Applied Quantitative Analysis II
Lab 7

Eric G. Zhou

Spring 2020

NYU Wagner School of Public Service
NYU Grossman School of Medicine

General reminder

Spring break is a great time to focus and work on your project

Spring break means Spring work for me at least for the past 3 years....

I know I owe many of you email replies, please bear with me and I will reply soon – I had midterms too

Cotents today

Omitted variable bias

Interaction effects in linear equation

Interaction effects in non-linear equation

Questions on the assignment

Omitted variable bias

AQA II is about what can go wrong with our assumptions

The omitted variable bias is notorious and common

Let's try to understand it what it could do to our estimations with some math

Remember this is how we derive the OLS estimator

$$\hat{\beta}_{ols} = (X'X)^{-1}X'Y \tag{1}$$

Since $Y = \beta' X + \epsilon$, we can rewrite equation 1 into

$$\hat{\beta}_{ols} = (X'X)^{-1}X'(\beta' X + \epsilon) \tag{2}$$

Condition for unbiasedness?

Omitted variable bias

Rearrange terms of equation (2) and we will have

$$\hat{\beta}_{ols} = \beta + \underbrace{(X'X)^{-1}\epsilon}_{\text{bias}} \tag{3}$$

We would say $\hat{\beta}_{ols}$ is an unbiased estimator of the true value β if $Cov(X, \epsilon) = 0$. But this does not tell us much about the omitted variable bias which is only one source for the bias in equation 3. Here is how we think about the omitted variable bias.

Assume a true data generating process is

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon \tag{4}$$

Omitted variable bias

However, due to either we forget to include X_2 or it is actually not measurable or observable, this is the estimating equation we have is

$$Y = \hat{\beta}_1 X_1 + \epsilon' \tag{5}$$

With this equation, the coefficient for X_1 we estimated will be

$$\begin{aligned}
\hat{\beta}_1 &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(\beta_1 X_1 + \beta_2 X_2 + \epsilon) \\
&= \frac{X'X\beta_1 + \beta_2 X_1' X_2 + X'\epsilon}{X'X}
\end{aligned}$$

Omitted variable bias

In equation (4), we assume the true error term and X_1 is still independent. So the last term is zero.

$$\hat{\beta}_1 = \beta_1 + (X'X)^{-1}\beta_2 X_1' X_2 \tag{6}$$

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

$$E[\hat{\beta}_1] = \underbrace{\beta_1}_{\text{true } \beta_1} + \underbrace{\beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)}}_{\text{omitted variable bias}} \tag{7}$$

Admittedly, we cannot quantity or identify the omitted variable bias (if X_2 is not available), but we could know its direction and theoretical magnitude, if we know the direction of β_2 and the value of $Cov(X_1, X_2)$.

Omitted variable bias

Questions:

In reality, knowing $Cov(X_1, X_2)$ is harder, but the sign of β_2 is easier. Why?

In what situations, it turns out to be fine if you leave out a relevant independent variable? (hint: look at equation 7)

p.s. general advice for OLS estimators, it's almost always right to use robust standard errors.

Interaction effects in linear equations

Assume we have an estimating equation of

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{8}$$

where the X_2 turns out to be $X_2 = X_1 \cdot X_3$, the equation now becomes

$$Y = \beta_1 X_1 + \beta_2 X_1 \cdot X_3 + \beta_3 X_3 + \epsilon \tag{9}$$

How do we interpret each coefficient? It's helpful to see the partial derivatives

Interaction effects in linear equations

The partial derivatives of Y:

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_2 X_3$$
$$\frac{\partial Y}{\partial X_3} = \beta_3 + \beta_2 X_1$$
$$\frac{\partial Y}{\partial X_1 \partial X_3} = \beta_2$$

Scenario 1, X_1 and X_3 are both binary variables

Scenario 2, one of X_1 and X_3 is binary while the other is continuous

Scenario 3, X_1 and X_3 are both continuous variables (too restricted, rarely used)

Interaction effects in non-linear equations

Assume we have an estimating equation of

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{10}$$

where the $Y = \log(\frac{P(x)|X}{(1-P(x))|X})$, and X_2 turns out to be $X_2 = X_1 \cdot X_3$, the equation now becomes (pay attention to the error term)

$$\log(\frac{p(x)}{1-p(x)}) = \beta_1 X_1 + \beta_2 X_1 \cdot X_3 + \beta_3 X_3 \tag{11}$$

How do we interpret each coefficient? It's helpful to see the partial derivatives

Interaction effects in non-linear equations

The partial derivatives of Y:

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_2 X_3$$
$$\frac{\partial Y}{\partial X_3} = \beta_3 + \beta_2 X_1$$
$$\frac{\partial Y}{\partial X_1 \partial X_3} = \beta_2$$

Note: the βs are still linear in log odds but not in the term most of us are familiar with, the probability of $Y = 1$ given the values of X_1, X_3. What can we do?

Interaction effects in non-linear equations

Since we have

$$\log(\frac{p(x)}{1-p(x)}) = \beta_1 X_1 + \beta_2 X_1 \cdot X_3 + \beta_3 X_3$$

we could derive the equation of $p(x)$ as a function of X_1, X_3.

$$p(x) = \frac{e^{\beta_1 X_1 + \beta_2 X_1 \cdot X_3 + \beta_3 X_3}}{1 + e^{\beta_1 X_1 + \beta_2 X_1 \cdot X_3 + \beta_3 X_3}} \tag{12}$$

This leads to an important question: what is the linear-equation equivalent of β_1 in here?

Or, in other words, what's the effect of one unit increase in my X on the probability increase or decrease of $Y = 1$?

Interaction effects in non-linear equations

Many people interpret the logit/probit coefficients directly as probability increase or decrease. That is just **sooooo wrong!**

Let us not do that, and here is how we can transform the interpretation of log odds to probability correctly.

If your X is binary, let's first run the model with your $X = 0$, save your $\hat{Y}|(X = 0)$, then calculate the associated probability $\Pr(Y = 1|X = 0)$ using equation (12). Next, we do the same thing to obtain $\Pr(Y = 1|X = 1)$.

Now do you see what is the marginal effect of the binary variable X on the probability of event $\Pr(Y = 1|X)$?

Interaction effects in non-linear equations

Now do you see what is the marginal effect of the binary variable X on the probability of event $\Pr(Y = 1|X)$?

$$\frac{\partial \Pr(Y=1|X)}{\partial X} = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)$$

Does this conclude the marginal effect for non-linear equations? What's missed here?

Interaction effects in non-linear equations

Does this conclude the marginal effect for non-linear equations? What's wrong here?

If X is binary, consider values 0 and 1 exhaust the measure(all possible values) of your independent variables. What if your X is continuous?

To make matter worse, remember the marginal effect in the Logit model is not constant.

Solutions of looking at marginal effects in probabilities?

Interaction effects in non-linear equations
Solutions of looking at marginal effects in probabilities?

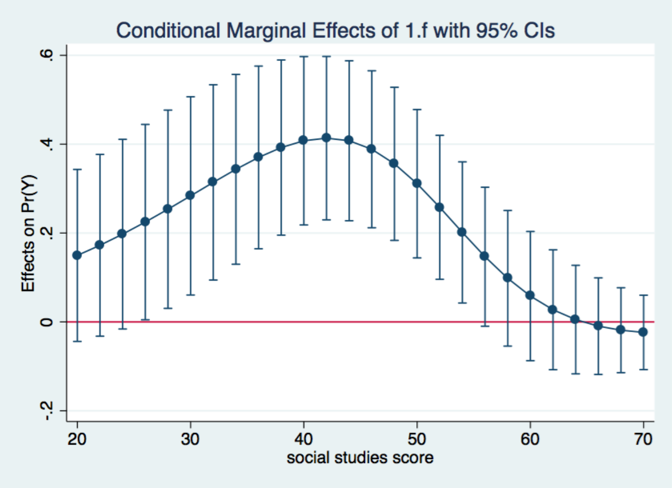
This might sound tedious, but you could calculate every possibility value $\text{Pr}(Y = 1|X)$ for every possible value of X then take expectation, which is something like this

$$E[\text{Pr}(Y = 1|X)] = \int f(x)p(x)\partial x \tag{13}$$

Or, we could take advantage of stata and use the command **margin**.

Plotting your marginal effect is also a very good way of interpreting your model.

Interaction effects in non-linear equations



Questions on lab assignment 3

Any questions on the lab assignment 3?