Homework 2

Advanced Econometrics

Due on Saturday 8th April (before 11:59pm)

Answer the questions below in whatever format you prefer (paper or electronic) as long it contains the code and the answer to all questions. Submit your work by mail to Laura Magazzini (laura.magazzini@santannapisa.it), Sebastiano Michele Zema (sebastianomichele.zema@santannapisa.it), and Tommaso Rughi (tommaso.rughi@santannapisa.it). For those questions requiring use of the software R, be sure that we can **replicate your results** (set the seed equal to your University ID).

Question 1

Consider the following linear model

$$y_i = \beta_1 + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i = \mathbf{x}_i \beta + \varepsilon_i \tag{1}$$

where $\varepsilon_i = \gamma_1 + \gamma_2 q_i x_{2i} + u_i$, q_i is an unobservable variable, $E(u_i|\mathbf{x}_i, q_i) = 0$ and $E(u_i^2|x_i) = \sigma^2$.

1.a) What is the partial effect of x_{2i} on $E(y_i|\mathbf{x}_i, q_i)$? What is the average partial effect?

1.b) Assume that $\gamma_2 = 0$. Is the OLS estimator unbiased and/or consistent for the partial effect you derived in part 1.a)?

1.c) Assume that $\gamma_2 \neq 0$ and $E(q_i|\mathbf{x}_i) = \kappa$, for some constant $\kappa \neq 0$ (to fix the ideas set $\kappa = 3$): what is OLS regression of y_i on $\mathbf{x}_i$ estimating under these assumptions?

1.d) What if in 1.c) we set $\kappa = 0$?

1.e) Suppose $E(u_i|\mathbf{x}_i) = f(\mathbf{x}_i)$ for some non constant function f, but $E(\mathbf{x}_i u_i) = 0$. Are your answers to 1.b)-1.e) still valid?

Question 2

In this question you have to write code to assess the effect of different distributions for the error in the regression on the quality of the CLT approximation. Start simulating the following linear model

$$y_i = \beta_1 + \beta_2 x_i + u_i \tag{2}$$

under the following three scenarios for the distribution of u_i and x_i (u_i are normalized to have mean zero and variance 1):

A. $u_i \sim N(0, 1)$, $x_i \sim N(0, 1)$

B. $u_i \sim (\chi^2_3 - 3)/\sqrt{6}$, $x_i \sim N(0, 1)$

where χ^2_3 is the chi-squared distribution with 3 degrees of freedom.
For each scenario simulate $\{y_i, x_i\}$, $i = 1, ..., n$ for $R = 5000$ times. For each simulation calculate the OLS estimators of β_1 and β_2 and their variances (under the assumption of homoskedasticity) and save these quantities. For the simulation you can set $\beta_1 = \beta_2 = 1$.
Propose a way to quantify the quality of the asymptotic approximation when $n = 100$ and $n = 250$ for the three scenarios (*Hint*: You have different options: (i) you can calculate the asymptotic variance of $(\hat{\beta}_1, \hat{\beta}_2)$; (ii) you know that a 95% CI should cover β_1 in 95% of the samples; (iii) explore relevant moment(s) of the distribution of $\hat{\beta}$).
Now consider the following distribution of u_i and x_i:

C. $x_i \sim N(0, 1)$, and $u_i \sim N(0, x_i^2)$

Simulate $\{y_i, x_i\}$, $i = 1, ..., n$ for $R = 5000$ times; and for each simulation calculate the OLS estimators of β_2 and its variance, both under the assumption of homoskedasticity and heteroskedasticity, and save these quantities. For the simulations, set $\beta_1 = \beta_2 = 1$ and $n = 250$.
Explore the 95% CI coverage with the two ways of computing the variances (under homoskedasticity and heteroskedasticity).
Write a general discussion of the results.