- **Independence of Irrelevant Alternatives (IIA) property**.
  - Recall that a critical assumption in the derivation of the multinomial logit model is that the disturbances (ε's) are independently and identically distributed.
  - When this assumption does not hold, a major specification error results.
  - This problem arises when only some of the functions, which determine possible outcomes, share unobserved elements (that show up in the disturbances).
  - If all outcomes shared the same unobserved effects, the problem would self correct because in the differencing of outcome functions common unobserved effects would cancel out.

- Because the common elements cancel in the differencing, a logit model with only two outcomes can never have an IIA violation.

- To illustrate the IIA problem, note the ratio of any two-outcome probabilities is independent of the functions determining any other outcome since

$$\frac{P(1)}{P(2)} = \frac{EXP[\beta_1 X_1]}{EXP[\beta_2 X_2]}$$

- **Problem**: consider the estimation of a model of choice of travel mode to work where the alternatives are to take a personal vehicle, a red transit bus, or a blue transit bus.

  - The red and blue transit buses clearly share unobserved effects that will appear in their disturbance terms and they will have exactly the same functions ($\beta_{rb}X_{rb} = \beta_{bb}X_{bb}$) if the only difference in their observable characteristics is their color.

- For illustrative purposes, assume that, for a sample commuter, all three modes have the same value of $\beta_i X_i$'s (the red bus and blue bus will, and assume that costs, time, and other factors that determine the likelihood of the personal vehicle being chosen works out to the same value as the buses).
- Then the predicted probabilities will give each mode a 33% chance of being selected.

- This is unrealistic since the correct answer should be a 50% chance of taking a personal vehicle and a 50% chance of taking a bus (both red and blue bus combined) and not 33.33% and 66.67% respectively as the MNL would predict.

- Most applications the IIA violation is more subtle than in the previous example.

- There are a number statistical tests that are conducted to test for IIA violations.

- One of the more common of these tests was developed by Small and Hsiao (1985). The procedure is to first split the data randomly into two samples (N^A and N^B) containing the same number of observations. Two separate models are then estimated producing parameter estimates β^A and β^B. A weighted average of these parameters is obtained from

$$\beta^{AB} = \left(1/\sqrt{2}\right)\beta^A + \left(1 - 1/\sqrt{2}\right)\beta^B$$

- Then, a restricted set of outcomes, *D*, is created as a sub-sample from the full set of outcomes. The sample N^B is then reduced to include only those observations in which the observed outcome lies in *D*.

- Two models are estimated with the reduced sample ($N^{B'}$) using *D* as if it were the entire outcome set (B' in superscripting denotes the sample reduced to observations with outcomes in *D*).

- One model is estimated by constraining the parameter vector to be equal to β^{AB} as computed above. The second model estimates the unconstrained parameter vector $\beta^{B'}$.

- The resulting log-likelihoods are used to evaluate the suitability of the model structure by creating a chi squared statistic with the number of degrees of freedom equal to the number of parameters in β^{AB} (also the same number as in $\beta^{B'}$). This statistic is

$$\chi^2 = -2[LL^{B'}(\beta^{AB}) - LL^{B'}(\beta^{B'})]$$

- The test is then repeated by interchanging the roles of the N^A and N^B sub-samples (reducing the N^A sample to observations were the observed outcomes lie in D and proceed). Using the same notation, Equation 11.39 becomes

$$\beta^{BA} = \left(1/\sqrt{2}\right)\beta^{B} + \left(1-1/\sqrt{2}\right)\beta^{A}$$

- and the chi-squared statistic is

$$\chi^2 = \text{-}2[LL^{A'}(\beta^{BA}) - LL^{A'}(\beta^{A'})]$$

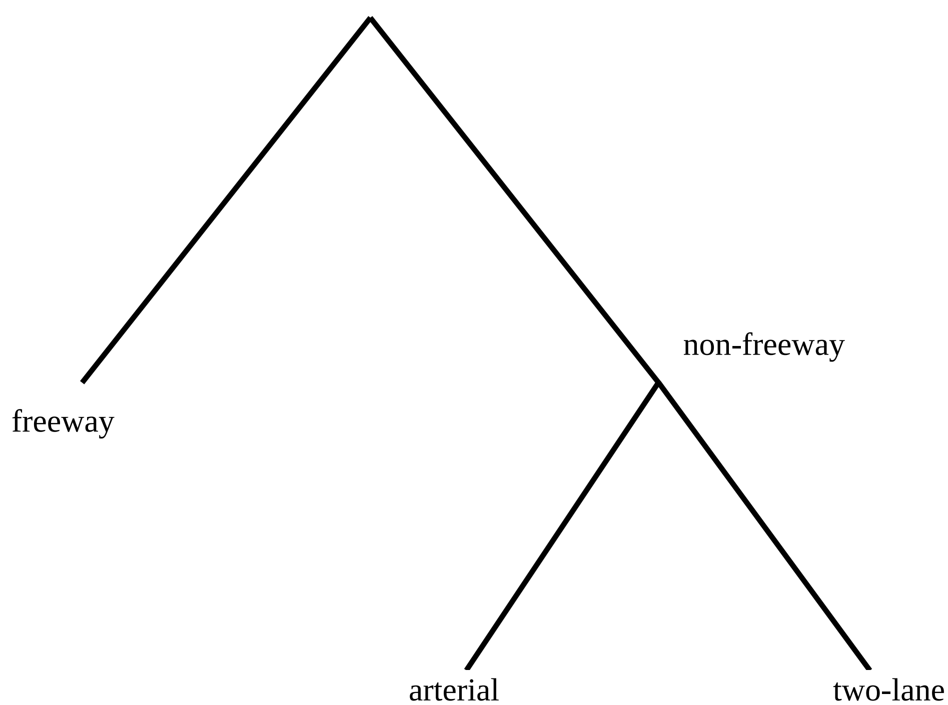
The Nested Logit Model (Generalized Extreme Value Models)

- To overcome the IIA problem, the idea behind a nested logit model is to group alternate outcomes suspected of sharing unobserved effects into nests (this sharing sets up the disturbance term correlation that violates the derivation assumption).

- Because the outcome probabilities are determined by differences in the functions determining these probabilities (both observed and unobserved), shared unobserved effects will cancel out in each nest providing that all alternatives in the nest share the same unobserved effects..

- This canceling out will not occur if a nest (group of alternatives) contains some alternative outcomes that share unobserved effects and others that do not (this sets up an IIA violation in the nest).

Suppose it is suspected that the arterial and two-lane road share unobserved elements (being lower level roads relative to the freeway with no access control, lower design speeds). When developing a nested structure to deal with the suspected disturbance term correlation, a structure shown visually in the Figure is used.



- By grouping the arterial and two-lane road in the same nest their shared unobserved elements cancel.
- Mathematically, McFadden (1981) has shown the GEV disturbance assumption leads to the following model structure for observation *n* choosing outcome *i*

$$P_n(i) = EXP[\beta_i X_{in} + \phi_i L_{in}] / \sum_{\forall I} EXP[\beta_I X_{In} + \phi_I LS_{In}]$$

$$P_n(j|i) = EXP[\beta_{j|i} X_n] / \sum_{\forall J} EXP[\beta_{J|i} X_{Jn}]$$

$$LS_{in} = LN[\sum_{\forall J} \exp(\beta_{J|i} X_{Jn})]$$

  - where

$P_n(i)$ is the unconditional probability of observation n having discrete outcome i,

X's are vectors of measurable characteristics that determine the probability of discrete outcomes,

β's are vectors of estimable parameters,

$P_n(j|i)$ is the probability of observation n having discrete outcome j conditioned on the outcome being in outcome category i (for example, for the nested structure shown in the Figure the outcome category i would be non-freeway) and

$P_n(j|i)$ would be the binary logit model of the choice between the arterial and two-lane road), J is the conditional set of outcomes (conditioned on i), I is the unconditional set of outcome categories (the upper two branches of Figure 11-5),

LS_{in} is the inclusive value (logsum), and

ϕ_i is an estimable parameter.

- Note that this equation system implies that the unconditional probability of having outcome j is,

$$P_n(j) = P_n(i) \times P_n(j|i)$$

- Estimation of a nested model is usually done in a sequential fashion.

1. Estimate the conditional model using only the observations in the sample that are observed having discrete outcomes J. In the example

illustrated in the Figure this is a binary model of commuters observed taking the arterial or the freeway.

2. Once these estimation results are obtained, the logsum is calculated (this is the denominator of one or more of the conditional models) for all observations, both those selecting *J* and those not (for all commuters in our example case).

3. These computed logsums (in our example there is just one logsum) are used as independent variables in the functions. Note that not all unconditional outcomes need to have a logsum in their respective functions (the example shown in the Figure would only have a logsum present in the function for the non-freeway choice).

- Caution needs to be exercised when using the sequential estimation procedure described above because results in variance-covariance matrices that are too small and thus *t*-statistics are inflated (typically by about 10-15%). This problem is resolved by estimating the entire model at once using full information maximum likelihood.

- It is important to note that the interpretation of the estimated parameter associated with logsums (ϕ_i's) has the following important elements:

  1. ϕ_i's must be greater than 0 and less than 1 in magnitude to be consistent with the nested logit derivation.

  2. If $\phi_i = 1$, the assumed shared unobserved effects in the nest are not significant and the nested model reduces to a simple MNL. Test with:

$$t = \frac{\beta \ - 1}{S.E.(\beta)}$$

3. If ϕ_i is less than zero then factors increasing the likelihood of an outcome being chosen in the lower nest will decrease the likelihood of the nest being chosen.

4. If ϕ_i is equal to zero then changes in nest outcome probabilities will not affect the probability of nest selection and the correct model is recursive.