Lecture 4b: Evaluating Treatment Effects Using the Regression Discontinuity Method

Reference: Hahn, Todd & Van der Klaauw EMA 2001.

As in the previous two lectures, we are interested in evaluating the effect of some binary treatment variable, here x_i, on outcome y_i.

$x_i = 1$ if *i* is treated and 0 if *i* is untreated. y_{1i} will represent the outcome for agent *i* if treated, and y_{0i} the outcome for *i* if untreated.

As before, our difficulty in identifying the true treatment effect is the missing counterfactual observation: y_{0i} if *i* is treated, y_{1i} if *i* is untreated.

<u>Applications of RD</u>

→Van der Klaauw (2002) uses an RD approach to estimate the effect of financial aid offers (note: not a binary treatment, aid offers vary) on students' decisions to accept admission to a given college (note: a binary outcome measure). He exploits multiple discontinuities in an administrative formula that determines aid based on SAT score, GPA, & other components.

→Angrist and Lavy (1999) estimate the effect of class size on student test scores, with identification coming from a rule requiring that one classroom be added in a school whenever average class size exceeds a predetermined threshold. Here class size is a discontinuous (and note: nonmonotonic) function of enrollment in the student's school.

→Black (1999) uses an RD approach to estimates parents' willingness to pay for school quality by comparing housing prices near school district boundaries. Clearly in this case school quality is discontinuous in geographic space at district borders, and housing prices are the outcome measure of interest.

We will consider two regression discontinuity designs, the sharp design and the fuzzy design.

What they have in common is that (the probability of) treatment depends on some underlying & observable variable *z*, which will be assumed continuous. Note that *z need not be independent of* y_i. The instrument for treatment here will be the discontinuity in the dependence of x_i on *z*, not *z* itself.

<u>Sharp Design</u>

With a sharp regression discontinuity design, x_i depends deterministically on *z*:

$$x_i = f(z_i)$$

The point z_0 where $f(z)$ is discontinuous is assumed known.

<u>Fuzzy Design</u>

With a fuzzy design, x_i remains random given z_i BUT

$$f(z) \equiv E[x_i \mid z_i = z] = \Pr[x_i = 1 \mid z_i = z]$$

is discontinuous at known value z_0.

The sharp and fuzzy designs differ in that in the sharp design the treatment assignment is deterministic given *z*, while in the fuzzy design the treatment assignment may depend on additional factors unobserved by the econometrician.

Their common feature is that treatment probability $\Pr[x_i = 1 | z_i = z]$ is a function of underlying continuous & observable variable *z*, and that function is discontinuous at z_0, i.e.

Assumption (RD): (i) Limits $x^{+} = \lim_{z \to z_0^{+}} E[x_i | z_i = z]$ and $x^{-} = \lim_{z \to z_0^{-}} E[x_i | z_i = z]$ exist. (ii) $x^{+} \neq x^{-}$.

Examples: An identifying assumption in Van der Klaauw (2002) would be that the financial aid award given to a student with an SAT score approaching (for example) 1300 from above differs from that given to a student with an SAT score approaching 1300 from below.

In Angrist and Lavy (1999), an identifying assumption would be that the class size for a student in a school with a number of pupils approaching (for example) 800 from above differs from that of a student in a school with a # of pupils approaching 800 from below.

We will look at identification under the fuzzy design, keeping the sharp design in mind as a special case.

Constant Treatment Effects

Rewrite outcome for individual *i* as

$$y_i = \alpha_i + x_i \beta_i, \text{ with } \alpha_i \equiv y_{0i} \text{ \& } \beta_i \equiv y_{1i} - y_{0i}.$$

Suppose treatment effect β is constant across individuals. This is the constant treatment effect case.

(Recall that in our previous discussion in the case of variable treatment effects we generally took as our goal the identification of the average treatment effect, or the average treatment effect among the treated, because the identification requirements for measures of the distribution of treatment effects were too great.

Additionally, we saw that when the treatment effect was constant across individuals the average treatment effect among the treated was equal to the average treatment effect.)

One assumption in addition to the discontinuity of *f(z)* at z_0 will be required to establish nonparametric identification of the treatment effect in this case:

Assumption (A1): $E[\alpha_i \mid z_i = z]$ is continuous in *z* at z_0.

This assumption is valid where we have reason to believe that persons close to threshold z_0 are similar, and thus would experience similar outcomes absent treatment.

Theorem 1: Suppose $\beta_i = \beta \ \forall i$, and (RD) and (A1) hold. Then

$$\beta = \frac{y^+ - y^-}{x^+ - x^-},$$

where $y^+ \equiv \lim_{z \to z_0^+} E[y_i \mid z_i = z]$ and $y^- \equiv \lim_{z \to z_0^+} E[y_i \mid z_i = z]$.

Proof: Let $e > 0$ be an arbitrarily small number. The mean difference in outcomes for those above and below the discontinuity point is

$$
\begin{aligned}
& E[y_i \mid z_i = z_0 + e] - E[y_i \mid z_i = z_0 - e] \\
& = \beta \cdot \left\{ E[x_i \mid z_i = z_0 + e] - E[x_i \mid z_i = z_0 - e] \right\} \\
& + \left\{ E[\alpha_i \mid z_i = z_0 + e] - E[\alpha_i \mid z_i = z_0 - e] \right\}.
\end{aligned}
$$

Under (A1),

$$
\begin{aligned}
& \lim_{z \to z_0^+} E[y_i \mid z_i = z] - \lim_{z \to z_0^-} E[y_i \mid z_i = z] \\
& = \beta \cdot \{ \lim_{z \to z_0^+} E[x_i \mid z_i = z] - \lim_{z \to z_0^-} E[x_i \mid z_i = z] \}.
\end{aligned}
$$

The conclusion follows from here. The denominator is nonzero by (RD).

For the sharp design, $x^+ = 1$ and $x^- = 0$. Here the treatment effect is identified simply by

$$\beta = y^+ - y^-.$$

Variable Treatment Effects

This is the case in which β_i need not be common across individuals.

See Hahn, Todd and Van der Klaauw for a discussion of the additional assumption required for nonparametric identification of the treatment effect at z_0 in this case (actually, 2 possible assumptions are presented, each of which suffices to identify the average treatment effect at z_0).

An important result for the variable treatment effects case is that only treatment effects at z_0 can be identified nonparametrically using the regression discontinuity method.

Estimation

In both the sharp & fuzzy designs, the treatment effect is identified by

$$\frac{y^+ - y^-}{x^+ - x^-}.$$

Therefore with consistent (& nonparametric) estimators of the four one-sided limits $\hat{y}^+, \hat{y}^-, \hat{x}^+, \hat{x}^-$, the treatment effect can be estimated consistently (& nonparametrically) by

$$\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-}.$$

Consider one possibility for the nonparametric estimator for the limits, the one-sided uniform kernel. Here the estimates of the limits are equivalent to

$$\hat{y}^{+} = \frac{\sum_{i \in \Lambda} y_i w_i}{\sum_{i \in \Lambda} w_i}, \quad \hat{y}^{-} = \frac{\sum_{i \in \Lambda} y_i (1 - w_i)}{\sum_{i \in \Lambda} (1 - w_i)},$$
$$\hat{x}^{+} = \frac{\sum_{i \in \Lambda} x_i w_i}{\sum_{i \in \Lambda} w_i}, \quad \hat{x}^{-} = \frac{\sum_{i \in \Lambda} x_i (1 - w_i)}{\sum_{i \in \Lambda} (1 - w_i)},$$

where Λ is the subsample such that $z_0 - h < z_i < z_0 + h,\ h > 0$ is the bandwidth, & $w_i \equiv 1(z_0 < z_i < z_0 + h)$.

An interesting note is that this estimator is numerically equivalent to an IV estimator for the regression of y_i on x_i in subsample Λ & using $w_i \equiv 1(z_0 < z_i < z_0 + h)$ as an instrument.

Hahn, Todd & Van der Klaauw note that the one-sided uniform kernel estimator for the limits causes the treatment effect estimate based on these limits to have undesirable properties, and propose an alternative approach estimating the limits by local linear regression.

Application: "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach," Wilbert van der Klaauw, IER 2002

Research question: What is the effect of the level of offered financial aid on the propensity of an admitted student to enroll in a selective college (as opposed to a competing college)?

Why not run a Probit of enrollment on aid?

For these reasons, Leslie & Brinkman in a 1988 survey find large variation in estimates of tuition and aid effects on enrollment.

Aid effects are generally positive and significant. However, magnitures vary from:

Negligible Seneca & Taussig 1987; Parker & Summers 1993

Considerable Ehrenberg & Sherman 1984; Moore et al 1991

Moreover, aid effects for different institutions may differ based on their tuition, characteristics of their applicant pools, and characteristics of their competitors.

Several studies (Leslie & Brinkman 1988; Schwartz 1985 7 6; St. John 1990; McPherson & Shapiro 1991) have found high school grads from higher income families to be significantly less sensitive to financial aid than grads from lower income families.

van der Klaauw argues that this literature suffers from a severe omitted variables problem. The studies listed above derived their disparate estimates from regression specifications that included very different information on students' characteristics and options.

His solution:

(i) Look at ONE selective East Coast college.

(ii) Turn to the college's discontinuous financial aid formula for plausibly exogenous variation in aid offers.

(iii) Promote desirable features of this estimation approach

- It is valid no matter what student characteristic and option regressors are omitted.

- It does not rely on arbitrary exclusion restrictions.

- It does not rely on functional form assumptions or distributional assumptions on error terms [My note: despite dealing with a *discrete dependent variable*!]

Regarding (iii), it is clear by now that this paper is not only about financial aid effects. It is an early regression discontinuity paper among the recent RD wave in economics, and some of its purpose is the sale of the estimator.

To this end, the paper also demonstrates that Angrist & Krueger's well-known 1991 evaluation of the returns to schooling based on variation in legal drop-out ages & children's birth dates can be interpreted as a Regression Discontinuity estimation.

<u>Model</u>

Let F be discretionary aid offered by college X and F^o be the aid offered by the next best competing college.

Then the change in utility for individual I associated with enrolling in college X is

$$EN_i^* = \delta_i(F_i - F_i^o) + \nu_i,$$

where ν_i represents all utility differences between choosing the two colleges not based on financial aid.

Since F_i^o is unobserved, rewrite

$$EN_i^* = \delta_i F_i + u_i,$$

where $u_i = \nu_i - \delta_i F_i^o$.

Note that F_i^o is likely correlated with F_i , and therefore F_i^o is likely correlated with u_i. F_i may also be correlated with v_i.

EN_i^* is a latent variable. The observed decision is

$$EN_i = 1 \text{ if } EN_i^* > 0;$$
$$EN_i = 0 \text{ otherwise.}$$

Therefore

$$\Pr(EN_i = 1) = \Pr(\delta F_i + u_i > 0)$$
$$\Pr(EN_i = 1) = 1 - \Pr(\delta F_i + u_i > 0)$$

<u>Financial Aid Policy</u>

College *X* first establishes an index score

$$S_i = \phi_0 \times (\text{first 3 digits of total SAT score}_i) + \phi_1 \times \text{GPA}_i.$$

The college then sets 3 cutoffs, $\overline{S_1} < \overline{S_2} < \overline{S_3}$.

Aid is allocated by index score region, with students above $\overline{S_3}$ eligible for the most aid.

Aid is then modified to account for family income, minority status, athletic skill, and, finally, the overall strength of the student's admissions package.

The latter may itself depend on SAT & GPA.

Thus we can write

$$F_i = E[F_i \mid S_i] + e_i = f(S_i) + \gamma_1 1\{S_i \geq \overline{S_1}\} + \gamma_2 1\{S_i \geq \overline{S_2}\} + \gamma_3 1\{S_i \geq \overline{S_3}\} + e_i \quad (1.1)$$

<u>Data & Estimation</u>

Based on assumption (RD), it is appropriate to apply the method to data which look something like Figure 1.

Figure 2 illustrates the difference between the sharp RD approach and the fuzzy RD approach.

[What version of the estimation approach do you expect will be applied?]

One way of describing the difference is that the sharp design applies to selection on observables only,

$$\Pr(T = 1 \mid S) = 1(S \geq \overline{S}),$$

where the fuzzy design applies to selection on both observables AND unobservables, so that

$$\Pr(T = 1 \mid S) = f(S, \overline{S})$$

with f a step function, as in Figure 2.

<u>Data Set</u>

Table 1 gives means among those who do and don't enroll in admissions categories 1-4, as determined by the cutoffs.

Filers and non-filers are those who do and do not submit the Free Application for Federal Student Aid (FAFSA).

FAFSA: This form is needed to receive US government grants nad loans for college. Non-filers tend to have family income that is high enough to preclude eligibility for federal aid, or international students.

Table 1 enroll v. not enroll averages suggest an effect of aid on enrollment, though we have not yet controlled for other factors.

Figure 3 shows the raw data on filers' *S* and *F*'s. The raw data already reveal something of a step function.

To summarize the shape of *F(S)* more clearly, van der Klaauw applies the following spline smooth to the data:

$$g(S) = \min_{\Gamma} \sum_{i=1}^{n} (F_i - g(S_i))^2 + \lambda \int (g''(S))^2 dS ,$$

where Γ represents the class of all twice differentiable functions over the observed domain of *S*.

The choice of smoothing parameter λ:

van der Klaauw trades off

- less smoothing (lower values of λ) but sharper step jumps

against

- more smoothing (higher values of λ) but less pronounced step jumps.

Figure 4 shows the same relationship as Figure 3, but for the non-filers. Here the step function is even clearer.

Estimation

Fuzzy RD with multiple cutoffs.

Table 2 shows parameter estimates for the aid function. The γ's are the cutoff parameters from (1.1) above. The ψ's are parameters on higher-order polynomials in S and the cutoffs.

So there is dependence of financial aid on the cutoffs based on these γ's.

A first shot at the dependence of enrollment on financial aid as indicated by behavior around the cutoffs is a graph of enrollment rates over the S range.

Figure 7 shows *EN(S)*, based on both a piecewise cubic regression and a nonparametric spline smooth of the data.

Behavior around the cutoffs?

Figure 8 does the same for non-filers.

Estimation equation

$$\widehat{\delta}(\overline{S}) = \frac{\lim_{S\downarrow\overline{S}} E[EN\,|\,S] - \lim_{S\uparrow\overline{S}} E[EN\,|\,S]}{\lim_{S\downarrow\overline{S}} E[F\,|\,S] - \lim_{S\uparrow\overline{S}} E[F\,|\,S]} \qquad (1.2)$$

identifies the *local* average treatment effect at an admissions score of $\overline{S}$.

Local Wald estimator uses a one-sided uniform Kernel estimator for each of the limits in (1.2).

Table 3 reports estimates for the effect of financial aid *at* each of the 3 discontinuity points for both filers and non-filers.

Bandwidth choice and size of standard errors?

<u>Validity of estimation assumptions</u>

Figure 3 demonstrated the validity of assumption (*RD)* for these data. Is assumption (*A1*) valid?

Do students *know* the values of $\overline{S}_1$, $\overline{S}_2$ & $\overline{S}_3$?

<u>Sensitivity</u>
Here largely about sensitivity to financial aid functional form, bandwidth choices, etc. See paper.

Next we continue our discussion of financial aid with

Carneiro & Heckman 2002

Brown, Mazzocco, Scholz, & Seshadri 2006 / Brown, Scholz, & Seshadri 2006