Pchosen2, Pchosen3
- LL = Ln(Pchosen1 * Pchosen2 * Pchosen3)
  - Ensures that the number of observations are appropriately considered – in the panel case the number of observations to be counted for Standard Error computation is the same as number of DMs not number of total records; this will ensure correct parameter estimation – otherwise we over-estimate parameter significance (because we think LL is from lot more records)

LL IN PANEL MODELS - MNL

- $y^*_{qit} = (\boldsymbol{\beta} + \boldsymbol{\alpha}_{qi})\boldsymbol{X}_{qit} + \varepsilon_{qit},$
- $Pr\left(y_{qit} = i \middle| \boldsymbol{\alpha}_{qi}\right) = \exp((\boldsymbol{\beta}+\boldsymbol{\alpha}_{qi})\boldsymbol{X}_{it}) / \sum_j \exp((\boldsymbol{\beta}+\boldsymbol{\alpha}_{qj})\boldsymbol{X}_{jt})$
- $\pi_{qit} = Pr\left(y_{qit} = i \middle| \boldsymbol{\alpha}_{qi}\right)$
- $\mathrm{L_q}|\boldsymbol{\alpha}_{qi}, = \prod_{t=1}^{T} \pi_{qit}$
- $\mathcal{L} = \sum_{q=1}^{Q} L_q$

- For other systems – the probability term will need to be appropriately changed.

82

MULTIPLE DEPENDENT VARIABLES

MULTIPLE DEPENDENT VARIABLES

- What does it mean?
  - Vehicle choice (number and type) and residential location are tied
  - We cannot really consider residential location as a simple exogenous variable
  - Unobserved factors that affect choice of residence will affect vehicle choices
- How to study this?

POSSIBLE APPROACHES

- Combo Models (or combination packages)
  - Lets say we have 3 choices in vehicle type and 5 residential location choices – we create 15 combo alternatives and study them
  - Wont work for continuous variables
- Joint distribution
  - One equation per choice and all are tied together as a multivariate distribution – for example bivariate normal distribution
  - Not all choice scenarios will have a closed form joint distribution
- Simultaneous equation
  - One equation per choice and all are tied together using unobserved components
  - This is an expansion of mixed model approach
- Self-selection or endogeneity
  - Similar to above but we also have one choice variable as an exogenous variable in the second choice

COPULA BASED MODELS

FRACTIONAL SPLIT MODEL

OBSERVED GOP

REFERENCES

- Ben-Akiva, M. E. and S. R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, The MIT Press.
- Ortuzar, J. D. and L. G. Willumsen (2011). Modelling transport, Wiley. 4th ed.
- Koppelman, F. S. and C. Bhat (2006). "A self instructing course in mode choice modeling: multinomial and nested logit models“
- Train, K. (2003). Discrete choice methods with simulation, Cambridge University Press (Available for free online at http://www.econ.berkeley.edu/books/choice2.html)
- Bhat, C. R., N. Eluru, R. Copperman (2008). "Flexible model structures for discrete choice analysis." Handbook of transport modelling 1: 71–90.
- Bhat, C. R. (1998). "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling." Transportation Research Part A: Policy and Practice 32(7): 495-507.
- Cameron A. C. and P.K. Trivedi “Microeconometrics: Methods and Applications” Cambridge University Press 2005.
- Bhat, C.R. (2003), "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences", Transportation Research Part B, Vol. 37, No. 9, pp. 837-855
- Bhat, C.R. (2001), "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model", Transportation Research Part B, Vol. 35, No. 7, pp. 677-693.