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

Lab Session 5 Assignment
Multinomial Choice Models

==
This assignment will consist of some exercises with various forms of the multinomial logit model. The data for the first set of exercises is

mnc.lpj
==

1. **<u>Test for functional form of utility functions</u>**. 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 test. What do you find?

```
?-----------------------------------------------------------------------
? Part 1.  Functional Form
? Test for significance of squared term and illustrate specification
?-----------------------------------------------------------------------
CLOGIT    ; Lhs = Choice ; Choices=Brand1,Brand2,Brand3,None
          ; Describe
          ; 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: $\beta_M = \beta_W$ for the vector of parameters in the model rejected or not?

```
?-----------------------------------------------------------------------
? Part 2. Structural change
?-----------------------------------------------------------------------
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**. 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?

```
?-----------------------------------------------------------------------
? Part 3. Partial Effects
? 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 (*) ; Full ; K&R = 25 $
```

4. **Impact of a price change**. 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%

```
?-----------------------------------------------------------------------
? Part 4.  Analysis of Market Shares
?-----------------------------------------------------------------------
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
          ; 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. **<u>Testing for IIA</u>**. 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?

```
?-----------------------------------------------------------------------
? Part 5.  Testing IIA
?-----------------------------------------------------------------------
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?

```
?-----------------------------------------------------------------------
? Part 6. Functional form and Marginal Impact
? Do men pay different attention to fashion than women?
? Is the difference statistically significant?
?-----------------------------------------------------------------------
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. **<u>Heteroscedastic extreme value model</u>**. Fit the MNL model while allowing the variancwes 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?

```
?-----------------------------------------------------------------------
? Part 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. **Constraints**. 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?

```
?--------------------------------------------------------------------
? Part 8. Testing for homoscedasticity with a Wald test
?--------------------------------------------------------------------
NLOGIT (het) ; Lhs = Choice ; Choices = Brand1,Brand2,Brand3,None
         ; Rhs = Fash,Qual,Price,ASC4
         ; Het ; Par ; Maxit=10$
MATRIX   ; c = b(5:7) ; vc = Varb(5:7,5:7) ; List ; WaldStat = c'<vc>c $
```

9. **Testing for variance heterogeneity**. Are age and sex significant determinants of the variances in the utility functions. Test the hypothesis that they are not using a likelihood ratio test. What do you find?

```
?--------------------------------------------------------------------
? Part 9. Heterogeneous. Does Gender 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;Maxit=10$
CREATE   ; LoglVar = Logl_Obs $
CREATE   ; V = LoglMean - LogLVar $
REJECT   ; brand < 4 $
CALC     ; List ; Vuong = sqr(n) * xbr(v) / sdv(v) $
SAMPLE   ; 1-12800 $
```

===

This assignment will use the mode choice, conditional logit data

===

1. Basic Model Forms: Multinomial Logit and Multinomial Probit

```
?--------------------------------------------------------------------
? 1. Multinomial Choice Models
?--------------------------------------------------------------------
SAMPLE    ; 1-840 $
NLOGIT    ; Lhs = Mode ; output=ic
          ; Choices = Air,Train,Bus,Car
          ; Rhs = TTME,INVC,INVT,GC; Rh2=One,Hinc
          ; Effects:GC(*) $
NLOGIT    ; Lhs = Mode ; RRM (Random Regret)
          ; 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(*) $
```

2. **Nested logit model**. We begin with a simple nested logit model.

```
?-----------------------------------------------------------------------
? 2. Basic nested logit
?-----------------------------------------------------------------------
Sample    ; 1 - 840 $
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 $
```

3. **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.

```
?-----------------------------------------------------------------------
? 3. 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$
```

4. **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.

```
?-----------------------------------------------------------------------
? 4. 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) $
```

5. **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.

```
?-----------------------------------------------------------------------
? 5. 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 $
```

6. **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?

```
?-----------------------------------------------------------------------
? 6. 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 ; Maxit=10$
```

7. 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.

```
?-----------------------------------------------------------------------
? 7. 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 ; Maxit=10 $
```

8. **<u>HEV Model</u>**. 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).

```
?-----------------------------------------------------------------------
? 8. 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) $
```

9. **<u>Income effect on mode choice model model</u>**. Does income 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

$$U_{ij} = \beta_1 TTME_{ij} + \beta_2 INVC_{ij} + \beta_3 INVT_{ij} + \beta_4 GC_{ij} + \alpha_j + \delta_j Income_i + \varepsilon_{ij}$$

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

$$U_{ij} = \beta_1 TTME_{ij} + \beta_2 INVC_{ij} + \beta_3 INVT_{ij} + \beta_4 GC_{ij} + \alpha_j + \varepsilon_{ij}$$

where $Var[\varepsilon_{ij}] = \sigma_j^2 \times \exp(\gamma\ Income_i)$. 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

$$m_i = LogL_{i0} - LogL_{i1}$$

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. 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 CAR. The CASC variable is a dummy variable which equals one for the CAR 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 Heterogeneity.  Does Income affect the means or the variances?
?-----------------------------------------------------------------------
SAMPLE        ; 1-840 $
NLOGIT        ; Lhs = Mode
              ; Choices = Air,Train,Bus,Car
                     ; Rhs = TTME,INVC,INVT,GC; Rh2=One,Hinc $
CREATE        ; LOGLMean = Logl_Obs $
NLOGIT        ; Lhs = Mode
              ; Choices = Air,Train,Bus,Car
                     ; Rhs = TTME,INVC,INVT,GC,one
                     ; Het ; Hfn = HINC $
CREATE        ; LoglVar = Logl_Obs $
CREATE        ; V = LoglMean - LogLVar $
REJECT        ; CASC = 0 $
CALC          ; List ; Vuong = sqr(n) * xbr(v) / sdv(v) $
```

<u>The next set of computations is based on the brand choices data, brandchoicesSP.lpj</u>

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.

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

1. **<u>Latent class model for brand choice</u>**. 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.
?------------------------------------------------------------------------
SAMPLE ; 1-12800 $
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 $

?------------------------------------------------------------------------
? (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. **<u>Random parameters models</u>**. 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) $

?------------------------------------------------------------------------
? (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 $
?------------------------------------------------------------------------
? (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 $
Kernel   ; Rhs = wtp_i ; Grid
         ; Title=Conditional Estimates of Willingness to Pay $
```

2. **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) $
```