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

Lab Session 4
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.