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

Lab Session 2 Assignment

Part 1. Binary Choice Modeling

This exercise will involve estimating and analyzing binary choice models. We will analyze the panel probit, manufacturing innovation data. The data set is PanelProbit.lpj. To begin, load these data. To save you some typing, most of the commands for this exercise are contained in the file LabAssignment-2.lim.

1. **Cluster Estimator.** This is a panel data set. Do the standard errors of the probit estimator need 'correction?' This exercise computes the standard covariance matrix and the 'cluster corrected' covariance matrix and compares them. Describe your findings.

```
Sample ; All $
Namelist ; X = One,IMUM,FDIUM,SP,LogSALES $
Probit ; Lhs = IP ; Rhs = X $
Matrix ; Var0 = Varb $  (Uncorrected covariance matrix)
Probit ; Lhs = IP ; Rhs = X ; Cluster = 5 $
Matrix ; VarPanel = Varb $ (Corrected covariance matrix)
? PCTDIFF is the percentage difference between the standard errors
Matrix ; SD0 = Diag(Var0)  ; Diff = Vecd(VarPanel) - Vecd(Var0)
     ; List ; PctDiff = 100*<SD0>*Diff$
```

2. **Robust Covariance Matrix**. You can also compute a 'robust,' sandwich style asymptotic covariance matrix. This estimator would only be robust to heteroscedasticity – though we are unsure what that would mean in the probit setting.

```
Probit ; Lhs = IP ; Rhs = X $
Matrix ; Var0 = Varb $  (Uncorrected covariance matrix)
Probit  ; Lhs = IP  ; Rhs = X ;  RobustVC $
Matrix  ; VarHet = Varb $
Matrix ; SD0 = Diag(Var0)  ; Diff = Vecd(VarHet) - Vecd(Var0)
     ; List ; PctDiff = 100*<SD0>*Diff$ - Init(5,1,100) $
```

3. **Marginal Effect for a Quadratic.** Marginal effects in the binary choice models are complicated functions of the parameters and the data. They are more so when the index function contains complex functions of the data. Suppose, for example,

$$P = \Phi(\boldsymbol{\beta}'\mathbf{x} + \alpha_0 \text{logSales} + \alpha_1 \text{logSales}^2).$$

The marginal effect of logSales, which is the effect on the probability of a one percent change in sales is

$$\partial P/\partial \text{logSales} = \phi(\boldsymbol{\beta}'\mathbf{x} + \alpha_0 \text{logSales} + \alpha_1 \text{logSales}^2) \times (\alpha_0 + 2\alpha_1 \text{logSales})$$

It is possible to program this computation into the WALD command. But, it is easier to use the built in function to obtain the result.

```
Create   ; LogS2  = LogSales^2 $
Namelist ; X2     = One,IMUM,FDIUM,SP,LogSALES,LogS2 $
Probit   ; Lhs    = IP ; Rhs = X2 ; Mar $
Wald     ; Start  = b ; Var = Varb
       ; Labels = beta0,beta1,beta2,beta3,a0,a1
       ; Fn1    = n01(beta0'X2)*(a0+2*a1*Logsales) ? partial wrt logsales
       ; Fn2    = n01(beta0'X2)* beta2  ? partial wrt fdium
       ; Average (computes average partial effects) $
? Easier way
probit   ; Lhs = ip ; rhs = One,IMUM,FDIUM,SP,logsales,logsales^2$
partial  ; effects: logsales / fdium $
? Extension to look more closely at partial effects
Partial  ; effects: logsales & fdium = .05(.05)1 ; plot(ci) $
```

4. **Heteroscedasticity**. The following suggests how to incorporate heteroscedasticity in the binary logit (or probit – by changing the command) model:

```
Logit ; Lhs = IP ; Rhs = X ; Het ; Hfn = RAWMTL; Marginal Effects $
```

(1) Note the effect on the coefficients and how the marginal effects are decomposed.
(2) Repeat the computation with **;Hfn = LogSales**. Note the effect on the estimates and significance levels. The difference between the reported marginal effects and the results from the PARTIALS command is that the former is computed at the means of the data while the second is averaged over all observations. To reproduce the results at the means, we add ;Means to the PARTIALS command.

```
Namelist ; X = one,imum,fdium,sp,logsales $
Logit   ; Lhs = IP ; Rhs = X ; Het
      ; Hfn = RAWMTL; Marginal Effects $
Partial ; Function= lgp((b1+b2*IMUM+b3*FDIUM+b4*SP+b5*LogSALES)/exp(c1*rawmtl))
      ; Labels  = b1,b2,b3,b4,b5,c1
      ; parameters = b
      ; covariance = varb
      ; effects: rawmtl / logsales $
Logit   ; Lhs = IP ; Rhs = X ; Het
      ; Hfn = LogSales; Marginal Effects $
Partial ; Function=lgp((b1+b2*IMUM+b3*FDIUM+b4*SP+b5*LogSALES)
                /exp(c1*LogSales))
      ; Labels   = b1,b2,b3,b4,b5,c1
      ; parameters = b
      ; covariance = varb
      ; effects: logsales ; means$
```

5. **Nonparametric and Semiparametric Estimation**. There are numerous alternative estimators you can use for analyzing binary choices. Interpretation of the results of these models requires some careful thought – but estimation is very straightforward. Estimation of these is generally very computer intensive, so we use only a subset of the sample – one year of the data. Compare the table of correct and incorrect predictions produced by **PROBIT** and **MSCORE**. (The other estimators do not produce enough information to generate predictions for individual observations.)

```
Namelist ; X0 = IMUM,FDIUM,SP,LogSALES $
Namelist ; X = One,X0 $
Reject ; New ; T > 1 $ (Use only first year of data)
?  Fully Parametric
PROBIT ; Lhs = IP ; Rhs = X $
?  Semiparametric:  Maximum Score
MSCORE ; Lhs = IP ; Rhs = X $
```

```
Semiparametric ; LHS = IP ; Rhs = X0 $ (Klein and Spady.)
? Nonparametric, Kernel density regression estimator
? Note, the nonparametric estimator can only have one RHS variable
NPREG ; LHS = IP ; Rhs = LogSales $
```

6. **Creating a Plot of Probabilities**. The following will demonstrate how to use NLOGIT to produce the plot shown in the class discussion.

```
Reject    ; New ; T > 1 $
Probit    ; Lhs = IP ; Rhs = one,IMUM,FDIUM,SP,logsales $
Calc      ; Low = .5*Min(LogSales) ; High = 1.5*Max(LogSales)
        ; inc = .05*(high-low) $$
Simulate  ; Scenario : logsales & logsales = Low(inc)high
   ;plot(ci)  ;title=Simulation of Innovation Probabilities vs. Log Sales$
```

7. **Testing for Structural Change**. It might be interesting to test whether underlying structural of the model has changed over the five year period of the data. Consider the structure

$$P_{it} = F(\boldsymbol{\beta}_t'\mathbf{x}_{it}),\ i = 1,\ldots,1270,\ t = 1,\ldots,5\ (1993 \text{ to } 1997)$$

which allows for different coefficient vectors in each year. We are interested in testing the hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5$$
$$H_1: \text{ not } H_0.$$

In a linear regression context, this would be a 'Chow' test and would be tested with an F test. Since this is not a linear regression model, we can't use the F test here. The easiest way to do this test is with a likelihood ratio test. The strategy is to fit the restricted model (pool the 5 years of data) and the unrestricted model (estimate the model separately for each year), and compare the log likelihoods. The log likelihood for the unrestricted model is the sum of the five years. Here is how you can automate this computation. The last part of the last CALC displays the 95% critical value from the chi-squared table.

```
Sample ; All $
Namelist ; X = One,IMUM,FDIUM,SP,LogSALES $
Probit ; Lhs = IP ; Rhs = X ; quietly $  (We suppress the model results)
Calc ; Logl0 = Logl ; Logl1 = 0 ; i = 0 $
Procedure
Include ; New ; T = i $
Probit ; Lhs = IP ; Rhs = X ; Quietly $
Calc ; Logl1 = Logl1 + Logl $
EndProc $
Execute ; i = 1,5  $
Calc ; List ; Chisq = 2*(Logl1 - Logl0) ; Df = 4*Col(X) ; Ctb(.95,df) $
```

Carry out the test. What do you conclude? Should the null hypothesis be rejected? Repeat the test using a logit model instead of a probit model. Does the conclusion change? Try the exercise again while adding the sector dummy variables to the model. To do these, it is only necessary to change the model name from PROBIT to LOGIT, or the NAMELIST command by adding variables to it.

8. **Hypothesis Tests:** This exercise will illustrate the three methods of carrying out hypothesis tests.

```
Reject   ; New ; T < 5 $
Namelist ; X = One,IMUM,FDIUM,LogSales $
Namelist ; Sectors = RawMtl,InvGood$
Probit   ; Lhs = IP ; Rhs = X $
Calc     ; Logl0 = LogL $
Probit   ; Lhs = IP ; Rhs = X ; Het ; Hfn = Sectors
       ; Start = b,0,0 ; Maxit = 0 $
Probit   ; Lhs = IP ; Rhs = X,Sectors ; Parameters ; test: sectors = 0$
Calc     ; KX = Col(X) ; K1 = KX + 1 ; Kc = Col(Sectors); K = KX + KC$
Matrix   ; c = B(K1:K) ; vc = Varb(K1:K , K1:K) $
Matrix   ; List ; Wald = c'<vc>c  $
Calc     ; List ; Ctb(.95,2) $
Wald     ; start = b ; Var = Varb ; labels=KX_d,Kc_c
          ; fn1 = c1 - 0 ; Fn2 = c2 - 0 $
```

The model is $y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$
$\varepsilon \sim N[0,\sigma_i^2]$, $\sigma_i = \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$.
$y_i = 1(y_i^* > 0]$

The various testing procedures shown estimate γ and test whether $\boldsymbol{\gamma} = 0$, in which case $\sigma_i^2 = 1$. Carry out the tests, and determine whether the null hypothesis, H_0:$\boldsymbol{\gamma} = 0$, should be rejected.

9. **Simulation:** Using the binary choice model simulator, examine how an increase in LOGSALES of 50% would affect the probability of innovation.

```
Probit       ; Lhs = IP ; Rhs=one,logsales,imum,fdium $
BinaryChoice ; Lhs = IP ; Rhs = one,logsales,imum,fdium
             ; model=probit  ; start=b
             ; scenario: logsales * = 1.5 ;plot : logsales $
```

Part 2. Extensions of the Probit Model

This exercise uses the data file panelprobit.lpj

1. **Bivariate probit model**. In this exercise, we will fit a bivariate probit model. The model is

$$y_1^* = x_1'\beta_1 + \varepsilon_1$$
$$y_2^* = x_2'\beta_2 + \varepsilon_2$$
$$\varepsilon_1,\varepsilon_2 \sim N_2[(0,0),(1,1,\rho)].$$

The model is fit by maximum likelihood. You can use the following commands to treat the 1984 and 1985 observations as a bivariate probit outcome:

```
Sample ; 1 - 1270$
Namelist ; x84 = one,imum84,fdium84,prod84$
Namelist ; x85 = one,imum85,fdium85,prod85$
Bivariate; Lhs = ip84,ip85 ; Rh1 = x84; Rh2 = x85 $
```

Notice that if $\beta_1 = \beta_2$, that this becomes a two period random effects model. You can constrain the slope parameters to be equal by using

```
Bivariate; Lhs = ip84,ip85 ; Rh1 = x84; Rh2 = x85
         ; Rst = b1,b2,b3,b4,b1,b2,b3,b4,corr$
```

Do the results change substantially when the restriction is imposed? Does the estimate of ρ change? The hypothesis of interest is H_0:$\beta_1 = \beta_2$. You can test this hypothesis using these models with a likelihood ratio test. Compute twice the difference in the log likelihoods.

Recall that we fit a random effects model for all 5 periods in Exercise 3. Go back to that exercise and examine the results you obtained. Does the value of ρ change when the five years of data are used?

2. **Multivariate probit model**. We can fit the "panel probit model" as a multivariate probit model by extending the model above. We will use a limited form, with three periods. The following commands can be used. Note, since this is a very slow estimator, we have used only 5 simulation points and limited it to 10 iterations. How do the results here compare to those in part 3? Is the correlation matrix what you would expect? Do the coefficients vary across periods?

```
Namelist ; x86 = one,imum86,fdium86,prod86$
Mprobit  ; lhs = ip84,ip85,ip86
         ; eq1 = x84
         ; eq2 = x85
         ; eq3 = x86
         ; Pts = 5 ; Maxit=10 $
```

This exercise uses the data file labor.lpj

3. We consider two standard applications of the probit model. The first is Heckman's classic model of sample selection, estimated by the two step least squares method proposed in the early paper in Econometrica. When you fit the model, is there evidence of sample "selection?" That is, is the estimate of ρ significantly different from zero. For the two step method, this is determined by examine the coefficient on "lambda" in the second step least squares results. Later, it was established that this model could be fit by maximum likelihood. The second estimator below uses MLE instead of two step least squares. Do the results change much?

```
?-------------------------------------------------------
? (3)  Sample selection Model
?-------------------------------------------------------
Namelist ; XLFP = One,KL6,K618,WA,FAMINC $
Namelist ; XHRS = One,WA,WE,WW,HW$
Probit   ; Lhs  = LFP ; Rhs = XLFP ; Hold $
Select   ; Lhs  = WHrs ; Rhs = XHRS ; Marginal Effects$
Select   ; Lhs  = WHrs ; Rhs = XHRS ; Marginal Effects ; MLE$
```

4. The next model considers the possibility of an endogenous variable on the right hand side of a probit equation.

$$y_1^* = x_1'\beta_1 + \gamma y_2 + \varepsilon_1,\ y_1 = 1[y_1^* > 0]$$
$$y_2 = x_2'\beta_2 + \varepsilon_2$$
$$\varepsilon_1,\varepsilon_2 \sim N_2[(0,0),(1,1,\rho)].$$

This model is estimated using maximum likelihood and the "control function" approach. In the labor supply model below, the husband's weekly earnings are treated as endogenous in the wife's labor force participation equation. The hypothesis seems a bit dubious. Do the results suggest that the husband's earnings are endogenous?

```
?-------------------------------------------------------------------------------
? (4) Endogenous right hand variable - husband's earnings
?-------------------------------------------------------------------------------
Namelist ; Hwork = one,ha,he $
Create   ; Hearn = hhrs*hw $
Probit   ; Lhs = lfp,hhrs
         ; Rhs = one,kl6,k618,wa,faminc,Hearn
         ; Rh2 = Hwork $
```

The two specifications are rather sparse. Are there other variables in the data set that might improve the specification? Try fitting a fuller specification of the model.

Part 3. Binary Choice Modeling with Panel Data

This assignment will extend the models of binary choice and ordered choice to panel data frameworks. These exercises will use the health care data, healthcare.lpj
Since these are a panel data set, we begin by identifying it as one

```
SAMPLE ; All $
SETPANEL ; Group = id ; Pds = ti $
```

1. **Logit conditional and unconditional fixed effects estimation**. For the binary logit model, the Chamberlain form of the fixed effects estimator is consistent while the unconditional (brute force) fixed effects estimator is inconsistent. (This is the incidental parameters problem that arises when T is small. In our unbalanced panel here, the largest group size is 7, and most groups have less than that. Thus, T is small here.) Fit the logit model by the two approaches, and compare the results. Are they very different? To see if we can't highlight the effect, let's look at the standard case, with T = 2. How different are the results now? Remember, in the T=2 case, plim $\mathbf{b}_{MLE} = 2\beta$ while plim $\mathbf{b}_C = \beta$. Do the results seem to bear this out?

```
SAMPLE ; All $
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Panel $ (Conditional)
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Panel ; Fixed $ (Unconditional)
REJECT ; ti > 2 $
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Panel $ (Conditional)
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Panel ; Fixed $  (Unconditional)
```

2. **Test for fixed effects**. In order to test for the need for fixed effects in the logit model, we can't use the likelihood ratio test because the unrestricted estimator is inconsistent. We can use the Hausman test, instead. This uses the chi-squared statistic

$$H = (\mathbf{b}_C - \mathbf{b}_R)' [\mathbf{V}_C - \mathbf{V}_R]^{-1}(\mathbf{b}_C - \mathbf{b}_R)$$

where 'C' refers to the Chamberlain, conditional estimator and 'U' refers to the 'restricted' estimator which has only a single constant term. Note that b_R is the subvector of the restricted estimator that strips off the overall constant term – it keeps on ly the slope coefficients. Using the model suggested in the commands below, carry out the test. What is the result? Do you reject the hypothesis? (What is the null hypothesis?) Note, it is not guaranteed that the difference matrix in the statistic is positive definite. To find out if it is, we will look at the characteristic roots. They must all be positive. Are they?

```
Sample ; All $
Logit ; Lhs = Doctor ; Rhs = hhninc,educ,hhkids ; panel $
Matrix ; bfe = B ; Vfe = VARB $
Logit ; Lhs = Doctor ; Rhs = hhninc,educ,hhkids,one $
Matrix ; db = bfe - b(1:3) ; dV = Vfe - Varb(1:3,1:3) $
Matr;list;root(dv)$
Matrix ; List ; Hausman = db'<dv>db $
```

3. **Fixed and Random Effects**. The fixed and random effects estimators are competing estimators for the panel model. Each has its virtues and shortcomings. (We use Hpt=8 to speed up the quadrature in the random effects model. Normally you would use the default of 64.)

```
Sample ; All $
Probit ; lhs = hospital ; Rhs = hhninc,educ,hhkids,one ; random ; panel ; maxit = 10 ; Hpt=8$
```

```
Probit ; lhs = hospital  ; Rhs = hhninc,educ,hhkids,one ; Fixed ; panel $
```

4. **Mundlak's approach**. The disadvantage of the random effects estimator is that it requires an assumption that the individual effects are uncorrelated with the included variables. If that assumption is not met, the estimator is inconsistent. The fixed effects estimator is inconsistent when T is not large. Thus, both estimators have problems. Chamberlain's conditional estimator provides a way to estimate the logit fixed effects model consistently. An approach often used in the random effects case is to add to the model the group means of the independent variables (those that vary over time, that is.) We'll try that approach here.

```
Sample ; All $
Create ; incbar=GroupMean (hhninc, Pds=_Groupti)  $
Create ; educbar=GroupMean (educ, Pds=_Groupti)  $
Create ; kidsbar=GroupMean (hhkids, Pds=_Groupti)  $
Logit ; lhs = Doctor ; Rhs = hhninc,educ,hhkids ; pds = _groupti $
Logit ; lhs = Doctor ; Rhs = hhninc,educ,hhkids,incbar,educbar,kidsbar
        ; Random ; panel  $
```

5. **Random effects probit models**. The random effects probit model can be fit using the Butler and Moffitt method, using quadrature, or using simulation by treating it as a random parameter model. Compute the estimator both ways and see how close the two estimators are. Note, the Butler and Moffitt estimator reports RHO in the output – this equals the squared correlation between observations in a group. The simulation estimator reports SIGMA, the standard deviation of the common individual effect. To compare the two estimates of ρ, you must compute $\rho^* = \sigma^2 / (1 + \sigma^2)$ from the random parameters estimates. What do you find? Are the estimates of the other slopes nearly the same?

```
?  This estimator is time consuming.  To speed things up, we use only
?  a subset of the data and a small number of draws.
Sample ; All $
Reject ; _Groupti < 7 $
Namelist ; X = hhninc,educ,hhkids,one $
Probit ; lhs = hospital ; Rhs = x ;  panel  ; maxit=10; random effects $
Probit ; ; lhs = hospital ; Rhs = x  ; pds = _groupti  ;maxit=10
        ; RPM ; Fcn = One(n) ; Pts = 20 ; panel $
Calc   ; K1 = Col(X) + 1 $
Calc   ; List ; SRP = B(K1) ; RhoRP = SRP^2 / (1 + SRP^2) $
```

Part 4. Binary Choice Modeling with Panel Data and Heterogeneity

This assignment will extend the models of binary choice to modeling heterogeneity in panel data frameworks. These exercises will use the German manufacturing innovation data, panelprobit.lpj

1. **Random parameters models** In the original study that used these data, the coefficients on IMUM and FDIUM were of particular interest. In his followup studies, Greene treated these two parameters as randomly distributed across firms. Here, you can partially replicate that study by reestimating the random parameters model. Three models are fit: (1) The two parameters are treated as independent normally distributed; (2) The two parameters are allowed to be freely correlated; (3) The two random parameters are assumed to have a mean that varies by industry. In this case, we specify $\beta_k = \beta_{0k} + \delta_{k1}$InvGood + β_{k2}Consgood + β_{k3}Food + $\sigma_k v_k$. We could also specify that the two parameters remain correlated. That is left for an exercise. The final PROBIT command contains ; PARAMETERS. This creates a matrix BETA_I that contains the firm specific conditional means of the random parameters. You can double click this matrix to see the values. The remaining commands manipulate this matrix to explore the distribution of parameter values across firms. The kernel density estimator in the last command does this exercise for the coefficient on IMUM. By changing BIMUM to BFDIUM in the KERNEL command, you can repeat the exercise for the coefficient on FDIUM.

```
Sample ; All $
Probit ; Lhs = Ip ; Rhs = X ; RPM ; Fcn = imum(n),fdium(n) ; Pts = 25 ; Pds = 5 $
Probit ; Lhs = Ip ; Rhs = X ; RPM ; Correlated
     ; Fcn = imum(n),fdium(n) ; Pts = 25 ; Pds = 5 $
Probit ; Lhs = Ip ; Rhs = X ; RPM = InvGood,ConsGood,Food
     ; Fcn = imum(n),fdium(n) ; Pts = 25 ; Pds = 5 ; Parameters $
Create ; Bimum = 0 ; Bfdium = 0 $
Namelist ; Bi = Bimum,Bfdium $
Sample ; 1 - 1270 $
Create ; Bi = Beta_i $
Kernel ; Rhs = Bimum $
```

2. **Latent class model**. In this exercise, we fit a three class latent class LOGIT model. This is an alternative method of building heterogeneity into the panel data model. In the third command, we produce a listing of the estimated conditional class probabilities, with a listing of the best guess as to which class each firm is in. (The number of observations is reduced just for purpose of a compact example.) You might try changing the specification of the equation, and examining how the results change. To see what happens when the model is overspecified, change **;pts=2** to **;pts=3** in the last command and refit the model.

```
Sample ; All $
Logit  ; Lhs = IP ; Rhs = X ; LCM ; Pts = 3 ; Pds = 5  ; Parameters $
Logit  ; Lhs = IP ; Rhs = X ; LCM = InvGood,Consgood,Food
                  ; Pts = 2 ; Pds = 5  ; Parameters $
Sample;1-500$
Logit  ; Lhs = IP ; Rhs = X ; LCM = InvGood,Consgood,Food
                  ; Pts = 2 ; Pds = 5  ; Parameters ; List $
```