Escola de Economia de São Paulo - Fundação Getulio Vargas

Course: Microeconometria 1

Instructor: Bruno Ferman and Vitor Possebom

Problem Set: Lectures 7-8

Total = 430 points

Question 1 (Lecture 7: Average Treatment Effect on the Untreated - 150 points)

Using the Inverse Probability Weighting approach, find an estimand that point-identifies the Average Treatment Effect on the Untreated,

$$ATU := \mathbb{E}\left[Y(1) - Y(0) \middle| D = 0\right],$$

and prove your result.

Question 2 (Lecture 8: CEF, IPW and DR Estimators - 240 points)

Using Monte Carlo Simulations ($M = 1,000$ repetitions), we will analyze the average relative bias of four estimators of the unconditional average treatment effect (ATE).

We will discuss the performance of our estimators in three different data generating processes.

DGP 1. Define:

$$U \sim Unif\left[0,1\right]$$

$$\epsilon \sim Unif\left[0,1\right]$$

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

$X_1 = \Phi\left(\tilde{X}_1\right)$, *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*

$$D = \mathbf{1}\left\{\frac{X_1}{2} + \frac{(X_1)^2}{2} \geq U\right\}$$

$$Y(0) = 1 + X_1 + \epsilon$$

$$Y(1) = Y(0) + 2 + X_1$$

$$Y = Y(1) \cdot D + Y(0) \cdot (1 - D)$$

$$Sample = \{D, X_1, X_2, Y\}_{i=1}^{N},\ where\ N = 10,000$$

DGP 2. Define:

$$U \sim Unif\,[0,1]$$

$$\epsilon \sim Unif\,[0,1]$$

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

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

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

$$D = \mathbf{1}\left\{\frac{X_1}{2} + \frac{X_2}{2} \geq U\right\}$$

$$Y(0) = 1 + X_1 + X_2 + \epsilon$$

$$Y(1) = Y(0) + 2 + X_1 + X_2$$

$$Y = Y(1) \cdot D + Y(0) \cdot (1 - D)$$

$$Sample = \{D, X_1, X_2, Y\}_{i=1}^{N},\ where\ N = 10,000$$

DGP 3. Define:

$$U \sim Unif\,[0,1]$$

$$\epsilon \sim Unif\,[0,1]$$

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

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

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

$$D = \mathbf{1}\left\{\frac{X_1}{2} + \frac{(X_1)^2}{2} \geq U\right\}$$

$$Y(0) = 1 + X_1 + X_2 + \epsilon$$

$$Y(1) = Y(0) + 2 + X_1 + X_2$$

$$Y = Y(1) \cdot D + Y(0) \cdot (1 - D)$$

$$Sample = \{D, X_1, X_2, Y\}_{i=1}^{N}, \text{ where } N = 10{,}000$$

Our four estimators are described below.

1. *Conditional Expectation Function estimator: Without using X_2, estimate $\mathbb{E}[Y|X_1]$ using OLS and estimate the ATE based on this regression.*

2. *Inverse Probability Weighting estimator: Using all observable variables, estimate the propensity score function and use it to estimate the ATE with an IPW estimator.*

3. *Doubly Robust estimator: Combining your first estimator and your second estimator, use a doubly robust estimator to estimate the ATE. Importantly, the CEF part of this estimator cannot include X_2.*

4. *Naive comparison of means between the treatment and control groups.*

The average relative bias of estimator j is given by

$$AVR := \mathbb{E}\left[\frac{\widehat{ATE}_j - ATE}{ATE}\right].$$

Discuss which estimators should be consistent in each DGP. Does your Monte Carlo Simulation confirm your consistency arguments?

Question 3 (Lecture 8: Quantile Regression - 40 points)

To illustrate quantile regressions, we will use Engel's (1857) original dataset. Engel collected data consisting of 235 observations on income and expenditure on food for Belgian working class households. There are two variables in this dataset:

- `income`: *annual household income in Belgian francs;*

- *`foodexp`: annual household food expenditure in Belgian francs.*

1. *Load the dataset using the command `data(engel)`.*

2. *Estimate a median regression of food expenditures on income and interpret the estimated coefficients.*

3. *Estimate quantile regressions for all ventiles. Plot all estimated intercepts in one graph and all estimated slopes in another graph.*