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

Lab Session 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

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?

```
Load the health care project
SAMPLE ; All $
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Pds = _Groupti $ (Conditional)
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Pds = _Groupti ; Fixed $
(Unconditional)
REJECT ; _Groupti > 2 $
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Pds = _Groupti $ (Conditional)
LOGIT ; Lhs = Doctor ; Rhs = hhninc,educ ; Pds = _Groupti ; 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 ; pds = _groupti $
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.

```
Sample ; All $
Probit ; lhs = hospital ; Rhs = hhninc,educ,hhkids,one ; random
       ; pds = _groupti  ;maxit=10$
Probit ; lhs = hospital  ; Rhs = hhninc,educ,hhkids,one ; Fixed
       ; pds = _groupti$
```

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 ; pds = _groupti $
```

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 ;  pds = _groupti  ;maxit=10
             ; random effects $
Probit ; ; lhs = hospital ; Rhs = x  ; pds = _groupti  ;maxit=10
       ; RPM ; Fcn = One(n) ; Pts = 20 ; Pds = _groupti $
Calc   ; K1 = Col(X) + 1 $
Calc   ; List ; SRP = B(K1) ; RhoRP = SRP^2 / (1 + SRP^2) $
```