Econometrics-2022-2023
Home assignment 2. Multiple Regression Model. Variables Transformations
To be submitted by October 23, 23:55 using github classroom.

1. [30 marks] Let the regression equation be:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t \quad t = 1,\ 2,\ \ldots,\ T.$$

1.1. [10 marks] □ Outline how you would test the following restrictions a) $\beta_5 = 0$, b) $\beta_3 = -1$ (in turn).
□ What are the assumptions and clarify their role?

1.2. [10 marks] Outline how you would test the following restrictions $\beta_3 = 3$, $\beta_4 = 4$ and $\beta_5 = 5$ simultaneously.

1.3. [10 marks] □ Outline how you would test the following restriction $\beta_2 - \beta_3 = \beta_4 - \beta_5$ against $\beta_2 - \beta_3 \neq \beta_4 - \beta_4$.
□ Outline how you would test the restriction $\beta_2 - \beta_3 = \beta_4 - \beta_5$ against $\beta_2 - \beta_3 < \beta_4 - \beta_5$.

2. [20 marks] Two students of the course "Elements of Econometrics", say A and B, decided to investigate the effects of time spent studying on the examination marks earned by students at the final exam. They surveyed 100 students of the current 4th year who took internal econometrics exam at ICEF, asking them to indicate the number of hours H_i spent studying econometrics during the main course (regular study), and the number of hours R_i spent reviewing the material in the last month, specially allocated for revision. Variable S_i is defined as their sum $S_i = H_i + R_i$. The exam results Y_i are from 0 to 100).
Both estimated multiple regression equations used the same data but include different sets of expanatory variables: student A decides to regress Y_i on H_i and R_i and fitted the following regression (A):

$$\hat{Y}_i = 13.3 + 0.17H_i + 0.25R_i \qquad \text{(A)}$$

while student B used total hours of study S_i and additionally stressed on main course hours H_i - regression (B)

$$\hat{Y}_i = 13.3 + 0.25S_i - 0.08H_i \qquad \text{(B)}$$

The results seemed somewhat strange to them and they asked to one of the fourth-year students C for advice.

2.1. [10 marks] □ Help students give the correct interpretation to the coefficients of both regressions.
□ The Student C suggested that the coincidence of some coefficients in equations (A) and (B) is accidental and can be explained by too rough rounding of estimation results, which are actually different. Do you agree with him? Give reasons for your answer.
□ Student C also said that a negative sign of the coefficient for a variable H_i in equation (B) confirms his belief that classes during the academic year even prevent a good exam result, and that it is better to study properly only during the revision period. And if, instead of regression (B), one evaluates regression $\hat{Y}_i = b_1 + b_2 S_i + b_3 R_i$, the sign of the coefficient b_3 will be definitely positive. Comment on these statements.

2.2. [10 marks] Unfortunately, the students A and B did not include neither R^2 values nor standard errors of coefficients in the regression results.
□ The student A states that R^2 in regression (A) should be greater than in (B) because this equation is more meaningful: it includes variables that show hours of study at different periods, while in (B) they are mixed up due to inclusion of the variable S_i that partly duplicates H_i. Comment on this.
□ For the same reason, according to student A, the standard error of the coefficient of variable $S_i = H_i + R_i$ in equation (B) must be greater than the standard error of coefficient of R_i of in equation (A), although the coefficients themselves are equal. Comment on this.

3. [**50 marks]** For your data set **ha02_data__** estimate two simple linear regression models ***EARNINGS*** on ***S***; ***EARNINGS*** on ***WEXP,*** and then run multiple regression model ***EARNINGS*** on ***S*** and ***WEXP*** **(**refer to **ha02_data description.pdf).**

3.1. [10 marks] □ Give interpretation to the coefficients of all equations.
□ Explain why the values of the coefficients for the same variables in simple and multiple regressions differ and try to explain this difference. Do additional calculations if necessary.
□ Briefly describe the significance of estimated coefficients and equations.

3.2. [10 marks] □ The coefficient of ***S*** in multiple regression from 3.1 is approximately twice as big as coefficient of ***WEXP*** so the year devoted to study is twice more efficient for future earnings than the year devoted to working. Do you agree with him? If no explain, if yes how this can be tested?
□ Add to the multiple regression model new explanatory variables: ***TENURE*** and ***HOURS.*** Are these variables singnificant individually and as a group?
□ Repeat this with some other variables. Are there any signs of multicollinearity?.

3.3. [10 marks] Add to regression ***EARNINGS*** on ***S*** and ***WEXP,*** in turn quadratic terms ***S*^2** and ***WEXP*^2** and then repeat estimation.
□ Comment the meaning of two regressions and their coefficiens. How to measure now marginal effects of schooling and work experience?
□ Are these marginal effects significant?

3.4. [10 marks] Based on regression ***EARNINGS*** on ***S*** and ***WEXP,*** use logarithms to get new specifications
EARNINGS on **log*S*** and **log*WEXP***
log*EARNINGS* on ***S*** and ***WEXP***
log*EARNINGS* on **log*S*** and **log*WEXP***
□ Give interpretation to the coefficients of all equations.
□ For non-linear regressions, briefly explain why your interpretation is correct.

3.5. [10 marks] □ Which of the regressions in 3.4 are directly comparable in their statistical quality? On what indicators is the comparison made? What are the results of comparing your regressions?
□ How can one compare the quality of regressions that are not directly comparable? Perform the necessary procedure to convert the data to match the quality of the regression and run the necessary tests. What are your final conclusions?