Applied Quantitative Analysis II
Lab 8

Eric G. Zhou

Spring 2020

NYU Wagner School of Public Service
NYU Grossman School of Medicine

Cotents today

Lab assignment 3 walk through

Direct/indirect effects

More on omitted variable bias

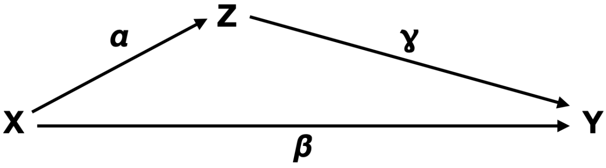
Lab assignment 3 walk through

See Jupyter notebook

AQA II: Lab 8

Direct/indirect effects

Remember the example in class, and let's go over it again:



$$y = \beta_0 + \beta X + \gamma Z + \epsilon$$

$$Z = a_0 + aX + \zeta$$

Direct/indirect effects

Remember the example in class, and let's go over it again:

$$Y = \beta_0 + \beta X + \gamma Z + \epsilon \tag{1}$$

$$Z = \alpha_0 + \alpha X + \xi \tag{2}$$

To make this work, let's assume equation (1) and (2) holds, i.e.,

- (β, γ, α) are the "true" coefficient parameters
- (β_0, α_0) are constant intercepts
- and (ϵ, ξ) are non-informative disturbance term (non-informative relative to their dependent variable)

Direct/indirect effects

Plug equation (2) to (1) and get

$$
\begin{aligned}
Y &= \beta_0 + \beta X + \gamma(\alpha_0 + \alpha X + \xi) + \epsilon && (3)\\
&= (\beta_0 + \gamma\alpha_0) + (\underbrace{\beta}_{\text{direct effect}} + \underbrace{\gamma\alpha}_{\text{indirect effect}})X + \gamma\xi + \epsilon && (4)\\
&= \lambda + \underbrace{\beta'}_{\text{total effect}} X + \gamma Z' + \epsilon && (5)
\end{aligned}
$$

The effect proportion $= \frac{\beta}{\beta'} = \frac{\beta}{\beta+\gamma\alpha}$, what does this measure?

This process is also called "partitioning" or "decomposition", if you understand this well then you are solid on regression basics.

Direct/indirect effects

We just showed that equation (1) and (3) are identical but having different coefficient vectors.

In other words, if we could break down Z into those components, we could run two regressions that are identical in nature, while unbiasedly estimating all coefficients. What about in reality? Can we do this easily?

More on omitted variable bias

Remember last week, we talked about the omitted variable bias (OVB), which looks like this:

True model being $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, however, we could only run model $Y = \beta_0 + \hat{\beta}_1 X_1 + \epsilon'$.

Then take expectation on $\hat{\beta}_1$, we will have

$$E[\hat{\beta}_1] = \underbrace{\beta_1}_{\text{true } \beta_1} + \underbrace{\beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)}}_{\text{omitted variable bias}} \tag{6}$$

More on omitted variable bias

Why do we bring this up again? Because this is tightly related to the direct/indirect effects part.

The catch from equation (6) is that it turns out $\frac{Cov(X_1,X_2)}{Var(X_1)}$ is a very familiar friend

$$\frac{Cov(X_1,X_2)}{Var(X_1)} = E[(X_1'X_1)^{-1}X_1'X_2] = \beta_{x_1,x_2} = \alpha \tag{7}$$

So, go back to the previous example, if we estimate $Y = \text{constant} + \hat{\beta}X + \hat{\epsilon}$, can you guess the value of $\hat{\beta}$?

More on omitted variable bias

So, go back to the previous example, if we estimate $Y = \text{constant} + \hat{\beta}X + \hat{\epsilon}$, can you guess the value of $\hat{\beta}$?

Invoke the formula of OVB, it becomes

$$E[\hat{\beta}] = \beta + \gamma\frac{Cov(X,Z)}{Var(X)} = \beta + \gamma\alpha = \beta'$$

Everything connects in some way! Now, let me ask you again what effects proportion represent?

$$\begin{aligned}\text{The effect proportion} &= \frac{\beta}{\beta'} = \frac{\beta}{\beta+\gamma\alpha} = \frac{\text{True estimate}}{\text{True estimate plus OVB}} \\ &= \frac{\text{True estimator}}{\text{Estimator presented with OVB}}\end{aligned}$$

Exercise

All of the issues mentioned above are due to:

1. Correlation among regressors
2. Omitted variable bias

Why don't we do some exercises to even understand this better, combining direct/indirect effects and OVB?

Exercise

Tell me if all the hat estimators are biased or not?

1. $Y =$ constant $+ \hat{\beta} X + \hat{\gamma} Z + e$
2. $Z =$ constant $+ \hat{\alpha} X + e$
3. $Y =$ constant $+ \hat{\beta} X + e$
4. $Y =$ constant $+ \hat{\gamma} Z + e$
5. $Y =$ constant $+ \hat{\beta} (\frac{50 * X + 10}{20}) + \hat{\gamma} Z + e$
6. $Y =$ constant $+ \hat{\beta} (\frac{50 * X + 10}{20}) + e$
7. $Y =$ constant $+ \hat{\beta} \log(X) + e$