Copenhagen Business School Fall 2022

Ph.D. course: Applied Econometrics for Researchers

Workshop 3

Stata tutorials can be found on the Stata webpage: http://www.stata.com/links/video-tutorials/

There are lots of other tutorials available on the WWW (of varying technical quality).

For logistic regression, you may find this LSE tutorial useful: https://www.youtube.com/watch?v=0C_Hlh_jNq8

For the count data exercises, you might find the following material helpful to supplement the lecture material on Count Data Models (note that they follow the same examples, but provide extra notes that help you interpret the outputs):

https://stats.idre.ucla.edu/stata/dae/poisson-regression/

https://stats.idre.ucla.edu/stata/dae/negative-binomial-regression/

1. **Motivation:** This workshop continues investigating the determinants of the innovative performance of a firm. For the binary response models, we will depart from the basic specification from Workshop 2 (without a possible interaction between external sources and own R&D). We will focus on the use of a discrete measure of innovative performance as the dependent variable in the econometric analysis.

   The second part of the workshop investigates the determinants of firm-level patent applications (count variable). In particular, we consider the role of firm size, cooperation with external parties, the intensity of a firm's own R&D, and characteristics of the firm's market. We also consider how to model the relationship between such determinants and the number of patents applied for by a firm.

2. **Data:**

   Use the do-file that you produced for Workshop 2 to set up the data set. Alternatively, you can download WORKSHOP2_final.do from the Canvas page.

3. **Variables:**

The data we are going to use are from the UK Innovation Survey. We looked up the variable definitions from the questionnaire in Workshop 1. We will introduce some new variables both for the DVs (***prodinov*** and

patapply) and also as IVs (***pcoop***, ***market***). Make sure that you know what these variables measure/how they are constructed. You can look up the variable definitions from the questionnaire.

You need to add the following conditions to select your regression sample:

```
drop if expint > 1

drop if rdint > 1 & rdint != .
```

4. **Logit and probit regression:**

Consider our basic model with ***extsource*** and ***rdintpct*** as explanatory variables and ***inconst*** and ***lempl00*** as controls. Now, instead of modelling ***prodnew***, we will consider the alternative measure of innovative performance, ***prodinov***, as our dependent variable.

a. What type of variable is ***prodinov****?* What would be a proper model for this variable?

b. Estimate the model as a logistic regression. How do you interpret the parameter estimates? Which variables are significant in the logit regression?

c. Calculate the average partial effects of the variables using the `margins command`.

d. Calculate the marginal effects of the variables when measured at their mean values (option `atmeans`).

e. Produce a plot illustrating these results using the Stata command *marginsplot*.

f. Compare the two sets of marginal effects that you have calculated for the logit model. Also compare to the results of a linear regression on the same set of variables.

g. Calculate the probability that a firm is product innovating if the regressors are set at their mean values.

h. Calculate the probability that a firm is product innovating for different values (of your choice) of internal R&D. Illustrate how this probability changes for different values of R&D in a plot.

i. Comment on your calculations.

Redo the analysis (points b, c, d, g) with the probit estimator.

- What is the difference between the logit and the probit model?

- Compare the parameter estimates and the marginal effects from the two estimators.

>>>>>>>>> Follow-up on question 4 in class <<<<<<<<<<<<<<<<<<<<<

5. **Count data models:**

The data you are going to use for this question are again from the UK Innovation Survey. The dependent variable (***patapply***) is a count of the number of patents a firm applied for during the period 1998 to 2000. The independent variables are ***lempl00, rdintpct***, ***pcoop***, and ***market***.

- Examine the dependent variable. Comment on its main characteristics. *Hint: use the summarize and hist commands to get an overview of this variable. Consider only a part of the range of* **patapply** *if you find the graphics a bit hard to interpret.*
- Do you find any problems with the data on the DV? How would you correct these problems?
- Consider the variable ***market***. How would you include a variable like this in a regression model?

Poisson model for the number of patent applications (Follow the Lecture slides and references therein for guidance)

Consider a Poisson regression with the DV and IV as listed above.

- Look up the syntax for Poisson regression (`poisson` command) in Stata Help and run the regression.
- How do you interpret the regression output?
- Which variables are significantly determining the number of patent applications for a firm? What hypotheses are you considering here?
- Calculate the expected number of patents for a firm that has average values for all explanatory variables. *Hint: Use "`margins`" with the "`atmeans`"* option.
- For firms that cooperated with other firms for innovation-related activities (pcoop=1), calculate and illustrate the different expected number of patents for a firm in each market type, holding average values for lempl00 and rdintpct. *Hint: Use "`margins`" with the "`atmeans`" and "at()" options.*

Another commonly used estimation method for count data is Negative Binomial Regression.

Therefore, let us estimate a Negative Binomial model for the number of patent applications (Follow the Lecture slides and references therein for guidance)

- Look up the syntax for `nbreg` regression in Stata Help and run the regression.

- How do you interpret the regression output?
- Which variables are significantly determining the number of patent applications for a firm?

>>>>>>>>> Follow-up on question 5 in class <<<<<<<<<<<<<<<<<<<<