HUL315: Econometric Methods

Harshit Goyal: 2021MT10143

April 19, 2024

1 Observing Omitted Variable Bias

1.1 $\rho_{x_1,x_3} > 0$

$$x_1, x_2 \sim \mathcal{U}(-1, 1)$$
$$u, v \sim \mathcal{N}(0, 1)$$

$$x_3 = 0.60 + 0.95x_1 + 0.00x_2 + v$$
$$y = 3.00 + 2.00x_1 + 1.00x_2 + 1.50x_3 + u$$

Variables	ρ
x_1 x_2	-0.012
x_2 x_3	0.001
x_1 x_3	0.506

Table 1: Correlation coefficients

True Values		Regressors	
		x_1 x_2 x_3	x_1 x_2
β_0	3.00	3.034	3.984
β_1	2.00	1.911	3.278
β_2	1.00	0.909	0.927
β_3	1.50	1.451	-

Table 2: Regression coefficients

Table 3: x_3 positively correlated with x_1

In table Table 1 the $\rho_{x_1,x_3} \approx 0.5$ i.e. positively correlated. Hence, in Table 2, we see when, x_3 is not one of the regressors, the coefficient of x_1 i.e. β_1 increases significantly, while β_2 is almost unchanged, as $\rho_{x_2,x_3} \approx 0$.

1.2 $\rho_{x_1,x_3} < 0$

$$x_1, x_2 \sim \mathcal{U}(-1, 1)$$
$$u, v \sim \mathcal{N}(0, 1)$$

$$x_3 = 0.60 - 0.95x_1 + 0.00x_2 + v$$
$$y = 3.00 + 2.00x_1 + 1.00x_2 + 1.50x_3 + u$$

In table Table 4 the $\rho_{x_1,x_3} \approx -0.55$ i.e. negatively correlated. Hence, in Table 5, we see when, x_3 is not one of the regressors, the coefficient of x_1 i.e. β_1 decreases significantly, while β_2 is almost unchanged, as $\rho_{x_2,x_3} \approx 0$.

The class quiz problem is analogous to the data discussed in Table 6. We can treat the latent *ability* as x_3, the $log(wage)$ as y and *train* as x_1. It's given that $\rho_{train,x_3} < 0$ as workers with less *ability* are more likely to be selected. Also $\rho_{log(wage),ability} > 0, \rho_{log(wage),train} > 0$, we can see that NOT using *ability* as a regressor will decrease β_1 i.e. underestimate the effect of traning program.

Variables	ρ
x_1 x_2	-0.012
x_2 x_3	0.044
x_1 x_3	-0.546

Table 4: Correlation coefficients

True Values		Regressors	
		x_1 x_2 x_3	x_1 x_2
β_0	3.00	2.941	3.795
β_1	2.00	2.080	0.406
β_2	1.00	1.082	1.196
β_3	1.50	1.545	-

Table 5: Regression coefficients

Table 6: x_3 negatively correlated with x_1

1.3 Conclusion

With some other settings of data generation, I verified,

1. When $\rho_{x_1,x_3} > 0$, if $\rho_{y,x_1}, \rho_{y,x_3}$ have the same sign, removing x_3 will increase magnitude of β_1 but if $\rho_{y,x_1}, \rho_{y,x_3}$ have the opposite signs, removing x_3 will decrease magnitude of β_1.
2. When $\rho_{x_1,x_3} < 0$, if $\rho_{y,x_1}, \rho_{y,x_3}$ have the same sign, removing x_3 will decrease magnitude of β_1 but if $\rho_{y,x_1}, \rho_{y,x_3}$ have the opposite signs, removing x_3 will increase magnitude of β_1.

2 Proxy Variables to solve OVB

2.1 $E[x_3|z_3] = E[x_3|x_1, x_2, z_3]$ is satisfied

$$z_3 \sim \mathcal{U}(-1, 1)$$
$$\epsilon_1, \epsilon_2, \epsilon_3, u \sim \mathcal{N}(0, 1)$$

$$\begin{aligned} x_3 &= 1.40 + 0.95z_3 + \epsilon_3 \\ x_1 &= 1.10 + 0.40x_3 + \epsilon_1 \\ x_2 &= 1.50 + 0.60x_3 + \epsilon_2 \\ y &= 3.00 + 2.00x_1 + 1.00x_2 + 15.0x_3 + u \end{aligned}$$

Variables	ρ
x_3 z_3	0.481
x_1 x_2	0.229
x_2 x_3	0.566
x_1 x_3	0.413
x_2 z_3	0.263
x_1 z_3	0.195

Table 7: Correlation coefficients

True Values		Regressors		
		x_1 x_2 x_3	x_1 x_2 z_3	x_1 x_2
β_0	3.00	3.006	3.384	-0.272
β_1	2.00	1.995	5.940	6.647
β_2	1.00	0.995	6.988	8.028
β_3	15.0	15.006	9.440	-

Table 8: Regression coefficients

Table 9: Proxy variable condition is satisfied

In this case, the proxy condition is not satisfied. From table Table 8 we see, the coefficients are almost same as the true values, when x_1, x_2 and x_3 are used as regressors. Coefficients β_1 and β_2 are much larger from the true values when only x_1 and x_2 are regressors (as both $\rho_{x_1,x_3}, \rho_{x_2,x_3} > 0$, see Table 7). When a proxy for x_3 i.e. z_3 is used (Table 7 shows that $\rho_{x_3,z_3} = 0.481$), the coefficients β_1 and β_2 are closer to their true value.

2.2 $E[x_3|z_3] \neq E[x_3|x_1, x_2, z_3]$ satisfied

$$x_1, x_2, z_3 \sim \mathcal{U}(-1, 1)$$
$$\epsilon, u \sim \mathcal{N}(0, 1)$$

$$x_3 = 1.40 + 0.50x_1 + 0.70x_2 + 0.95z_3 + \epsilon$$
$$y = 3.00 + 2.00x_1 + 1.00x_2 + 15.0x_3 + u$$

Variables	ρ
x_3 z_3	0.442
x_1 x_2	0.007
x_2 x_3	0.323
x_1 x_3	0.234
x_2 z_3	-0.001
x_1 z_3	-0.000

Table 10: Correlation coefficients

True Values		Regressors		
		x_1 x_2 x_3	x_1 x_2 z_3	x_1 x_2
β_0	3.00	3.008	24.163	24.174
β_1	2.00	1.997	9.436	9.430
β_2	1.00	1.001	11.364	11.356
β_3	15.0	14.997	14.236	-

Table 11: Regression coefficients

Table 12: Proxy variable condition is not satisfied

From table Table 11 we see, the coefficients are almost same as the true values, when x_1, x_2 and x_3 are used as regressors. When x_1 x_2 are used the β_1 and β_2 increase as $\rho_{x_1,x_3}, \rho_{x_2,x_3} > 0$ (see Table 10). When, x_1 x_2 z_3 as regressors, there is no significant improvement in the coefficients obtained, i.e. using the proxy variable doesn't solve the bias on coefficients β_1 and β_2. This is because the proxy condition is not satisfied!