**Discrete Choice Modeling**
**Count Data**
**Exercise 6**

# Part I. Models for Count Data

This exercise will use the full panel data set, healthcare.lpj

As usual, after loading the data set, identify the panel with

**SETPANEL ; Group = id ; Pds = ti $**

## A. Poisson and NB Models for Doctor Visits.

We will model the doctor visits variable, DOCVIS. To set up the estimation of the count data models, define your model specification with

**NAMELIST ; XC = the variables in your model $**

(Be sure to include the constant term, ONE, in XC.) Now, fit pooled Poisson and negative binomial models. Is there evidence of overdispersion in the reported statistics?

As usual, partial effects are more interesting than coefficients. Obtain estimates of the partial effects for your model.

NLOGIT Tip: The sample is fairly large. The delta method can take a while if you use the average partial effects. You can use a subsample as follows. Also, the command below shows a useful shorthand:

**PARTIALS ; If [ ti = 7 ] ; Effects: XC ; Summary $**

You can also reduce computation time close to zero by using the means, instead. Add **;Means** to the **PARTIALS** command to do this.

Note that you estimated the model using the full unbalanced panel. You care computing the average partial effects using the subset of groups that have 7 years of data. It might be useful to see if the partial effects at the means are nearly the same as the average partial effects. You can get a comparison by adding ;Means to the PARTIALS command. Do the results change very much?

## B. Two Part Models

The dependent variable in many 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. These are competing behavioral models. The latter is more common in the health economics literature. You want to compute both the hurdle and ZIP models. The models are not nested, so there is no natural (Neyman-Pearson) test for one against the other. The Vuong test seems like it might be helpful.

NLOGIT Tip. Your command stream might look like the following:

```
NAMELIST     ; X2 = your variables $   (Choose 2 or 3, not including female)
                                        We used X2=one,married,hhkids.)
HISTOGRAM ; Rhs=DocVis $  (Just to see the problem)
POISSON      ; Lhs = DocVis ; Rhs = X ; Hurdle ; Rh2 = X2 ; Table=Hurdle$
CREATE       ; lh = logl_obs $
POISSON      ; Lhs = DocVis ; Rhs = X; Zip      ; Rh2 = X2;  Table=zip$
CREATE       ; lz = logl_obs $
MAKETABLE ; hurdle,zip $
CALC         ; Tsd ( lh,lz)  $
```

What do you conclude? Which model seems to be the better specification?

## Part II. Random Parameter Models

### A. Latent Class Hurdle Model

This exercise will replicate some of the flavor of the Bago d'Uva and Jones study of health care utilization. This will involve fitting a two latent class Poisson/Hurdle model for doctor visits. Some preliminaries: As usual, if you have not already done so, define the panel with

**SETPANEL ; Group = id ; Pds = ti $**

The outlying values of DOCVIS impede the analysis, so we will censor them for purpose of the application. We top code the values at 12 with

**CREATE ; docvis12 = docvis ~ 12 $**

(The ~ operator produces the minimum of the two operands.). Finally, define the parts of your model with

**NAMELIST ; XP = … the variables in your main equation $**

Use 3 or 4, including ONE. And,

**NAMELIST ; XH = … the variables in your hurdle equation $**

Use 3, including ONE. We used one,hhkids,working in our own experiment.

1. We start by fitting a two class latent class Poisson model to the DocVis12 variable. To fit latent class count data models with NLOGIT, you must first fit the pooled model, then fit the identical model with latent classes. Your starting point will look like

**POISSON ; Lhs = DocVis12 ; Rhs = XP $**
**POISSON ; Lhs = DocVis12 ; Rhs = XP ; LCM ; Pts = 2 ; Panel $**

An LCM for a count variable will always consist of these two commands. Try fitting a three class and a 4 class model. (Remember to fit the pooled model first each time.)

2. We will now fit a two class Hurdle model. The generic hurdle model command is

**POISSON ; Lhs = DocVis12 ; Rhs = XP ; Rh2 = XH ; Hurdle $**

Fit a two latent class hurdle model and observe the results.

3. Jones and Bago d'Uva identified their two classes with 'light' users and 'heavy' users. Let's see if our data display that feature. We are going to try to identify the class membership. To do that, we are going to compute the posterior class probabilities. Fit the two class hurdle model, and add **;Parameters** to the second POISSON command. The model results will be the same as in part 2. Go to the Project window and open the Matrices folder. Among other results, you will see a matrix named CLASSP_I. Double click that name. Note that it is a 7,293 rows and 2 columns matrix. In each row, the two values are the estimated posterior class probabilities. We are going to manipulate these to do the classification. First, we move the probabilities to a variable with

**CREATE ; pclass1 = classp_i (group,1) $**
**CREATE ; class1 = pclass1 > .5 $**

Class1 is a new variable that equals 1 for individuals estimated to be in class 1. When class1 equals zero, that means we estimate the individual to be in class 2. Now, do our data look like Jones and Bago d'Uva's?

**DSTAT ; Rhs = Docvis12 ; Str = class1 $**
**CALC ; Tst(docvis12,class1) $** (Tests whether means are different in the two groups)

What do you find?

4. Jones and Bago d'Uva fit a negative binomial model rather than a Poisson model. You can replicate that part of the analysis by going back to the beginning, and fitting a NEGBIN model rather than a POISSON model. The LCM NegBin model is actually quite difficult to fit. For purpose of your exploration, you can add **;MAXIT=20** to your latent class command. Also, be sure to use DOCVIS12, not DOCVIS for the analysis.

## B. Fixed Effects Model – The Curious HHG FE Model

### 1. Fixed Effects Poisson Model

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.) We will use your NAMELIST definition for XP. However, if your list includes FEMALE, you must take that out – FEMALE is time invariant. Note also, your XP list contains ONE. Don't worry about that – NLOGIT will take that out. The following fits the model using the conditional estimator derived by Hausman, Hall and Griliches (1984).

**POISSON ; Lhs = docvis12 ; Rhs = your XP list $**
**POISSON ; Lhs = docvis12 ; Rhs = your XP list ; Panel ; Fixed $**

To compute the unconditional estimator, change Fixed to FEM in the second command. Now, refit the model. Notice that you get the same results (though the second method is much faster).

### 2. Fixed Effects Negative Binomial Model

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). Their model puts the fixed effect in the scale parameter, rather than the conditional mean. These two approaches produce different results, as we examine here. To demonstrate this, just change POISSON to NEGBIN in the model commands, and fit the models with **;Fixed** (for the HHG model) and **;FEM** for the unconditional estimator.

### 3. FE with a Time Invariant Variable

It is generally understood that parameters on time invariant variables in a fixed effects model are not identified – FE models cannot contain time invariant variables. But, the HHG FENB model can. To see this, just redo part b, but rather than **;Rhs=XP**, use **;Rhs=XP,female** in your command. (The model command can be stated in terms of a namelist and variables.)

## C. A Random Parameters Model

The following template shows how to specify a random parameters model, and do an interesting exercise with the results. The template is general – you can use it with a probit, Poisson, linear regression, etc. The ;Halton is a useful option that speeds up and improves the simulation. ;Pts=50 is less than usually used in practice, but it will make the example work fairly quickly.

```
? Set the panel as usual if you have not done so already in this session.
? This step is necessary for the Count data and ordered choice models
MODEL  ;  Lhs = … ; Rhs = a set of variables $  (Fit the pooled model, no options.)
?
MODEL  ; Lhs = …  ; Rhs = a set of variables
          ; RPM ; Halton ; Pts = 50 ; Panel ; Parameters ; Maxit = 20
          ; Fcn = variable (n), variable (n), …  $
```

This fits the model. After estimation, as requested by the ;Parameters, there will be a new matrix named BETA_I which has 1 row for each individual in the sample and one column for the conditional expected parameters specified in the FCN list. Since our sample has 7,293 individuals, say we used FCN=one(n),female(n), then BETA_I would be 7,293 by 1. A kernel estimator of the distribution of means across individuals is interesting. You can do something like the following:

```
MATRIX  ; beta_fem = beta_i(1:7293,2:2) $
KERNEL  ; Rhs = beta_fem ; Title=Distribution of E[beta_fem] Across Sample $
```

1. Fit a probit model for DOCTOR model that contains EDUC, then describe the distribution of EDUC across individuals.

2. Fit a Poisson regression model for DOCVIS12 that contains FEMALE as shown above.