

## Homework 2: OLS and Probit

Due on Feb 4<sup>th</sup> 2022 at 11 PM

Your final output should consist of raw code and a pdf file with answers. Output should be uploaded on Github before the due date/hours. For this assignment,

This exercise follows Homework 1 using the French SRCV data. We will focus on the data from 2009 (datind2009) and use the following variables

- empstat: individual's participation in the labor market
- wage: wage of the individual
- age: age of the individual

### Exercise 1 OLS estimate

In this exercise, you are expected to use matrix operations to derive the solutions. DO NOT USE `lm()`!

Consider the following model

$$Y = X\beta + \epsilon \quad (1)$$

where  $X$  is the age of individuals plus intercept, and  $Y$  is the wage.

- Calculate the correlation between  $Y$  and  $X$ .
- Calculate the coefficients on this regression. remember,  $\hat{\beta} = (X^T X)^{-1} X^T Y$
- Calculate the standard errors of  $\beta$ 
  - Using the standard formulas of the OLS.
  - Using bootstrap with 49 and 499 replications respectively. Comment on the difference between the two strategies.

## Exercise 2      Detrend Data

Consider the same application as exercise 1 but using a pooled version of individual data from 2005 to 2018.

- Create a categorical variable  $ag$ , which bins the age variables into the following groups: "18-25", "26-30", "31-35", "36-40", "41-45", "46-50", "51-55", "56-60", and "60+".
- Plot the wage of each age group across years. Is there a trend?
- Consider  $Y_{it} = \beta X_{it} + \gamma_t + e_{it}$ . After including a time fixed effect, how do the estimated coefficients change?

## Exercise 3      Numerical Optimization

We are interested in the effect of age on labor market participation. We consider this problem using the data from 2007. Consider a probit model.

- Exclude all individuals who are inactive.
- Write a function that returns the likelihood of the probit of being employed.  
You might want to write  $X\beta$  first. Then, calculate  $F(X\beta)$  and the log likelihood.  
Remember, for the probit model,  $F(x)$  is the standard normal distribution function.
- Optimize the model and interpret the coefficients. You can use pre-programmed optimization packages.
- Can you estimate the same model including wages as a determinant of labor market participation? Explain.

## Exercise 4      Discrete choice

We are interested in the effect of age on labor market participation. Use the pooled version of the data from 2005 to 2015. **Additional controls include time-fixed effects.**

- Exclude all individuals who are inactive.
- Write and optimize the probit, logit, and the linear probability models.  
Remember, for the logit model,  $F(x)$  is the logistic function  $\frac{\exp(x)}{(1+\exp(x))}$
- Interpret and compare the estimated coefficients. How significant are they?

## Exercise 5      Marginal Effects

- Compute the marginal effect of the previous probit and logit models.
- Construct the standard errors of the marginal effects. *Hint:* Bootstrap may be the easiest way.