Lecture Note 19 — Causal Inference Using the Regression Discontinuity Design

David Autor, Massachusetts Institute of Technology

The Regression Discontinuity Design

During the *in-class* component of this lecture/topic, we will study the 2010 paper on securitization and subprime mortgages by Keys, Mukherjee, Seru, and Vig. In this lecture note, we'll review the framework that Keys et al. use for causal inference, which is called the Regression Discontinuity design (RD). You can now add RD to your toolbox along with Difference-in-Differences (DD), Randomized Control Trials (RCT), and Instrumental Variables (IV).

As usual, we seek to estimate the causal effect of a treatment. We posit that for each individual i, there exists a pair of potential outcomes: Y_{i1} for what would occur if i were exposed to the treatment and Y_{i0} if i were not exposed. The causal effect of the treatment is represented by the difference $T = Y_{i1} - Y_{i0}$. The fundamental problem of causal inference (FPCI) is that we cannot observe both Y_{i1} and Y_{i0}.

We have so far handled the FPCI using three primary techniques: randomization into treatment and control groups (RCT), difference-in-difference estimation, and instrumental variables estimation. Each method attempts to find treated and control units that are in expectation comparable—that is, their potential outcomes if treated (or if untreated) are expected to be the same—and then contrasts outcomes among those treated relative to those not treated to estimate the average effect of treatment on the treated (ATT).

The Regression Discontinuity (RD) estimator takes a fresh approach to identifying a causal relationship when the treatment and control groups do *not* have potential outcomes that are identical in expectation. It instead looks for units that are *arbitrarily close* in terms of their potential outcomes and yet are treated differently (one assigned to treatment, the other assigned to control) due to some bright line rule that determines assignment. This situation occurs more commonly than one might expect. For example, the result of a national election can be decided by a single vote, or the cutoff for which children are allowed to enter 1st grade in a given year may depend on whether they were born before or after midnight on September 1 six years earlier. Arbitrary cutoffs are inevitable for administrative purposes. A driver either is or is not speeding. A potential candidate for office either does or does not have the requisite number of signatures to get on the ballot. A library book is not overdue until the moment that it is.

While arbitrary cutoffs are necessary for administration, why are they useful for economists? Glad you asked. Define a variable X that is used to determine the cutoff above/below which a person (or unit) i is or is not assigned to treatment. For example, X could be the percentage of voters for candidate A or X could be the exact hour of birth. We will refer to X as the *run variable*, and we'd like that variable to be continuous.

Imagine there are two underlying relationships between potential outcomes and treatment, represented by $E[Y_{i1}|X_i]$ and $E[Y_{i0}|X_i]$. Thus at each value of X_i, the causal effect of treatment is $E\left[T|X_i = x\right] = E[Y_{i1}|X_i = x] - E[Y_{i0}|X_i = x]$. Let's say that individuals to the right of a cutoff c (e.g., $X_i \geq 0.5$) are exposed to treatment, while those to the left ($X_i < 0.5$) are denied treatment. We therefore observe $E\left[Y_{i1}|X_i\right]$ to the right of the cutoff and $E\left[Y_{i0}|X_i\right]$ to the left of the cutoff.

As we consider units i that are arbitrarily close to the threshold, it may be reasonable to assume that:

$$
\begin{aligned}
\lim_{\varepsilon \downarrow 0} E\left[Y_{i1}|X_i = c + \varepsilon\right] &= \lim_{\varepsilon \uparrow 0} E\left[Y_{i1}|X_i = c + \varepsilon\right], \\
\lim_{\varepsilon \downarrow 0} E\left[Y_{i0}|X_i = c + \varepsilon\right] &= \lim_{\varepsilon \uparrow 0} E\left[Y_{i0}|X_i = c + \varepsilon\right].
\end{aligned}
$$

That is, for units that are *almost identical*, we may be willing to assume that had both been treated (or not treated), their outcomes would have been arbitrarily similar. If this assumption is plausible, we can form a Regression Discontinuity estimate of the causal effect of treatment on outcome Y using the contrast:

$$
\hat{T} = \lim_{\varepsilon \downarrow 0} E\left[Y_i|X_i = c + \varepsilon\right] - \lim_{\varepsilon \uparrow 0} E\left[Y_i|X_i = c + \varepsilon\right],
$$

which in the limit is equal to:

$$
T = E\left[Y_{i1} - Y_{i0}|X_i = c\right].
$$

The RD estimator estimates the causal effect of a treatment as the 'jump' in an outcome variable, Y, as near-identical units on one side of a discontinuity, c, are allocated to treatment while those on the other side are allocated to non-treatment. Note that while RD estimation does estimate the treatment effect given that $x_i = c$, if the treatment effect is not the same for everyone, it will not give you the average treatment effect on the treated. For example, imagine you are studying the effect of a scholarship on student grades. If you randomly assign scholarships, you would get the average treatment effect for the entire sample of students (i.e the average treatment effect on the treated). If, instead, scholarships are given to students with SAT scores above 2100, and you use an RD design, you will get the treatment effect on those students with SAT scores of 2100, but not the average treatment effect for all students who received the scholarship.

Keys et al. use the RD estimator to test for adverse selection in the market for subprime mortgages.