Panel Data Analysis of Microeconomic Decisions Assignment Part II

Due date: December 2, 2019

Assignment

Your assignment should include the relevant Stata code and output, as well as your answers to the questions (maximum 8 pages in total). Each student must submit individual answers that differ from the answers of the other students. Please submit your assignment by email to m.husiatynski@uvt.nl

Data

For this assignment we use a US sample from the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. These individuals were interviewed annually until 1994 and once every two years after that. The last available round of the survey was conducted in 2016.

The sub-sample you are given comprehends 11,257 individuals (N) observed through 12 waves of survey (T), for a total of 135,084 observations ($N \cdot T$). Only the period from 1994 to 2016 is considered. You can find the data in the file data.dta. The variables in the data set are given in Table 1 below.

The students will work on different samples composed of eight consecutive waves. The sample you have to use depends on the last digit of your student number (SNR):

- if the last digit is 0 or 1 you only use waves from 1 to 8 (i.e. from 1994 to 2008)

- if the last digit is 2 or 3 you only use waves from 2 to 9 (i.e. from 1996 to 2010)

- if the last digit is 4 or 5 you only use waves from 3 to 10 (i.e. from 1998 to 2012)

- if the last digit is 6 or 7 you only use waves from 4 to 11 (i.e. from 2000 to 2014)

- if the last digit is 8 or 9 you only use waves from 5 to 12 (i.e. from 2002 to 2016)

Variable Name	Description
ID	Identification number
YEAR	Year of the survey
WAVE	Survey wave
SEX	1 for women, 0 for men
AGE	Age at the time of the survey
MARRIED	1 for married, 0 otherwise
WHITE	1 for white people, 0 otherwise
INCOME	Total income from wage and salary
SP_INC	Total income from wage and salary of the spouse
EMPL	1 for employed, 0 otherwise
YOUNG_CH	1 for child less than 11, otherwise
NUM_CH	Number of children
EDU_12	1 if highest degree is high school level, 0 otherwise
EDU_13_15	1 if highest degree is undergrad level, 0 otherwise
EDU_15	1 if highest degree is grad level, 0 otherwise
JOB_SAT	Job satisfaction on a scale from 1 (like it very much) to 4 (dislike it very much)

Table 1: Variables

Questions

1. Binary choice models

1. Present a table with descriptive statistics of the variables of interest (not ID, YEAR, WAVE) and briefly comment on what you think are the most interesting numbers in the table (1 page max).

2. Use a static pooled logit model to explain employment (EMPL) from age, marital status, presence of young children, number of children, education, and a dummy for white people. Use people with at most a high school degree as the reference group for the education dummies. Comment on the results.

3. Estimate a static random effects logit model using the same specification as in the previous question and comment on the results. What are the main differences compared to the results of the pooled model?

4. Which of the two models estimated so far is more appropriate? Support your answer using a statistical test. Mention all steps of your test.

5. Start from the model that you selected in the previous question. Investigate whether there is a non-monotonic effect of age on employment status using a formal test.

6. Using the model of the previous question, compute the marginal effect of a one year increase in age on the probability to be employed for an observation with probability 0.5 to be employed and age 40 years old. At which age would the marginal effect be equal to zero?

7. Considering the specification used in question 3, run a static fixed effects model and comment on the results. Carry out a test to choose between the random and the fixed effects specification.

8. Estimate a quasi fixed effects model that allows the individual effects to be correlated with education (you can assume that the parameters that determine the relation between education and the individual effects are the same across years). Comment on the results.

9. Test the quasi fixed effects model versus the fixed effects model.

10. Unemployment can be quite persistent over time and you wonder whether state dependence plays any role. Estimate a dynamic version of the random effects logit model using the Wooldridge approach and comment on the results. What can you say about the importance of unobserved heterogeneity and state dependence?

2. Tobit models

1. Estimate a static random effects tobit model to explain INCOME from sex, age, white dummy, marital status and education dummies. Comment on what you think are the main results.

2. What can you say about the importance of unobserved heterogeneity according to this model?

3. Investigate whether the effect of marital status is different for males and females.

4. Estimate the dynamic version of the model you estimated in question 1 using the Wooldridge approach. What can you say about the importance of unobserved heterogeneity and state dependence?

5. Test whether the individual effects are correlated with the initial value of INCOME.

3. Ordered response models

1. Estimate a static random effects ordered probit model to explain job satisfaction (JOB_SAT) from age, income and education. In this case we are only interested in the sub-sample of employed individuals.

2. Why does the output not include an estimate for the constant term?

3. Comment on the likelihood ratio test reported in the bottom part of the Stata output: what are the null and the alternative hypothesis, the test statistic and the conclusion.

4. What is the marginal effect of EDU_15 on the probability that JOB_SAT=1 for an observation with probability of JOB_SAT=1 equal to 0.5? And on the probability that JOB_SAT=4 for an observation with probability of JOB_SAT=4 equal to 0.5?