

## SOC 910: Assignment 2

Using `CPS2020.dta` or a dataset of your choice, estimate a series of OLS regression models as follows and interpret the results. You can download the data set from our course website.

You need to develop at least three hypotheses and three models.

1. For the first hypothesis, try to develop a question for which the dependent variable is log transformed, but the main independent variable is not. Here you compare Base Model and Full Model. Your task is to measure how the coefficient of the main independent variable,  $X$ , ( $= \beta$ ) changes as you control for an additional set of control variables ( $= Z$ ) in Full Model. Don't forget to discuss fitness of statistics (i.e., F-statistic and R-squared).

- Base Model

$$\ln y = \alpha_0 + \beta X + \sum \gamma C + e \quad (1)$$

- Full Model

$$\ln y = \alpha_0 + \beta X + \sum \gamma C + \sum \theta Z + e \quad (2)$$

2. For the second hypothesis, develop a question for which you utilize log-transformed variables for both dependent and independent variables. Your equation may look similar as follows:

$$\ln y = \alpha_0 + \beta \ln X + \sum \gamma Z + e \quad (3)$$

3. For the final hypothesis, you need to develop a question that test a quadratic term and its interaction terms with another variable. For example, we can test the following model:

$$\ln y = \alpha_0 + \beta_1 D_1 + \beta_2 age + \beta_3 age^2 + \beta_4 (age \times D_1) + \beta_5 (age^2 \times D_1) + \sum \gamma Z + e \quad (4)$$

where  $D_1$  refers to a group dummy (e.g., female==1 and male==0).

Your written assignment should have the following contents:

1. A brief description of your data.
2. A brief discussion of your models and hypotheses. I strongly recommend you to write equations.
3. One or two main regression tables and the discussion of the results.
4. One or two graphs which show the change in the net effect of your main independent variable after controlling for other covariates.

Due date: February 26th.