Econometrics 1
TD 3: inference, Monte-Carlo simulations

Updated notation to be closer to the course's one

This version: October 20$^{\text{th}}$, 2022

We consider the model (i.e., the data-generating process we use to generate the data in the simulations)

$$Y = \delta_{01} + \delta_{02} D + (\gamma_{01} + \gamma_{02} D) U,$$

where Y, D, and U are real random variables with D and U independent, $D \sim \mathcal{N}(0,1)$ and U is distributed according to a continuous Uniform distribution on $[-a, a]$ while δ_{01}, δ_{02}, γ_{01}, γ_{02}, and a are real parameters. Unless otherwise specified, we set $\delta_{01} = \delta_{02} = \gamma_{01} = \gamma_{02} = a = 1$.
We denote by $(\widehat{\beta}_1, \widehat{\beta}_2)$, or $(\widehat{\beta}_{1,n}, \widehat{\beta}_{2,n})$ to stress the dependence on the sample $(Y_i, D_i)_{i=1,\dots,n}$ of size n, the Ordinary Least Squares (OLS) estimators of $(\delta_{01}, \delta_{02})$ in the linear regression of Y on D and a constant.[1]

1. Simulate data from the model, then estimate the linear regression of Y on D and a constant for samples of different sizes: $n = 20, 100, 10^3, 10^5$. What do you remark for $(\widehat{\beta}_{1,n}, \widehat{\beta}_{2,n})$? Which property of the OLS estimator do we illustrate here?

2. Same question[2] when we draw U in a Cauchy distribution, that is, with a density with respect to Lebesgue's measure equal to $f \colon \mathbb{R} \longrightarrow \mathbb{R}_+$, $f(x) := \frac{1}{\pi(1+x^2)}$.

3. Henceforth, we take $U \sim$ Uniform again. Draw $S = 10^3$ samples of size $n = 20$ and stock the obtained OLS estimates of the slope $(\widehat{\beta}_{2,n}^{(1)}, \dots, \widehat{\beta}_{2,n}^{(S)})$, then plot the empirical density of those estimates. Do the same for samples of size $n = 1,000$, and compare the results. What do you remark? Which property of the OLS estimator do we illustrate here?

[1] In the simulations, we know the realizations of U since we will use them to generate the data. But, in the OLS estimation, we do as if U is an unobserved variable, hence we can only perform the linear regression of Y on D and a constant.

[2] Additional question 2.bis: same question when U is distributed according to a non-centered random variable with finite second-order moment, for instance $U \sim \mathcal{N}(\mu, 1)$. Compare the results when $\mu = 0$ and when $\mu \neq 0$.

4. In this question, we assume $\gamma_{02} = 0$. Fix a sample size n. Compare the empirical density of $\widehat{\beta}_{2,n}$ when $D \sim \mathcal{N}(0,1)$ and when $D \sim \mathcal{N}(0,4)$. Likewise, when $D \sim \mathcal{N}(0,1)$, compare the two cases: $a = 1$ and $a = 2$. In particular, comment on the variation of $\mathrm{V}(\widehat{\beta}_{2,n})$, the variance of $\widehat{\beta}_{2,n}$ Did you expect those results?

5. We consider the simple bilateral test of the null hypothesis $H_0 : \delta_{02} = 1$ against the alternative $H_1 : \delta_{02} \neq 1$. To do so, we examine two different test statistics:

   - on the one hand, the one based on the estimator of the asymptotic variance robust to heteroscedasticity,
   - on the other, the one based on the estimator of the asymptotic variance assuming homoscedasticity.

   For each of S samples of respective sizes $n = 20$ and $n = 1,000$, compute and stock the p-values for those two test statistics.

   For the two tests and the two sample sizes, what is the probability of rejecting the null hypothesis for a nominal level $\alpha = 1\%$? Same question at 5% and 10%. Did you expect the results?

6. Same question for the joint/multiple test of null hypothesis $H_0 : \delta_{01} = \delta_{02} = 1$ against the alternative $H_1 : \delta_{01} \neq 1$ or $\delta_{02} \neq 1$.

7. In the previous two questions, what do you obtain if we assume $\gamma_{02} = 0$ instead of $\gamma_{02} = 1$?