Assignment: Potential Outcome and Control for Observables

Due: April 4, 2022

1 Identification

1.1 Heuristic Identification

Recall the identification terminology and evaluate the following statements.

1. "We don't have enough sample size to identify the causal effects of the problem."
2. "We don't have a good identification strategy so I need to use a structural model."
3. "Because I have a structural model, I don't need to think about identification."
4. "Because I can use the maximum likelihood estimator, I can identify that."

1.2 Identification of OLS

Consider the model:

$$y_i = \beta x_i + \epsilon_i$$
$$\epsilon_i \perp X_i$$
$$\epsilon_i \sim N(0,1)$$

Show that β is identified with the definition we used during the lecture. (Hint: Think about OLS.)

1.3 Identification of a Factor Model

Consider the following model:

$$y_{it} = \nu_{it} + \epsilon_{it}$$
$$\nu_{it} = \rho \nu_{it-1} + \zeta_{it}$$

Think of y_{it} as income for individual i at time t. ϵ_{it} is the idiosyncratic income shock, e.g., lottery income. ν_{it} denotes the permanent component that follows an AR(1) process. For example, ν_{it} could be the unobserved productivity. The only observed data is $\{y_{it}\}$.

We assume that $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $\zeta_{it} \sim N(0, \sigma_\zeta^2)$. Both are assumed to be iid across individuals, over time, and of each other. We assume $0 < |\rho| < 1$ so that the process is stationary.

1. Show that ρ is identified. (Hint: What is y_{it-1}?)
2. Show that σ_ϵ^2 is identified.
3. Show that σ_ζ^2 is identified.
4. How do we estimate these parameters? Write down an estimator.

1.4 Simulation of MLE

$$y_i = \epsilon_i^1 + \epsilon_i^2$$
$$\epsilon_i^1 \sim N(0, \sigma_1^2)$$
$$\epsilon_i^2 \sim N(0, \sigma_2^2)$$
$$\epsilon_i^1 \perp \epsilon_i^2$$

The data is $\{y_i\}_{i=1}^N$. Assume y_i to be iid accross i.

1. Write down the likelihood function.
2. Let's draw just $N = 2$, use `optim` function to get maximum likelihood estimates.
3. Now stare at the model, can one separately identify σ_1^2 from σ_2^2? Show that it is identified or it is not identified.
4. How about $\sigma_1^2 + \sigma_2^2$? Show that it is identified or it is not identified.
5. Does the procedure in question 2 make sense?

2 Potential Outcome Framework

Recall our Roy model.

$$w_0 = \mu_0 + \epsilon_0$$
$$w_1 = \mu_1 + \epsilon_1$$

1. Use the exact potential outcome notations in class to write out the model.
2. What is $Y_i(0)$? What is $Y_i(1)$?
3. What is D_i?

3 Control for Observables

3.1 Rosenbaum and Rubin

Write down propensity score proof (Rosenbaum and Rubin). Remark at each equality, write down the reason why that is true. Convince yourself.

3.2 Propensity Score

$$w_0 = \mu_0 + \beta_1 X_1 + \epsilon_0$$
$$w_1 = \mu_1 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_1$$

1. Pick you favorite $\beta_1, \beta_2 \neq 0$. Go back to your simulation in the first homework and re-simulate the model.
2. What is an example of X_1? Is β_1 identified?
3. Define the propensity score using the notation set up here.
4. Derive the propensity score analytically.
5. Create a column in your simulated data for the estimated propensity score using the derived formula above.
6. Use logit to estimate the propensity score.

7. What is the correlation coefficient of the above two types of propensity scores?

8. Use both types of propensity score to conduct IPW estimates.

9. People regress w_i on X_i. Can you recover the parameters?

10. Now estimate by adding "control variables." Does that work?