Regression and controls

Felipe Balcazar

NYU

September, 2022

NEW YORK UNIVERSITY

Recapitulating and further instructions

- Don't forget the deadlines.
- You should discuss with me your idea and data in the next two weeks.
  - Book office hours ASAP.
  - If no slots write me an email.
- Last class: We covered the basic use of stata.
- This class: basics of ensambling a database and intro to regression.

Recapitulating and further instructions

- Don't forget the deadlines.
- You should discuss with me your idea and data in the next two weeks.
  - Book office hours ASAP.
  - If no slots write me an email.
- Last class: We covered the basic use of stata.
- This class: basics of ensambling a database and intro to regression.

Recapitulating and further instructions

- Don't forget the deadlines.
- You should discuss with me your idea and data in the next two weeks.
  - Book office hours ASAP.
  - If no slots write me an email.
- Last class: We covered the basic use of stata.
- This class: basics of ensambling a database and intro to regression.

Recapitulating and further instructions

- Don't forget the deadlines.
- You should discuss with me your idea and data in the next two weeks.
  - Book office hours ASAP.
  - If no slots write me an email.
- Last class: We covered the basic use of stata.
- This class: basics of ensambling a database and intro to regression.

We are interested in causal questions

What is the impact of D (your independent variable, the treatment) on Y (the outcome of interest or dependent variable)?

or...

Does D (your independent variable, the treatment) cause Y (the outcome of interest or dependent variable)?

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.

The method for causal inference: regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

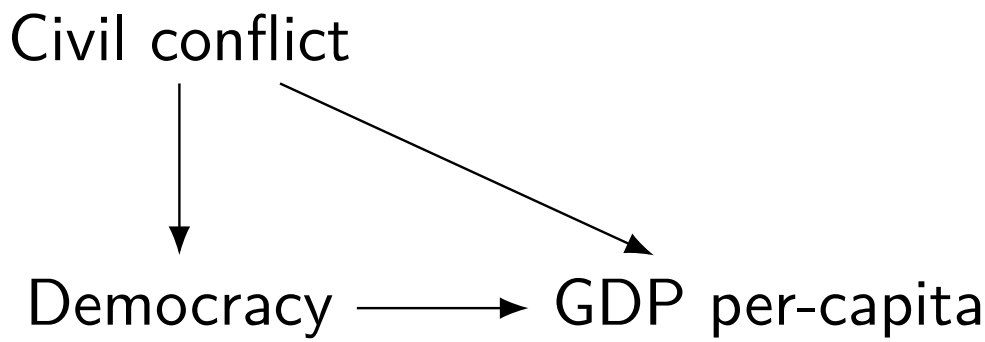
- Y_i is the *outcome* for individual i.
- D_i is the the value of the *treatment* for i.
  - i is the *unit of observation*—the level at which your outcome is defined.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- ε is an error term with mean zero $E(\varepsilon) = 0$; in it there could be other things that explain Y and random noise.

Does democracy lead to higher GDP?

Democracy $\longrightarrow$ GDP per-capita

- We ask whether democracy causes GDP per-capita.
- But conflict causes both.
- Conflict is a confounder.

Does democracy lead to higher GDP?



- We ask whether democracy causes GDP per-capita.
- But conflict causes both.
- Conflict is a confounder.

The method for causal inference: regression

$$GDP_{it} = \alpha + \beta Democracy_{it} + \delta Conflict_{it} + \varepsilon_{it}.$$

- i is the country (unit of analysis); t is the year. Pair it is unit of observation.

The method for causal inference: regression

$$GDP_{it} = \alpha + \beta Democracy_{it} + \delta Conflict_{it} + \varepsilon_{it}.$$

- i is the country (unit of analysis); t is the year. Pair it is unit of observation.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- δ is the coefficient for conflict.
  - coefficients for confounders should not be interpreted.

The method for causal inference: regression

$$GDP_{it} = \alpha + \beta Democracy_{it} + \delta Conflict_{it} + \varepsilon_{it}.$$

- i is the country (unit of analysis); t is the year. Pair it is unit of observation.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- δ is the coefficient for conflict.
  - coefficients for confounders should not be interpreted.

The method for causal inference: regression

$$GDP_{it} = \alpha + \beta Democracy_{it} + \delta Conflict_{it} + \varepsilon_{it}.$$

- i is the country (unit of analysis); t is the year. Pair it is unit of observation.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- δ is the coefficient for conflict.
  - coefficients for confounders should not be interpreted.

The method for causal inference: regression

$$GDP_{it} = \alpha + \beta Democracy_{it} + \delta Conflict_{it} + \varepsilon_{it}.$$

- i is the country (unit of analysis); t is the year. Pair it is unit of observation.
- α is the intercept, the constant.
- β is the slope; the *effect of the treatment*.
- δ is the coefficient for conflict.
  - coefficients for confounders should not be interpreted.