Escola de Economia de São Paulo - Fundação Getulio Vargas

Course: Microeconometria 1

Instructor: Bruno Ferman and Vitor Possebom

Problem Set: Lecture 11

Total = 190 points

Question 1 (Lecture 11: Nonparametric and Semiparametric Regressions - 190 points)

Using only one simulated dataset, we will analyze the performance of nonparametric and semiparametric estimators.[1]

We will discuss the performance of our estimators using the following data generating processes:

$\epsilon_1 \sim N(0,1)$

$\epsilon_2 \sim N(0,1)$

$X = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \right)$

$X_1 = \Phi\left(\tilde{X}_1\right)$, *where* $\Phi(\cdot)$ *is the CDF of the Standard Normal Distribution*

$X_2 = \Phi\left(\tilde{X}_2\right)$, *where* $\Phi(\cdot)$ *is the CDF of the Standard Normal Distribution*

$Y_1 = f(X_1) + \epsilon_1 = \sin(\beta_1 \cdot X_1) + \epsilon_1$, *where* $f(x) = \sin(\beta_1 \cdot x)$

$Y_2 = g(X_1, X_2) = f(X_1) + \beta_2 \cdot X_2 + \epsilon_2 = \sin(\beta_1 \cdot X_1) + \beta_2 \cdot X_2 + \epsilon_2$

$Sample = \{X_1, X_2, Y_1, Y_2\}_{i=1}^{N}$, *where* $N = 10,000$

Our target parameters are the function $f(\cdot)$ *and the coefficient* β_2*, where we impose that* $\beta_1 = 4$ *and* $\beta_2 = 2$*.*

1. *(10 points) Simulate one dataset according to the data generating process described above.*

2. *(10 points) Plot the function* $f(\cdot)$ *in the support of* X_1*.*

3. *(20 points) Using data on* X_1 *and* Y_1*, estimate the function* f *using a nonparametric local-linear regression.*

[1] For fun, you can create a Monte Carlo Simulation based on this question.

4. *(30 points) Plot the estimated function f and the true function f in the same graph. Include the bias-corrected 95%-confidence interval in the same graph. Is the estimated function f close to the true function f? Should these two objects be close to each other?*

5. *(20 points) Using data on X_1 and Y_2 only, estimate the function f using a nonparametric local-linear regression.*

6. *(30 points) Plot the estimated function f and the true function f in the same graph. Include the bias-corrected 95%-confidence interval in the same graph. Is the estimated function f close to the true function f? Should these two objects be close to each other?*

7. *(20 points) Using data on X_1, X_2 and Y_2, estimate the function f and the coefficient β_2 using a semiparametric local-linear regression.*

8. *(30 points) Plot the estimated function f and the true function f in the same graph. Include the bias-corrected 95%-confidence interval in the same graph. Is the estimated function f close to the true function f? Should these two objects be close to each other?*

9. *(20 points) What is the value of the estimated β_2? Is it close to the true β_2? Should these two objects be close to each other?*