Bad Control

Alex

September 30, 2021

What goes wrong with bad controls? Everything!

A Very Simple Example

Let's make some data in just the same way that we typically make data.

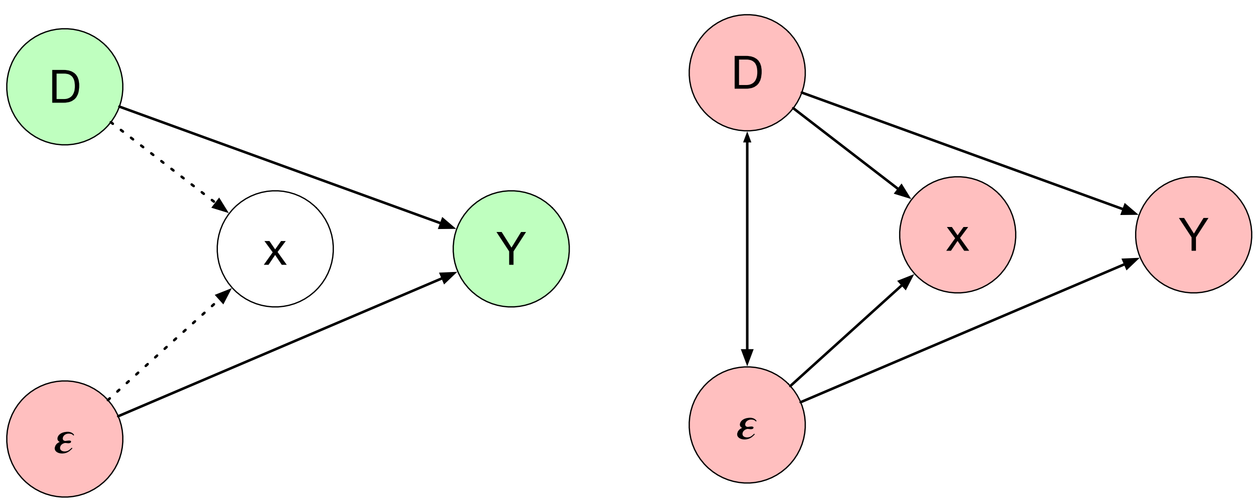
First, let's make and `id` variable, an `epsilon` variable that represents everything in the world that we haven't measured, a `D` variable that represents the assignment to treatment, and a `tau` variable that represents each individuals response to treatment if he or she or it is exposed to the treatment.

```
d <- data.table(
  id  = 1:1000,
  epsilon = rnorm(1000),
  D   = sample(0:1, 1000, replace = TRUE),
  tau = rnorm(n = 1000, mean = 2, sd = 1)
)
```

Now, lets make the *slightly* more complicated quantities, `x`, the bad control, and the potential outcomes. We're going to take some care to create the x-variable. As we show in the figure below, the x-variable is **caused** by both the treatment and other things in the system.

For example, consider the case of the white collar and blue collar jobs. If you're born the heir of a real-estate empire in NY (in this case represented by ϵ), you might be likely, no matter your college attendance or draft status, to end up in a white collar job (in this case represented in X). But, it may also be the case in the data that having been assigned to attend college (in this case represented in D) increases the probability that you get a white collar job.

```
knitr::include_graphics("./coding_bad_controls_diagram.pdf")
```



Suppose that we measure four variables, D, x, Y, ϵ that all have a positive relationship between one another, shown in the figures above. The one exception, and this is important to make this example easy-ish to understand, is that x actually has *no* relationship on Y.

The way we've built this data, if someone is in treatment ($D = 1$), then they're more likely to have a larger value for X and also a larger value for Y. Similarly, if someone has a high ϵ value, then they're more likely to have a higher value for X and also a higher value for Y.

In the figure on the left, where we've included dotted-line relationships between D, X, and ϵ, we're acknowledging that these relationships exist, but we're not going to condition on them. As a result, there is no covariance between D and ϵ, which we represent with no line. In the figure on the right, consider that we've conditioned on the X variable. As a result, there is a solid line relationship between ϵ and X; but so too is there a solid line relationship between ϵ and D **after we condition on** X.

Perhaps think of it this way – if we don't condition on X then any value of ϵ is possible, and so too is any value of D. But, after we set the X to a value by conditioning on it, some arrangements are more likely to have been the case than others.

- If we don't condition on X, and if ϵ is low, it could still be the case that the unit was in treatment or was in control. We don't have any conditional information.
- But, if we condition so that the units we're considering have a $X = 1$, and if ϵ is low, then *we can know* that the person was more likely to be in the treatment group than the control group.

Let's make these, carefully. Create the variable x, the realization of the variable X to be a binary indicator that takes the values of either 1 or 0. And, suppose that the probability of being a 1 increases if you're in treatment and also if you've got a high value of epsilon. (I'm going to scale these probabilities using a `pnorm` call, just so that the values for the `prob` argument to `rbinom` conforms to be a valid probability statement.

```
d[ , x := rbinom(n = 1000, size = 1, prob = pnorm(2*D + epsilon))]
```

Questions for Understanding

1. If we examine the relationship between x and D with a regression, should the relationship be positive, negative, or we don't know?
2. If we examine the relationship between x and ϵ with a regression, should the relationship be positive, negative, or we don't know?
3. **Challenge**: If we examine the regression of `x ~ D + epsilon` what will be the nature of each variables' relationship with x? Will they be the same, different, and will they be positive, negative, or we don't know?

```
mod_d    <- d[ , lm(x ~ D)]
mod_x    <- d[ , lm(x ~ epsilon)]
mod_both <- d[ , lm(x ~ D + epsilon)]
```

```
## I've commented this out so that you can uncomment it after you've answered the
## questions for understanding.
##
## stargazer(mod_d, mod_x, mod_both, type = 'text')
```

Make Potential Outcomes

To this point, we haven't actually built any outcomes, we've been principally concerned with building the *upstream* data for our causal system. Let's make those potential outcomes now.

For everybody, their potential outcomes to control are a combination of some individual idiosyncratic value (represented by the random draw in this code), plus their x value, plus all the other features in the world that we haven't measured, their ϵ.

```r
d[ , y0 := runif(1000, min = 0, max = 10) + epsilon]
```

Question for understanding

Do we **need** in a strict sense to define the potential outcomes to treatment on this data? Or, would that information simply be a linear combination of some data that we already have in hand?

Finally, let's produce a measured outcome value for Y.

```r
d[ , Y := as.numeric(NA)]
d[D==0, Y := y0]
d[D==1, Y := y0 + tau]
```

Finally, Estimate Causal Effects

In the world that we've occupied to this point, whether we *do* or *do not* condition on a variable, an experiment that we've executed that has successfully randomized the treatment should produce a reliable causal estimate. That is, in this setup, **if we estimate a model, it should provide us with an unbiased estiamte of τ, or 2.**.

```r
unconditional_model <- d[ , lm(Y ~ D)]
```

But, what is going to happen if we use a *cough cough bad...* control variable to "clean up our estimate". After all, if controls only *increase the accuracy of our predictions*, then this will improve our model right?

```r
bad_control_model <- d[ , lm(Y ~ D + x)]
```

What are other ways that we could see this relationship? Well, earlier we talked about "setting x to be equal to 1". Let's try that on this data. Among the people who have an x value that is one, what is the causal relationship between D and Y?

```r
x_equals_one_model  <- d[x==1, lm(Y ~ D)]
x_equals_zero_model <- d[x==0, lm(Y ~ D)]
```

```r
stargazer(
  unconditional_model, bad_control_model, x_equals_one_model, x_equals_zero_model,
  type = 'latex', header = FALSE, table.placement = 'h',
  digits = 2,
  add.lines = list(c("Data Subset", "All", "All", "$x==1$")),
  title = 'Good and Bad Control Models'
  )
```

Table 1: Good and Bad Control Models

	Dependent variable:			
	Y			
	(1)	(2)	(3)	(4)
D	2.07***	1.51***	1.63***	0.98*
	(0.20)	(0.23)	(0.26)	(0.53)
x		1.22***		
		(0.25)		
Constant	4.88***	4.31***	5.46***	4.38***
	(0.14)	(0.18)	(0.21)	(0.18)
Data Subset	All	All	$x == 1$	
Observations	1,000	1,000	686	314
R^2	0.09	0.12	0.05	0.01
Adjusted R^2	0.09	0.11	0.05	0.01
Residual Std. Error	3.20 (df = 998)	3.17 (df = 997)	3.22 (df = 684)	3.04 (df = 312)
F Statistic	104.58*** (df = 1; 998)	65.62*** (df = 2; 997)	39.79*** (df = 1; 684)	3.47* (df = 1; 312)
Note:				*p<0.1; **p<0.05; ***p<0.01