# Multinomial Choice Models

# Two data sets

$$U_{nj} = \alpha T_{nj} + \beta M_{nj} + \varepsilon_{nj}^{B} \;\forall n \text{ in Boston}$$
$$U_{nj} = \alpha T_{nj} + \beta M_{nj} + \varepsilon_{nj}^{C} \;\forall n \text{ in Chicago,}$$

where the variance of $\varepsilon_{nj}^{B}$ is not the same as the variance of $\varepsilon_{nj}^{C}$. Label the ratio of variances as $k = \text{Var}(\varepsilon_{nj}^{C})/\text{Var}(\varepsilon_{nj}^{B})$. We can divide the utility for travelers in Chicago by $\sqrt{k}$; this division doesn't affect their choices, of course, since the scale of utility doesn't matter. However, doing so allows us to rewrite the model as

$$U_{nj} = \alpha T_{nj} + \beta M_{nj} + \varepsilon_{nj} \;\forall n \text{ in Boston}$$
$$U_{nj} = (\alpha/\sqrt{k}) T_{nj} + (\beta/\sqrt{k}) M_{nj} + \varepsilon_{nj} \;\forall n \text{ in Chicago.}$$

where now the variance of $\varepsilon_{nj}$ is the same for all $n$ in both cities (since $\mathrm{Var}(\varepsilon^C_{nj}/\sqrt{k}) = (1/k)\mathrm{Var}(\varepsilon^C_{nj}) = [\mathrm{Var}(\varepsilon^B_{nj})/\mathrm{Var}(\varepsilon^C_{nj})]\mathrm{Var}(\varepsilon^C_{nj}) = \mathrm{Var}(\varepsilon^B_{nj})$. The scale of utility is set by normalizing the variance of $\varepsilon_{nj}$. The parameter $k$, which is often called the scale parameter, is estimated along with $\beta$ and $\alpha$. The estimated value $\hat{k}$ of $k$ tells the researcher the variance of unobserved factors in Chicago relative to that in Boston. For example, $\hat{k} = 1.2$ implies that the variance of unobserved factors is twenty percent greater in Chicago than in Boston.

# Overview

- Model definition
- Model properties
- Utility functions for MNL models
  - Attributes of alternatives and individuals
  - Alternative-specific constants
- Independence from irrelevant alternatives (IIA)
  - Definition
  - The red bus/blue bus paradox
  - Avoiding IIA consequences
  - Introducing new modes
- Selecting choice sets

# MNL Model Definitions

- Three modes

$$\Pr(1) = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2) + \exp(V_3)}$$

- N modes

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)}$$

- Example

| Mode i | $V_i$ | $\exp(V_i)$ | Pr(i) |
|---|---|---|---|
| Drive Alone | 1.5 | 4.48 | 0.31 |
| Carpool | 1.9 | 6.69 | 0.46 |
| Bus | 1.2 | 3.32 | 0.23 |
| Totals | | 14.49 | 1.00 |

# MNL Model Properties

- Each probability depends on the deterministic components of the utilities of all alternatives ($V_j$, all j)
- Each probability i increases as $V_i$ increases, and decreases as $V_j$ ($j \neq i$) increases
- Can be used for any number of alternatives
- Relatively easy to understand and apply
- Probability that alternative i is chosen depends only on the values ($V_j - V_i$) for all alternatives j, except i

# MNL Utility Functions

## Attributes of Alternatives and Individuals

- Similar to the previos examples

***Example***

**Deterministic component of utility for mode j**

$V_j = -T_j - 5C_j/Y$

| Mode | Time | Cost | Y = 15 | | | Y = 30 | | |
|---|---|---|---|---|---|---|---|---|
| | | | V | exp(V) | Pr | V | exp(V) | Pr |
| Drive Alone | 0.50 | 2.00 | -1.17 | 0.31 | 0.33 | -0.83 | 0.44 | 0.38 |
| Carpool | 0.75 | 1.00 | -1.08 | 0.34 | 0.34 | -0.92 | 0.40 | 0.34 |
| Bus | 1.00 | 0.75 | -1.25 | 0.29 | 0.29 | -1.13 | 0.32 | 0.28 |
| Totals | — | — | — | 0.94 | 1.00 | | 1.16 | 1.00 |

# MNL Utility Functions

## Alternative Specific Constants

- Previous example with equal times and costs for all modes (T = 0.75, C = 1.00; Y = 20)

| Mode | V | exp(V) | Pr |
|---|---|---|---|
| **Drive Alone** | **-1.00** | **0.37** | **0.33** |
| **Carpool** | **-1.00** | **0.37** | **0.33** |
| **Bus** | **-1.00** | **0.37** | **0.33** |
| **Total** | | **1.10** | **1.00** |

- Would we expect this result?
- Accounting for other modal factors
  – Include more variables if possible; and/or
  – Add constants to N – 1 modes

# MNL Utility Functions

## Alternative Specific Constants (continued)

**Example of constants –**

$V_{da} = 0.8 - T_{da} - 5^*C_{da}/Y$

$V_{cp} = 0.2 - T_{cp} - 5^*C_{cp}/Y$

$V_b = -T_b - 5^*C_b/Y$

Choice of base mode is arbitrary

| Mode | V | exp(V) | Pr |
|---|---|---|---|
| Drive Alone | -0.20 | 0.82 | 0.50 |
| Carpool | -0.80 | 0.45 | 0.28 |
| Bus | -1.00 | 0.37 | 0.23 |
| Total | | 1.64 | 1.00 |

# Taste Variation

- Logit models can capture taste variations that relate to observed characteristics of the decision maker but not random taste variation
- Example:
  - Household choosing among make and models of cars
  - Two attributes:
    - PP – purchase power
    - SR – shoulder room (interior size of the car)

$$U_{nj} = \alpha_n \mathrm{SR}_j + \beta_n \mathrm{PP}_j + \varepsilon_{nj},$$

# Taste variation - continue

- Suppose SR vary with number of people in the HH Mn, and importance of purchase price in inverse to income

$$\alpha_n = \rho M_n$$

$$\beta_n = \theta / I_n$$

$$U_{nj} = \rho(M_n \mathrm{SR}_j) + \theta(\mathrm{PP}_j / I_n) + \varepsilon_{nj}$$

# But if random – more problematic

- If there addition random effects on these parameters, for example they vary with size of people which we don't observe…..

$$\alpha_n = \rho M_n + \mu_n$$

$$\beta_n = \theta / I_n + \eta_n$$

$$U_{nj} = \rho(M_n \mathrm{SR}_j) + \mu_n \mathrm{SR}_j + \theta(\mathrm{PP}_j / I_n) + \eta_n \mathrm{PP}_j + \varepsilon_{nj}$$

$$U_{nj} = \rho(M_n \mathrm{SR}_j) + \theta(\mathrm{PP}_j / I_n) + \tilde{\varepsilon}_{nj}$$

$$\tilde{\varepsilon}_{nj} = \mu_n \mathrm{SR}_j + \eta_n \mathrm{PP}_j + \varepsilon_{nj}$$

- This error term can't possibly be i.i.d.

# The IIA Property

Definition

- The independence from irrelevant alternatives property
  - "For any individual, the ratio of the probabilities of choosing two (available) alternatives is independent of the availability or attributes of any other alternative"
- Mathematically

$$\frac{\Pr(i)}{\Pr(k)} = \frac{\exp(V_i)}{\exp(V_k)} = \exp(V_i - V_k)$$

**No dependence on $V_j$ ($j \neq i$ or $k$)**

# The IIA Property

## Red and Blue

- Scenario 1
  - Available modes are (da) and red buses (rb); red buses have plenty of seats for all passengers
  - $V_{da} = V_{rb}$
  - MNL model says Pr(da) = Pr(rb) = 0.5
  - Is this reasonable?
- Scenario 2
  - A new bus operator exactly duplicates red bus service using blue buses (bb)
  - MNL model says Pr(da) = Pr(rb) = Pr(bb) = 0.33
  - How has service changed for the passengers?
  - What new mode shares would we expect?
  - What is the MNL prediction if we say that red and blue buses are the same mode?

# The IIA Property

- A more realistic example – light rail fares increase

| Mode | Base V | Case Pr(i) | Fare Increase V | Fare Increase Pr(i) | Change in Pr(i) |
|---|---|---|---|---|---|
| Drive Alone | -0.20 | 0.458 | -0.20 | 0.467 | +0.009 |
| Carpool | -0.80 | 0.251 | -0.80 | 0.256 | +0.005 |
| Bus | -1.53 | 0.121 | -1.53 | 0.123 | +0.002 |
| Light Rail | -1.19 | 0.170 | -1.31 | 0.154 | -0.016 |

**Is this realistic?**

# The IIA Property

Avoiding its Consequences

- The source of the problem: dependency between the error terms
- Include additional variables
- Use other choice models
  - Nested logit
  - Probit
  - Mixed logit

# Panel Data

- If the unobserved factors that affect decision makers are independent over repeated choices, logit is fine…
- Any dynamic related to observed factors that enter the decision process (e.x., person past choice influence current choice, or lagged response to change in attribute) can be handled.
- Dynamic associated with unobserved factors can’t be handled, since the unobserved factors are assumed to be unrelated over choices.
- The dependent variable in previous periods can also be entered as explanatory variable, as long as we assume that the errors are independent over time

# Panel Data - Continue

- However, in many cases one would expect there be some factors that are not observed by the researcher that affect each of the decision maker's choice
- In such cases other model structure may be more appropriated
- Or, if possible, re specify the model to bring the source of the unobserved dynamic into the model explicitly such that the remaining errors are independent over time.

# Non Linear Parameters

- In some context, we may want to allow non-linear parameters
- However, estimation is more difficult since the log likelihood function may not be globally concave, and
- Computer routine are not widely available, so one may need to write his own code.

# Adding New Modes

- Transfer the deterministic component of utility (V) from an existing mode (except the modal constant) to the new mode
- Use judgment to specify the modal constant, guided by experience where the new mode exists
- Result – uncertain forecasts for new modes

# Selecting Choice Sets

- Only consider modes which are practically significant assume others are never chosen
  - Should walk and bike modes be included?
- Tailor available modes to individuals and trips
  - Children cannot drive -> no drive alone mode
  - Households without autos -> no drive alone or drive to transit modes
  - Transit farther than 2 miles at origin or destination -> no transit with walk access mode
  - Others?

# Home-Based Work Mode Choice Model Coefficients From Selected Cities

| City | Survey Year | Auto In-Vehicle Time | Auto Out-of Vehicle Time | Auto Operating Cost | Parking Cost |
|---|---|---|---|---|---|
| Baltimore | 1993 | -0.034 | -0.044 | -0.143 | -0.143 |
| Dallas | 1996 | -0.055 | | -0.558 | -0.558 |
| Denver | 1985 | -0.018 | -0.093 | -0.350 | -0.950 |
| Detroit | 1996 | -0.052 | | -0.410 | -0.410 |
| Houston | 1985 | -0.022 | | -0.614 | -1.540 |
| Los Angeles | 1991 | -0.021 | | -0.296 | -0.296 |
| Milwaukee | 1991 | -0.016 | -0.041 | -0.450 | -0.450 |
| Philadelphia | 1986 | -0.042 | | -0.260 | -0.260 |
| Pittsburgh | 1978 | -0.047 | -0.069 | -2.100 | -2.100 |
| Portland | 1985 | -0.039 | -0.065 | -1.353 | -1.353 |
| Sacramento | 1991 | -0.025 | -0.038 | -0.279 | -0.279 |
| St. Louis | 1965 | -0.023 | -0.057 | -1.170 | -1.170 |
| Tucson | 1965 | -0.034 | | -0.184 | -0.184 |

# Home-Based Work Mode Choice Model Coefficients From Selected Cities

| City | Survey Year | Transit In-Vehicle Time | Transit Walk Time | Transit 1st Wait Time | Transit Transfer Time | Transit Cost | Number of Transfers |
|---|---|---|---|---|---|---|---|
| Baltimore | 1993 | -0.034 | -0.044 | -0.029 | -0.016 | -0.053 | -0.268 |
| Dallas | 1996 | -0.025 | -0.064 | -0.064 | -0.064 | -0.550 | |
| Denver | 1985 | -0.018 | -0.054 | -0.028 | -0.059 | -0.440 | |
| Detroit | 1996 | -0.009 | -0.019 | -0.019 | -0.019 | -0.410 | |
| Houston | 1985 | -0.022 | -0.057 | -0.057 | -0.057 | -0.614 | -0.088 |
| Los Angeles | 1991 | -0.021 | -0.053 | -0.053 | -0.053 | -0.296 | |
| Milwaukee | 1991 | -0.016 | -0.041 | -0.041 | -0.041 | -0.450 | |
| Philadelphia | 1986 | -0.042 | -0.032 | -0.051 | -0.051 | -0.115 | |
| Pittsburgh | 1978 | -0.047 | -0.069 | -0.069 | -0.069 | -2.100 | |
| Portland | 1985 | -0.039 | -0.065 | -0.040 | -0.090 | -1.353 | |
| Sacramento | 1991 | -0.025 | -0.038 | -0.038 | -0.038 | -0.279 | |
| St. Louis | 1965 | -0.023 | -0.057 | -0.057 | -0.057 | -1.170 | |
| Tucson | 1965 | -0.034 | -0.040 | -0.040 | -0.040 | -0.184 | |

# Home-Based Work Mode Choice Model Coefficient Relationships From Selected Cities

| City | Survey Year | Ratio: Walk to In-Vehicle Time | Ratio: Wait to In-Vehicle Time | Value of Time (Auto) | Value of Time (Transit) |
|---|---|---|---|---|---|
| Baltimore | 1993 | 3.55 | 2.33 | $14.16 | $14.16 |
| Dallas | 1996 | 2.56 | 2.56 | $5.91 | $2.73 |
| Denver | 1985 | 3.00 | 1.57 | $3.09 | $2.45 |
| Detroit | 1996 | 2.00 | 2.00 | $7.61 | $1.36 |
| Houston | 1985 | 2.58 | 2.58 | $2.15 | $2.15 |
| Los Angeles | 1991 | 2.50 | 2.50 | $4.25 | $4.25 |
| Milwaukee | 1991 | 2.62 | 2.62 | $2.09 | $2.09 |
| Philadelphia | 1986 | 2.97 | 4.80 | $9.66 | $5.53 |
| Pittsburgh | 1978 | 1.47 | 1.47 | $1.33 | $1.33 |
| Portland | 1985 | 1.64 | 1.01 | $1.75 | $1.75 |
| Sacramento | 1991 | 1.52 | 1.52 | $5.39 | $5.39 |
| St. Louis | 1965 | 2.50 | 2.50 | $1.17 | $1.17 |
| Tucson | 1965 | 2.25 | 2.25 | $5.78 | $5.78 |

# Incremental Logit

$$p(k) = \frac{e^{U_k}}{\sum_x e^{U_x}}$$

when each utility has changed by $\Delta U_k$

$$p'(k) = \frac{e^{(U_k + \Delta U_k)}}{\sum_x e^{(U_x + \Delta U_x)}} \qquad (4.4)$$

$$i.e. p'(k) = \frac{e^{U_k} * e^{\Delta U_k}}{\sum_x [e^{U_x} * e^{\Delta U_x}]} \qquad (4.5)$$

# Incremental Logit - continue

Now dividing numerator and denominator by $\sum_x e^{U_x}$ we have,

$$p'(k) = \frac{\dfrac{e^{U_k}}{\sum_x e^{U_x}} * e^{\Delta U_k}}{\sum_x \left[ \dfrac{e^{U_x}}{\sum_x e^{U_x}} * e^{\Delta U_x} \right]} \tag{4.6}$$

$$\boxed{\therefore\ p'(k) = \frac{p(k) * e^{\Delta U_k}}{\sum_x [p(x) * e^{\Delta U_x}]}} \tag{4.7}$$

# Consumer Surplus

$$E(\mathrm{CS}_n) = \frac{1}{\alpha_n} E[\max_j (V_{nj} + \varepsilon_{nj})],$$

**If all errors are i.i.d extreme value ad utility is linear in income, this expectation become:**

$$E(\mathrm{CS}_n) = \frac{1}{\alpha_n} \ln\left(\sum_{j=1}^{J} e^{V_{nj}}\right) + C$$

$$\Delta E(\mathrm{CS}_n) = \frac{1}{\alpha_n}\left[\ln\left(\sum_{j=1}^{J^1} e^{V_{nj}^1}\right) - \ln\left(\sum_{j=1}^{J^0} e^{V_{nj}^0}\right)\right]$$

# Derivatives

$$\begin{aligned}
\frac{\partial P_{ni}}{\partial z_{ni}} &= \frac{\partial\left(e^{V_{ni}} / \sum_j e^{V_{nj}}\right)}{\partial z_{ni}} \\
&= \frac{e^{V_{ni}}}{\sum e^{V_{nj}}} \frac{\partial V_{ni}}{\partial z_{ni}} - \frac{e^{V_{ni}}}{\left(\sum e^{V_{nj}}\right)^2} e^{V_{ni}} \frac{\partial V_{ni}}{\partial z_{ni}} \\
&= \frac{\partial V_{ni}}{\partial z_{ni}} \left(P_{ni} - P_{ni}^2\right) \\
&= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni}).
\end{aligned}$$

If representative utility is linear in $z_{ni}$ with coefficient $\beta_z$, the derivative becomes $\beta_z P_{ni}(1 - P_{ni})$. This derivative is largest when $P_{ni} = 1 - P_{ni}$,

# Cross Derivatives

$$\frac{\partial P_{ni}}{\partial z_{nj}} = \frac{\partial\left(e^{V_{ni}} / \sum_k e^{V_{nk}}\right)}{\partial z_{nj}}$$

$$= -\frac{e^{V_{ni}}}{\left(\sum e^{V_{nk}}\right)^2} e^{V_{nj}} \frac{\partial V_{nj}}{\partial z_{nj}}$$

$$= -\frac{\partial V_{nj}}{\partial z_{nj}} P_{ni} P_{nj}.$$

# When an observed variable change, the changes in the choice probabilities sum to zero

$$
\begin{aligned}
\sum_{i=1}^{J} \frac{\partial P_{ni}}{\partial z_{nj}} &= \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj}(1 - P_{nj}) + \sum_{i \neq j} \left( -\frac{\partial V_{nj}}{\partial z_{nj}} \right) P_{nj} P_{ni} \\
&= \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj} \left[ (1 - P_{nj}) - \sum_{i \neq j} P_{ni} \right] \\
&= \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj} [(1 - P_{nj}) - (1 - P_{nj})] \\
&= 0.
\end{aligned}
$$

# Elasticities

$$
\begin{aligned}
E_{iz_{ni}} &= \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} \\
&= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni}) \frac{z_{ni}}{P_{ni}} \\
&= \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni}(1 - P_{ni}).
\end{aligned}
$$

If representative utility is linear in $z_{ni}$ with coefficient $\beta_z$, then $E_{iz_{ni}} = \beta_z z_{ni}(1 - P_{ni})$.

# Cross Elasticities

$$E_{iz_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}}$$
$$= -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj},$$

which in the case of linear utility reduces to $E_{iz_{nj}} = -\beta_z z_{nj} P_{nj}$

The cross elasticity is the same for all i, a change in an attribute of alternative j changes the probability for all other alternative by the same percent.
This manifests the IIA property