"A Course in Applied Econometrics"

Lecture 12

Regression Discontinuity Designs

Guido Imbens

IRP Lectures, UW Madison, August 2008

1

Outline

1. Introduction

2. Basics

3. Graphical Analyses

4. Local Linear Regression

5. Choosing the Bandwidth

6. Variance Estimation

7. Specification Checks

1

---

1. Introduction

A Regression Discontinuity (RD) Design is a powerful and widely applicable identification strategy.

Often access to, or incentives for participation in, a service or program is assigned based on transparent rules with criteria based on clear cutoff values, rather than on discretion of administrators.

Comparisons of individuals that are similar but on different sides of the cutoff point can be credible estimates of causal effects for a specific subpopulation.

Good for internal validity, not much external validity.

Long history in Psychology literature (Thistlewaite and Campbell, 1960), early work by Goldberger (1972), recent resurgence in economics.

2

---

2. Basics

Two potential outcomes $Y_i(0)$ and $Y_i(1)$,
causal effect $Y_i(1) - Y_i(0)$,
binary treatment indicator W_i, covariate X_i,
and the observed outcome equal to:

$$Y_i = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases} \tag{1}$$

At $X_i = c$ incentives to participate change.

Two cases, **Sharp Regression Discontinuity**:

$$W_i = 1\{X_i \geq c\}. \tag{SRD}$$

and **Fuzzy Regression Discontinuity Design**:

$$\lim_{x \downarrow c} \Pr(W_i = 1 | X_i = x) \neq \lim_{x \uparrow c} \Pr(W_i = 1 | X_i = x), \tag{FRD}$$

3

Sharp Regression Discontinuity

Example (Lee, 2007)

What is effect of incumbency on election outcomes? (More specifically, what is the probability of a Democrat winning the next election given that the last election was won by a Democrat?)

Compare election outcomes in cases where previous election was very close.

SRD

Key assumption:
$\mathbb{E}[Y(0)|X = x]$ and $\mathbb{E}[Y(1)|X = x]$ are continuous in x.

Under this assumption,

$$\tau_{\mathsf{SRD}} = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]. \text{ (FRD estimand)}$$

The estimand is the difference of two regression functions at a point.

Extrapolation is unavoidable.

Fuzzy Regression Discontinuity

Examples (VanderKlaauw, 2002)

What is effect of financial aid offer on acceptance of college admission.

College admissions office puts applicants in a few categories based on numerical score.

Financial aid offer is highly correlated with category.

Compare individuals close to cutoff score.

FRD

What do we look at in the FRD case: ratio of discontinuities in regression function of outcome and treatment:

$$\tau_{\mathsf{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[W_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[W_i|X_i = x]}.$$

(FRD Estimand)

Interpretation of FRD (Hahn, Todd, VanderKlaauw)

Let $W_i(x)$ be potential treatment status given cutoff point x, for x in some small neigborhood around c (which requires that the cutoff point is at least in principle manipulable)

$W_i(x)$ is non-increasing in x at $x = c$.

A complier is a unit such that

$$\lim_{x \downarrow X_i} W_i(x) = 0, \quad \text{and} \quad \lim_{x \uparrow X_i} W_i(x) = 1.$$

Then

$$\frac{\lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[W_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[W_i | X_i = x]}$$

$$= \mathbb{E}[Y_i(1) - Y_i(0) | \text{unit } i \text{ is a complier and } X_i = c].$$

External Validity

The estimatand has little external validity. It is at best valid for a population defined by the cutoff value c, and by the sub-population that is affected at that value.

FRD versus Unconfoundedness

$$Y_i(0), Y_i(1) \ \perp\!\!\!\perp \ W_i \ \Big| \ X_i. \qquad \text{(unconfoundedness)}$$

Under this assumption:

$$\mathbb{E}[Y_i(1) - Y_i(0) | X_i = c] = \mathbb{E}[Y_i | W_i = 1, X_i = c] - \mathbb{E}[Y_i | W_i = 0, X_i = c].$$

This approach assumes that differences between treated and control units with $X_i = c$ have a causal interpretation, without exploiting the discontinuity.

Unconfoundedness is fundamentally based on units being comparable if their covariates are similar. This is not an attractive assumption in the current setting where the probability of receiving the treatment is discontinuous in the covariate.

Even if unconfoundedness holds, under continuity of potential outcome regression functions FRD approach will be consistent for the average effect for compliers at $X_i = c$.

---

3. Graphical Analyses

A. Plot regression function $\mathbb{E}[Y_i | X_i = x]$

B. Plot regression functions $\mathbb{E}[Z_i | X_i = x]$ for covariates that do not enter the assignment rule Z_i

C. Plot density $f_X(x)$.

In all cases use estimators that do not smooth around the cutoff value. For example, for binwidth h define bins $[b_{k-1}, b_k]$, where $b_k = c - (K_0 - k + 1) \cdot h$, and average outcomes within bins.

4. Local Linear Regression

We are interested in the value of a regression function at the boundary of the support. Standard kernel regression

$$\widehat{\mu_l(c)} = \sum_{i|c-h<X_i<c}^{N} Y_i \bigg/ \sum_{i|c-h<X_i<c}^{N} 1 \tag{2}$$

does not work well for that case (slower convergence rates)

Better rates are obtained by using local linear regression. First

$$\min_{\alpha_l,\beta_l} \sum_{i|c-h<X_i<c}^{N} (Y_i - \alpha_l - \beta_l \cdot (X_i - c))^2, \tag{3}$$

The value of lefthand limit $\mu_l(c)$ is then estimated as

$$\widehat{\mu_l(c)} = \widehat{\alpha}_l + \widehat{\beta}_l \cdot (c - c) = \widehat{\alpha}_l. \tag{4}$$

Similarly for righthand side. Not much gained by using a non-uniform kernel.

Alternatively one can estimate the average effect directly in a single regression,

$$Y_i = \alpha + \beta \cdot (X_i - c) + \tau \cdot W_i + \gamma \cdot (X_i - c) \cdot W_i + \varepsilon_i$$

thus solving

$$\min_{\alpha,\beta,\tau,\gamma} \sum_{i=1}^{N} 1\{c - h \leq X_i \leq c + h\}$$

$$\times (Y_i - \alpha - \beta \cdot (X_i - c) - \tau \cdot W_i - \gamma \cdot (X_i - c) \cdot W_i)^2,$$

which will numerically yield the same estimate of τ_{SRD}.

This interpretation extends easily to the inclusion of covariates.

Estimation for the FRD Case

Do local linear regression for both the outcome and the treatment indicator, on both sides,

$$\left(\widehat{\alpha}_{yl}, \widehat{\beta}_{yl}\right) = \arg\min_{\alpha_{yl},\beta_{yl}} \sum_{i:c-h\leq X_i<c} \left(Y_i - \alpha_{yl} - \beta_{yl} \cdot (X_i - c)\right)^2,$$

$$\left(\widehat{\alpha}_{wl}, \widehat{\beta}_{wl}\right) = \arg\min_{\alpha_{wl},\beta_{wl}} \sum_{i:c-h\leq X_i<c} (W_i - \alpha_{wl} - \beta_{wl} \cdot (X_i - c))^2,$$

and similarly $(\widehat{\alpha}_{yr}, \widehat{\beta}_{yr})$ and $(\widehat{\alpha}_{wr}, \widehat{\beta}_{wr})$. Then the FRD estimator is

$$\widehat{\tau}_{\mathsf{FRD}} = \frac{\widehat{\tau}_y}{\widehat{\tau}_w} = \frac{\widehat{\alpha}_{yr} - \widehat{\alpha}_{yl}}{\widehat{\alpha}_{wr} - \widehat{\alpha}_{wl}}.$$

Alternatively, define the vector of covariates

$$V_i = \begin{pmatrix} 1 \\ 1\{X_i < c\} \cdot (X_i - c) \\ 1\{X_i \geq c\} \cdot (X_i - c) \end{pmatrix}, \quad \text{and} \quad \delta = \begin{pmatrix} \alpha_{yl} \\ \beta_{yl} \\ \beta_{yr} \end{pmatrix}.$$

Then we can write

$$Y_i = \delta' V_i + \tau \cdot W_i + \varepsilon_i. \tag{TSLS}$$

Then estimating τ based on the regression function (TSLS) by Two-Stage-Least-Squares methods, using

W_i as the endogenous regressor,
the indicator $1\{X_i \geq c\}$ as the excluded instrument
V_i as the set of exogenous variables

This is is numerically identical to $\widehat{\tau}_{\mathsf{FRD}}$ before (because of uniform kernel)

Can add other covariates in straightfoward manner.

5. Choosing the Bandwidth (Imbens-Kalyanaraman)

We wish to take into account that (i) we are interested in the regression function at the boundary of the support, and (ii) that we are interested in the regression function at $x = c$.

IK focus on minimizing

$$\mathbb{E}\left[(\hat{\mu}_l(c) - \mu_l(c))^2 + (\hat{\mu}_r(c) - \mu_r(c))^2\right]$$

Both $\hat{\mu}_l(c)$ and $\hat{\mu}_r(c)$ are based on local linear estimators, with the same bandwidth h.

16

Optimal Bandwidth

$$h_{\text{opt}} = \left(\frac{C_2}{C_1}\right)^{1/5} \cdot \left(\frac{\frac{\sigma_r^2(c)}{p \cdot f_r(c)} + \frac{\sigma_l^2(c)}{(1-p) \cdot f_l(c)}}{\left(\frac{\partial^2 m_r}{\partial x^2}(c)\right)^2 + \left(\frac{\partial^2 m_l}{\partial x^2}(c)\right)^2}\right)^{1/5} \cdot N^{-1/5}$$

p is share of observations above threshold.

$$C_1 = \frac{1}{4} \cdot \left(\frac{\nu_2^2 - \nu_1 \nu_3}{\nu_2 \nu_0 - \nu_1^2}\right)^2 \qquad C_2 = \frac{\int_0^\infty (\nu_2 - u\nu_1))^2 K^2(u) du}{\left(\nu_2 \nu_0 - \nu_1^2\right)^2}$$

$$\nu_j = \int_0^\infty u^j K(u) du$$

If $K(u) = 1_{|u|<0.5}$, then $(C_2/C_1) = 5.40$

17

Bandwidth for FRD Design

1. Calculate optimal bandwidth separately for both regression functions and choose smallest.

2. Calculate optimal bandwidth only for outcome and use that for both regression functions.

Typically the regression function for the treatment indicator is flatter than the regression function for the outcome away from the discontinuity point (completely flat in the SRD case). So using same criterion would lead to larger bandwidth for estimation of regression function for treatment indicator. In practice it is easier to use the same bandwidth, and so to avoid bias, use the bandwidth from criterion for SRD design or smallest.

18

6. Variance Estimation

$$\sigma_{Yl}^2 = \lim_{x \uparrow c} \text{Var}(Y_i | X_i = x), \qquad C_{YWl} = \lim_{x \uparrow c} \text{Cov}(Y_i, W_i | X_i = x),$$

$$V_{\tau y} = \frac{4}{f_X(c)} \cdot \left(\sigma_{Yr}^2 + \sigma_{Yl}^2\right), \qquad V_{\tau w} = \frac{4}{f_X(c)} \cdot \left(\sigma_{Wr}^2 + \sigma_{Wl}^2\right)$$

The asymptotic covar of $\sqrt{Nh}(\hat{\tau}_y - \tau_y)$ and $\sqrt{Nh}(\hat{\tau}_w - \tau_w)$ is

$$C_{\tau y, \tau w} = \frac{4}{f_X(c)} \cdot (C_{YWr} + C_{YWl}).$$

Finally, the asymptotic distribution has the form

$$\sqrt{Nh} \cdot (\hat{\tau} - \tau) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\tau_w^2} \cdot V_{\tau y} + \frac{\tau_y^2}{\tau_w^4} \cdot V_{\tau w} - 2 \cdot \frac{\tau_y}{\tau_w^3} \cdot C_{\tau y, \tau w}\right).$$

This asymptotic distribution is a special case of that in HTV, using the rectangular kernel, and with $h = N^{-\delta}$, for $1/5 < \delta < 2/5$ (so that the asymptotic bias can be ignored).

Can use plug in estimators for components of variance.

19

TSLS Variance for FRD Design

The second estimator for the asymptotic variance of $\hat{\tau}$ exploits the interpretation of the $\hat{\tau}$ as a TSLS estimator.

The variance estimator is equal to the robust variance for TSLS based on the subsample of observations with $c-h \leq X_i \leq c+h$, using the indicator $1\{X_i \geq c\}$ as the excluded instrument, the treatment W_i as the endogenous regressor and the V_i as the exogenous covariates.

7. Concerns about Validity

Two main conceptual concerns in the application of RD designs, sharp or fuzzy.

Other Changes
Possibility of other changes at the same cutoff value of the covariate. Such changes may affect the outcome, and these effects may be attributed erroneously to the treatment of interest.

Manipulation of Forcing Variable
The second concern is that of manipulation of the covariate value.

Specification Checks

A. Discontinuities in Average Covariates

B. A Discontinuity in the Distribution of the Forcing Variable

C. Discontinuities in Avareage Outcomes at Other Values

D. Sensitivity to Bandwidth Choice

E. RD Designs with Misspecification

7.A Discontinuities in Average Covariates

Test the null hypothesis of a zero average effect on pseudo outcomes known not to be affected by the treatment.

Such variables includes covariates that are by definition not affected by the treatment. Such tests are familiar from settings with identification based on unconfoundedness assumptions.

Although not required for the validity of the design, in most cases, the reason for the discontinuity in the probability of the treatment does not suggest a discontinuity in the average value of covariates. If we find such a discontinuity, it typically casts doubt on the assumptions underlying the RD design.

7.B A Discontinuity in the Distribution of the Forcing Variable

McCrary (2007) suggests testing the null hypothesis of continuity of the density of the covariate that underlies the assignment at the discontinuity point, against the alternative of a jump in the density function at that point.

Again, in principle, the design does not require continuity of the density of X at c, but a discontinuity is suggestive of violations of the no-manipulation assumption.

If in fact individuals partly manage to manipulate the value of X in order to be on one side of the boundary rather than the other, one might expect to see a discontinuity in this density at the discontinuity point.

7.C Discontinuities in Avareage Outcomes at Other Values

Taking the subsample with $X_i < c$ we can test for a jump in the conditional mean of the outcome at the median of the forcing variable.

To implement the test, use the same method for selecting the binwidth as before. Also estimate the standard errors of the jump and use this to test the hypothesis of a zero jump.

Repeat this using the subsample to the right of the cutoff point with $X_i \geq c$. Now estimate the jump in the regression function and at $q_{X,1/2,r}$, and test whether it is equal to zero.

7.D Sensitivity to Bandwidth Choice

One should investigate the sensitivity of the inferences to this choice, for example, by including results for bandwidths twice (or four times) and half (or a quarter of) the size of the originally chosen bandwidth.

Obviously, such bandwidth choices affect both estimates and standard errors, but if the results are critically dependent on a particular bandwidth choice, they are clearly less credible than if they are robust to such variation in bandwidths.

7.E RD Designs with Misspecification

Lee and Card (2007) study the case where the forcing variable variable X is discrete. In practice this is of course always true. This implies that ultimately one relies for identification on functional form assumptions for the regression function $\mu(x)$.

They consider a parametric specification for the regression function that does not fully saturate the model and interpret the deviation between the true conditional expectation and the estimated regression function as random specification error that introduces a group structure on the standard errors.

Lee and Card then show how to incorporate this group structure into the standard errors for the estimated treatment effect. Within the local linear regression framework discussed in the current paper one can calculate the Lee-Card standard errors and compare them to the conventional ones.

Illustration Based on David Lee Election Data

Forcing variable is difference in dem vs rep vote share in last election.

Outcomes are dem vote share in next election, and indicator for democrats winning the next election.

Covariate for testing is dem vote share in prior election

6558 congressional elections.

Uniform kernel with support $[-0.5, 0.5]$

Outcome	IK Bandwidth	Estimate	(s.e.)
Dem Win Next Elect	0.36	0.082	(0.010)
Demt Margin Next Election	0.27	0.412	(0.039)
Dem Margin Prev Election	0.28	-0.003	(0.013)

Fig 1: Regression Function for Margin

Fig 1: Regression Function for Winning

Fig 1: Regression Function for Covariate