Discrete Choice Modeling
William Greene
Stern School of Business, New York University

Lab Session 5
Multinomial Choice

===
This assignment will consist of some exercises with the multinomial logit and nested models. The data for the exercises is the multinomial choice file

mnc.lpj

This project file contains both the brand choices data used in the first set of exercises and the travel mode data used later. Altogether, there are 12,800 observations in the brand choices data. The travel mode data appear in the first 840 rows of the data area.
===

Part I. Conditional Logit, Nested Logit and Other Models

1. Test for Functional Form of Utility Functions

The discrete choice model we will use in this exercise is

$$U(\text{brand}) = \beta_1\text{Fashion} + \beta_2\text{Quality} + \beta_3\text{Price} + \beta_4\text{ASC4} + \varepsilon_{\text{brand}},$$

for brand = brand1, brand2, brand3 and none. Fashion, Quality and Price are all zero for NONE, while ASCNONE is 1 for NONE and zero for the others. This is a convenient way to consider the 'none of the above' choice. We are interested in testing the hypothesis that the price enters the utility function quadratically, rather than linearly. Thus, we test for significance of an additional variable, PriceSq = Price^2. The commands below can carry out the LR test. What do you find? Utility functions can become extremely involved. Sometimes the generic form

$$U(\text{alt}) = \boldsymbol{\beta}'\mathbf{x}_{\text{alt}}$$

is not flexible enough. A second form is available. The second **CLOGIT** command illustrates. The one shown produces results identical to the first. As another example, suppose it were known that the utility of brand 3 did not depend on quality. The utility function for brand3 might be different from those of brands 1 and 2. The third command illustrates.

```
?  Test for significance of squared term and illustrate specification
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4 $
CALC            ; L0 = logl $
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Model: U(brand1,brand2,brand3) = bf*fash + bq*qual + bp*price /
                         U(none) = ascnone $
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Model: U(brand1,brand2) = bf*fash + bq*qual + bp*price /
                         U(brand3) = bf*fash + bp*price /
                         U(none) = ascnone $
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Pricesq,Asc4 $
CALC            ; L1  = logl ; list ; chisq = 2*(L1 - L0) $
```

2. Structural Change.

Are men's preferences the same as women's? We carry out the equivalent of a Chow test for structural change. What do you find. Is the null hypothesis: H_0: $\boldsymbol{\beta}_M = \boldsymbol{\beta}_W$ for the vector of parameters in the model rejected or not? We carry out the test using the likelihood ratio and Wald tests. The final instruction shows a way to automate the test.

```
CLOGIT          ; For[Male = 0] ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4 $
CALC            ; LoglF = LogL $
MATRIX          ; db = b ; dv = varb $
CLOGIT          ; For[ Male = 1] ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4 $
CALC            ; LoglM = LogL $
MATRIX          ; db = db - b ; dv = dv + varb $
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4 $
CALC            ; LoglMF = LogL $
? Likelihood Ratio Test
CALC            ; List ; Chisq = 2*(LoglM + LogLF - LOGLMF) ; Ctb(.95,8) $
? Wald test
MATRIX          ; List ; Wald = db'<dv>db $
? A convenient way to compute all three models.
CLOGIT          ; For[(test) Male = *,1,0] ; Lhs = Choice
                ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4 $
```

3. Marginal Effects of Price on Brand Choice

We estimate a marginal effect (of price) in the MNL model. What are the estimates of the own and cross elasticities across the three brands? What is the evidence of the IIA assumption in these results?

```
?  3.  Examine the marginal effect of price on brand choice
CLOGIT          ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,Asc4
                ; Effects : Price (*) $
```

4. Impact of a Price Change on Market Shares

What would happen to the market shares of the three brands if the price of Brand 1 of shoes rose by 50%. What would happen to the market shares if the prices of all three brands rose by 50%

```
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4
          ; Effects : Price (*) $
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4
          ; Simulation = *
          ; Scenario: Price (Brand1)  = [*] 1.5 $
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4
          ; Simulation = *
          ; Scenario: Price (Brand1,Brand2,Brand3)  = [*] 1.5  $
```

5. Testing for IIA

Is Brand3 an irrelevant alternative in the choice model? Given the way the data are constructed, one wouldn't think so. Here we investigate. Carry out the Hausman to test the IIA assumption using Brand 3 as the omitted alternative. What do you find?

```
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4  $
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4
          ; IAS = Brand3  $
```

6. Functional Form and Marginal Impact.

Do men pay more attention to fashioni than women? To investigate, we fit the choice model with a different coefficient on fashion for men and women. Then, simulate the model so as to see what happens when the variable which carries this effect into the model is zero'd out. What are the results? How do you interpret your findings?

```
CREATE          ; MaleFash = Male*Fash $
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4,MaleFash $
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Rhs = Fash,Qual,Price,Asc4,MaleFash
          ; Simulation = * ; Scenario: MaleFash(*) = [*] 0 $
```

7. Heteroscedastic Extreme Value Model

Fit the MNL model while allowing the variances to differ across the utiity functions. First, fit the basic MNL model. Then, allow the variances to vary. Finally, allow the variances to vary across utililties and with age and sex. In each case, obtain the marginal effects with respect to price. Does the change in the model specification produce changes in the impacts? The log likelihoods for the heteroscedasticity models involves an integral that is computed using Laguerre quadrature. The default setting is to use 64 points to compute the integrals. To speed up the computations, we reduce this to 8, and stop after 10 iterations.

```
?
? 7.  Building heterogeneity into the model
?
NLOGIT (none)   ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,ASC4
                ; Effects: Price(Brand1,Brand2,Brand3)  $
NLOGIT (het)    ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,ASC4
                ; Het  ; Effects: Price(Brand1,Brand2,Brand3)
                ; Maxit=10 ; Lpt = 8$
NLOGIT (variance, het) ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,ASC4
                ; Het ; Hfn = Male,age25,age39  ; Effects: Price(Brand1,Brand2,Brand3)
                ; Maxit=10 ; Lpt = 8$ $
```

8. Constraint

Test the hypothesis that the variances in the four utility functions are all equal. Since one of them is normalized to one, this is done by testing whether the first J-1 are equal. In NLOGIT's HEV model, the first set of values reported are $(\sigma_j/\sigma_J - 1)$, so the desired test can be carried out by testing the hypothesis that these three (J-1) coefficients are zero. Carry out the test using the brand choice data. What do you find?

```
?
? 8.  Testing for homoscedasticity with a Wald test
?
NLOGIT (het)    ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
                ; Rhs = Fash,Qual,Price,ASC4
                ; Het ; Par $
MATRIX          ; c = b(5:7) ; vc = Varb(5:7,5:7)  ; List ; WaldStat = c'<vc>c $
```

9. Gender Effect on Mode Choice Model Model

Does gender affect the means in the utility functions of the mode choice model, or the variances? We will use a Vuong test to explore the question. The initial random utility model has

$$U_{ij} = \beta_1 FASH_{ij} + \beta_2 QUAL_{ij} + \beta_3 PRICE_{ij} + \beta_4 ASC4_{ij} + \delta_j Male_i + \varepsilon_{ij}$$

where $Var[\varepsilon_{ij}] = \sigma^2$, the same for all utilities. The second form of the model is

$$U_{ij} = \beta_1 FASH_{ij} + \beta_2 QUAL_{ij} + \beta_3 PRICE_{ij} + \beta_4 ASC4_{ij} + \varepsilon_{ij}$$

where $\text{Var}[\varepsilon_{ij}] = \sigma_j^2 \times \exp(\gamma \text{ Male}_i)$. These models are not nested, so we cannot use a likelihood ratio test to test one against the other. We use a Vuong test, instead. We fit each model, then for each, we retrieve LogL_i, the contribution of each individual to the log likelihood. The Vuong statistic is computed by first obtaining

$$m_i = \text{LogL}_{i0} - \text{LogL}_{i1}$$

where LogL_{i0} and LogL_{i1} are the contributions to the log likelihood for the null model and the alternative model, respectively. We then compute the Vuong statistic,

$$V = \frac{\sqrt{n}\ \bar{m}}{s_m}$$

The limiting distribution of the Voung statistic is standard normal. Large positive values (using 1.96 for 95% confidence) favor the null hypothesis, large negative values favor the alternative hypothesis. The travel mode data are not well suited for this model,

Note, in the calculations below, for a MNL model with J alternatives, NLOGIT stores the individual log likelihoods with the last alternative, in this case the 4th brand, none. The brand variable is a counter variable which equals 4 for the 4th alternative, so it is a convenient device to restrict our sample to the observations we want for our computation. Carry out the test. What do you conclude?

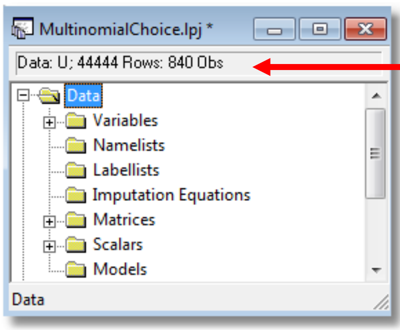
```
?
? 9. Heterogeneous.  Does Income affect the means or the variances?
?
SAMPLE          ; 1-12800 $
NLOGIT          ; lhs=choice;choices=b1,b2,b3,none
                ; rhs=fash,qual,price,asc4;rh2=male$
CREATE          ; LOGLMean = Logl_Obs $
NLOGIT          ; lhs=choice;choices=b1,b2,b3,none
                ; rhs=fash,qual,price,asc4;het;hfn=male$
CREATE          ; LoglVar = Logl_Obs $
CREATE          ; V = LoglMean - LogLVar $
REJECT          ; brand < 4 $
CALC            ; List ; Vuong = sqr(n) * xbr(v) / sdv(v) $
SAMPLE          ; 1-12800 $
```

==

This part of the assignment will use the mode choice, conditional logit data. In what follows, be sure that you are only using the first 840 rows in the combined data set. The command to set this data set is

```
SAMPLE ; 1 – 840 $
```

You can see the number of observations in the current sample at the top of the project window, as shown below.



If this value is not 840 at any time, you can just issue the sample command to reset the sample. Note, it is possible to enforce this ‘on the fly’ in the estimation commands by including ‘;If[_Obsno <= 840]’ at the beginning of the command.

==

1. Nested Logit Model.

We begin with a simple nested logit model.

```
?
? 1.  Basic nested logit
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Private(Air,Car),Public(Train,Bus)
                ; Show Tree $
CALC            ; LOGLU = LOGL $
```

2. RU1 and RU2

These are different formulations of the model. They are not simple reparameterizations of the model, so they will not give identical results in a finite sample. Which is the appropriate to use is up to the analyst. There is no way to test the specification as a hypothesis.

```
?
? 2.  Different normalization
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Private(Air,Car),Public(Train,Bus)
                ; RU2
                ; CrossTab$
```

3. Constrained Nested Logit Model

Constraining the IV parameters to equal 1 returns the original multinomial logit model. Use this device to test the restriction. Note that this specification test is whether the MNL is appropriate, against the alternative of the nested logit model.

```
?
? 3.  Constrain IV parameters to produce MNL model
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Private(Air,Car),Public(Train,Bus)
                ; IVSET:(Private,Public)=[1] $
CALC            ; LOGLR = LOGL $
CALC            ; List ; LRTEST = 2*(LOGLU - LOGLR) $
```

4. Degenerate Branch.

A branch that contains only one alternative is labeled ‘degenerate’ (for reasons lost to antiquity). The RU1 and RU2 normalizations produce different results for such models. Fit the two and examine the effect.

```
?
? 4.  Degenerate branch.  Two normalizations
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Fly(Air),Ground(Car,Train,Bus) $
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Fly(Air),Ground(Car,Train,Bus) ; RU2 $
```

5. Alternative Approaches to Reveal Scaling

The nested logit model can be modified to act like the heteroscedastic extreme value buy making all branches contain one alternative. This will allow a different scale parameter in each branch. The HEV model is another way to do this. Are the results similar?

```
?
? 5. Use nested logit to reveal scaling.
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Fly(Air),Drive(Car),Rail(Train),Ride(Bus)
                ; IVSET: (Ride) = [1]
                ; Par $
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; HET ; SDV = SA,ST,1.0,SC $
```

6. Generalized Nested Logit Model

The GNL model is a fairly exotic formulation (not yet in the mainstream) of the nested logit model in which alternatives may appear in more than one branch. The model allocates a portion of the alternative to the various branches. We fit one here, and leave the interpretation of the resulting model to the analyst.

```
?
? 6.  A Generalized Nested Logit Model
?
NLOGIT          ; Lhs = Mode ; Choices=Air,Train,Bus,Car
                ; Rhs = One,GC,TTME,INVT,INVC
                ; Tree = Fast(Air,Car,Train),Public(Train,Bus)
                ; GNL $
```

7. HEV Model

Fit an HEV model with these data, allowing the variances of the utilities to differ across alternatives. Use a likelihood ratio test to test for equal variances. Examine the impact of the heteroscedasticity on the marginal effect of IN VEHICLE TIME (INVT).

```
?
? 7.  Homoscedastic vs. Heteroscedastic Extreme Value
?
NLOGIT          ; Lhs = Mode
                ; Choices = Air,Train,Bus,Car
                ; Rhs = TTME,INVC,INVT,GC,One
                ; Effects: INVT(*) $
CALC            ; LR = LogL $
NLOGIT          ; Lhs = Mode
                ; Choices = Air,Train,Bus,Car
                ; Rhs = TTME,INVC,INVT,GC,One
                ; Het
                ; Effects: INVT(*) $
CALC            ; LU = LogL $
CALC            ; List ; LRTEST = 2*(LU - LR); Ctb(.95,3) $
```

This assignment involves a sampling of latent class models. Though there are, of course, many aspects of the underlying models, latent class modeling, itself, is fairly uncomplicated. That is, beyond the underlying models, latent class modeling involves a small number of straightforward principles. In this exercise, we will fit a handful of latent class models to different kinds of choice variables.

8. Multinomial Probit Model

Do the multinomial logit and multinomial probit models give similar results? You can't tell directly from the coefficient estimates because of scaling and normalization, so you have to rely on other indicators such as marginal effects. Fit a multinomial probit and a multinomial logit model, and compare the results. Note, estimation of the MNP model is ***extremely*** slow, so we have set it up with a very small number of replications and stopped the iterations at 5. This particular model would take 30-50 iterations, and an hour or two, to finish.

```
?
? 8. Multinomial Probit Model
?
NLOGIT          ; Lhs = Mode ; output=ic
                ; Choices = Air,Train,Bus,Car
                ; Rhs = TTME,INVC,INVT,GC; Rh2=One,Hinc
                ; Effects:GC(*) $
NLOGIT          ; Lhs = Mode ; MNP ; PTS = 5 ; Maxit = 5 ; Halton
                ; Choices = Air,Train,Bus,Car
                ; Rhs = TTME,INVC,INVT,GC; Rh2=One,Hinc
                ; Effects:GC(*) $
```

Part II. Latent Class and Random Parameter Models

The next set of computations is based on the shoe brand choices data.

Be sure that the sample setting uses all the data. Use

SAMPLE ; 1 - 12800 $

to set the sample correctly. Note that in these simulated data, the true underlying model really is a latent class data generating mechanishm, with three classes.

1. Latent Class Model for Brand Choice.

First, fit a simple three class model with constant class probabilities. Then, fit the same model, but allow the class probabilities to very with age and sex. Finally, since we know that the true model is a three class model, we explore what happens when the model is over fit by fitting a four class model.

```
?
? (1) Basic 3 class model.
?
Nlogit   ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
         ; Rhs = Fash,Qual,Price,ASC4
         ; LCM ; Pds = 8 ; Pts = 3 $
?
? (2) 3 class model. Class probabilities depend on covariates
?
Nlogit   ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
         ; Rhs = Fash,Qual,Price,ASC4
         ; LCM=Male,Age25,Age39 ; Pds = 8 ; Pts = 3 $
?
? (3) Overspecified model.  4 class model.  The true model
?    underlying the data has three classes
?
Nlogit   ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
```

```
; Rhs = Fash,Qual,Price,ASC4
; LCM ; Pds = 8 ; Pts = 4 $
```

2. Attribute Nonattendance

One aspect of choice theory that has attracted some attention is the possibility that in a choice study, some individuals in the sample may be ignoring some of the attributes. This is called ‘attribute nonattendance’ in the recent literature. There are two ways to handle this phenomenon, depending on the available information. Consider an example: The utility function for shoe brands for a generic consumer is

$$U_{ij} = \beta_1 FASH_{ij} + \beta_2 QUAL_{ij} + \beta_3 PRICE_{ij} + \beta_4 ASC4_{ij} + \varepsilon_{ij}$$

However, if an individual is known not to consider FASH in their choice, then the appropriate utility function is

$$U_{ij} = 0 \times FASH_{ij} + \beta_2 QUAL_{ij} + \beta_3 PRICE_{ij} + \beta_4 ASC4_{ij} + \varepsilon_{ij}$$

That is, with a zero in the first position. (Simply coding FASH to be zero is not the solution.) NLOGIT automatically handles this case with it’s “-888” special code. If the program finds the special missing value -888 for an attribute, it switches to this special model. The more interesting case is that in which individuals appear to be revealing this king of behavior, but there is no definitive observed indicator. The latent class model provides a way to analyze this possibility. In the example below, we fit a model in which two attributes, fashion and quality are allowed to be separately or jointly nonattended. Note that the nonzero coefficients are the same in the 4 classes. The second model is more general in that it allows the 4 types of choosers to have different coefficients. However, as can be seen in this case, the more general model is overspecified.

```
?
? (4) A model for attribute nonattendance
?
NLOGIT      ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
            ; Rhs = Fash,Qual,Price,ASC4
            ; LCM ; Pds = 8 ; Pts = 4
            ; RST = b1,b2,b3,b4,    b1, 0,b3,b4,    0,b2,b3,b4,   0, 0,b3,b4 $
NLOGIT      ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
            ; Rhs = Fash,Qual,Price,ASC4
            ; LCM ; Pds = 8 ; Pts = 4
             ; RST = b1,b2,b3,b4,    c1, 0,c3,c4,    0,d2,d3,d4,   0, 0,e3,e4  $
```

3. Random Parameters Model

We fit two specifications of a random parameters model. We also test the null hypothesis that the parameters are nonrandom.

```
?
? (5) Random parameters model
?
Nlogit   ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
         ; Rhs = Fash,Qual,Price,ASC4 $
CALC     ; logl0 = logl $
Nlogit   ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
         ; Rhs = Fash,Qual,Price,ASC4
         ; RPL  ; Fcn= Fash(n),Price(n)
```

```
        ; Pds = 8 ; Pts = 25 $
CALC    ; logl1 = logl $
CALC    ; List ; chisq = 2*(logl1 – logl0) $
```

How many degrees of freedom are there for this test? Is the null hypothesis rejected?

```
?
? (6) Correlated parameters
?
Nlogit  ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
        ; Rhs = Fash,Qual,Price,ASC4
        ; RPL  ; Fcn= Fash(n),Price(n) ; Correlated
        ; Pds = 8 ; Pts = 10 $
```

4. Willingness to Pay

Random parameters, latent class, and other complicated models produce individual specific estimates of the expectations of some parameters. The 'willingness to pay' implied by a choice model is measured by a ratio such as $-\beta_{attribute} / \beta_{cost}$. If these estimates differ by individual, as they might in a mixed model, this produces an opportunity to study the distribution of willingness to pay in the population. The following produces a set of estimates. Note that the random parameter on quality is type '(o)'. This specifies a one sided triangular distribution, so that all WTP estimates for quality will be positive.

```
?
? (7) Willingness to Pay
?
RPLogit     ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
            ; Rhs = Fash,Qual,Price,ASC4
            ; Fcn = Fash(n),Qual(o) ; Pds=8 ; Pts=10
            ; WTP = Qual / Price ; Par $
Matrix      ; wtp_i = -1*wtp_i $
Kernel      ; Rhs = wtp_i
            ; Title=Conditional Estimates of Willingness to Pay $
```

5. Error Components logit Model

Fit the simple brand choice model with the addition of a person specific random effect. Note that here, we will take advantage of the fact that this is a panel. The same person is observed 8 times in each choice situation, so we assume that the effect does not change from one choice setting to the next. To speed this up, for purpose of the exercise, we use only 10 points in the simulation estimator. After obtaining the estimates, interpret your estimated model.

```
?
? (8)  An Error Components Logit model
?
ECLOGIT ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
    ; Rhs = Fash,Qual,Price,ASC4
    ; Pts = 10 ; Pds = 8
    ; ECM = (Brand1,Brand2,Brand3),(none) $
```

Part I. Combining Revealed and Stated Preference Data

This short exercise consists of estimation of a model using a data set that combines stated and revealed preference data. The different scaling needed to accommodate the two parts of the data set can be built into the model by using a nested logit specification. The specification below embodies some of the more advanced features of the conditional logit model, including the nesting with degenerate branches to reveal the scaling and choice based sampling in the revealed preference data. The data set is also complicated by having the choice sets vary across individuals, with each individual choosing from a possibly different subset of the master choice set.

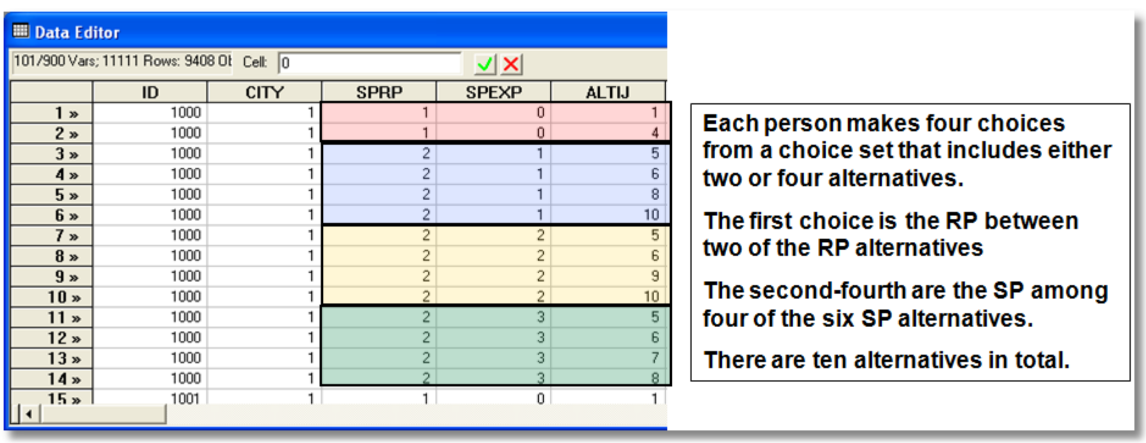
Data for this application are in sprp.lpj

The data set is a survey sample of 2,688 trips in Sydney, AustraliA, 2 or 4 choices per situation. The sample consists of 672 individuals, 9408 observations in total. The choice situations are a revealed choice case in which the choices are as follows:

Revealed choice experiment:
Revealed: Drive,ShortRail,Bus,Train

The revealed choice is followed by one or two hypothetical choice situations in which the individual chooses from among 4 of six experimental choices:

Hypothetical choice experiment:
Drive,ShortRail,Bus,Train,LightRail,ExpressBus


Data Editor
101/900 Vars; 11111 Rows: 9408 Ot   Cell: 0

	ID	CITY	SPRP	SPEXP	ALTIJ
1 »	1000	1	1	0	1
2 »	1000	1	1	0	4
3 »	1000	1	2	1	5
4 »	1000	1	2	1	6
5 »	1000	1	2	1	8
6 »	1000	1	2	1	10
7 »	1000	1	2	2	5
8 »	1000	1	2	2	6
9 »	1000	1	2	2	9
10 »	1000	1	2	2	10
11 »	1000	1	2	3	5
12 »	1000	1	2	3	6
13 »	1000	1	2	3	7
14 »	1000	1	2	3	8
15 »	1001	1	1	0	1

The choice attributes in the model are

Cost –Fuel or fare
Transit time
Parking cost
Access and Egress time

The revealed preference are a choice based sample. The sample market shares for the RP choices differ systematically from the known shares, .592,.208,.089,.111. The role of these in the models is seen in the statement of the list of the choices in the model commands below.

```
?
? 1.  Nested logit to reveal scaling difference between RP and SP choices
```

```
?
? Sample is choice based, as shown by weights.  Choice variable is CHOSEN
? Number of choices in choice set is CSET.
? Specific choices from master set given by ALTIJ
? FCOST = fuel cost
? AUTOTIME = time spent commuting by car.
? Numerous other variables in the data set are not used here.
?
NLOGIT  ;lhs=chosen,cset,altij
        ;choices=RPDA,RPRS,RPBS,RPTN,SPDA,SPRS,SPBS,SPTN,SPLR,SPBW
                /.592,.208,.089,.111, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
        ;tree=Commute [  rp  (RPDA,RPRS,RPBS,RPTN),
               spda(SPDA), sprs(SPRS), spbs(SPBS),
               sptn(SPTN), splr(SPLR), spbw(SPBW)]
        ;ivset: (rp)=[1.0]  ;ru1 ;maxit=50
        ;model:
  U(RPDA) = rdasc  + invc*fcost + tmrs*autotime /
  U(RPRS) = rrsasc + invc*fcost + tmrs*autotime /
  U(RPBS) = rbsasc + invc*mptrfare  + mtpt*mptrtime/
  U(RPTN) =          cstrs*mptrfare + mtpt*mptrtime/
  U(SPDA) = sdasc  + invc*fueld + tmrs*time + cavda*carav /
  U(SPRS) = srsasc + invc*fueld + tmrs*time/
  U(SPBS) =          invc*fared + mtpt*time + acegt*spacegtm/
  U(SPTN) = stnasc + invc*fared + mtpt*time + acegt*spacegtm/
  U(SPLR) = slrasc + invc*fared + mtpt*time + acegt*spacegtm/
  U(SPBW) = sbwasc + invc*fared + mtpt*time + acegt*spacegtm $

?
? 2. Using only Revealed Preference Data. Simple MNL
?
NLOGIT ; if[sprp = 1] ? Using only RP data
       ; lhs=chosen,cset,altij ; choices=RPDA,RPRS,RPBS,RPTN
       ; model:
  U(RPDA) = rdasc  + fl*fcost     + tm*autotime/
  U(RPRS) = rrsasc + fl*fcost     + tm*autotime/
  U(RPBS) = rbsasc + ptc*mptrfare + mt*mptrtime/
  U(RPTN) =          ptc*mptrfare + mt*mptrtime$

?
? 3. Using only Stated Preference Data.  Simple MNL
?
SAMPLE  ; all$
NLOGIT  ; if[sprp = 2] ? Using only SP data
        ; lhs=chosen,cset,alt ; choices=SPDA,SPRS,SPBS,SPTN,SPLR,SPBW
        ; crosstab
        ; model:
  U(SPDA) = dasc + cst*fueld + tmcar*time + prk*parking
                 + pincda*pincome +cavda*carav/
  U(SPRS) = rsasc+cst*fueld + tmcar*time + prk*parking/
  U(SPBS) = bsasc+cst*fared +  tmpt*time + act*acctime + egt*eggtime/
  U(SPTN) = tnasc+cst*fared +  tmpt*time + act*acctime + egt*eggtime/
  U(SPLR) = lrasc+cst*fared +  tmpt*time + act*acctime + egt*eggtime/
  U(SPBW) =       cst*fared +  tmpt*time + act*acctime + egt*eggtime$


?
? 4.  Using all data. Nested logit reveals scaling
?
SAMPLE ; All$
NLOGIT ; lhs=chosen,cset,altij
       ; choices=RPDA,RPRS,RPBS,RPTN,SPDA,SPRS,SPBS,SPTN,SPLR,SPBW
                /.592,.208,.089,.111, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
       ; tree=mode[rp(RPDA,RPRS,RPBS,RPTN),spda(SPDA),
```

```
              sprs(SPRS),spbs(SPBS),sptn(SPTN),splr(SPLR),spbw(SPBW)]
       ; ivset: (rp)=[1.0] ; ru1
       ; maxit = 50
       ; model:
    U(RPDA) = rdasc  + invc*fcost     + tmrs*autotime
                     + pinc*pincome   + CAVDA*CARAV/
    U(RPRS) = rrsasc + invc*fcost     + tmrs*autotime/
    U(RPBS) = rbsasc + invc*mptrfare  + mtpt*mptrtime/
    U(RPTN) =          cstrs*mptrfare + mtpt*mptrtime/
    U(SPDA) = sdasc  + invc*fueld     + tmrs*time+cavda*carav
                     + pinc*pincome/
    U(SPRS) = srsasc + invc*fueld + tmrs*time/
    U(SPBS) =          invc*fared + mtpt*time + acegt*spacegtm/
    U(SPTN) = stnasc + invc*fared + mtpt*time + acegt*spacegtm/
    U(SPLR) = slrasc + invc*fared + mtpt*time + acegt*spacegtm/
    U(SPBW) = sbwasc + invc*fared + mtpt*time + acegt*spacegtm$


?
? 5.  Using all data. Random parameters model connects choice
?     situations.
?
SAMPLE ; All$
NLOGIT ; lhs=chosen,cset,altij
       ; choices=RPDA,RPRS,RPBS,RPTN,SPDA,SPRS,SPBS,SPTN,SPLR,SPBW
                /.592,.208,.089,.111, 1.0,1.0, 1.0, 1.0, 1.0, 1.0
       ; rpl ; pds=4 ; halton ; pts=25 ; fcn=invc(n)
; model:
   U(RPDA)  = rdasc  + invc*fcost     + tmrs*autotime
                     + pinc*pincome   + CAVDA*CARAV/
    U(RPRS) = rrsasc + invc*fcost     + tmrs*autotime/
    U(RPBS) = rbsasc + invc*mptrfare  + mtpt*mptrtime/
    U(RPTN) =          cstrs*mptrfare + mtpt*mptrtime/
    U(SPDA) = sdasc  + invc*fueld     + tmrs*time+cavda*carav
                     + pinc*pincome/
    U(SPRS) = srsasc + invc*fueld + tmrs*time/
    U(SPBS) =          invc*fared + mtpt*time +acegt*spacegtm/
    U(SPTN) = stnasc + invc*fared + mtpt*time+acegt*spacegtm/
    U(SPLR) = slrasc + invc*fared + mtpt*time+acegt*spacegtm/
    U(SPBW) = sbwasc + invc*fared + mtpt*time+acegt*spacegtm$
```