Applied Econometrics
Måns Söderbom
University of Gothenburg
September 2021
mans.soderbom@economics.gu.se

Computer Exercise

Part I: Difference-in-Differences estimation

In this part of the exercise we will use the Bacon-Goodman data on divorce law reform and female suicides. The Stata file is called bacon_example_extended.dta and can be obtained from my website. The dataset contains the following variables:

```
Contains     data     from     C:\Users\xsodem\Dropbox\teaching\Applied     econometrics     fall
2021\bacon_example_extended.dta
  obs:         1,617
 vars:            15                          20 Sep 2021 22:49
                                              (_dta has notes)
-------------------------------------------------------------------------------------------
storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------
stfips          byte    %10.0g                State FIPS code
year            float   %9.0g                 Year
_nfd            int     %10.0g                No-fault divorce onset
post            float   %9.0g                 Treatment dummy
asmrs           float   %9.0g                 Suicide Mortality
pcinc           double  %10.0g                Per-Capita Income, BEA
asmrh           float   %9.0g                 Homicide Mortality
cases           double  %10.0g                AFDC cases
weight          float   %9.0g                 Population weight
copop           double  %9.0g                 Population
Npost           float   %9.0g                 Number of treatment years
nonreform       float   %9.0g                 Dummy = 1 for Non-reform states
prereform       float   %9.0g                 Dummy = 1 for Pre-1964 reform states
reformyr        float   %9.0g                 Year of reform for state
rperiod         float   %9.0g                 Reform period (same as year up to 1977, then 10 for
                                               1980, 11 1984 and 12 for 1985
-------------------------------------------------------------------------------------------
Sorted by: stfips  year
```

Question 1. In my lecture, I showed event study estimates comparing the 1973 reform states to the non-reform states. Replicate the results in my lecture, and interpret the results. Comment on whether the results are informative about whether the assumption of common trends in the pre-reform period holds.

Question 2. Figure 5 in Goodman-Bacon shows event study estimates graphically, for the entire sample.

(i) Obtain regression results that replicate the results in this graph (i.e. for the entire sample).

(ii) Would you say that the graph suggests that treatment effects happen instantaneously once the reform has been undertaken, or would you say that the event study results suggest that treatment effects develop gradually over time? Discuss possible implications of your conclusions for the TWFEDD estimator.

Question 3. Replicate the TWFEDD estimate (-3.08) using the Stata command xtreg (with the FE option added). Check if the effect remains statistically significant once you cluster standard errors at the state level. Comment on your findings.

Question 4. Use the nonreform states as a control group and choose one 'timing' group as treatment group (but do not use the 1973 group since we did that in class earlier). Obtain 2 x 2 DD estimates. Propose a way of testing for a common time trend distinguishing the pre-reform period and the post-reform period. Carry out the test and interpret the result.

Question 5. Have a look at Figure 6 in Goodman-Bacon. How does the inclusion in the sample of the non-reform states and the pre-1964 states affect the DD estimate? Investigate how the DD estimate changes if you exclude these states and discuss the results.

Question 6. Add controls for per-capita income, female homicide rates and per-capita welfare caseloads to the specification. How do these additional control variables affect female suicide rate? How is the DD estimate affected by the addition of these control variables to the specification? Why?

Part II: Count responses and tobit estimation

In this part of the exercise we will use a dataset on the number of visits made a patients to the doctor. The Stata file is called randdata.dta and can be obtained from my website.

Question 7. Explain briefly the dataset and provide some summary statistics: What variables are included? What are their mean values? How many observations? Show a histogram of the mdvis variable and comment on the distribution of this variable.

Question 8. We start by using OLS: Run the following in Stata,

reg mdvis logc idp lpi fmde physlm disea hlthg hlthf hlthp linc lfam xage female child femchild black educdec, robust

What is the estimated effect of income on the expected number of visits to the doctor?

Question 9. Now use a Poisson regression to estimate the model, with the same explanatory variables as in question 8. What is the estimated effect of income on the expected number of visits to the doctor?

Question 10. Use negative binomial regression type 1 and type 2 to estimate the model. Compare the effects of income to those obtained in question 9 and 10.

Question 11a (optional; only do this if you have time). You have just estimated 4 different econometric models. For each model (including OLS) compute the predicted probabilities that the number of visits to the doctor is equal to 0, 1, 2,…,15. Discuss how these predicted probabilities fit the actual distribution of the mdvis variable.

Question 11b (optional; only do this if you have time). Use tobit to estimate the model. Compare the effects of income to those obtained previously.

Part III: Heckit estimation

In this part of the exercise we will use the "Heckit" estimator to analyse the relationship between experience and wage offers amongst women in the US. The Stata file is called MROZ.dta and can be obtained from my website.

Question 12. Explain briefly the dataset and provide some summary statistics. For what proportion of the sampled women do we have information about their wage?

Question 13. First, let's investigate if we can obtain reliable estimates of the Heckit model without exclusion restrictions. Estimate a Heckit model where lwage depends on educ and exper, and where selection depends on the same two variables. What's the effect of experience on log wage? Discuss the results.

Question 14 (optional). Replicate the results obtained in question 13 by using probit to obtain an estimate of the inverse Mill's ratio, and then OLS on the selected sample with IMR added.

Question 15. Add the following variables to the selection model (but not to the wage model): nwifeinc age kidslt6 kidsge6. How do the Heckit results change as a result?