Bias and Fixed Effects

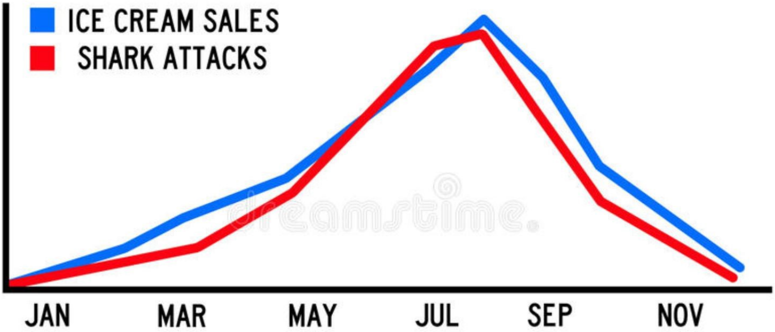
Felipe Balcazar

NYU

October, 2022

NEW YORK UNIVERSITY

Correlation does not imply causation



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

Shark attacks

Ice cream $\longrightarrow$ Shark Attack

- Does eating icecream cause shark attacks.

Shark attacks

$$\begin{tikzcd}
\text{Temperature} \arrow[d] \arrow[dr] & \\
\text{Ice cream} \arrow[r] & \text{Shark Attack}
\end{tikzcd}$$

- Does eating icecream cause shark attacks.

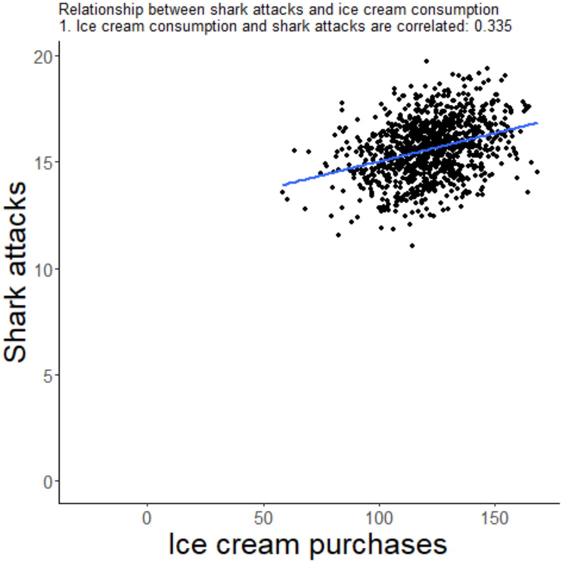
The method for causal inference: regression

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

When we estimate a regression

$$Y_i = \widehat{\alpha} + \widehat{\beta} D_i.$$

Correlation and a regression line



What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).

What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).

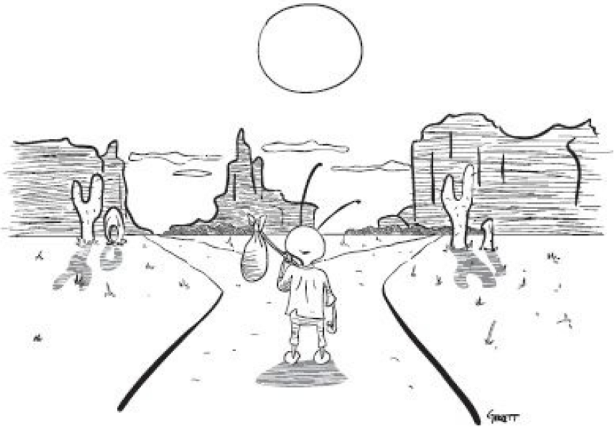
What did we see in the previous graph?

- The graph suggests a positive correlation to begin with.
- The regression line had a positive slope:

$$SharkAttacks_i = \widehat{\alpha} + \widehat{\beta} IceCream_i, \quad \widehat{\beta} > 0$$

- Once we control for confounders the slope is zero; i.e., no actual relationship (i.e., $\beta = 0$).

Other things equal



The ideal case scenario: no confounders

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment,} \\ 0 & \text{otherwise.} \end{cases}$$

and

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

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned}
(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1 \\
E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\
=& \alpha + \beta + E(\varepsilon_i|D_i = 1) \\
=& \alpha + \beta + E(\varepsilon_i) \\
E(Y_i|D_i = 1) =& \alpha + \beta
\end{aligned}$$

Let's unpack the previous expression

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment,} \\ 0 & \text{otherwise.} \end{cases}$$

and

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

Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned} (Y_i | D_i = 0) =& \alpha + \varepsilon_i | D_i = 0 \\ E(Y_i | D_i = 0) =& E(\alpha + [\varepsilon_i | D_i = 0]) \\ =& \alpha + E(\varepsilon_i | D_i = 0) \\ =& \alpha + E(\varepsilon_i) \\ E(Y_i | D_i = 0) =& \alpha \end{aligned}$$

Let's unpack the previous expression

We obtain a simple system of equations with two unknowns!

$$
\begin{aligned}
E(Y_i|D_i = 1) =& \alpha + \beta \\
E(Y_i|D_i = 0) =& \alpha
\end{aligned}
$$

Remember

- $\widehat{\alpha} = E(Y_i|D_i = 0)$.
- $\widehat{\beta} = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$.

Selection bias is omitted variable bias

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
D \arrow[r] & Y
\end{tikzcd}
$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

Selection bias is omitted variable bias

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
D \arrow[r] & Y
\end{tikzcd}
$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

- It is no longer the case that $E[\varepsilon_i | D_i = 1] \neq E[\varepsilon_i] = 0$ or $E[\varepsilon_i | D_i = 0] \neq E[\varepsilon_i] = 0$.

Selection bias is omitted variable bias

$$\begin{tikzcd} X \arrow[d] \arrow[dr] & \\ D \arrow[r] & Y \end{tikzcd}$$

$$Y_i = \alpha + \beta D_i + \underbrace{\delta X_i + \epsilon_i}_{\varepsilon_i}.$$

- It is no longer the case that $E[\varepsilon_i|D_i = 1] \neq E[\varepsilon_i] = 0$ or $E[\varepsilon_i|D_i = 0] \neq E[\varepsilon_i] = 0$.
- Selection bias only comes from confounders (X).

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$
\begin{aligned}
(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1 \\
E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\
=& \alpha + \beta + E(\varepsilon_i|D_i = 1)
\end{aligned}
$$

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned}(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1 \\ E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\ =& \alpha + \beta + E(\varepsilon_i|D_i = 1)\end{aligned}$$

Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned}(Y_i|D_i = 0) =& \alpha + \varepsilon_i|D_i = 0 \\ E(Y_i|D_i = 0) =& E(\alpha + [\varepsilon_i|D_i = 0]) \\ =& \alpha + E(\varepsilon_i|D_i = 0)\end{aligned}$$

What happens if we don't control for confounders?

Let us restrict our attention to those individuals whose $D_i = 1$.

$$\begin{aligned}
(Y_i|D_i = 1) =& \alpha + \beta + \varepsilon_i|D_i = 1 \\
E(Y_i|D_i = 1) =& E(\alpha + \beta + [\varepsilon_i|D_i = 1]) \\
=& \alpha + \beta + E(\varepsilon_i|D_i = 1)
\end{aligned}$$

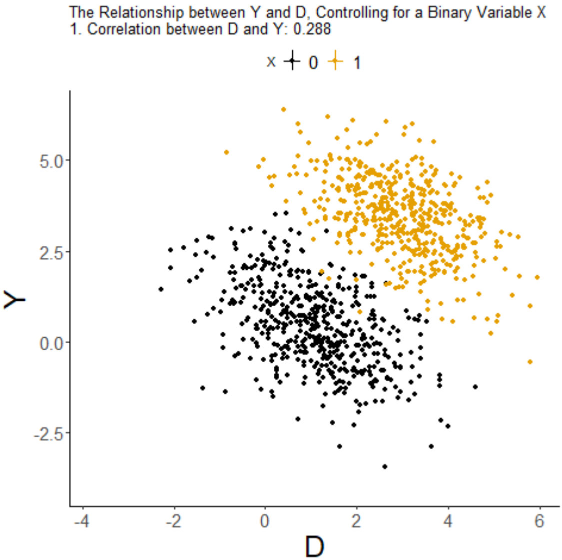
Let us restrict our attention now to those individuals whose $D_i = 0$.

$$\begin{aligned}
(Y_i|D_i = 0) =& \alpha + \varepsilon_i|D_i = 0 \\
E(Y_i|D_i = 0) =& E(\alpha + [\varepsilon_i|D_i = 0]) \\
=& \alpha + E(\varepsilon_i|D_i = 0)
\end{aligned}$$

Solving the system of equations:

$$\widehat{\beta} = \underbrace{E(Y_i|D_i = 1) - E(Y_i|D_i = 0)}_{ATE} + \underbrace{[E(\varepsilon_i|D_i = 1) - E(\varepsilon_i|D_i = 0)]}_{\text{selection bias}}.$$

The effect of a confounder: illustration



What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$
\begin{aligned}
(Y_i|D_i = 1, X_i) &= \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) &= \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
&= \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) &= \alpha + \beta E(D_i = 1|X_i)
\end{aligned}
$$

What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$
\begin{aligned}
(Y_i|D_i = 1, X_i) =& \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
=& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i)
\end{aligned}
$$

Let us restrict our attention now to those individuals whose $D_i = 0|X_i$.

$$
\begin{aligned}
(Y_i|D_i = 0, X_i) =& \alpha + \beta(D_i = 0|X_i) + \epsilon_i|(D_i = 0|X_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i|(D_i = 0|X_i)) \\
=& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i)
\end{aligned}
$$

What happens if we control for ALL confounders?

Let us restrict our attention to those individuals whose $D_i = 1|X_i$.

$$\begin{aligned}
(Y_i|D_i = 1, X_i) =& \alpha + \beta(D_i = 1|X_i) + \epsilon_i|(D_i = 1|X_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i|(D_i = 1|X_i)) \\
=& \alpha + \beta E(D_i = 1|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 1, X_i) =& \alpha + \beta E(D_i = 1|X_i)
\end{aligned}$$

Let us restrict our attention now to those individuals whose $D_i = 0|X_i$.

$$\begin{aligned}
(Y_i|D_i = 0, X_i) =& \alpha + \beta(D_i = 0|X_i) + \epsilon_i|(D_i = 0|X_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i|(D_i = 0|X_i)) \\
=& \alpha + \beta E(D_i = 0|X_i) + E(\epsilon_i) \\
E(Y_i|D_i = 0, X_i) =& \alpha + \beta E(D_i = 0|X_i)
\end{aligned}$$

Solving the resulting system of equations we have

$$\widehat{\beta} = [E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i)].$$

The crux of observational studies

- The main challenge in observational studies are observed and unobserved confounders. This generates selection (omitted variables) bias!

- This is particularly problematic when we are making comparisons *between* different units (i.e. between different individuals, between countries or cities, etc).

- Cross-country regressions, are perhaps the most subject to this concern: lots of differences between countries and really hard to make "other things equal" comparisons!

Causal relationship of interest is:

$$\text{Economic prosperity} \longrightarrow \text{Democracy}$$

What does the data show?

Causal Effects?

- As usual, hard to make causal statements from these comparisons.
- There are many confounders of income/growth that may also explain why a country is more likely to become democratic.

$$\begin{tikzcd} \text{Income} \arrow[r] & \text{Democracy} \\ \text{Geography, Colonial History} \arrow[u, dashed] \arrow[ur, dashed] & \end{tikzcd}$$

Fixed-Effects Regressions

- Whenever we have several observations for a given unit, we can include fixed effects that will account for all time-invariant characteristics.

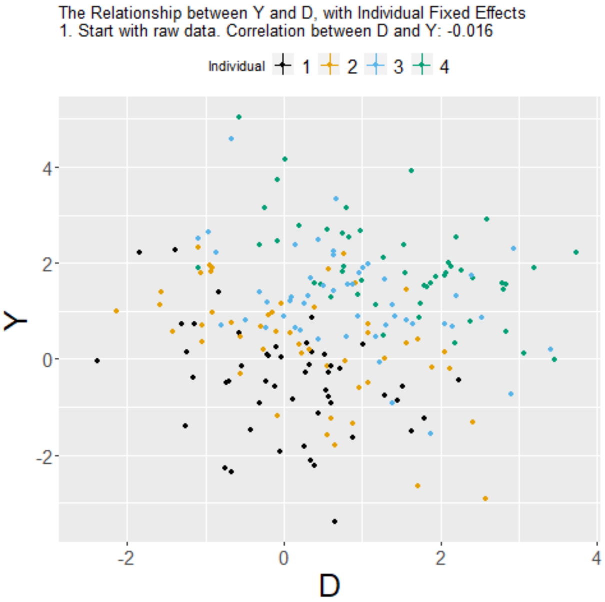
- The regression takes the form:

$$Y_{it} = \alpha + \beta D_{it} + \mu_i + \varepsilon_{it}.$$

- The inclusion of fixed effects essentially transforms all observations as deviations from their mean.

- This means that the analysis now compares treated units before and after treatment, relative to the change over time for not treated units.

What do fixed effects do?



In regression form

$$\begin{tikzcd} \text{GDP} \arrow[r] & \text{Democracy} \\ \text{i} \arrow[u] \arrow[ur] & \end{tikzcd}$$

$$Dem_{it} = \alpha + \beta GDP_{it} + \mu_i + \varepsilon_{it}.$$

In regression form

$$\begin{tikzcd} \text{GDP} \arrow[r] & \text{Democracy} \\ \text{i} \arrow[u] \arrow[ur] & \end{tikzcd}$$

$$Dem_{it} = \alpha + \beta GDP_{it} + \mu_i + \varepsilon_{it}.$$

If you control for fixed effects this occurs:

$$Dem_{it} - \overline{Dem}_i =$$
$$(\alpha - \alpha) + \beta(GDP_{it} - \overline{GDP}_i) + (\mu_i - \mu_i) + (\varepsilon_{it} - \overline{\varepsilon}_i).$$

In regression form

$$
\begin{tikzcd}
GDP \arrow[r] & Democracy \\
i \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

$$Dem_{it} = \alpha + \beta GDP_{it} + \mu_i + \varepsilon_{it}.$$

If you control for fixed effects this occurs:

$$Dem_{it} - \overline{Dem}_i =$$
$$(\alpha - \alpha) + \beta(GDP_{it} - \overline{GDP}_i) + (\mu_i - \mu_i) + (\varepsilon_{it} - \overline{\varepsilon}_i).$$

Therefore

$$\tilde{Dem}_{it} = \beta(\tilde{Income}_{it}) + \tilde{\varepsilon}_{it}.$$

Adding year fixed-Effects

- Whenever we have several observations for a given unit, we can include time fixed effects that will account for all variables that generate common effects to all units.
- The regression takes the form:

$$Y_{it} = \alpha + \beta D_{it} + \mu_i + \gamma_t + \varepsilon_i t.$$

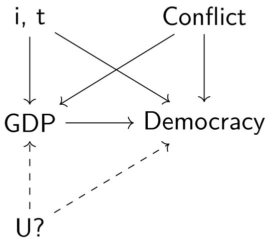
- The inclusion of time fixed effects essentially captures non-monotonic time trends.

Limitations of fixed effects

- You still have to worry about time-varying observable confounders (X).
- You still have to worry about time-varying unobservable confounders (U).
- But if your most worrisome source of confounding is fixed at the time of the study, you can present compelling results.

Limitations of fixed effects

- You still have to worry about time-varying observable confounders (X).
- You still have to worry about time-varying unobservable confounders (U).
- But if your most worrisome source of confounding is fixed at the time of the study, you can present compelling results.



The crux of the empirical researcher

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

The crux of the empirical researcher

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

The crux of the empirical researcher

- When we use observational data we may not observe all confounders.
  - We may be able to control for observable/measurable ones.
- We can find clever ways to provide a very good sense of the direction of the causal effect.
- To interpret the magnitude is not always possible although many do, with a grain of salt.

Ways to address omitted variable bias

- Randomized control trial.
- Natural experiments.
  - Exogenous source of variation: policy experiment, or natural occurrence.
  - Instrumental variable design.
- Differences-in-differences.
- Regression discontinuity design.
  - Standard.
  - Spatial.
  - Event studies.

Ways to address omitted variable bias

- Randomized control trial.
- Natural experiments.
    - Exogenous source of variation: policy experiment, or natural occurrence.
    - Instrumental variable design.
- Differences-in-differences.
- Regression discontinuity design.
    - Standard.
    - Spatial.
    - Event studies.

Ways to address omitted variable bias

- Randomized control trial.
- Natural experiments.
  - Exogenous source of variation: policy experiment, or natural occurrence.
  - Instrumental variable design.
- Differences-in-differences.
- Regression discontinuity design.
  - Standard.
  - Spatial.
  - Event studies.

Ways to address omitted variable bias

- Randomized control trial.
- Natural experiments.
  - Exogenous source of variation: policy experiment, or natural occurrence.
  - Instrumental variable design.
- Differences-in-differences.
- Regression discontinuity design.
  - Standard.
  - Spatial.
  - Event studies.

Temporary page!

LaTeX was unable to guess the total number of pages correctly. A
there was some unprocessed data that should have been added t
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.