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

Lab Session 4
Ordered Choice and Count Data Modeling

This assignment will explore a few variants of the ordered probit/logit model and some of the extensions of the Poisson regression model for count data..

All of the applications will use the health care data, healthcare.lpj

Part I. Ordered Choice Models

1. Preliminaries

For current purposes, the interesting variable in the file is 'Health Satisfaction,' HSAT, coded 0,1,…,10. In order to make our applications a little more compact, and estimate faster, we will collapse the variable from 11 to 5 cells as follows. We also define the data set as a panel.

```
?
? 1.  Preliminaries
?
Setpanel        ; Group = id ; Pds = ti $
Recode          ; HSAT -> HS ; 0/4=0;5/6=1;7/8=2;9/9.1=3;9.5/10=4$
Histogram       ; Rhs = HSAT,HS ; All $
```

2. Ordered Probit vs. Ordered Logit.

The same scaling effect occurs in the ordered choice setting as we see in the binary choice case. Fit ordered probit and ordered logit models and compare the results. Does the functional form appear to matter in the results? Compare both coefficients and marginal effects.

```
?
? 2. Ordered probit vs. ordered logit
?
Namelist        ; Demogrfc = age,income,female,married,hhkids $
OProbit         ; If[year = 1988] ; Lhs = HS ; Rhs = one,demogrfc ; Table = ocprobit $
Partials        ; If[year = 1988] ; Effects : demogrfc ; Summary  ; Means
                ; Table=ME_Probt $
OLogit          ; if[year = 1988] ; Lhs = HS ; Rhs = one,demogrfc ; Table = oclogit $
Partials        ; If[year = 1988] ; Effects : demogrfc ; Summary  ; Means
                ; Table=ME_Logit $
Maketable       ; ocprobit, oclogit $
Maketable       ; ME_Probt,ME_Logit $
```

3. Brant Specification Test

The Brant test is a specification test for the ordered logit model. The test is based on this observation. In the ordered choice model, for any j, j=0,…,J-1,

$$\text{Prob}(y \geq j) = (\alpha + \mu_j) + \beta'x.$$

This binary logit model can be computed for the J-1 specific binary variables defined by the condition. In principle, each logit model should estimate the same β, subject to sampling variation. If the estimated coefficient vectors appear to be systematically different, this casts doubt on the specification of the ordered logit model.

```
?
? 3. The Brant specification test
?
OLogit          ; if[year = 1988] ; Lhs = HS ; Rhs = one,income,hhkids,age ; Brant$
```

4. Explorations of the ordered choice models

The following fits an ordered logit model then simulates the probabilties.

```
?
? 4. Simulate an ordered logit model
?
OLogit          ; if[year = 1988] ; Lhs = HS ; Rhs = one,demogrfc ; Margin ; full$
Simulate        ; if[year = 1988]
                ; outcome=*;scenario:&income=.1(.1)2 | hhkids=0,1 ;means$
```

5. Partial Effects

The partial effects in the ordered choice models are complicated functions of the parameters and the data, as in the binary choice case. The issue is made worse here because the model implies a full set of probabilities, and each of them might be interesting. Here, we examine the effect of age and income on the probabilities for a model with interactions. Sometimes, a figure will be helpful in examining the impact of changes in a variable on the probabilities. The first figure below examines the behavior of the probability for the top cell. The second one looks at all the probabilities at the same time.

```
?
? 5. Partial effects in an ordered probit model
?
Oprobit         ; if[year = 1988] ; Lhs = hs
                ; Rhs = one,income,income*income,educ,age*educ $
Partials        ; if[year = 1988] ; outcome=* ; effects: income & age=25(3)65 ; Means $
Simulate        ; if[year = 1988] ; Scenario: & age=25(3)65 ; Outcome=4 ; Plot(ci) ; Means $
Simulate        ; if[year = 1988] ; Scenario: & age=25(3)65 ; Outcome=* ; Plot ; Means $
```

6. Different Margins

Since there is a full set of probabilities to be examined, there are various combinations of partial effects that might be of interest.

```
?
? 6. Different margins for an ordered choice model
?
OLogit          ; if[year = 1988] ; Lhs = HS ; Rhs = one,demogrfc ; Margin ; full$
```

7. HOPIT model

There have been numerous proposed extensions of the ordered choice models, some to allow the thresholds to vary with covariates in the observed data. The HOPIT model specifies

$$\mu_j = \exp(\theta_j + \boldsymbol{\delta}'\mathbf{z})$$

```
?
? 7. Generalized model with variable thresholds
?
ORDERED      ; if[year = 1988] ; Lhs = hs ; Rhs = one,age,educ,female,income ;partials $
CREATE       ; insuranc = public + addon $
ORDERED      ; if[year = 1988] ; Lhs = hs ; Rhs = one,age,educ,female,income
             ; HO1 = hhkids,insuranc ; Partial Effects $
```

8. A Model with Sample Selection

An ordered probit model for the transformed health status variable is specified below. Selection is based on whether the individual has public insurance. Is there evidence of 'selectivity?' The model is fit by full information maximum likelihood, not by any type of two step method.

```
?
? 8. Sample Selection Model
?
Probit       ; if[ti=7] ;Lhs = Public ; Rhs = One,Income,Hhkids ; Hold $
OProbit      ; if[ti=7] ;Lhs = HS ; Rhs = one,age,educ,income,married
             ; Selection ; Maxit=20$
```

9. Ordered Probit with Fixed and Random Effects.

Fit the ordered probit with no effects, with fixed effects and with random effects. Does the incorporation of the individual effects lead to large changes in the estimates?

```
?
? 9.  Panel Data Models
?
Ordered      ; Lhs = HS ; Rhs = one,income,hhkids,age ;table=pooled$
Ordered      ; Lhs = HS ; Rhs = one,income,hhkids,age ; panel ; FEM ;table=fem $
Ordered      ; Lhs = HS ; Rhs = one,income,hhkids,age ; panel ; Random ; Hpt=8
             ; Maxit=25 ; table=rem $
Oprobit      ; lhs = HS ; Rhs = one,demogrfc ; panel
             ; RPM ; Fcn = one(n),female(n),income(n) ; Pts=20
             ; Halton ; Maxit=10 ; Table = RPM$
Maketable    ; pooled,fem,rem $
Maketable    ; rem,rpm $
```

Part II. Models for Count Data

1. Effect of Censoring

We are going to model the number of doctor visits. This would be a conventional variable for analysis using Poisson regression except that it has a fair number of quite extreme values. One might be tempted to remove those observations. We will censor them at 10 instead. There is a predictable effect of censoring. It generally pushes coefficients toward zero. The following investigates. Use the original variable DOCVIS to fit a Poisson regression model. Then, use DOCVIS10, the censored variable and refit the model. What happens to the estimated coefficients?

```
Histogram     ; Rhs = DocVis $
Namelist      ; X=One,Age,Educ,Income,hhkids,Married$
Poisson       ; Lhs = DocVis ; Rhs = X ; Table = Full$
Create        ; DocVis10 = Docvis ~ 10 $
Poisson       ; Lhs = DocVis10 ; Rhs = X ; Table = Censored$
MakeTable     ; Full,Censored $
```

2. Poisson model with sample selection.

We fit a model for the number of visits to the hospital by people who have public insurance. This is a time consuming estimator (uses Hermite quadrature) so we reduce the sample size to speed things up. What do you conclude about the 'selectivity?' Is it significant? Interpret the estimated model.

```
Probit        ; if[Ti=7] ; Lhs = Public ; Rhs = One,Income,Hhkids ; Hold $
Poisson       ; if[Ti=7] ; Lhs = HospVis ; Rhs = X ; Selection ; MLE ; Hpt=8 $
```

3. Poisson and NB Models for Doctor Visits.

Fit the two models and compare the results. Is there evidence of overdispersion in the results? Fit the Poisson Hurdle model with the same specification. Do the results change significantly?

```
Poisson       ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Table = Poisson$
Negbin        ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Table = NB2$
Negbin        ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Model = NB1 ; Table=NB1$
Negbin        ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Model = NBP ; Table=NBP$
MakeTable     ; Poisson,NB2,NBP $
MakeTable     ; NB* $
```

The last three commands fit three forms of the negative binomial model, NB1, NB2, then NBP which nests NB1 and NB2. Based on the log likelihood functions, does NBP reject NB1 and NB2? Which of the two does NBP suggest is preferred, if one must choose between NB1 and NB2? If one of the NB models is preferred to the Poisson model, the Poisson estimator is still consistent, but a better covariance matrix estimator is called for. We examine this issue.

```
Poisson       ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Robust ; Table = Robust$
MakeTable     ; Poisson,Robust ; StandardErrors$
```

4. Two Part Models

The dependent variable in these applications has an extreme number of zeros. In these cases, one often employs a two part model, either a zero inflation model or a hurdle model. The latter is more common in the health economics literature. The following compares the two, though the comparison is a little dubious – they are different models.

```
Histogram    ; if[Ti=7] ; Rhs=DocVis $
Poisson      ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Hurdle ; Table=Hurdle$
Partials     ; Effects : income / married $
Poisson      ; if[Ti=7] ; Lhs = DocVis ; Rhs = X ; Rh2 = X ; Zip ; Table=zip$
Partials     ; Effects : income / married $
MakeTable    ; hurdle,zip $
```

5. Latent Class Models

We start by fitting a three class latent class logit model to the DOCTOR variable (visited the doctor at least once) variable. We then fit a three class model to the actual count of visits to the doctor. One of the post-estimation results for an LC model is the posterior probabilities, $\pi_{j|i}$ = Prob(class=j|all data for individual i). This provides the best information available for estimating which individual is in which class. After estimating the Poisson model, you will find a matrix named CLASSP_I with 887 rows (one for each person) and 3 columns (one for each class). The predictor of the class is the cell with the largest probability. It is usually easy to see which is which in the matrix.

```
Logit        ; if[Ti=7] ;Lhs=doctor ; Rhs=one,age,educ,income,hhkids,married
             ; LCM ; pts=3 ; panel ; maxit=25 $
Poisson      ; if[Ti=7] ;Lhs=docvis ? This model must be fit first for start values.
             ; Rhs=one,age,educ,income,hhkids,married$
Poisson      ; if[Ti=7] ;Lhs=docvis ; Rhs=one,age,educ,income,hhkids,married
             ; LCM ; pts=3 ; panel ; maxit=25 ; mar ; par$
Poisson      ; if[Ti=7] ; Lhs = docvis
             ; Rhs=one,age,educ,income,hhkids,married
             ; LCM ; pts=5 ; panel ; maxit=25 ; mar ;par$
Poisson      ; if[Ti=7] ; Lhs = docvis
             ; Rhs=one,age,educ,income,hhkids,married
             ; LCM = female ; pts=3 ; panel ; maxit=25 ; mar;par$
```

Researchers are sometimes interested in characterizing the different classes. For example, some studies in health economics of health care utilization measures such as DocVis that we are examining here, are interested in whether the two classes in a two class model can be labeled ‘heavy users’ and ‘light users.’ To do this, we would fit our two class model, compute the posterior class probabilities, use the probabilities to particition the sample into the two classes, then examine the mean usage in the two classes. The following sequence of commands does this for our data.

```
?-------------------------------------------------------------------
? Do the classes appear different?
?-------------------------------------------------------------------
Poisson     ; if[Ti=7] ; Lhs = docvis
            ; Rhs=one,age,educ,income,hhkids,married
            ; LCM=female,working ; Panel ; Pts = 2 ; Parameters $
Reject       ; Ti < 7 $
Create      ; group7 = trn(7,0)$
Create      ; pclass1 = classp_i(group7,1)$
Create      ; class = 1*(pclass1 > .5) + 2*(pclass1 <= .5) $
Dstat       ; if[Ti = 7] ;Rhs = Docvis ; Str = class $
Sample      ; All$
```

Fitting LCMs with large numbers of classes is ambitious. Though technically, the model is identified for any number of classes – with the RPM being the limiting model, in practice, identification is fairly weak once the number of classes gets even moderately large. Try fitting a five class model to the DOCVIS variable. Note, we have restricted the sample to those who have a full 7 observations. The last model is a three class model in which gender is assumed to influence the class probabilities. Do the results suggest that this is an appropriate model? (The estimator assumes that the LCM variables are time invariant, since the classes and the class probabilities are time invariant.)

```
Poisson     ; if[Ti=7] ; Lhs = docvis ; Rhs=one,age,educ,income,hhkids,married $
Poisson     ; if[Ti=7] ; Lhs = docvis ; Rhs=one,age,educ,income,hhkids,married
            ; LCM ; pts=5 ; panel ; maxit=25 ; mar  ;par$
Poisson     ; if[Ti=7]  ; Lhs = docvis ; Rhs=one,age,educ,income,hhkids,married $
Poisson     ; if[Ti=7]  ; Lhs = docvis ; Rhs=one,age,educ,income,hhkids,married
            ; LCM = female ; pts=3 ; panel ; maxit=25 ; mar;par$
```

6. Fixed Effects Models

We consider estimation of fixed effects models for count data. We start with the Poisson regression. There are two approaches, brute force and conditional estimation. For the Poisson model, these two approaches produce numerically identical results. (As for the LC models, each begins with a pooled version to deliver starting values.)

```
Poisson     ; if[Ti=7]  ; Lhs=docvis  ; Rhs=one,age,educ,income,hhkids,married $
Poisson     ; if[Ti=7]  ; FEM ; Panel ; Lhs=docvis
            ; Rhs=one,age,educ,income,hhkids,married$
Poisson     ; if[Ti=7]  ; Lhs=docvis  ; Rhs=one,age,educ,income,hhkids,married $
Poisson     ; if[Ti=7]  ; Fixed effects ; Panel ; Lhs=docvis
            ; Rhs=one,age,educ,income,hhkids,married$
```

The preceding result does not apply to the negative binomial model. There is what we call the 'True Fixed Effects' model, which is based on a conditional mean function, $\lambda_{it} = \exp(\boldsymbol{\beta}'\mathbf{x}_{it}+\alpha_i)$. The second approach was developed by Hausman, Hall and Griliches (1984). This model puts the fixed effect in the scale parameter, rather than the conditional mean. These two approaches produce different results, as we examine hhere.

```
?
? Fixed effects with no time invariant variables
?
Negbin          ; if[Ti=7]  ; Lhs=docvis  ; Rhs=one,age,educ,income,hhkids,married $
Negbin          ; if[Ti=7]  ; FEM ; Panel ; Lhs=docvis ; Table=TrueFE
                ; Rhs=one,age,educ,income,hhkids,married$
Negbin          ; if[Ti=7]  ; Lhs=docvis ; Rhs=one,age,educ,income,hhkids,married $
Negbin          ; if[Ti=7]  ; Fixed Effects ; Panel ; Lhs=docvis
                ; Rhs=one,age,educ,income,hhkids,married ; Table=HHGFE$
MakeTable ; TrueFE,HHGFE $
```

The HHG model has the peculiar feature that it allows time invariant variables in the FE model. This is unfamiliar, and is unique to the negative binomial model. We explore this below.

```
?
? Fixed effects with a time invariant variable, FEMALE
?
Negbin          ; if[Ti=7]  ; Lhs=docvis; Rhs=one,age,educ,income,hhkids,married,female$
Negbin          ; if[Ti=7]  ; Fixed Effects ; Panel ; Lhs=docvis
                ; Rhs=one,age,educ,income,hhkids,married,female$
Negbin          ; if[Ti=7]  ; Lhs=docvis; Rhs=one,age,educ,income,hhkids,married,female$
Negbin          ; if[Ti=7]  ; FEM ; Panel ; Lhs=docvis
                ; Rhs=one,age,educ,income,hhkids,married,female$
```

7. A Random Parameters Model

The same set of random parameters approaches that we applied to the binary choice models earlier can also be applied to the count data models. The estimator below is fairly elaborate, with several random parameters. The last two commands examine the distributions of the expected values of one of the coefficients.

```
?
? 7. A Random Parameters Count Data Model
?
Poisson         ; if[Ti=7]  ; Lhs=docvis
                ; Rhs=one,age,educ,income,hhkids,married,female,hs $
Poisson         ; if[Ti=7]  ; Lhs=docvis
                ; Rhs=one,age,educ,income,hhkids,married,female,hs
                ; RPM ; Halton ; Pts = 20 ; Panel ; maxit=10
                ; Fcn = one(n),income(n),female(n),hs(n) ; Parameters $
Matrix  ; beta_fem = beta_i(1:887,3:3) $
Kernel  ; Rhs = beta_fem ; Title=Distribution of E[beta_fem] Across Sample $
```