ISEG – Lisbon School of Economics and Management
ECONOMETRICS
First Semester 2016/2017
Problem Set I – Estimation and Inference (I)

Question:	**1**	**2**	**3**	**4**	**5**	Total
Points:	4	4	4	4	34	50

Justify all your answers. You are required to show your work on each problem on this exam. **Organize your work**. Work scattered all over the page will receive very little credit. A correct answer in a multiple choice question worths 4 points; an incorrect one worths -1 point.

(4) **1**. Consider the following linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$ and the corresponding OLS residuals, $\hat{u}_i$. Assume that the assumptions MLR.1 to MLR.5 hold. Which of the following statements is **true**?

- ◯ The OLS residuals, $\hat{u}_i$, are the same as the random variable u_i, in the observed sample.
- ◯ The random variable u_i is not observed, even though it exists a random sample for the dependent variable y_i and the explanatory variables $x_{i1}, x_{i2}, \ldots, x_{ik}$.
- ◯ When the constant term, β_0, is included in the model, the sum of squared residuals (SSR) is equal to zero.
- ◯ Although the random variable u_i is observed, we prefer to consider the OLS residuals $\hat{u}_i$, once they are obtained substituting $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ by the corresponding OLS estimates, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$.

(4) **2**. Which of the following statements is **false**?

- ◯ The R-squared, R^2, increases as the sum of squared residuals (SSR) reduces for a given dependent variable.
- ◯ Adding an irrelevant variable to the model may result in a reduction of the adjusted R-squared, $\bar{R}^2$.
- ◯ The R^2 and the $\bar{R}^2$ are equal if the dependent variable is logarithmic.
- ◯ The R^2 should not be used to compare models where the number of explanatory variables is not the same.

(4) **3**. Suppose that assumptions MLR.1 to MLR.4 hold. For the OLS estimator, choose the **correct** statement:

- ◯ Under heteroscedasticity, the OLS estimator is still the Best Linear Unbiased Estimator (BLUE).
- ◯ Can be obtained as the minimization of the explained sum of squares (SSE).
- ◯ Is not applicable to models with a logarithmic dependent variable.
- ◯ Under heteroscedasticity, the OLS estimator is unbiased.

(4) **4**. The omission of a relevant variable in a given model:

- ◯ Can be detected through a t test of individual significance.
- ◯ Implies that assumption MLR.4 does not hold, if the omitted variable is correlated with at least one explanatory variable included in the model.
- ◯ Implies that the OLS estimator is necessarily biased.
- ◯ Increases the variance of the OLS estimates.

5. Using the data set CollegeDistance.WF1, you will investigate the effects of community colleges on educational attainment.

(5) (a) Estimate the following regression by OLS:

$$ED_i = \beta_0 + \beta_1 \log(dist_i) + \beta_2 bytest_i + \beta_3 tuition_i + u_i$$

where *ED* are the years of completed education, *dist* is the distance to the nearest college (in miles), *bytest* is a base year test score and tuition is the average state college *tuition* (in thousand dollars). Write the estimated equation with the corresponding standard errors.

(4) (b) Interpret the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$. Discuss the signs of these estimates.

(5) (c) Rachel lives 5 miles distant from the nearest college in a state with average college tuition of 1 thousand dollars and had a 50 points score in the base year test. How many years of completed education do you predict for Rachel? If her completed education is 10 years, what can you conclude about Rachel's education compared to the average of the population with her characteristics?

(2.5) (d) Do the explanatory variables explain a large fraction of the variance in the years of completed education across individuals? Explain.

(5) (e) Construct a 99% confidence interval for β_3. What can you conclude? Can it be statistically equal to zero?

(5) (f) Test the individual significance of β_1 against a one-sided alternative at 5% significance level. What can you conclude? Can it be statistically equal to zero?

(2.5) (g) Repeat part (f) against a two-sided alternative at 10% significance level. Will you change your answer?

(2.5) (h) What is the estimated variance of the error term?

(2.5) (i) Suppose that we suspect that *bytest* is correlated with the error term. Will the OLS estimates be biased?