Problem Set 3

November 24, 2020

Rules

- Solutions to problem sets must be completed and submitted individually.
- The deadline for submission is **23.59 on 6 December 2020**. Submissions after the deadline will not be accepted.
- The datasets for the STATA exercises can be found at SUCourse.
- **Submissions must include two files: 1) one word file including your answers 2) one STATA log file.** Results in the log file and interpreted results must match. Otherwise, you will get zero points for those exercises. **Do not upload zipped files.**
- You are expected to answer all questions very clearly. Therefore, please be careful about use of language and writing. **For instance; if you are asked to interpret a coefficient, you need to interpret it in a way that someone who cannot see the data and does not know econometrics can understand what you mean.**
- Double check that you write your name/surname and student ID number.
- Failure in fulfilling any of these will result in a FAIL grade for that homework.

Impact of restaurant smoking restrictions on smoking rate (55 points)

In this problem set, you are going to use SMOKE.DTA data. In this data set, you have information on individuals' smoking behaviour and some other individual and locational characteristics for a random sample of single adults from the United States.

1) (10 points) What is the share of people in the sample who smokes (you need to generate a **new binary variable** indicating whether the person is "smoking")? What is the share of people who resides in a state with restaurant smoking restrictions? What is the difference in average smoking probability between states with restaurant smoking restrictions and states without restrictions?

2) (10 points) Using the information available in the data set, estimate a linear probability model that examines the determinants of smoking probability (note that the dependent variable will be a binary variable). Include **all potentially relevant variables** that might affect smoking behavior (you need to decide which variables might affect smoking behavior). Interpret the signs and magnitudes of the coefficients that are significant at 10% significance level.

3) (10 points) Estimating a new specification, check whether the effect of age on smoking probability is quadratic or not (assuming that you included "age" in question 2 in linear form). At what age does the impact of age on smoking probability becomes negative?

4) (5 points) Now, estimate your model including the **income** and **cigarette prices** in logarithmic form instead of level form (assuming that you included them in question 2 in level form). Interpret the signs and magnitudes of the estimated coefficients for these two variables.

5) (10 points) You might think that the impact of restaurant smoking restrictions on smoking probability might be different for white and non-white individuals. Using an interaction variable approach, test whether this hypothesis is true at 10% significance level.

6) (10 points) Can we consider the coefficient for *restaurn* as the causal effect of restaurant smoking restrictions on individuals' smoking probability? Discuss whether there might be an endogeneity problem here. Provide an example of omitted state specific factor that could lead to a bias in the estimated effect of smoking restrictions and discuss the direction of the potential bias that might arise because of this omitted factor.

Impact of Job Training Grant (25 points)

Use the data from JTRAIN.DTA for this exercise.

7) (10 points) Consider the simple regression model

$$log(scrap) = \beta_0 + \beta_1 grant + u,$$

where *scrap* is the firm scrap rate and *grant* is a dummy variable indicating whether a firm received a job training grant. Can you think of some reasons why the unobserved factors in u might be correlated with *grant*? Provide example.

8) (5 points) Estimate the simple regression model using the data **for only 1988** (You should have 54 observations.) Does receiving a job training grant significantly lower a firm's scrap rate?

9) (10 points) Now, add an additional explanatory variable (to the model in question 8) indicating the log scrap rate of the company in year 1987 ($lscrap_1$). Interpret the coefficient on *grant*. Is it statistically significant at the 10% significance level? How do you explain the change in the coefficient of grant between two models (from question 8 to 9)?

Marijuana usage and Wage (20 points)

Suppose you collect data from a survey on wages, education, experience, and gender. In addition, you ask for information about marijuana usage. The original question in the survey is: "On how many separate occasions last month did you smoke marijuana?"

10) (5 points) Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, "Smoking marijuana five more times per month is estimated to change wage by x%."

11) (5 points) Write a model that would allow you to test whether marijuana usage has different effects on wages for men and women. How would you test that there are no differences in the effects of marijuana usage on wage for men and women?

12) (5 points) Suppose you think it is better to measure marijuana usage by putting people into one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and

heavy user (more than 10 times per month). Now, write a model that allows you to estimate the effects of marijuana usage on wage by using this categorical variable.

13) (5 points) Discuss whether it is possible to estimate the causal effect of marijuana usage on wage based on this survey data? What might be the problem in identifying the causal effect here? Provide an example.