Review of Omitted Variable Bias and Regression Discontinuity Designs

Andrew M. Camp

Updated 2021-11-16

Omitted Variables Bias (OVB)

OVB occurs when a regressor X is correlated with some omitted variable that is a determinant of Y.

OVB presents a challenge to identification and produces a biased result because it is a violation of the OLS zero conditional mean assumption ($\mathbb{E}[u|x] = 0$). Often, OVB is mentioned with the implication that longer regression equations with more regressors are less subject to bias and can somehow be viewed as "more causal." While these notions of "short" and "long" equations are misleading, those are the terms that are commonly used and will be used here. Stealing an example from *Mostly Harmless Econometrics*, assume that individuals' wages can be fully explained by years of schooling, ability, and a stochastic error term.

$$W_i = \beta_0 + \beta_1(Schooling_i) + \beta_2(Ability_i) + e_i$$

Ability is difficult to measure, and so we may be tempted to estimate the returns to schooling using a shorter regression:

$$\tilde{W}_i = \tilde{\alpha}_0 + \tilde{\alpha}_1(Schooling_i) + u_i$$

Results from this second regression are biases because as schooling increases, so does ability, and we might expect $\mathbb{E}[u_i|Schooling_i]$ increases as schooling increases. The short and long equations shown above can be related as:

$$\frac{Cov(W_i, Schooling_i)}{V(Schooling_i)} = \tilde{\alpha}_1 = \beta_1 + \beta_2(\delta)$$

This is often referred to a the OVB formula[1] and is sometimes memorized as "*short equals long plus the effect of omitted times the regression of omitted on included.*"

From this, bias in the estimate of $\tilde{\alpha}_1$ can be algebraically derived as $\text{Bias}(\tilde{\alpha}_1) = \beta_2(\delta)$ where the parameter δ is obtained through the regression of $Ability_i$ on $Schooling_i$ and is directly related to the correlation of the included and omitted variable. This gives us two insights:

1. When $\beta_2 = 0$ or $\tilde{\delta} = 0$, there is no bias.

[1] See Andrew Crane-Droesch's class notes for a concise derivation of the OVB formula.

2. Direction of bias can be predicted if we are willing to assume both the sign of correlations between schooling, ability, and wages.

Given prior knowledge, we might assume that schooling and ability are positively correlated and that individuals with more ability earn higher wages. We would then predict the estimate of α_1 to be positively biased because both β_2 and δ are positive. The table below summarizes this relationship.

Table 1: Direction of bias from correlations

	$Corr(S_i, A_i) > 0$	$Corr(S_i, A_i) < 0$
$Ability_i$ has a + effect on W_i	+ Bias	− Bias
$Ability_i$ has a − effect on W_i	− Bias	+ Bias

Regression Discontinuity Designs (RDDs)

RDDs rely upon plausibly random sorting across some threshold value to estimate .

Public and private policies often rely upon strict thresholds for implementation. Drivers over 0.08 BAC receive a DUI, laws take effect on particular dates, and winners of elections may be decided by just a fraction of a percentage point. For events close to the threshold, which side they land on can conceivably be the result of randomness. Additionally, RDDs can be used when some functional relationship between the running variable and outcome is known or can be anticipated from institutional/economic knowledge.

Terms

Continuity Assumption

Treatment Effects

Even in cases where sorting across the threshold is exogenous, the relationship between the running variable and outcome variable can serve to confound identification of causal effects. We'll consider Cody Tuttle's 2019 paper[2] looking at the effect of a welfare reform on recidivism. In 1996, Florida changed regulations governing food stamp eligibility. If someone was convicted for a crime on or after August 23rd, 1996 they became permanently ineligible to relieve food stamps[3]. This ineligibility may result in economic hardship after release and increase the probability that an individual re-offends.

Here, the running variable (date a crime was committed) and the outcome variable (recidivism rate) have a natural pre-existing relationship. We see recidivism rates near zero towards the right of the graph

[2] https://www.aeaweb.org/articles?id=10.1257/pol.20170490

[3] Food stamps are now referred to as SNAP benefits. I'll probably use the terms interchangeably and apologize if that's confusing.

because individuals who have committed crimes recently will still be incarcerated and thus unable to re-offend. For causal identification, we need to model this relationship in our RDD.

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```
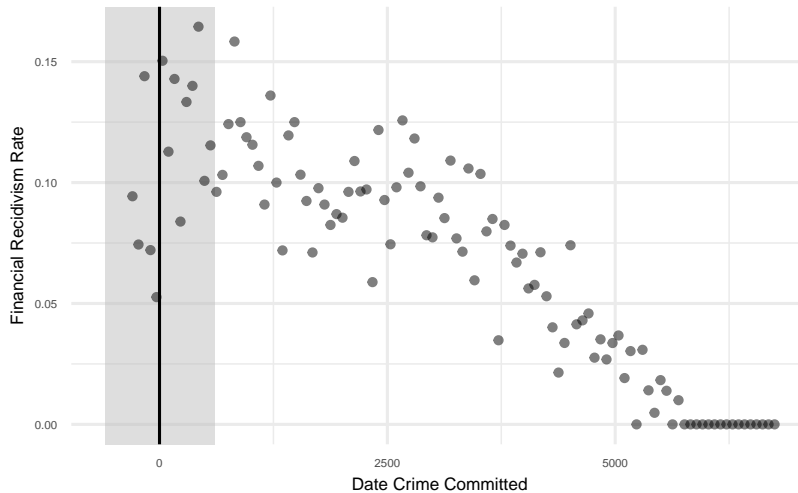

The other thing to remember when thinking about treatment effects during RDD is **who** estimates are valid for. Part of the typical estimation process is to restrict the data used to a range or window of observations surrounding the cutoff. In this example, that window is shaded in gray. Restricting the window is essential to minimize bias from unobservables and imperfect specification of the functional relationship between the running and outcome variables. However, this reduces the *external validity* of our estimate. Treatments effects estimated though RDD are **local average treatment effects (LATEs)**.

Specification and Estimation via OLS

To estimate treatment effects in this scenario, we can begin with a relatively simple regression equation. Let $Pr(Y_i)$ represent the probability that individual i commits a financially-motivated crime (such as selling drugs), D_i be a dummy variable taking a value of 1 if individual i's original crime occurred on or after August 23, 1996 and 0 otherwise, and $T_i = \text{CrimeDate}_i - \text{August } 23, 1996$ so that negative values indicate the crime was committed prior to the policy change and positive values otherwise.

$$Pr(Y_i) = \beta_0 + \beta_1 D_i + \beta_2 T_i + e_i$$
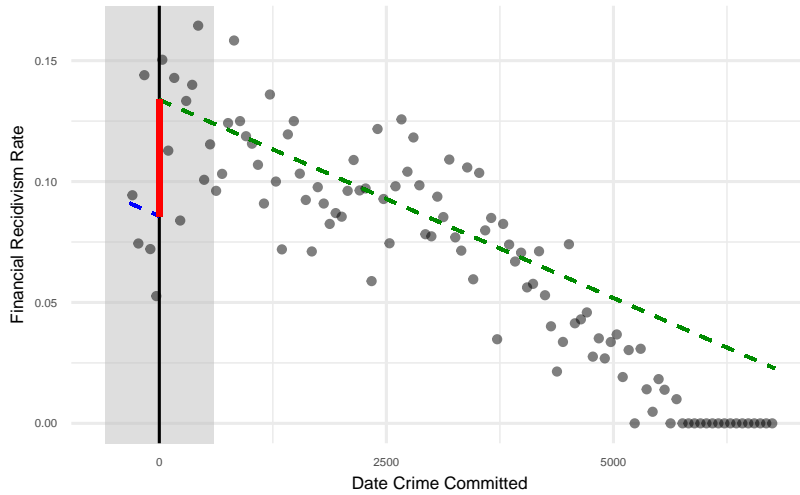
A problem with the RDD model specified above is that it structurally imposes the same slope (β_2) for both the treated and untreated groups. This is fine if we expect treatment to **only** shift outcomes but would fail to capture dynamic treatment effects. We can allow for separate slopes on either side of the cutoff with the following specification.

$$Pr(Y_i) = \beta_0 + \beta_1 D_i + \beta_2 T_i + \beta_3 (D_i \times T_i) + e_i$$