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

Lab Session 3

I. Bivariate and Multivariate Probit

This exercise uses the data file panelprobit.lpj

Some preliminaries after the file is loaded.

```
Setpanel      ; Group = firm ; pds = ti $
Namelist      ; x84 = one,imum84,fdium84,prod84$
Namelist      ; x85 = one,imum85,fdium85,prod85$
Namelist      ; x86 = one,imum86,fdium86,prod86$
```

1. The Bivariate Probit Model.

In this exercise, we will first 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: (Later, we will apply the model to two distinct decisions.)

```
Bivariate     ; if[t = 1] ; Lhs = ip84,ip85 ; Rh1 = x84; Rh2 = x85 $
Calc          ; lu = logl $
```

Notice that if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, that this becomes a two period random effects model. You can constrain the slope parameters to be equal by using

```
Bivariate     ; if[t = 1] ; Lhs = ip84,ip85
              ; Rh1 = x84
              ; Rh2 = x85
              ; Rst = b1,b2,b3,b4,b1,b2,b3,b4,corr$
Calc          ; lr = logl $
Calc          ; list ; lrstat = 2*(lu - lr) $
```

Do the results change substantially when the restriction is imposed? Does the estimate of ρ change? The hypothesis of interest is H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. You can test this hypothesis using these models with a likelihood ratio test. Compute twice the difference in the log likelihoods. What are the degrees of freedom for the test? What is the 95% critical value from the chi squared table?

(It is also possible to reproduce the bivariate probit model with

```
Probit        ; if[t <= 2] ; lhs=ip;rhs=one,imum,fdium,prod ; pds=2;random ; halton $
```

2. Recursion

Many applications involve simultaneous equations sorts of binary choice models. The bivariate probit model is analogous to the seemingly unrelated regressions, save, of course for the discrete dependent variables.

The recursive bivariate probit model has made some recent appearances in the literature. A two period panel version might appear as follows:

$$y_{i1} = 1[\boldsymbol{\beta}_1'\mathbf{x}_{i1} + \varepsilon_{i1} > 0],$$
$$y_{i2} = 1[\boldsymbol{\beta}_2'\mathbf{x}_{i2} + \gamma_2 y_{i1} + \varepsilon_{i1} > 0]$$

One might model insurance take up with such a model. Here, we model innovation in 1985 as a function of 1985 covariates and 1984 innovation. An interesting feature of the recursive bivariate probit model is that variables in the second may have two effects on y_{i2}. A direct effect in $\mathbf{x}_{i2}$ and an indirect effect if they appear in the y_{i1} equation and effect y_{i1} which then affects y_{i2}. Here is an example.

```
Bivariate     ;  if[t=1]   ; lhs=ip84,ip85
              ;  rh1=x84,invgood,consgood
              ;  rh2=x85,invgood,consgood,ip84
              ;  Marginal effects $
Partials      ;  if[t = 1] ; effects: invgood / consgood ; summary $
```

3. Identification and Incoherence

There is a temptation sometimes to specify a fully simultaneous model, as in

$$y_{i1} = 1[\boldsymbol{\beta}_1'\mathbf{x}_{i1} + \gamma_1 y_{i2} + \varepsilon_{i1} > 0],$$
$$y_{i2} = 1[\boldsymbol{\beta}_2'\mathbf{x}_{i2} + \gamma_2 y_{i1} + \varepsilon_{i1} > 0]$$

This model is not identified and is inestimable. In the common language used for discrete choice modeling, the model is ‘incoherent.’ That does not mean that one cannot try to fit the model. Unlike linear models, it is sometimes possible to obtain numbers for unidentified nonlinear models. For this particular one, NLOGIT will refuse to try – the first command produces an error message. But, suppose we try to bypass this control, and program and estimate our own unidentified model.

```
? This produces an error
Bivariate     ; if[t=1]   ; lhs=ip84,ip85;rh1=x84,ip85;rh2=x85,ip84$
? Try to program around the built in control
? Get starting values
Probit        ; if[t=1]   ; lhs = ip84 ; rhs=x84,ip85 $
Matrix        ; {c84=b(5) } ; b84=b(1:4) $
Probit        ; if[t=1]   ; lhs = ip85 ; rhs=x85,ip84 $
Matrix        ; {c85=b(5) } ; b85=b(1:4)$
```

```
? Try our own MLE. We use a familiar device to constrain rho to (-1,+1)
Maximize        ; if[t=1] ; maxit=10  ; start = b84,c84,b85,c85,0
                ; labels= a1,a2,a3,a4,a5,c1,c2,c3,c4,c5,tt
                ; fcn   = bx1=(2*ip84-1)*(a1'x84+a5*ip85)    |
                          bx2=(2*ip85-1)*(c1'x85+c5*ip84)    |
                           r = (exp(tt)-1) / (exp(tt)+1)     |
                          r12=(2*ip84-1)*(2*ip85-1)*r        |
                          log(bvn(bx1,bx2,r12)) $
```

This does not work very well.

```
? What is the sign of the correlation of these two variables?
CALC            ; List ; r12  =  (exp(tt)-1)/(exp(tt)+1) $
Crosstab        ; if[t=1] ; lhs = ip84 ; rhs = ip85 $
Bivariate       ; if[t=1] ; lhs=ip84,ip85 ; rh1=one;rh2=one $
Bivariate       ; if[t=1]; lhs=ip84,ip85;rh1=x84;rh2=x85$
```

The estimator appears to try to estimate something. But, the results are nonsense.

4. Sample Selection

A sample selection model for binary outcomes would work the same as in the linear case. However, the model is fit by maximum likelihood, and there is no inverse Mills ratio (‘lambda’) involved in the estimation. Here, we select on a particular industry, investment goods, and fit a model for innovation.

```
Bivariate       ; if[t=1] ;lhs = ip84,invgood
                ; rh1 = one,imum,fdium,sp,logsales
                ; rh2 = one,prod,im
                ; selection $
```

One occasionally reads that the sample selection model requires for identification that there be at least one variable in the selection (probit) equation that is not in the main equation. Strictly, this is not true for the bivariate probit model – in fact, it is not even true for the linear model. The following illustrates. In fact, the estimator is improved when the model is ‘identified’ in this fashion, but statistically, it is not necessary.

```
Bivariate       ; if[t=1] ;lhs = ip84,invgood
                ; rh1 = one,imum,fdium,sp,logsales
                ; rh2 = one,prod,im  ? With exclusion restrictions
                ; selection $
Bivariate       ; if[t=1] ;lhs = ip84,invgood
                ; rh1 = one,prod,imum,fdium,sp,logsales,im  ? No exclusion restrictions
                ; rh2 = one,prod,imum,fdium,sp,logsales,im
                ; selection $
```

5. Multivariate Probit Model

A multivariate probit model with more than two equations must be estimated by simulation. Currently, this is done using the GHK simulator. 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?

```
Mprobit        ; If [ _obsno <= 1270 ]
               ; lhs = ip84,ip85,ip86
               ; eq1 = x84
               ; eq2 = x85
               ; eq3 = x86
               ; Pts = 5 ; Maxit=10 $
```

Note, this is not the usual application of a bivariate or multivariate probit model. In the more common case, one would model two or more distinct binary decisions. We'll reconsider that possibility when we reexamine the healthcare data.

Part II. 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 continue to 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 (RP) model. Several models are fit:

(1) We start with the basic fixed parameters probit model.

(2) The second probit model is a random effects (random constant term) model fit using quadrature (the Butler and Moffitt method). We then fit several RP models.

- (a) This is the equivalent of the random effects model, formally fit as an RP model.
- (b) This is the relatively common form of the RP model in which the two coefficients of interest are treated as random normally distributed and independent.
- (c) This is the same as (b) except the two parameters are allowed to be freely correlated;
- (d) This is a much more general form in which the distribution of the random parameters can vary systematically with covariates. Here, the two random parameters are assumed to have a mean that varies by industry. In this case, we specify $\beta_k = \beta_{0k} + \delta_{k1}\text{RawMtl} + \beta_{k2}\text{InvGood} + \sigma_k v_k$.

We could also specify that the two parameters remain correlated and/or heteroscedastic. That is left for an exercise.

One of the interesting computations one can do is examine the distribution of the parameters across the sample observations (firms). To do this, we compute for each firm, the conditional expection of the parameter, given the information about the firm in the sample.

(e) 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. It also plots a normal distribution with the same mean and variance. By changing BIMUM to BFDIUM in the commands, you can repeat the exercise for the coefficient on FDIUM.

```
Namelist        ; x=one,logsales,sp,imum,fdium,prod,rawmtl,invgood$
Probit          ; lhs = ip ; rhs = x $
Probit          ; lhs=ip ; rhs = x ; panel ; random ; hpt=8$
? (a) Random constant model
Probit          ; Lhs = ip ; rhs=x ; Panel ; RPM ; fcn=one(n) ; Halton ; pts=25 ; Maxit = 10 $
? (b)  Uncorrelated random parameters
Probit          ; Lhs = Ip ; Rhs = x ; Panel ; RPM ; Maxit = 10
                ; Halton ; Fcn = imum(n),fdium(n) ; Pts = 25 $
? (c)  Correlated random parameters
Probit          ; Lhs = Ip ; Rhs = X ; Panel ; RPM ; Correlated
                ; Halton ; Fcn = imum(n),fdium(n) ; Pts = 25 ; Maxit = 10 $
? (d) Random parameters with covariates in the means of the random parameters.
Namelist        ; xit = one,logsales,sp,imum,fdium,prod $
Namelist        ; Sector = rawmtl,invgood$
Probit          ; Lhs = Ip ; Rhs = xit ; Panel ; RPM = Sector
                ; Correlated ; Fcn = imum(n),fdium(n) ; Pts = 25 ; Maxit=10$
? (e) Examine conditional means of random parameters.
Probit          ; Lhs = Ip ; Rhs = X ; Panel ; RPM ; Maxit=10
                ; Fcn = imum(n),fdium(n) ; Halton ; Pts = 25 ; Parameters $
Kernel          ; rhs = beta_i[1,B_IMUM]  ; Grid
                ; Title=Population Distribution of Probit Coefficient on IMUM $
Kernel          ; rhs = beta_i[2,B_FDIUM] ; Grid
                ;Title=Population Distribution of Probit Coefficient on FDIUM$
```

7. Latent Class Model.

In this exercise, we fit a latent class LOGIT model. This is an alternative method of building heterogeneity into the panel data model. The procedure looks for an appropriate specification by computing the MLE, then assembling the AIC and reporting the number of classes and the AIC. The optimal model has 6 classes. The model with 7 classes is clearly overspecified. An interesting computation from a latent class model is to derive the posterior class probabilities for each individual. With those in hand, we compute a prediction for the class for each person. The last Logit command displays the predictions from a 2 class model.

```
Namelist        ; xit    = one,logsales,sp,imum,fdium,prod $
Sample          ; All $
? Two class latent class logit model
Logit           ; Lhs = IP ; Rhs = xit ; LCM ; pts = 2 ; panel  ; Parameters $
? Constraint across classes
Logit           ; Lhs = IP ; Rhs = xit ; LCM ; pts = 2 ; Panel
                ; rst = b1,b2,b3,a1,a2,b4, c1,c2,c3,a1,a2,c4,p1,p2 $
? Specification Search Using AIC
Proc = LCM $
Logit           ; Lhs = IP ; Rhs = Xit ; LCM ; Pts = NumClass ; panel  ; quietly $
Calc            ; List ; Numclass ; ModelAIC $
EndProc$
Exec            ; Proc = LCM ; Numclass = 2,7 $
? An Overspecified model - too many classes
Logit           ; Lhs = IP ; Rhs = Xit ; LCM ; Pts = 7 ; panel  $
? Listing of Posterior Class Assignments
Logit           ; if[_obsno <= 500]
                ; Lhs = IP ; Rhs = Xit ; LCM = InvGood,Consgood,Food
                ; Pts = 2 ; panel  ; List $
```