

### Notes:

- This exercise uses the datafile TrainExer45 and requires a computer.
- The dataset TrainExer45 is available on the website.

### Questions

In this exercise we reconsider the example of lecture 4.5. In this lecture we related the Grade Point Average [GPA] of learners in an engineering MOOC to the participation in a preparatory course. The dataset contains the following variables

- GPA: Grade Point Average in Engineering MOOC
  - Gender: 0/1 dummy for gender (1: male, 0: female)
  - Participation: 0/1 dummy for participation in a preparatory mathematics course (1: did participate, 0: did not participate)
  - Email: 0/1 dummy for receiving an email invitation to take the preparatory course (1: received invitation, 0: did not receive invitation)
- (a) Redo the OLS estimation of the coefficients in a model that explains GPA using a constant, gender and preparatory course participation. Also calculate standard errors and t-values. Confirm that you obtain the same results as mentioned in the lecture.
- (b) Use the email dummy as an instrument to perform 2SLS estimation. First do the first-stage regression

$$\text{Participation} = \gamma_1 + \gamma_2 \text{Gender} + \gamma_3 \text{Email} + \eta.$$

Next calculate the predicted values according to this regression and perform OLS on the model

$$\text{GPA} = \beta_1 + \beta_2 \text{Gender} + \beta_3 \widehat{\text{Participation}} + \varepsilon.$$

Confirm that the parameter estimates that you obtain are the same as reported in the lecture.

- (c) Obtain the standard errors that correspond to the final regression in the previous part. These do **not** match with the standard errors reported in the lecture! Why are the standard errors from part (b) wrong?
- (d) Calculate the ratio between the standard errors in part (b) and those reported in the lecture. Why is the obtained ratio the same for all parameters? Explain how we can also obtain this ratio using different residual series.