Homework 1 for ECON 31340

T. Lamadon

September 23, 2021

In this problem we are going to look at the effect of changing the conditioning set in a regression framework. Consider the following model for some data (Y_i, X_i, T_i):

$$Y_i = \gamma T_i + \beta X_i + \epsilon_i$$

where T_i is a treatment of interest, X_i is a covariate, ϵ_i is a residual and $\beta \neq 0$. Assume further that T_i, X_i, ϵ_i are jointly normally distributed according to:

$$\begin{bmatrix} \epsilon_i \\ X_i \\ T_i \end{bmatrix} \sim \mathcal{N}\left(0, \Omega\right) \quad \text{with} \quad \Omega = \begin{bmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{bmatrix}$$

We are interested in the parameter γ that measures the treatment effect.

1. (15 points) What is the joint distribution of (T_i, ϵ_i) unconditional of X_i? Does this tell us that $T_i \perp \epsilon_i$? Does this mean that $(T_i \perp \epsilon_i)|X_i$? Explain.

First we are interested in the result of the regression of Y_i on T_i only, excluding X_i. This requires us to express $\mathbb{E}[Y_i \mid T_i]$. We are going to use the result that if some variables $X_{i,1}, X_{i,2}, X_{i,3}$ are joint normal then one can write each as linear combination of the two others. Hence there exist α_1, α_2 such that $X_{i,3} = \alpha_1 X_{i,1} + \alpha_2 X_{i,2} + u_i$ where u_i is a random variable, normally distributed and independent of $X_{i,1}, X_{i,2}$.

2. (15 points) Given that T_i and X_i are jointly normal, we can write $X_i = aT_i + u_i$ for some scalar a and a normal random variable u_i independent of T_i. Use the entries of the Ω matrix and express $\text{Cov}(T_i, X_i)$ as a function of a to show that $a = \rho_2$.

3. (15 points) Use the previous expression for X_i to show that $\mathbb{E}[Y_i \mid T_i] = \clubsuit \cdot T_i$. Report the expression for $\clubsuit$ as a function of the parameters γ, β, ρ_2.

4. (15 points) Under what condition does the regression coefficient $\clubsuit$ provide an unbiased estimate of γ? How can you interpret this condition as a condition on the relationship between X_i and T_i?

We are now interested in the result of the regression of Y_i on T_i, X_i jointly. This requires us to express $\mathbb{E}(Y_i|T_i, X_i)$.

5. (15 points) Given that T_i,X_i and ϵ_i are jointly normal, we can write $\epsilon_i = b \cdot T_i + c \cdot X_i + v_i$ for some scalars b, c and a normal random variable v_i independent of T_i, X_i. Similar to the previous part, show that using $Cov(X_i, \epsilon_i)$ and $Cov(T_i, \epsilon_i)$ together with the entry of Ω we can establish that:

$$0 = b + c\rho_2$$
$$\rho_1 = b\rho_2 + c$$

leading to the following expression $\mathbb{E}[\epsilon_i \mid T_i{=}t, X_i{=}x] = \frac{\rho_1}{1-\rho_2^2} \cdot x - \frac{\rho_1\rho_2}{1-\rho_2^2} \cdot t$ (which you are not asked to derive).

6. (10 points) We are then ready to find our regression expression of Y_i on T_i, X_i. To do so find the missing parts in the following equation:

$$\mathbb{E}[Y_i|X_i = x, T_i = t] = \blacksquare \cdot t + \blacklozenge \cdot x.$$

7. (15 points) Under what sufficient conditions does $\blacksquare$ provide an unbiased estimate of γ? How can you interpret this condition as a condition on the relationship between X_i and T_i or between ϵ_i and X_i.

8. (10 points) [Bonus] If a researcher then tells you "It is always better to condition on all possible observables available to you", in the light of the previous question, what would you respond? Is it possible for $\blacksquare$ to be more strongly biased than $\clubsuit$?