Homework 7: Multichotomous Dependent Variables III

Due Date:

Exercise 1: Event Count Models

For this exercise use vetos.dta, which includes information on presidential vetoes and veto overrides. If you type 'describe', you will see what all the variables refer to.

1. Estimate the following Poisson model of presidential veto overrides:

   - poisson nover hmargin smargin congexpr govexpr reelect popvote

   Report the results in the first column of the following table.

Table 1: The Determinants of the Number of Presidential Veto Overrides

	Dependent Variable: Number of Presidential Veto Overrides				
Regressor	Poisson	Poisson (exposure)	Poisson (lnvetoes)	Negative Binomial	Poisson (bootstrap)
MarginHouse					
MarginSenate					
CongressExperience					
GubernatorialExperience					
Reelect					
PresidentVote					
Exposure(NumberVetos)					
ln(PresidentVetos)					
Constant					
α					
Log likelihood					
Observations					

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses (columns 1-4)

2. Interpret the coefficients on congressional experience (congexpr) and gubernatorial experience (govexpr).

3. Manually, calculate the expected number of veto overrides for the following situation: hmargin = -9; smargin = 2; congexp = 0; govexp = 0; reelect = 1; popvote=52. Provide confidence intervals and interpret.

4. Manually, calculate the expected number of veto overrides for the same situation but where govexp = 1

5. Calculate the change in expected counts between these two scenarios. Provide confidence intervals and interpret.

6. Calculate the marginal effect of an increase in smargin in the following situation: hmargin = -9; smargin = 2; congexp = 0; govexp = 0; reelect = 1; popvote=52. Provide confidence intervals and interpret.

7. Calculate an incident rate ratio if we change govexp from 0 to 1. Use STATAs 'irr' command and interpret. Remember to show confidence intervals.

8. The number of presidential veto overrides that is possible is limited by the number of presidential vetoes that there have been. As a result, there is an issue with 'exposure' that needs to be dealt with.

    - Estimate the same model as before but take account of exposure using STATA's automatic **exposure** and **offset** commands. They should produce the same estimates. Put one of the sets of estimates in the second column of the table above.
    - Write one or two sentences explaining why we need to take account of exposure.
    - Based on these results, did we need to take account of exposure or not? How do you know?
    - Compare these results from those in column 1. What is the same and what is different?
    - Create a variable that is the natural log of the number of presidential vetoes. Instead of using STATA's automatic commands for exposure, now just include this new variable in the model. Put the results in the third column of the table above. Test to see whether we needed to take account of exposure. How might we interpret the coefficient on this new variable?

9. Now estimate the original specification with a Negative Binomial model and put the results in the fourth column of the table above. Do we have overdispersion? How do you know?

10. One thing you may have noticed is that there are only 26 observations in the data set. This means that the asymptotic assumptions of MLE are not met. As I suggested a few weeks ago, one way around this might be to bootstrap the standard errors. Estimate the original Poisson specification that you did in column 1 using STATA's **bootstrap** command with 100 draws (this will take a little time). Put the results based on the N or the P confidence interval in the last column of the table above. How do these results differ from the Poisson model estimates?

Exercise 2: Zero-Truncated Event Count Models

For this exercise use biochem.dta, which includes information on the number of publications produced by Ph.D. biochemists. If you type 'describe', you will see what all the variables refer to.

1. First drop all those people who have not produced an article in order to artificially truncate the sample.
2. Estimate a zero-truncated negative binomial model with the following independent variables: FEMALE MARRIED KID5 PHDPRESTIGE MENTOR3. Put the results in Table 2 below.

Table 2: The Determinants of the Number of Publications

Dependent Variable: Number of Publications	
Regressor	ZNTB
Female	
Married	
Kid5	
PhD Prestige	
Mentor3	
Constant	
α	
Log likelihood	
Observations	

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses

3. Interpret the coefficient on FEMALE and MENTOR3.
4. Calculate the predicted probability of various counts for the situation where FEMALE = 0, PHDPRESTIGE = 1, MENTOR3 = 0, and the other variables are at their mean by using the PRVALUE command from Long's SPOST ado file.
    - prvalue, x(female=0 phdprestige=1 mentor3=0) rest(mean) all maxcnt(19)
5. Interpret the expected and unexpected rates that are shown in the output.
6. Why is there no conditional probability for 0 publications?
7. Interpret the predicted probability of 0 publications.

Exercise 3: Hurdle Models

For this exercise use the same biochem.dta as before but don't drop observations of no publications.

1. Estimate a hurdle model using a logit and a zero-truncated negative binomial model with the following independent variables – FEMALE MARRIED KID5 PHDPRESTIGE MENTOR3 – in both equations. Put the results in Table 3 below.

Table 3: The Determinants of the Number of Publications

	Dependent Variable: Number of Publications	
Regressor	Logit	ZNTB
Female		
Married		
Kid5		
PhD Prestige		
Mentor3		
Constant		
α		
Log likelihood		
Observations		

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses

2. Calculate the predicted probability of various counts for the situation where FEMALE = 0, PHDPRESTIGE = 1, MENTOR3 = 0, and the other variables are at their mean by using the code in the class notes.
3. Interpret the results. What is the probability of not publishing? What is the probability of publishing 2 articles?

Exercise 4: Zero-Inflated Models

For this exercise use the same biochem.dta as before.

1. Estimate a zero-inflated poisson and negative binomial model with the following independent variables – FEMALE MARRIED KID5 PHDPRESTIGE MENTOR3 – in both equations. Use a logit for the first stage. Put the results in Table 4 below.

Table 4: The Determinants of the Number of Publications

Dependent Variable: Number of Publications				
Regressor	ZIP		ZINB	
	logit	count	logit	count
Female				
Married				
Kid5				
PhD Prestige				
Mentor3				
Constant				
α				
Log likelihood				
Observations				

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses

2. Interpret the coefficient on MENTOR3 in the logit equation from the ZINB model. Interpret the coefficient on MENTOR3 in the count equation from the ZINB model.

3. Interpret the exponentiated coefficient on MENTOR3 in the logit equation from the ZINB model. Interpret the exponentiated coefficient on MENTOR3 in the count equation from the ZINB model.

4. Suppose we want to compare the predicted probabilities for a married female scientist with young children who came from a weak graduate program with those for a married male from a strong department who had a productive mentor using the results from the ZINB model. To do this, type:

   - quietly prvalue, x(female=0 married=1 kid5=3 phdprestige=3 mentor3=10) save;
   - prvalue, x(female=1 married=1 kid5=3 phdprestige=1 mentor3=0) diff;

5. Why are there two predicted probabilities for 0 counts? Interpret each of them.

6. Interpret some of the other results.

Exercise 5: Article Assignment

Find a substantive article that uses methods covered in the class so far.

1. In a paragraph summarize the substance of the article.
2. In a second paragraph lay out the methods used. If they use more than one method, choose the most interesting one.
3. In terms of what we have discussed, what does the article do methodologically right? What good practice do you see?
4. What bad practices would you correct?
5. What is unclear in the methods presentation that you would have liked to have seen discussed?

Please try to keep this to about two pages. Over three pages and you are doing too much. Much under two and you are doing too little.

Be prepared to discuss in class if called upon.