EC 508, Econometrics *Jean-Jacques Forneron*
Boston University

Midterm

due Friday March 17, 2023 – 9:30am

Problem 1: True or False? Explain (40 points)

State whether the following statements are True or False, provide a brief explanation for your answer (2 sentences maximum). The questions will refer to the following regression model:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + u_i$$

i. OLS with multiple regressors always provides a coefficient with a causal interpretation.

ii. The t and Wald tests require that the u_i are independent and $\mathcal{N}(0,1)$ distributed.

iii. We can use the t-test with $k = 2$, $n = 8$ and $c_{0.95} = 1.96$.

iv. The adjusted-R^2 always increases with the number of regressors k.

v. $\hat{\beta}$ is consistent even if there are outliers (infinite fourth moments).

vi. $\hat{\beta}$ is not consistent if the regressors are perfectly collinear.

vii. $\hat{\beta}$ is biased if the regressors are close to, but not exactly, perfectly collinear.

viii. $\hat{\beta}$ is unbiased even if there are outliers (infinite fourth moments).

Problem 2: Determinants of Housing Prices (40 points)

For a sample of $n = 506$ communities in the Boston area, we estimate two models relating median housing price (**price**) in the community to various community characteristics: **nox** is the amount of nitrous oxide in the air, in parts per million; **dist** is a weighted distance of the community from five employment centers, in miles; **rooms** is the average number of rooms in houses in the community; **stratio** is the average student-teacher ratio of schools in the community; **crime** measures crimes committed per capita; **proptax** is the property tax per $1,000. The estimated regression models are reported below, standard errors are in parentheses:

Model 1:

$$price_i = \beta_0 + \beta_1 \times nox_i + u_i$$

Model 2:

$$price_i = \beta_0 + \beta_1 \times nox_i + \beta_2 \times rooms_i + \beta_3 \times dist_i + \beta_4 \times crime_i + \beta_5 \times proptax_i + u_i$$

i. Load the dataset *hprice2.dta* in R. Create a scatterplot with *nox* on the x-axis, and *price* on the y-axis. Do you think LSA 1 and LSA 3 hold for Model 1?

ii. Regardless of your previous answer, estimate Model 1 in R and report your estimates and heteroskedasticity-robust (HC) standard errors. Is the effect of *nox* positive or negative? Is this what you would expect?

iii. Is the effect of *nox* on *price* statistically significant at the 5% significance level? How about the 1% significance level? Report a 95% confidence interval for β_1.

iv. Should you be concerned about potential omitted variables in Model 1? Why and how would it affect your estimates (do you over/under-estimate β_1)?

v. Estimate Model 2, report the estimates and the heteroskedasticity-robust standard errors. Report a 95% confidence interval for β_1. Is the estimated effect of *nox* larger/smaller than you had previously found in ii-iii?

vi. Compute and report the R^2 and $\bar{R}^2$ for Models 1 and 2. Based on these measures, which specification do you prefer, why?

vii. The average price in the data is $\$22,000$, the standard deviation of price is $\$9,000$. Using Model 2, determine the effect of decreasing the nox level from 5.5 (this is the sample average) to 3.5 (the smallest value in the sample). Report a 95% confidence interval for your estimate. Discuss the statistical significance of the effect and its magnitude.

viii. Using Model 2, determine the effect of decreasing the level of crime from 3.5 (this is the sample average) to 0 (the smallest value in the sample). Report a 95% confidence interval for your estimate. Discuss the statistical significance of the effect but also its magnitude. From the perspective of housing price only, would you prefer a policy which reduces *crime* or pollution (*nox*)?

ix. The mayor is planning measures to reduce nox emission from 5.5 to 3.5 and to finance these with an increase in property tax from 40 (sample average) to 70 (maximum in the sample). What is the expected effect of the policy change on housing prices?

x. We want to test the null hypothesis that the change computed in v. is zero. Write down H_0 the usual way and in matrix notation (explicitly write down R and c). What is q (number of restrictions)?

xi. Compute the Wald statistic for the test in the previous question.[1] Can you reject H_0 at the 95% confidence level?

xii. What is the largest level in the table at which you cannot reject H_0? What does this suggest about the p-value?

Table 1: Quantiles of the χ^2_q distribution

Quantiles / q	1	2	3	4	5
0.1	0.0	0.2	0.6	1.1	1.6
0.3	0.1	0.7	1.4	2.2	3.0
0.5	0.5	1.4	2.4	3.4	4.4
0.8	1.6	3.2	4.6	6.0	7.3
0.95	3.8	6.0	7.8	9.5	11.1
0.99	6.6	9.2	11.3	13.3	15.1

Remark: rows correspond to the quantiles $(1-\alpha)$ and the columns to the degrees of freedom (q)

[1] Hint: if you are not sure what to do, look at the sample R code for lecture 10.