The International College of Economics and Finance
Econometrics-2021-2022.
Home assignment 1. Simple Linear Regression Model.
To be submitted by September 26, 23:59 (following instructions in HA Rules)

1. [30 marks] The 3rd year student of HSE who have just started to study econometrics, is triyng to estimate linear regression model describing the work of food deliverers during the COVID-19 epidemic. The model

$$M_i = \beta K_i + u_i, \; i = 1, 2, ..., n.$$

discribes the relationship between the number of orders K_i completed by the deliverer per day and money M_i earned. The model does not include intercept as there is no constant part of earnings for delivery. Variable K_i is assumed nonstochastic, disturbance term has zero expectation, constant variance σ_u^2 with different values not correlated each other.
Being inexperienced in econometrics, he asked a fourth-year student who had already studied econometrics to tell him how to estimate this regression, and the latter advised him to use the well known OLS estimator

$$\beta_{OLS} = \frac{\text{Cov}(M, K)}{\text{Var}(K)}.$$

1.1. [10 marks] Comment on the idea of the fourth year student. Help a third year student derive the OLS estimator β_{OLS}^* for the model $M_i = \beta K_i + u_i$. Show that the resulting estimator β_{OLS}^* is linear and unbiased. Does β_{OLS} also possess these properties?

1.2. [10 marks] Obtain the expression for the variance of β_{OLS}^*. Show that the variance of the estimator β_{OLS} is generally speaking greater than the variance of estimator β_{OLS}^* (except for some special cases). What conclusions about the properties of the estimators under consideration can be drawn from here? Can this result be obtained from the Gauss-Markov theorem from the lecture?

1.3. [10 marks] Assume that under assumptions of model A the number of observations n of the sample tends to infinity. Use sufficient condition for consistency to show that the estimator β_{OLS}^* is consistent.

2. [30 marks]. A teacher of Economics estimated (OLS) a relationship between average score of home assignment X_i submitted during the year of study and the score obtained at the final exam Y_i for a sample of *n* students

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

2.1. [10 marks]. After performing the calculation, she noticed that she forgot to include one more student in the sample, but she did not want to recalculate everything and decided to simply enter a small correction up or down for each indicator. Tell the teacher how TSS would change if she did include an additional student in the sample. Give reasons for your answer.

2.2. [10 marks]. In the situation of 2.1 explain the teacher how RSS, R^2, *F*-statistic and *t*-statistic for the coefficient β_2 would change if she did include an additional student in the sample. Give reasons for your answer.

Computer Practice. *Consult* ***'Dustribution of files for all HA'*** *to find your personal file*

3. **[50 marks]** For your data set **ha01_data__** use simple linear regression models to investigate whether earnings of respondents (variable ***EARNINGS***) are significantly dependent on the following variables - ***AGE, ASVABC, HIGHT, HOURS, S, SIBLIBGS, TENURE, WEIGHT, WEXP*** **(**refer to **ha01_data description.pdf).** See detailed questions below.

3.1. [10 marks] **Preliminary analysis of the relationship between earnings and tenure.**

□ Consider data on ***EARNINGS*** and ***TENURE.*** Construct a scatter-diagram for the data under consideration without a regression line and comment on it. Add regression line on the scatter diagram and give additional comments.

□ Build a residual graph. What information can you get out of it?

□ Regress ***EARNINGS*** on ***TENURE*** **– regresssion (1)**. Present the resulting equation in a standard form (with standard errors under each coefficient and specifying the value of R-squared).

□ Comment on the obtained value of R-square?

□ Give interpretation of coefficients of estimated equation.

□ Estimate the significance of the coefficients and the equation as a whole, based on p-values from the regression printout. Explain the logic of analysis.

3.2. [10 marks] **Detailed analysis of the statistical quality of the regression (1).**

□ Perform statistical tests on significance of regression coefficients using *t*-statistics. Explain the logic of t-test (pair of hypotheses, statistic, critical values, conclusion).

□ Construct confidence intervals for the regression coefficients. Use these confidence intervals to find whether regression coefficients are significant.

□ Perform statistical test on significance of regression using *F*-statistics. Comment on.

□ Compare the conclusions obtained with analysis results in 3.1.

3.3. [10 marks] **Analysis of the relationship between earnings and schooling.**

□ Regress ***EARNINGS*** on ***S*** **– regression (2).** Give interpretations to the coefficients of estimated equation.

□ Generate new variable ***H* = *S*–12**, explain its meaning, regress ***EARNINGS*** on ***H*** **– regression (3)** and give interpretations to the coefficients of estimated equation.

□ Compare regressions **(2)** and **(3)** on the base of their printouts: explain mathematically why some values are the same while some others are different.

3.4. [10 marks] **Analysis of the relationship between earnings and other available variables.**

□ Regress *EARNINGS* in turn on ***ASVABC, HOURS, WEXP AGE, SIBLIBGS, HIGHT, WEIGHT*** **(only simple linear regressions both with and without intercept).** Present and discuss briefly **only two or three of them** that seem most interesting to you because they demonstate some strange or anusual econometric phenomena. Suggest some explanation to the observed phenomena.

3.5. [10 marks] **Relationships between the elements of regression output table.**

□ In the printout table of the **regression (1)** evaluated in 3.1 all data was deleted except b_1, b_2, TSS, RSS, $\overline{Y}$ and n? Can you restrore the values of $s.e.(b_1)$, $s.e.(b_2)$, *t*-statistic, *F*-statistic and R^2 on the base of available data?