Statistical Modeling and Causal Inference with R

Week 3: Revisiting regression estimators of causal effects

Manuel Bosancianu

Max Schaub

September 21, 2020

Hertie School of Governance

Today's focus

- Recap: Motivation for causal inference (CI)
- Ordinary Least Square Regression (OLS)
- Regression and the potential outcomes framework (POF)
- Regression as a tool to reduce omitted variable bias (OVB)

Recap: Motivation for causal inference (CI)

Motivation for CI

- ✓ How to "get from raw numbers to reliable causal knowledge"? (Angrist & Pischke, 2015, xiii)
- ✓ How to get to *ceteris paribus* / all other things equal condition?

Motivation for CI

- More formally, we are after estimating the difference between a control and a treatment group
- Due to the 'fundamental problem of causal inference' (Holland, 1986), we only ever observe individuals either in their treated $Y_{1i}|D_i = 1$ or their untreated $Y_{0i}|D_i = 0$ state
- We do not observe the counterfactuals $Y_{1i}|D_i = 0$ (the potential outcome of the treated had they not received the treatment) and $Y_{0i}|D_i = 1$, the outcome of the non-treated had they been treated

Motivation for CI

- The naive comparison (the NATE) in non-experimental settings is therefore very likely biased
- Using the constant-effect assumption $Y_{1i} = Y_{0i} + \kappa$ (i.e. assuming no heterogeneous treatment effect (HTE) bias) we can identify the selection bias as follows:

$$\begin{aligned} NATE &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= E[Y_{0i} + \kappa|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= \kappa + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \end{aligned} \tag{1}$$

Motivation for CI

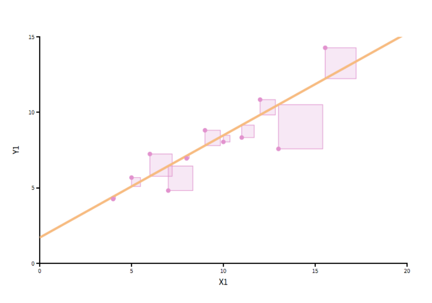
- ✓ The term $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is the selection bias: *everything* that distinguishes the treatment group from the control group in their non-treated states

- ✓ The principal goal of CI methods is the elimination of the selection bias

Ordinary Least Square Regression (OLS)

Ordinary Least Square Regression *aka* Regression

- What does regression do?
- Simple regression: look for the best linear approximation of the relationship between two variables ($\rightarrow$ link to interactive visualization)



Ordinary Least Square Regression *aka* Regression

- Best fitting line is found by minimizing the sum of the squared deviations from a line
- OLS regression is a solution to an optimization problem:

$$Y_i = \beta_0 + \beta_1 X + r_i$$
$$\sum_{i=1} r_i^2 = \min$$

- How does this relate to causal inference?

Regression as a descriptive tool

- Regression may be justified without any reference to causality: as a method for obtaining a best-fitting linear descriptive model of Y given X – the conditional expectation function (CEF) $E(Y|X)$
- The focus here is on predicting Y given specific realizations of X
- Under this use of regression, the outcome Y is not generally thought to be a function of potential outcomes associated with causal states
- No close link between this use of regression and CI; inappropriate to give a causal interpretation to any of the estimated regression coefficients

Regression as a tool for causal inference

- But often, regression models are used to do causal inference! When is this justified?
- Regression can be used to estimate quantities you know from the potential outcomes framework (POF)
- Notably,
  1. The individual treatment effect $ITE = Y_{1i} - Y_{0i}$ can be expressed as a regression equation
  2. The selection bias is intimately linked with the error term (often written as u_i or e_i)
  3. Controlling for potential confounders is a way to reduce selection bias

Regression and the potential outcomes framework (POF)

Regression and the POF

1. The individual treatment effect (ITE) and the regression equation:

$$
\begin{aligned}
Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} && \text{switching equation} \\
&= D_i Y_{1i} + Y_{0i} - D_i Y_{0i} && \text{rearrange} \\
&= Y_{0i} + D_i (Y_{1i} - Y_{0i}) && \text{add } E[Y_{0i}] - E[Y_{0i}]\text{, rearrange} \\
&= Y_{0i} + D_i (Y_{1i} - Y_{0i}) + E[Y_{0i}] - E[Y_{0i}] && \text{use } Y_{1i} = Y_{0i} + \kappa\text{, rearrange} \\
&= E[Y_{0i}] + \kappa D_i + Y_{0i} - E[Y_{0i}] && \text{simplify terms} \\
Y_i &= \underbrace{\alpha}_{E[Y_{0i}]} + \underbrace{\kappa}_{Y_{1i} - Y_{0i}} D_i + \underbrace{u_i}_{Y_{0i} - E[Y_{0i}]} && \text{the classic regression equation!}
\end{aligned}
$$

Regression and the POF

2. The error term and selection bias:

Evaluating the conditional expectation of this equation with treatment switched off and on allows us to see the connection between the error term of the regression and selection bias:

$$E[Y_{1i}|D = 1] = \alpha + \kappa + E[u_i|D = 1]$$
$$E[Y_{0i}|D = 0] = \alpha + E[u_i|D = 0]$$

after deducting the above from each other, we get:

$$\underbrace{E[Y_{1i}|D = 1] - E[Y_{0i}|D = 0]}_{\text{NATE}} = \underbrace{\kappa}_{\text{ATE}} + E[u_i|D = 1] - E[u_i|D = 0]$$

Regression and the POF

2. The error term and selection bias:

Since $u_i = Y_{0i} - E[Y_{0i}]$, we can rewrite the last term:

$$E[u_i|D=1] - E[u_i|D=0] = E[Y_{i0}|D=1] - E[Y_{i0}|D=0]$$

which gives us:

$$\underbrace{E[Y_{1i}|D=1] - E[Y_{0i}|D=0]}_{\text{NATE}} = \underbrace{\kappa}_{\text{ATE}} + \underbrace{E[Y_{i0}|D=1] - E[Y_{i0}|D=0]}_{\text{Selection bias}}$$

This means the regression equation recovers the ATE if it's not affected by selection bias.

Regression and the POF

2. The error term and selection bias:

Going back to the regression equation notation of the causal outcomes, i.e. $Y_i = \alpha + \kappa D_i + u_i$, the unbiasedness requirement can also be expressed as the the conditional mean zero requirement for the error term: $E[u_i|D_i] = 0$.

This means that for κ to capture the causal effect of the treatment D_i, the error has to be independent from the treatment, or, in other words, treatment assignment has to be unaffected by selection bias.

Regression and the POF

3. Controlling for confounders as a way to reduce selection bias

- ✓ Why bother using the regression framework for causal inference, if the problem of the regression estimator is the same as with the naive estimator?
- ✓ Multiple regression allows to control for other variables: this can help to reduce (or even remove) selection bias
- ✓ The multiple regression function can be written as

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \ldots + \beta_n Z_i + u_i$$

or sticking with our notation as:

$$Y_i = \alpha + \kappa D_i + \beta W_i + u_i$$

Regression and the POF

3. Controlling for confounders as a way to reduce selection bias

Think of W_i as an additional factor that we know or assume to influence the relationship between the treatment and the outcome.

Example 1: In an experimental setting, we might know that one group had a higher chance of receiving a treatment. E.g. an experiment that involved randomly distributing loans where the implementing micro-finance company insisted that 70% of participants had to be female.

Regression and the POF

3. Controlling for confounders as a way to reduce selection bias

Example 2: In an observational study, where we are interested in the effect of contact on prejudice, W_i could stand for education, as it is plausible that the educated are both more likely to seek out contact and to hold lower prejudice.

Using the benchmark of the ideal experiment, we can think of these factors as devitations from random assignment: women/the educated are more likely to receive the treatment.

If we don't take into account this deviation, this would mean that treatment assignment is no longer independent of the error term $E[u_i|D_i] \neq 0$ – our estimate for the causal effect is biased.

Regression and the POF

3. Controlling for confounders as a way to reduce selection bias

How does controlling for W_i affect our estimates?

Conditional on W_i, the treatment assignment becomes independent of the error term again:

$E[u_i|D_i, W_i] = E[u_i|W_i]$

(but note that the same is not true for the confounder W_i, i.e. $E[u_i|W_i] \neq 0$ – this is one reason why you should only ever focus strictly on one variable during a given analysis).

Regression and the POF

3. Controlling for confounders as a way to reduce selection bias

We can see this by evaluating the conditional expectation of the multiple regression equation:

$$
\begin{aligned}
Y_i &= \alpha + \kappa D_i + \beta W_i + u_i \\
E[Y_i|D=1] &= \alpha + \kappa + \beta W_i + E[u_i|D_i=1, W_i] \\
E[Y_i|D=0] &= \alpha + \beta W_i + E[u_i|D_i=0, W_i]
\end{aligned}
$$

deduct from each other and resolve:

$$
\begin{aligned}
E[Y_i|D=1] - E[Y_i|D=0] &= \kappa + E[u_i|D_i=1, W_i] - E[u_i|D_i=0, W_i] \\
E[Y_i|D=1] - E[Y_i|D=0] &= \kappa + E[u_i|W_i] - E[u_i|W_i] \\
E[Y_i|D=1] - E[Y_i|D=0] &= \kappa
\end{aligned}
$$

Regression as a tool to reduce omitted variable bias (OVB)

Regression as a tool to reduce OVB

Under a causal inference perpective, the use of regression is to minimize the influence of confounders that can cause selection bias in our estimate of the causal effect.

In other words, we aim to make our treatment variable condionally independent of the error term.

In yet other words, we aim to make our treatment ignorable.

And in yet other words, in CI we use regression to minimize omitted variable bias (OVB).

Regression as a tool to reduce OVB: Example

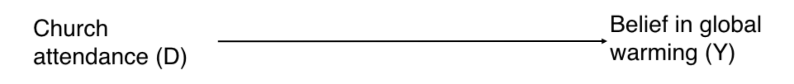
Your starting points for a causally minded regression analysis are:

- ✓ A clear hypothesis that formulates an outcome and a treatment
- ✓ Data including measures for treatment, outcome and potential confounders
- ✓ A causal model detailing how all factors interact, e.g. in the form of a directed acyclic graph (DAG) – much more on this in the next session
- ✓ A statistical model / regression equation

Regression as a tool to reduce OVB: Example

Here:

- Hypothesis: Church attendance reduces belief in anthropogenic global warming
  - Doubtful of humans as cause of global developments
  - Influence of preachers critical of global warming
  - ...
- Data from Egan and Mullin (2012)
- Causal model:



- Statistical model: $Y_i = \alpha + \kappa D_i + u_i$

Regression as a tool to reduce OVB: Example

Note a few things:

- The setting does not approximate an experiment in any way
  - no random assignment, but individual choice
  - hard to know who goes to church and why
  - a lot of potential confounders
- This should make you extra cautious and pay careful attention to the explanation as to how the causal effect is to be recovered
- Similarly difficult-to-test hypotheses are tested all the time in published research – some of them with good identification strategies, some (many) with bad ones

Regression as a tool to reduce OVB: Example

Let's first look at the data:

	mean	sd	min	max
Earth getting warmer?	0.79	0.40	0	1
Attends church	0.89	0.32	0	1
Sex: Male	0.48	0.50	0	1
Party ID: Republican	0.30	0.46	0	1
Departure from normal local temperature (in 10 °F) in week prior to survey	0.38	0.59	-2	4

Summary statistics for selected variables from Egan and Mullin (2012)

Regression as a tool to reduce OVB: Example

Regression results estimating $Y_i = \alpha + \kappa D_i + u_i$:

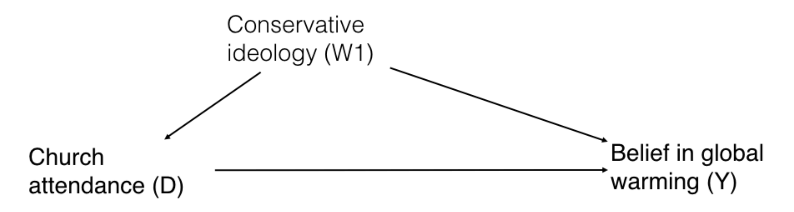
	Naive Model
(Intercept)	0.950^{***}
	(0.030)
Attends church	-0.082^{***}
	(0.016)
R^2	0.004
Num. obs.	6219

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Effect of church attendance on belief in anthropogenic climate change

Regression as a tool to reduce OVB: Example

How credible is this result? Is 8.2% a good estimate for the causal effect? Is the estimate free from confounding? That is, is 'treatment assignment' uncorrelated with the error term, i.e. $E[u_i|D_i] = 0$? Many possible confounders, e.g. ideology:



How does the failure to control for ideology bias our results?

Regression as a tool to reduce OVB

To find out, we analyze omitted variable bias – the difference in the causal effect when controlling for the confounder (the 'long regression') as compared to when not controlling for it (the 'short regression'), which is captured by the following formula (Angrist & Pischke, 2015):

OVB = {Relationship between potential confounder and proposed treatment variable} × {Effect of potential confounder in long regression}

Given the three regressions

$$
\begin{aligned}
Y_i &= \alpha^s + \kappa^s D_i + u_i^s \text{ (short)} \\
Y_i &= \alpha^l + \kappa^l D_i + \beta W_i + u_i^l \text{ (long)} \\
W_i &= \theta + \gamma D_i + e_i \text{ (relationship confounder and treatment)}
\end{aligned}
$$

the OVB is calculated as:

$$OVB = \kappa^s - \kappa^l = \gamma \times \beta$$

Regression as a tool to reduce OVB

The OVB formula is a heuristic tool for gauging the presence and direction of the selection bias.

It's most useful when you cannot actually observe a confounder but want to determine its likely effect.

When does OVB become zero? I.e. $\kappa^s - \kappa^l = \gamma \times \beta = 0$?

1. $\beta = 0$ (no effect of confounder on outcome)
2. $cov(D_i, W_i) = 0$ (treatment and confounder actually independent)

Regression as a tool to reduce OVB

When is OVB negative? I.e. the estimate for the causal effect a) overestimates the true effect when the estimate for the causal effect is negative, b) underestimates the true effect when the estimate is positive: $\kappa^s - \kappa^l = \gamma \times \beta < 0$?

1. γ and β take different signs

When is OVB positive? I.e. the estimate for the causal effect a) underestimates the true effect when the estimate for the causal effect is negative, b) overestimates the true effect when the estimate is positive: $\kappa^s - \kappa^l = \gamma \times \beta > 0$?

1. γ and β both positive
2. γ and β both negative

Regression as a tool to reduce OVB: Example

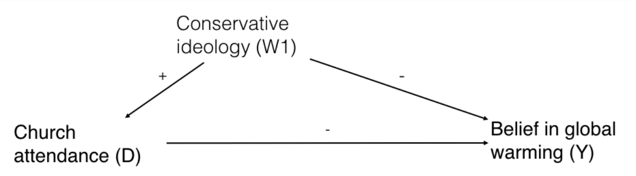
In our example, is the effect of ideology on the outcome zero? Highly unlikely. And is the correlation between church attendance and ideology zero? Also highly unlikely. Estimates likely affected by OVB.

What is the likely direction of the bias?

What are the likely signs on the coefficients for β and γ? (Hint: Use your DAG)

$$Y_i = \alpha^l + \kappa^l D_i + \beta W_i + u_i^l \text{ (long)}$$

$$W_i = \theta + \gamma D_i + e_i \text{ (relationship confounder and treatment)}$$



Regression as a tool to reduce OVB: Example

Our data provides a proxy for ideology, party ID, so we can actually check

	"Short" Model	"Long" Model	DV:RepID
(Intercept)	0.950^{***}	0.955^{***}	0.029
	(0.030)	(0.029)	(0.033)
Attends church	-0.082^{***}	-0.050^{**}	0.141^{***}
	(0.016)	(0.016)	(0.017)
Party ID: Republican		-0.228^{***}	
		(0.011)	
R^2	0.004	0.069	0.010
Num. obs.	6219	6219	6726

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Effect of church attendance on belief in anthropogenic climate change

Regression as a tool to reduce OVB: Example

Estimate of OVB bias: $\kappa^s - \kappa^l = \gamma \times \beta = -0.082 - (-0.050) = -0.228 \times 0.141 = -0.032$

Estimate for κ^s overestimates the effect of church attendance.

The strong change in coefficient is a clear warning sign: Ideology seems to be a highly important confounder, but is likely measured imperfectly with party ID.

This indicates that there are likely unobserved/unobservable variables influencing the estimate of κ. In other words, it remains doubtful that the ‘treatment’ is independent from the error term, even when controlling for ideology, i.e. $E[u_i|D_i, W_i] \neq 0$

Unclear if this can be addressed with controls at all.

Regression as a tool to reduce OVB: Example

Further controls further reduce effect size

	"Short" Model	"Long" Model	"Controlled" Model
(Intercept)	0.950^{***}	0.955^{***}	0.963^{***}
	(0.030)	(0.029)	(0.030)
Attends church	-0.082^{***}	-0.050^{**}	-0.036^{*}
	(0.016)	(0.016)	(0.015)
Party ID: Republican		-0.228^{***}	-0.202^{***}
		(0.011)	(0.012)
Party ID: leans Republican			-0.130^{***}
			(0.016)
Self-rating: very conservative			-0.166^{***}
			(0.021)
Self-rating: conservative			-0.074^{***}
			(0.012)
Self-rating: liberal			0.040^{**}
			(0.015)
R^2	0.004	0.069	0.098
Num. obs.	6219	6219	6219

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Regression as a tool to reduce OVB

Compare these results to the ones using the suggested treatment in Egan and Mullin (2012)

	"Short" Model	"Long" Model	"Controlled" Model
(Intercept)	0.784^{***}	0.853^{***}	0.888^{***}
	(0.006)	(0.007)	(0.008)
Departure from normal local temperature (in 10 °F) in week prior to survey	0.028^{**}	0.021^{*}	0.022^{**}
	(0.009)	(0.008)	(0.008)
Party ID: Republican		-0.230^{***}	-0.202^{***}
		(0.011)	(0.012)
Party ID: leans Republican			-0.130^{***}
			(0.016)
Self-rating: very conservative			-0.168^{***}
			(0.021)
Self-rating: conservative			-0.076^{***}
			(0.012)
Self-rating: liberal			0.041^{**}
			(0.015)
R^2	0.002	0.068	0.099
Num. obs.	6219	6219	6219

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Regression as a tool to reduce OVB: Final thoughts

Egan and Mullin (2012) have a much better claim that their proposed treatment – deviations from local temperatures – may be quasi-randomly distributed, conditional on things like geographical regions.

Other studies study the effect of factors that are due to choices. In such studies, using controls to make the 'treatment' independent of confounders (a strategy dubbed 'selection on observables') becomes much less credible (even though you should still read the study and judge for yourself).

The problem is that we are often interested in 'treatments' that depend on individual choices (e.g. effect of education, vaccination, micro-finance,...)!

Regression as a tool to reduce OVB: Final thoughts

In such cases, very careful analyses are necessary, ideally involving causal graphs and detailed discussions of the direction of the biases that might be effecting the estimated effect.

Because ‘selection into treatment’ will often depend on unobservable/unmeasurable confounders, it might not be possible to present a fully convincing argument based on ‘statistical control’ alone.

In such cases, we will seek to/should use some of the design-based techniques such as instrumental variables or RDD. These techniques seek to isolate the more plausibly random part of the variation in our treatment, and use that part of the variation to predict the outcome. Many of these techniques build upon regression.

Thank you for watching, and see you next Monday!

References

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

Egan, P. J., & Mullin, M. (2012, July). Turning Personal Experience into Political Attitudes: The Effect of Local Weather on Americans' Perceptions about Global Warming. *The Journal of Politics*, *74*(3), 796–809.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960.