# More regressions stuff!

Econ 140, Section 4

Jonathan Old

# Roadmap

1. Omitted Variable Bias
2. Bad Controls
3. Quadratic Terms
4. Interaction terms

Any questions?

... The midterm is getting closer: Now is the time for clarifying questions!

# Omitted Variable Bias

# Recap: OVB (Very important!)

We can summarize everything of OVB in three equations. Let $Y_i$ be the outcome variable, $X_i$ our regressor of interest, and $Z_i$ the "omitted" variable.

$$
\begin{aligned}
\text{[Long regression]} \quad & Y_i = c_1 + \beta_L X_i + \delta Z_i + e_i \\
\text{[Short regression]} \quad & Y_i = c_2 + \beta_S X_i + u_i \\
\text{[Auxiliary regression]} \quad & Z_i = c_3 + \gamma X_i + v_i
\end{aligned}
$$

Then, the **Omitted variable bias formula** states that:

$$\underbrace{\beta_S}_{\text{Short =}} = \underbrace{\beta_L}_{\text{Long +}} + \underbrace{\delta}_{\text{Omitted} \times} \cdot \underbrace{\gamma}_{\text{Included}}$$

We call $\delta\gamma$ the **omitted variable bias**. We can appraise the direction of the bias by multiplying our guesses for the signs of $\delta$ and $\gamma$. **If either $\delta = 0$ or $\gamma = 0$, then OVB is zero!**

# Regression: Basics, Interpretation, Quadratic Terms

See in `Datahub`: OVB example, OVB visualized

# Bad Controls

# Control variables

Control variables are additional variables (or covariates) **included in a regression**. We do this for **various reasons** (in decreasing order of importance):

- To remove selection bias / omitted variable bias
- To increase precision of our estimates
- To know about the (conditional/partial) correlation of other variables
- To better predict the outcome

Let's see graphically how control variables work! `Datahub`

# Bad controls

- Some controls are called ”bad controls”. These are:
  1. Variables that are themselves outcomes of a treatment: What happens if you control for the change in English test scores in the regression below?
  2. Variables that moderate the treatment effect, e.g. controlling for occupation choice in gender wage gap regression . . .
- Rule of Thumb: Good controls are either pre-determined or immutable characteristics.
- Another way to think about it: Controls help us make ”apples to apples” comparisons. Which apples matter?

Mathematically, good and bad controls are the same thing.

We need to use our 🧠 to distinguish them!

# Quadratic Terms

# Making OLS more interesting

- We can extend the simple OLS framework

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

to something richer:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

- All questions of the type *"how is $Y_i$ expected to change if we change $X_i$", keeping all other variables in the regression fixed* can be solved with **partial derivatives** – in this case:

$$\frac{\partial Y_i}{\partial X_i} =$$

# Making OLS more interesting

- We can extend the simple OLS framework

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

to something richer:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

- All questions of the type *"how is $Y_i$ expected to change if we change $X_i$", keeping all other variables in the regression fixed* can be solved with **partial derivatives** – in this case:

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2 \cdot \beta_2 \cdot X_i$$

# Example for Quadratic Terms

Let's see how to use quadratic terms on `Datahub`

# Interaction terms

# Interaction terms: Making OLS more interesting

Let us consider the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

where $Y_i$ is a country's GDP per capita, $X_{1i}$ the value of its natural resources, and $X_{2i}$ a measure of how democratic it is.

1. How do we interpret $\beta_1$?

# Interaction terms: Making OLS more interesting

Let us consider the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

where $Y_i$ is a country's GDP per capita, $X_{1i}$ the value of its natural resources, and $X_{2i}$ a measure of how democratic it is.

1. How do we interpret $\beta_1$?
   **Keeping democracy fixed, increasing the value of a country's natural resources by one unit is associated with $\beta_1$ higher GDP per capita.**
2. How do we interpret $\beta_2$?

# Interaction terms: Making OLS more interesting

Let us consider the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

where $Y_i$ is a country's GDP per capita, $X_{1i}$ the value of its natural resources, and $X_{2i}$ a measure of how democratic it is.

1. How do we interpret $\beta_1$?
   **Keeping democracy fixed, increasing the value of a country's natural resources by one unit is associated with $\beta_1$ higher GDP per capita.**
2. How do we interpret $\beta_2$?
   **Keeping natural resources fixed, increasing a country's democracy score by one unit is associated with $\beta_2$ higher GDP per capita.**

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$?

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$?

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 0.**

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 0.**
3. How do we interpret $\beta_2$?

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 0.**
3. How do we interpret $\beta_2$? **The effect of an additional unit of $X_{2i}$ , if $X_{1i}$ is equal to 0.**

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 0.**
3. How do we interpret $\beta_2$? **The effect of an additional unit of $X_{2i}$ , if $X_{1i}$ is equal to 0.**
4. How do we interpret $\beta_1 + \beta_3$?

# Interaction Terms (ii)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of $X_{1i}$ on $Y_i$? $\beta_1$ + $\beta_3 X_{2i}$
2. How do we interpret $\beta_1$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 0.**
3. How do we interpret $\beta_2$? **The effect of an additional unit of $X_{2i}$ , if $X_{1i}$ is equal to 0.**
4. How do we interpret $\beta_1 + \beta_3$? **The effect of an additional unit of $X_{1i}$ , if $X_{2i}$ is equal to 1.**

**Rule of thumb: Always use partial derivatives to make sure that you are right!**

# Interactions with only dummy variables

- Take any two binary variables, for example: Studies at UC Berkeley or not ($\text{UCB}_i$), and is from California or not ($\text{Cali}_i$).
- The regression with interactions looks like this:

$$Y_i = \underbrace{\beta_0}_{17.77} + \underbrace{\beta_1}_{2.28} \text{UCB}_i + \underbrace{\beta_2}_{0.95} \text{Cali}_i + \underbrace{\beta_3}_{-0.97} \text{UCB}_i \times \text{Cali}_i$$

- We can write this in a table, and get all group averages

| | Cali = 1 | Cali=0 |
|---|---|---|
| UCB = 1 | $\beta_0 + \beta_1 + \beta_2 + \beta_3$<br>=17.77+2.28+0.95-0.97 | $\beta_0 + \beta_1$<br>= 17.77 + 2.28 |
| UCB = 0 | $\beta_0 + \beta_2$<br>= 17.77 + 0.95 | $\beta_0$<br>=17.77 |

- Take differences between cells to get different effects!