Econometrics 1

TD 8

We consider the (causal) model where $D = (D_1, D_2)'$ and

$$Y(d) = \zeta_0 + d_1\delta_{01} + d_2\delta_{02} + \eta \tag{1}$$

and where we suppose that $\text{cov}(D_2, \eta) = 0$ but *a priori*, $\text{cov}(D_1, \eta) \neq 0$.

It is very important that you understand the difference between the 3 following theoretical equations, that are at the core of the Econometrics 1 course. Prs. d'Haultfœuille and Lapenta always use the same notations : the δs always refer to causal effects while the βs always refer to coefficients of (theoretical) linear regressions :

$$Y(d) = \zeta_0 + d_1\delta_{01} + d_2\delta_{02} + \eta, \qquad (1)$$

Equation (1) corresponds to the causal model. Here, $d = (d_1, d_2))$ isn't a random variable, so it doesn't make sense to talk about $\text{cov}(d_1, \eta)$ or $\text{cov}(d_2, \eta)$. Actually, in this equation, only η is random. Besides, since there is the constant ζ_0 in the model, we have that $E(\eta) = 0$.

$$Y(D) = \zeta_0 + D_1\delta_{01} + D_2\delta_{02} + \eta, \qquad (2)$$

Equation (2) still corresponds to the causal model but this time "evaluated" in D, i.e. in the realization of the random variable D that we observe in the data. Here, it makes sense to talk about $\text{cov}(D_1, \eta)$ and $\text{cov}(D_2, \eta)$.
Equation (3) is simply the theoretical linear regression of Y on D :

$$Y(D) = \alpha_0 + D_1\beta_{01} + D_2\beta_{02} + \varepsilon, \qquad (3)$$

In this last equation, by construction (coming from the first order conditions related to the Least Squares minimization + the presence of a constant in the model), we have $E(\varepsilon) = 0$, $\text{cov}(D_1, \varepsilon) = 0$ and $\text{cov}(D_2, \varepsilon) = 0$.

In this exercise, we want to answer the following question : does $\beta_{02} = \delta_{02}$ when $\text{cov}(D_2, \eta) = 0$ but *a priori* $\text{cov}(D_1, \eta) \neq 0$?

1. The coefficient of D_2 in the multiple linear regression of Y on D (and, implicitly, a constant as always) can be obtained by the simple linear regression of Y on some random variable R. Quote the famous theorem behind that result and give the expression of R.

The coefficient of D_2 in the regression of Y on D, β_{02}, can be obtained by a simple linear regression using the Frish-Waugh theorem. Let R be the residual from the theoretical linear regression of D_2 on D_1, i.e.

$$R = D_2 - \lambda_0 - \lambda_1 D_1 \text{ where } \lambda_1 = \frac{Cov(D_2, D_1)}{V(D_1)} \text{ and } \lambda_0 = E[D_2] - \lambda_1 E[D_1]$$

Then,

$$\beta_{02} = \frac{Cov(Y, R)}{V(R)}$$

In other words, β_{02} is the coefficient of R in the simple regression of Y on R.

2. Deduce that, even if D_2 is exogenous, the regression of Y on D cannot recover the causal parameter δ_{02} in general. Explicit the bias term.

In other words, we have to show that $\beta_{02} \neq \delta_{02}$. From Question 1, we know that :

$$\beta_{02} = \frac{\text{cov}(R, Y)}{V(R)}$$

With

$$D_2 = \lambda_0 + D_1\lambda_1 + R \text{ and } \begin{cases} \text{cov}(D_1, R) = 0 & (*) \\ E[R] = 0 \end{cases}$$

So

$$\beta_{02} = \frac{\text{cov}(R, \zeta_0 + D_1\delta_{01} + D_2\delta_{02} + \eta)}{V(R)} = \frac{\text{cov}(R, D_2\delta_{02} + \eta)}{V(R)} = \underbrace{\frac{\text{cov}(R, D_2\delta_{02})}{V(R)}}_{(**)} + \underbrace{\frac{\text{cov}(R, \eta)}{V(R)}}_{(***)}$$

Let's first look at $(**)$. From $(*)$, we know that $Cov(R, D_1) = Cov(R, 1) = 0$. So

$$\begin{aligned} \text{cov}(R, D_2\delta_{02}) &= \text{cov}(R, D_2\delta_{02}) - \delta_{02}\lambda_1 \overbrace{\text{cov}(R, D_1)}^{=0} - \delta_{02}\lambda_0 \overbrace{\text{cov}(R, 1)}^{=0} \\ &= \text{cov}(R, \delta_{02}(D_2 - \lambda_0 - \lambda_1 D_1)) \\ &= \text{cov}(R, \delta_{02} R) \\ &= \delta_{02} V(R) \end{aligned}$$

And therefore, $(**) = \delta_{02}$. Now, let's look at $(***)$:

$$\begin{aligned}\text{cov}(R, \eta) &= \text{cov}(D_2 - \lambda_0 - \lambda_1 D_1, \eta) \\ &= \underbrace{\text{cov}(D_2, \eta)}_{=0 \text{ by hyp.}} - \lambda_0 \underbrace{\text{cov}(1, \eta)}_{=0} - \lambda_1 \text{cov}(D_1, \eta) \\ &= -\lambda_1 \text{cov}(D_1, \eta)\end{aligned}$$

To conclude, when plugging in the results we got for $(**)$ and $(***)$, we obtain

$$\beta_{02} = \delta_{02} - \lambda_1 \frac{\text{cov}(D_1, \eta)}{V(R)}$$

3. Discuss the sign of the bias as a function of the signs of $\text{cov}(D_1, \eta)$ and of $\text{cov}(D_1, D_2)$.

Since λ_1 is the coefficient of D_1 on the simple linear regression of D_2 on D_1, we can re-write the previous formula as follows :

$$\beta_{02} = \delta_{02} - \frac{\text{cov}(D_1, D_2)\,\text{cov}(D_1, \eta)}{V(D_1)V(R)}$$

Therefore, the sign of the bias depends on both $\text{cov}(D_1, \eta)$ and $\text{cov}(D_1, D_2)$. If these two quantities have the same sign, the bias will be negative. If they are of the opposite sign, then it will be positive

4. Discuss the hypothesis and the sign of the bias you can expect in the following situation : Y is the hourly wage, D_1 is the number of years of education, and D_2 is the maximum possible professional experience, often called the "potential experience", i.e. the number of years elapsed from the end of studies [1]

First of all, we may expect $\text{cov}(D_1, D_2)$ to be negative for two potential reasons :

(a) There is a "cohort" effect. There has recently been a massive spread of higher education : the average 25-year old individual from the 2022 "cohort" has had more years of education than the average 25-year old individual of the 1995 "cohort". Nevertheless, by construction, the older is the individual, the greater is his/her "potential experience". So a 52-year old individual from the 2022 "cohort" (who used to be 25 in

1. For an individual i, D_{i2} is thus his or her age now, when surveyed, minus the age at the end of his or her studies.

1995) has, on average, less years of education and a greater "potential experience", hence the negative correlation.

(b) There is a "choice" effect. At a fixed age (let's say 20), if you choose to study an extra year, you necessarily loose a year of "potential experience". Hence, once again, the negative correlation.

Second, we may expect a positive correlation between D_1 and η. Remember, η corresponds to the unobserved characteristics that causally impact the individual's wage. η includes, as an omitted variable, the individual's unobserved "ability". More "able" individuals will usually study longer. Moreover, for a fixed number of years of education, more "able" individuals will have a greater wage as they are usually more productive. Hence, $\text{cov}(D_1, \eta) > 0$.

Since $\text{cov}(D_1, D_2)$ and $\text{cov}(D_1, \eta)$ are of the opposite sign, we get a positive bias. We are going to over-estimate the effect of potential experience.

5. We now consider the case where D_1 is also exogenous in model (1), namely, $\text{cov}(D_1, \eta) = 0$, but there is some measurement error. We do not observe D_1 but only $\widetilde{D}_1 = D_1 + \nu$, with $\text{cov}(\nu, D_1) = \text{cov}(\nu, D_2) = \text{cov}(\nu, \eta) = 0$. Show that we obtain a similar result as the one of Question 2.

We consider the same causal model :

$$Y(D) = \zeta_0 + D_1\delta_{01} + D_2\delta_{02} + \eta$$

with, this time, $\text{cov}(D_1, \eta) = \text{cov}(D_2, \eta)$=0. However, we only observe $\widetilde{D}_1 = D_1 + \nu$, where $\text{cov}(\nu, D_1) = 0$. We can only implement the following regression :

$$Y = \widetilde{\alpha}_0 + \widetilde{D}_1\widetilde{\beta}_{01} + D_2\widetilde{\beta}_{02} + \widetilde{\varepsilon}$$

Let's show that $\widetilde{\beta}_{02} \neq \delta_{02}$ in such setting. As in Questions 1-2, let's apply the Frisch-Waugh Theorem :

$$\widetilde{\beta}_{02} = \frac{\text{cov}(\widetilde{R}, Y)}{V(\widetilde{R})}$$

where

$$D_2 = \widetilde{\lambda}_0 + \widetilde{\lambda}_1\widetilde{D}_1 + \widetilde{R} \text{ with } \begin{cases} \text{cov}(\widetilde{D}_1, \widetilde{R}) = 0 & (\Delta) \\ E\left[\widetilde{R}\right] = 0 \end{cases}$$

Or similarly

$$\widetilde{R} = D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 \widetilde{D}_1 = D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 D_1 - \widetilde{\lambda}_1 \nu \qquad (\Delta\Delta)$$

Then

$$\begin{aligned}
\widetilde{\beta}_{02} &= \frac{\text{cov}(\widetilde{R}, Y)}{V(\widetilde{R})} \\
&= \frac{\text{cov}(\widetilde{R}, \zeta_0 + D_1\delta_{01} + D_2\delta_{02} + \eta)}{V(\widetilde{R})} \\
&= \frac{\text{cov}(\widetilde{R}, D_2\delta_{02})}{V(\widetilde{R})} + \frac{\text{cov}(\widetilde{R}, (\widetilde{D}_1 - \nu)\delta_{01} + \eta)}{V(\widetilde{R})} \\
&= \delta_{02}\frac{\text{cov}(\widetilde{R}, D_2)}{V(\widetilde{R})} + \delta_{01}\frac{\overbrace{\text{cov}(\widetilde{R}, \widetilde{D}_1)}^{=0 \text{ by } (\Delta)}}{V(\widetilde{R})} - \delta_{01}\frac{\text{cov}(\widetilde{R}, \nu)}{V(\widetilde{R})} + \frac{\text{cov}(\widetilde{R}, \eta)}{V(\widetilde{R})} \\
&= \delta_{02}\frac{\text{cov}(\widetilde{R}, D_2)}{V(\widetilde{R})} - \delta_{01}\frac{\text{cov}(\widetilde{R}, \nu)}{V(\widetilde{R})} + \frac{\text{cov}(\widetilde{R}, \eta)}{V(\widetilde{R})} \qquad (\Delta\Delta\Delta)
\end{aligned}$$

First, since $\text{cov}(\widetilde{R}, \widetilde{D}_1) = \text{cov}(\widetilde{R}, 1) = 0$, then

$$\begin{aligned}
\text{cov}(\widetilde{R}, D_2) &= \text{cov}(\widetilde{R}, D_2) - \widetilde{\lambda}_0 \text{cov}(\widetilde{R}, 1) - \widetilde{\lambda}_1 \text{cov}(\widetilde{R}, \widetilde{D}_1) \\
&= \text{cov}(\widetilde{R}, D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 \widetilde{D}_1) \\
&= \text{cov}(\widetilde{R}, \widetilde{R}) \\
&= V(\widetilde{R})
\end{aligned}$$

Second,

$$\begin{aligned}
\text{cov}(\widetilde{R}, \nu) &= \text{cov}(D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 \widetilde{D}_1, \nu) \\
&= \text{cov}(D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 (D_1 + \nu), \nu) \\
&= \underbrace{\text{cov}(D_2, \nu)}_{=0 \text{ by hyp.}} - \widetilde{\lambda}_0 \underbrace{\text{cov}(1, \nu)}_{=0} - \widetilde{\lambda}_1 \underbrace{\text{cov}(D_1, \nu)}_{=0 \text{ by hyp.}} - \widetilde{\lambda}_1 \text{cov}(\nu, \nu) \\
&= -\widetilde{\lambda}_1 V(\nu)
\end{aligned}$$

Third,

$$\begin{aligned}
\text{cov}(\widetilde{R}, \eta) &= \text{cov}(D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 \widetilde{D}_1, \eta) \\
&= \text{cov}(D_2 - \widetilde{\lambda}_0 - \widetilde{\lambda}_1 (D_1 + \nu), \eta) \\
&= \underbrace{\text{cov}(D_2, \eta)}_{=0 \text{ by hyp.}} - \widetilde{\lambda}_0 \underbrace{\text{cov}(1, \eta)}_{=0} - \widetilde{\lambda}_1 \underbrace{\text{cov}(D_1, \eta)}_{=0 \text{ by hyp.}} - \widetilde{\lambda}_1 \underbrace{\text{cov}(\nu, \eta)}_{=0 \text{ by hyp.}} \\
&= 0
\end{aligned}$$

Finally, plugging in all these results in ($\Delta\Delta\Delta$), we get :

$$\begin{aligned}\widetilde{\beta}_{02} &= \delta_{02}\frac{V(\widetilde{R})}{V(\widetilde{R})} + \delta_{01}\frac{\widetilde{\lambda}_1 V(\nu)}{V(\widetilde{R})} \\ &= \delta_{02} + \delta_{01}\widetilde{\lambda}_1\frac{V(\nu)}{V(\widetilde{R})}\end{aligned}$$

Moreover, $\widetilde{\lambda}_1 = \dfrac{\text{cov}(D_2, \widetilde{D}_1)}{V(\widetilde{D}_1)} = \dfrac{\text{cov}(D_2, D_1)}{V(D_1) + V(\nu)}$ since $\text{cov}(D_2, \nu) = \text{cov}(D_1, \nu) = 0$. So we finally get :

$$\widetilde{\beta}_{02} = \delta_{02} + \delta_{01}\frac{\text{cov}(D_2, D_1)V(\nu)}{(V(D_1) + V(\nu))V(\widetilde{R})}$$

Therefore, the sign of the bias depends on both δ_{01} and $\text{cov}(D_1, D_2)$. If these two quantities have the same sign, the bias will be positive. If they are of the opposite sign, then it will be negative

6. Application of Question 5 : what is the likely sign of the bias when D_1 is the initial (say, in grade 1) school level of a pupil, D_2 an indicator of grade repetition (say, in grade 2) ("redoublement"), and $Y(d)$ the potential final (say, in grade 3) school level ?

First, we expect a negative correlation between D_1 and D_2 : the greater the initial school level, the lower the probability of repeating the next school year. Second, we expect δ_{01} to be positive : if the pupil has good grades in grade 1, she will also likely have good grades in grade 3. Therefore, since the two relevant quantities are of opposite sign, the bias is likely to be negative. β_{02} is going to be smaller than the causal effect δ_{02}. In the worst case scenario, it is possible that we predict a significant negative effect of grade repetition on school level in the next year while actually, grade repetition has a neutral or even positive causal effect on school level.