Modern Mode Choice Analysis

Philip A. Viton

March 21, 2014

Introduction (I)

Transit’s mode share in the urban US (all trip purposes) and the 10 urban areas where it is most popular (2008 data):

<table>
<thead>
<tr>
<th>Area</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban US</td>
<td>1.6</td>
</tr>
<tr>
<td>New York</td>
<td>11.0</td>
</tr>
<tr>
<td>San Francisco</td>
<td>5.0</td>
</tr>
<tr>
<td>Washington DC</td>
<td>4.5</td>
</tr>
<tr>
<td>Chicago</td>
<td>3.9</td>
</tr>
<tr>
<td>Honolulu</td>
<td>3.8</td>
</tr>
<tr>
<td>Boston</td>
<td>3.3</td>
</tr>
<tr>
<td>Seattle</td>
<td>2.8</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>2.7</td>
</tr>
<tr>
<td>Portland</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Introduction (II)

Clearly, one major problem facing transit (and transportation planners) is that not many people use it.

Can anything be done about this? Is there a way to make transit more attractive to potential users?

To answer this question we need to look at what determines people’s choices of transportation modes, for a specific trip purpose.

This is the mode choice analysis problem.

Introduction (III)

Until recently, mode choice studies were dominated by aggregate approaches: researchers would try to explain the fraction (share) of people in a city who chose a given mode (for example, the share using transit) based on average observed modal characteristics in the city, average disposable personal income, etc.

But two developments have changed this: we now have:

1. Extensive surveys of individuals.
2. The computing power needed to analyze this individual data.

So nowadays it is usual to focus on modeling and understanding behavior at the individual level.
The Setting

- Consider one individual making a single trip between an origin and a destination.
- She has available a variety of modes on which she could make her trip: drive alone, carpool, bus, walk, etc. These are her Choice Set.
- From her list of available modes (her choice set) she must select exactly one to use for her trip. (This may involve redefinitions of the modes, in order to have a trip made on a single mode).
- This setting — select exactly one from a list of options — defines the discrete choice setting.
- Discrete choice is contrasted with the more usual setting (called continuous choice) in which individuals can make multiple selections from the alternatives facing them.

Discrete Choice

Though our interest here is in mode choice, the discrete choice setting is applicable in a wide variety of decision-making contexts. Examples:

- Choice of a principal place (location) to live, or for a vacation home.
- Choice of a vacation destination.
- Choice of a car/major appliance model to purchase.
- Choice of a university (from among those where you’ve been accepted) to attend.
- Appellate judge: vote to affirm or reverse the lower court decision.
- City council member: vote for or against a project/plan.
- State agency: decide where to build a new freeway, from among a list of alternative routes.

Individual Behavior

- We study the behavior of a utility-maximizing, price-taking individual $i$.
- The individual faces $J$ discrete alternatives $j = 1, 2, \ldots, J$.
- Conditional utility: if individual $i$ selects alternative $j$ he/she gets utility $u_{ij}$.

Structure of Utility

Individual choice depends on:

- Factors observable to the analyst (eg, prices of alternatives).
- Factors unobservable to the analyst, but known to the decision-maker (eg, whether the individual’s last experience with a particular alternative was good/bad).
The Unobservable Part of Utility

- Consider an individual making repeated choices in a setting where:
  - the choice set does not change
  - the observable factors do not change
- Are we surprised if the individual makes different choices each time?
  - Clearly not: the reason is the influence of the unobservables, which may be changing.
- But conditional on the observables, the individual’s decisions vary for no observable reason: they appear to an observer/analyst to have an element of randomness to them.
- We would like our models to reflect this seeming randomness.

Detailed Structure of Utility

We now assume: for individual $i$, alternative $j$’s total utility:
- Has an observable (systematic) component $v_{ij}$.
- Has an unobservable (idiosyncratic) component that we represent by a random variable $\eta_{ij}$ distributed independently of $v_{ij}$.
- So (total) utility $u_{ij}$ is represented as:
  $$u_{ij} = v_{ij} + \eta_{ij}$$
- The random variable ($\eta_{ij}$) represents the analyst’s ignorance of some of the factors influencing individual $i$: it does not imply that the individual behaves randomly (from his/her own perspective).

Implications for the Study of Choice

- It is the random component in utility that allows the analyst to account for the apparent randomness exhibited by individuals in situations where the observables do not change.
- Because of the random component $\eta_{ij}$, the total utility $u_{ij}$ is also a random variable.
- This means that we can analyze only the probability that total utility takes on a given value.
- Hence, we as analysts can study only the choice probability that the individual makes a given choice.

Choice Probabilities

- We therefore define:
  $$P_{ij} = \Pr[\text{individual } i \text{ chooses alternative } j]$$
- By utility maximization, $i$ will select alternative $j$ if it yields the highest utility.
- So we have, equivalently:
  $$P_{ij} = \Pr[\text{alternative } j \text{ has highest utility, for individual } i]$$
Choice Probabilities — The Binary Case (I)

- It may be easier to see what is going on if we first study the binary case, where the individual has only two modes, say auto (mode 1) and transit (mode 2).
- In this case the mode-1 choice probability is:

\[ P_{i1} = \Pr[\text{individual } i \text{ chooses alternative 1}] = \Pr[\text{alternative 1 has highest utility, for individual } i] \]

(and of course we know that \( P_{i2} = 1 - P_{i1} \)).
- Write this in terms of the utilities:

\[ P_{i1} = \Pr[u_{i1} \geq u_{i2}] \]

Choice Probabilities — The Binary Case (II)

- Now expand the utilities into their systematic and random components:

\[ P_{i1} = \Pr[v_{i1} + \eta_{i1} \geq v_{i2} + \eta_{i2}] \]

Reverse the direction of the inequality:

\[ P_{i1} = \Pr[v_{i2} + \eta_{i2} < v_{i1} + \eta_{i1}] \]

and collect the random terms on the left, and the systematic terms on the right:

\[ P_{i1} = \Pr[\eta_{i2} - \eta_{i1} \leq v_{i1} - v_{i2}] \]

Choice Probabilities — The Binary Case (III)

- So our conclusion is that in binary choice individual i’s choice probability for mode 1 can be written as:

\[ P_{i1} = \Pr[\eta_{i2} - \eta_{i1} \leq v_{i1} - v_{i2}] \]

- We recognize this as the cumulative distribution function of the compound random variable \( \eta_{i2} - \eta_{i1} \) evaluated at the “point” \( v_{i1} - v_{i2} \).

Choice Probabilities — The Binary Case (IV)

- Individual i’s mode-1 choice probability \( P_{ij} \) is the cumulative distribution function \( G(\eta_{i2} - \eta_{i1}) \) of the compound random variable \( \eta_{i2} - \eta_{i1} \) evaluated at the “point” \( v_{i1} - v_{i2} \).
Choice Probabilities — The General Case (I)

- The general case is done in the same way, though the expressions are more complicated, since there are now \( J \) modes/alternatives.
- Alternative \( j \) is better than alternative \( k \) if \( u_{ij} \geq u_{ik} \)
- Alternative \( j \) is best if:
  - alternative \( j \) is better than alternative 1 AND
  - alternative \( j \) is better than alternative 2 AND
  - \( \ldots \) AND \( \ldots \)
  - alternative \( j \) is better than alternative \( J \)
  (this is \( J - 1 \) conditions).

Choice Probabilities — The General Case (II)

- So

\[
P_{ij} = \Pr[\text{individual } i \text{ chooses alternative } j]
= \Pr[\text{alternative } j \text{ has the highest utility for individual } i]
= \Pr[u_{ij} \geq u_{i1} \& u_{ij} \geq u_{i2} \& \ldots \& u_{ij} \geq u_{iJ}]
\]

(\( J - 1 \) conditions).

Choice Probabilities — The General Case (III)

- Now expand each utility into its systematic and random components:

\[
P_{ij} = \Pr[v_{ij} + \eta_{ij} \geq v_{i1} + \eta_{i1}, \ldots, v_{ij} + \eta_{ij} \geq v_{iJ} + \eta_{ij}]
\]

Reverse the direction of the inequalities:

\[
P_{ij} = \Pr[v_{i1} + \eta_{i1} \leq v_{ij} + \eta_{ij}, \ldots, v_{iJ} + \eta_{ij} \leq v_{ij} + \eta_{ij}]
\]

and collect up terms (random on the left):

\[
P_{ij} = \Pr[\eta_{i1} - \eta_{ij} \leq v_{ij} - v_{i1}, \ldots, \eta_{iJ} - \eta_{ij} \leq v_{ij} - v_{iJ}]
\]

(\( J - 1 \) terms in each expression; decide ties in favor of alternative \( j \)).

Choice Probabilities — The General Case (IV)

- So we have:

\[
P_{ij} = \Pr[\eta_{i1} - \eta_{ij} \leq v_{ij} - v_{i1}, \ldots, \eta_{iJ} - \eta_{ij} \leq v_{ij} - v_{iJ}]
\]

- We see that the choice probabilities are the joint cumulative distribution function of the \( J - 1 \) random variables \( \eta_{ik} - \eta_{ij} \ (k \neq j) \) evaluated at the “points” \( v_{ij} - v_{ik} \).
- Any assumption about the joint distribution of the random variables (the \( \eta_{ij} \)) will in principle give rise to a parametric choice model.
The Logit Model

- But one model has dominated the literature.
- Suppose the $\eta_{ij}$ are independently and identically distributed (i.i.d) random variables with Type1-Extreme-Value distributions (see Appendix).
- Then the choice probabilities are given by:
  $$P_{ij} = \frac{e^{v_{ij}}}{\sum_{k=1}^{J} e^{v_{ik}}}$$

known as the Logit model of discrete choice.
- Note that once you know the $v_{ij}$ is is very easy to calculate the choice probabilities $P_{ij}$. So we turn now to that issue.

Structure of Systematic Utility

- Suppose alternative $j$ has $K$ observable characteristics (as perceived by individual $i$): $(x_{ij1}, x_{ij2}, \ldots, x_{ijk}) = x_{ij}$
- It is conventional to assume that systematic utility $v_{ij}$ is a linear-in-parameters function of the $x_{ij}$:
  $$v_{ij} = x_{ij1} \beta_1 + x_{ij2} \beta_2 + \cdots + x_{ijk} \beta_K$$
  $$= x_{ij} \beta$$

- So the logit choice probabilities become:
  $$P_{ij} = \frac{e^{x_{ij} \beta}}{\sum_{k=1}^{J} e^{x_{ik} \beta}}$$

  in which the only unknown is the weighting vector $\beta$.

Data (I)

- We want to estimate $\beta$ from data on a sample of individuals.
- The simplest assumption is that we have a random sample.
- For each individual we observe the complete set of characteristics (the $x_{ij}$) of the alternatives, whether chosen or not.
- We also observe a choice indicator, defined as:
  $$y_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses alternative } j \\ 0 & \text{otherwise} \end{cases}$$

Data (II)

<table>
<thead>
<tr>
<th>Indiv</th>
<th>Mode</th>
<th>$y$</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$y_{11}$</td>
<td>$x_1$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$y_{12}$</td>
<td>$x_{111}$</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$y_{1J}$</td>
<td>$x_{1J1}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$y_{21}$</td>
<td>$x_{211}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$y_{22}$</td>
<td>$x_{221}$</td>
</tr>
<tr>
<td></td>
<td>$J$</td>
<td>$y_{2J}$</td>
<td>$x_{2J1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
</tbody>
</table>
Estimation

- The model is usually estimated by maximum likelihood: choose \( \beta \) to maximize:

\[
\mathcal{L} = \prod_{i=1}^{I} \prod_{j=1}^{J} P_{ij}^{y_{ij}}
\]

where \( I \) is the sample size. (Or, to avoid numerical issues: maximize log likelihood, which is equivalent).

- Properties: in large samples, the MLE of \( \beta \) is
  - consistent (i.e., converges to true — but unknown — parameter value).
  - asymptotically normal (so we can use standard tests, like t-test).
  - asymptotically efficient (i.e., has minimum variance among all consistent estimators).

- Software: any stand-alone statistics package almost certainly includes estimation of the discrete-choice logit model. Examples include SYSTAT, LIMDEP, R, Stata, SAS, SPSS.

Example (I)

- Individuals face a discrete choice between 4-modes: auto-alone; bus + walk access; bus + auto access; carpool.
- Demand model: model 12 from McFadden + Talvitie (1978), estimated for SFBA work trips.
- Structure of utility: naive model in which choices are explained by only a few factors:
  - Cost (and the individual’s post-tax wage).
  - In-vehicle time.
  - Excess time: this is the time to access the mode, plus wait time for transit modes.
  - Dummies: these take the value 1 for the mode in question, 0 otherwise. Also known as alternative-specific constants.

Example (II)

- Results of estimating the logit model:

<table>
<thead>
<tr>
<th>Indep var.</th>
<th>Estd. Coef</th>
<th>t-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost ÷ post-tax-wage (( c / c / \text{min} ))</td>
<td>-0.0412</td>
<td>7.63</td>
</tr>
<tr>
<td>In-vehicle time (mins)</td>
<td>-0.0201</td>
<td>2.78</td>
</tr>
<tr>
<td>Excess time (mins)</td>
<td>-0.0531</td>
<td>7.54</td>
</tr>
<tr>
<td>Auto dummy</td>
<td>-0.892</td>
<td>3.38</td>
</tr>
<tr>
<td>Bus+Auto dummy</td>
<td>-1.78</td>
<td>7.52</td>
</tr>
<tr>
<td>Carpool dummy</td>
<td>-2.15</td>
<td>8.56</td>
</tr>
</tbody>
</table>

This model correctly predicted 58.5% of the choices actually made by the sample, which is considered good for these disaggregate models.

Example — Assumed Data

- We assume that for this individual the modal characteristics are:

<table>
<thead>
<tr>
<th>Raw costs, fares (( c / \text{min} ))</th>
<th>100</th>
<th>125</th>
<th>125</th>
<th>50</th>
</tr>
</thead>
</table>

We assume that car-pooling involves 2 people and they split the auto costs. Post-tax income: $50,000 per year = 40.8479 \( c / \text{min} \)
**Example — Systematic Utilities**

- Mode 1 (auto-drive-alone). Cost = 100\(c\); post-tax wage = 40.8479\(c\), so \([\text{cost} \div \text{post-tax wage}] = 100 \div 40.8479 = 2.448\).

- Then:
  
  \[
  v_{i1} = x_{i1}\beta = (2.448 \times (-0.0412)) + (20 \times (-0.0201)) \\
  + (0 \times (-0.0531)) + (1 \times (-0.892)) \\
  + (0 \times (-1.78)) + (0 \times (-2.15)) \\
  = -1.39486
  \]

- Calculations for the other modes follow the same logic.

**Example — Predicted Probabilities**

Here are the details for computing the mode choice probabilities based on the example data:

<table>
<thead>
<tr>
<th>Mode:</th>
<th>Auto</th>
<th>Bus+Walk</th>
<th>Bus+Auto</th>
<th>Carpool</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_{ij}\beta)</td>
<td>-1.39486</td>
<td>-1.26007</td>
<td>-2.77457</td>
<td>-3.13343</td>
</tr>
<tr>
<td>(e^{x_{ij}\beta})</td>
<td>0.24787</td>
<td>0.28363</td>
<td>0.06237</td>
<td>0.04357</td>
</tr>
<tr>
<td>(\sum_j e^{x_{ij}\beta})</td>
<td>0.63744</td>
<td>0.63744</td>
<td>0.63744</td>
<td>0.63744</td>
</tr>
<tr>
<td>(P_{ij})</td>
<td>0.38885</td>
<td>0.44495</td>
<td>0.09785</td>
<td>0.06835</td>
</tr>
</tbody>
</table>

**Marginal Effects (I)**

- How do the choice probabilities change as we change one of the observable characteristic? The answer depends on whether we change an own-characteristic or a cross-characteristic.

- For an own-characteristic we ask: how does \(P_{ij}\) change if we change the \(k\)-th characteristic \(x_{ijk}\) of mode \(j\) itself? (How does the probability of choosing COTA change if we change COTA’s fare?). In this case we have:
  
  \[
  \frac{\partial P_{ij}}{\partial x_{ijk}} = \beta_k P_{ij}(1 - P_{ij})
  \]

**Marginal Effects (II)**

- For a cross-characteristic we ask: how does \(P_{ij}\) change if we change the \(k\)-th characteristic \(x_{irk}\) of some other mode \(r\)? (How does the probability of choosing COTA change if we change the cost of using the automobile?). In this case we have:
  
  \[
  \frac{\partial P_{ij}}{\partial x_{irk}} = -\beta_k P_{ij} P_{ir}
  \]

- Note that in both cases the impact depends on the starting (pre-change) choice probabilities.
**Example — Marginal Effects**

- The impact on individual $i$ of a 1-unit (ie 1-cent) increase in the bus fare (characteristic 1) for the bus+walk mode (mode 2) is (remembering that in the specification of utility we divide by the post-tax wage):

$$
\frac{\partial P_{i2}}{\partial x_{i21}} = \frac{\beta_1}{w_i} P_{i2}(1 - P_{i2})
$$

$$= -0.0412 \cdot 0.46935 \cdot (1 - 0.46935)
$$

$$= -0.0025
$$

- The impact of the same change on the auto (mode 1) probability is:

$$
\frac{\partial P_{i1}}{\partial x_{i21}} = -\frac{\beta_1}{w_i} P_{i1} P_{i2}
$$

$$= -0.0412 \cdot 0.36354 \cdot 0.46935
$$

$$= 0.0017
$$

**Value of Time (I)**

Consider the following story:

- An individual has bus transit available at a price of $1.75/trip; and her trips take 20 minutes.
- Suppose that the transit provider reduces the travel time to 15 mins/trip; but at the same time, raises the fare to $2.10.
- And suppose that we see that our individual makes exactly the same number of transit trips in the two situations.

**Value of Time (II)**

What does this story tell us?

- We conclude that (because her behavior is unchanged) the $\Delta C = $2.10 $- $1.75 $= $0.35 cost increase exactly balances the $\Delta T = 20 - 15 = 5$ mins travel time improvement.
- So for her, a 5 minute reduction in travel time is worth $0.35.
- We say that her (unit) value of time is:

$$w_T = \frac{\Delta C}{\Delta T} = \frac{0.35}{5} = $0.07 / minute
$$

ie 7¢ per minute or $0.07 \times 60 = $4.20 per hour.

**Value of Time (III)**

Let’s translate this story into symbols.

- Assume that we can write the systematic utility as:

$$v_{ij} = \beta_C C_{ij} + \beta_T T_{ij} + \beta_E x_{ij}^E
$$

where:

- $C_{ij}$ is the cost to $i$ of alternative $j$
- $T_{ij}$ is the time taken, if alternative $j$ is chosen by $i$ (or more generally, any observed characteristic of alternative $j$)
- $x_{ij}^E$ is everything else

- At the original cost $C_{ij}^0$ and time $T_{ij}^0$, her utility is:

$$v_{ij}^0 = \beta_C C_{ij}^0 + \beta_T T_{ij}^0 + \beta_E x_{ij}^E
$$

- Now let cost change by $\Delta C$ and travel time by $\Delta T$. Then her new utility is:

$$v_{ij}^1 = \beta_C (C_{ij}^0 + \Delta C) + \beta_T (T_{ij}^0 + \Delta C) + \beta_E x_{ij}^E$$
Value of Time (IV)

- The change in utility is:
  \[ \Delta v_{ij} = v^1_{ij} - v^0_{ij} = \beta_C \Delta C + \beta_T \Delta T \]

- In order for behavior to be unchanged, we must have \( \Delta v_{ij} = 0 \). Now solve for the value of time: we find:
  \[ w_T = \frac{\Delta C}{\Delta T} = -\frac{\beta_T}{\beta_C} \]

and note that this can be read off directly from the results of our logit model, where we estimate the \( \beta \) coefficients.

Value of Time — Example (I)

- Our example empirical specification was:
  \[ v_{ij} = \beta_C \frac{C_{ij}}{m_i} + \beta_T T_{ij} + \beta_E x_{ij}^E \]
  where \( m_i \) is \( i \)'s post-tax wage. This differs from the model in slide 36. So the first thing to do is to re-write it to match slide 36.

- We have:
  \[ v_{ij} = \frac{\beta_C}{m_i} C_{ij} + \beta_T T_{ij} + \beta_E x_{ij}^E = \beta_C' C_{ij} + \beta_T T_{ij} + \beta_E x_{ij}^E \]
  where \( \beta_C' = \beta_C / m_i \) (that is we have a coefficient \( \beta_C' \) multiplying \( C_{ij} \)). In this form, it matches slide 36.

Value of Time (V)

- In general, for any formulation of the systematic utility function \( v_{ij} \) (including linear-in-parameters), \( i \)'s value of time is given by:
  \[ w_T = -\frac{\partial v_{ij}}{\partial C_{ij}} \]

- This represents \( i \)'s willingness to trade time savings for cost savings, while keeping utility (ie behavior) constant.

- Values of time represent one easy way to assess the impact on individuals of changes in travel time associated with usage of transit modes.

- Of course, exactly the same idea can be applied to any other characteristic in a utility function.

Value of Time — Example (II)

- Then, applying our earlier result, we immediately see that the value of time in this model is:
  \[ w_T = -\frac{\beta_T}{\beta_C} m_i = -0.0201 m_i = -0.0412 \]
  \[ w_T = -\frac{\beta_T}{\beta_C} m_i = -0.487 86 m_i \]

In the mode-choice context it is conventional to ignore the minus sign and say that the value of in-vehicle time is (about) 49% of the post-tax wage.
The Remainder of These Notes

The remainder of these notes briefly discuss a couple of additional topics in mode choice analysis which will not be needed for our discussion of the role of public transit. These topics are covered in more detail in CRP 5700.

IIA Property

For the logit model, a simple calculation shows:

\[
\frac{P_{ij}}{P_{ik}} = \frac{e^{v_{ij}}}{e^{v_{ik}}}
\]

We say: the odds of \(i\)'s choosing alternative \(j\) over alternative \(k\) \((P_{ij} / P_{ik})\) depends only on the observed characteristics of alternatives \(j\) and \(k\) (ie only on \(v_{ij}\) and \(v_{ik}\)), and not on the characteristics of any other ‘irrelevant’ alternatives.

Another interpretation: \(P_{ij} / P_{ik}\) is the same, no matter what other alternatives are present in individual \(i\)'s choice set.

This is the Independence of Irrelevant Alternatives (IIA) property.

Red Bus / Blue Bus (I)

- Consider a situation where a population faces 2 travel alternatives, say auto (A) and a transit bus system whose vehicles are painted red: call it the Red-Bus (RB) system.
- We interpret the choice probabilities as aggregate modal shares: suppose that \(P_A = 0.70\) while \(P_{RB} = 0.30\).
- Now suppose we add a new mode: a second transit system identical to the first in every respect except that its buses are painted blue (alternatively: paint half the red buses blue). Call this the Blue Bus (BB) system.
- What do we expect for the modal shares? Obviously, the two bus systems will split the transit-using population, so:

\[
\begin{align*}
P_A' &= 0.70 \\
P_{RB}' &= 0.15 \\
P_{BB}' &= 0.15
\end{align*}
\]

Red Bus / Blue Bus (II)

- **The logit model does not agree.** It says:

\[
\begin{align*}
\frac{P_A''}{P_{RB}''} &= \frac{P_A}{P_{RB}} = 0.70/0.30 = 2.333 \quad \text{(by IIA)} \\
P_{BB}'' &= P_{BB}'' \quad \text{(identical bus systems)} \\
P_A'' + P_{RB}'' + P_{BB}'' &= 1 \quad \text{(probabilities must sum to 1)}
\end{align*}
\]

- Hence we predict:

\[
\begin{align*}
P_{RB}'' &= \frac{1}{2.333} = 0.23079 \\
P_{BB}'' &= 0.23079 \\
P_A'' &= 0.53842
\end{align*}
\]

which we know to be incorrect. Conclusion: We should not rely on the logit model in cases where the alternatives are very similar (here, identical).
Testing for IIA

Before using the logit model it is important to know: are the choices made by the individuals consistent with the IIA property?

- If so, then it is safe to estimate a logit model and use it for prediction.
- If not, then the logit model may give wrong predictions.

There are several tests in the literature to decide whether observed choices satisfy IIA. Examples: the Hausmann-McFadden and the Small-Hsiao tests. (See advanced texts for references).

Models Without IIA

In advanced treatments we can formulate models that do not involve the IIA property, and hence will not mis-predict in the Red-Bus / Blue-Bus (or similar) cases. Some of these are:

- Nested logit.
- Heteroskedastic logit.
- Mixed logit: this is the most recent development, in which the coefficients (the $\beta$’s) of systematic utility themselves become random variables, that is, vary over the population.

Modern computational techniques also allow estimation of models in which the idiosyncratic elements of utility (the $\eta$’s) are (non-independently) Normally distributed: this is the Probit family of models.

Elasticity Interpretation (I)

- Consider the elasticity of $i$’s choice probability for mode $j$ with respect to the $k$-th characteristic of some other mode $r$: The elasticity is defined as
  $$E_{ijk} = \frac{\% \text{ change in } P_{ij}}{\% \text{ change in } x_{irk}}$$

- And for the logit model we have:
  $$E_{ijk} = \frac{\partial P_{ij}}{\partial x_{irk}} \frac{x_{irk}}{P_{ij}} = -\beta_k P_{ir} x_{irk}$$
  (using the expression for the marginal cross-effect previously derived).

Elasticity Interpretation (II)

What is interesting about the elasticity expression is that it is independent of mode $j$, or putting it differently, will take the same value for any mode $j$ other than $r$.

- To see why this could be a problem for the logit model, suppose mode 1 is COTA, mode 2 is a cheap Yugo subcompact, and mode 3 a high-end Lexus.
- Suppose COTA raises its fare by 10%.
- Then the equation implies that this will have an equal impact on the choice probability for the Yugo and for the Lexus.
- This is surely implausible: one would expect the impact of a COTA fare change to be major for the Yugo and relatively minor for the Lexus.
- But the logit model predicts differently. This is another way of stating the IIA problem with logit.
A random variable $\eta$ has a Type-1-Extreme-Value distribution (sometimes known as Weibull or Gnedenko distribution) if its probability density (frequency) function is

$$g(\eta) = e^{-\eta} e^{-e^{-\eta}}$$

and its cumulative distribution function is

$$G(a) = Pr(\eta \leq a) = e^{-e^{-a}}$$

This (plus independence) is the distribution of the random elements $(\eta_{ij})$ underlying the logit model.

References