The Econometric Evaluation of Policy Design:: Part II: Some Approaches

Edward Vytlacil,
Yale University

Renmin University
June 2017

Lectures primarily drawing upon:

- Heckman and Smith (1995), "Assessing the Case for Social Experiments"
- Heckman, Ichimura, and Todd, (1998) "Matching as an Econometric Evaluation Estimator"
- Hahn, Todd and Van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design"
- Lee and Lemieux (2013), "Regression Discontinuity Designs in the Social Sciences."
- Lee (2008), "Randomized experiments from non-random selection in U.S. House Elections."
- Van der Klaauw (2002), "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach."

and on . . .

Lectures primarily drawing upon:

- Heckman, James (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator."
- Imbens and Angrist (1994), “Identification and Estimation of Local Average Treatment Effects.”
- Angrist and Evans (1998), “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.”

Overview of Some Possible Approaches

- Randomized experiments
- Matching
- Instrumental Variables / Selection Model Approach
- Regression Discontinuity
- Bounding/Set-Identification

For each approach, will consider implications of heterogeneous treatment effects, and ask what parameter is actually being identified by the approach.

Randomized Experiments

- Suppose, among those who desire treatment ($D_i = 1$), can randomly deny treatment.
- Let $R_i = 1$ if randomly assigned to receive treatment, $R_i = 0$ if denied treatment.
- Suppose R_i properly randomized, so that $R_i \perp\!\!\!\perp Y_{0i}, Y_{1i} | D_i$.
- Suppose *full compliance*, all those with $R_i = 1$ receive treatment, and no one with $R_i = 0$ receives treatment,

Randomized Experiments (cont'd)

Then:

$$
\begin{aligned}
E(Y_i|D_i = 1, R_i = 1) - E(Y_i|D_i = 1, R_i = 0) \\
&= E(Y_{1i}|D_i = 1, R_i = 1) - E(Y_{0i}|D_i = 1, R_i = 0) \\
&\qquad \text{(by full compliance)} \\
&= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)) \\
&\qquad \text{(by randomization)} \\
&= E(Y_{1i} - Y_{0i}|D_i = 1)
\end{aligned}
$$

Recover treatment on the treated.

Can recover effect on distribution, but cannot recover distribution of effects without further assumptions.

If can randomly force treatment on those who do not wish to receive treatment, can recover treatment on the untreated.

Randomized Experiments (cont'd)

Potential Problem: Lack of compliance

- Some individuals assigned treatment might not actually take the treatment, while some of those randomized out of treatment might still take the treatment.
- Lack of compliance is typically addressed by either
  1. redefining object of interest to be "intention to treat", or
  2. using non-experimental methods
     - Use assignment to treatment as an instrument for receipt of treatment.
     - impose that compliance is exogenous conditional on observed covariates, use a matching approach.

Will further discuss use of random assignment as an instrumental variable for receipt of treatment after covering the LATE IV framework.

Randomized Experiments (cont'd)

Other Potential Issues:

- Imperfect Randomization
- Multiple hypothesis testing

See, e.g., "Inference with Imperfect Randomization: The Case of the Perry Preschool Program", by Heckman, Pinto, Shaikh and Yavitz (2011).

Randomized Experiments (cont'd)

Internal vs. External Validity

- RCTs focus on internal validity, effect of treatment within experimental sample.
- External validity often less clear
  - Does average effect among participants in trial equal average effect in population of interest?
- Threats to external validity include that the treatment effect may vary with observed and unobserved characteristics of individuals, and the distribution of those characteristics in experimental sample may be different from that in target population.

Randomized Experiments (cont'd)

RCT eligibliity criteria often explicitly restrict trial to small subset of target population in terms of observed covariates.

Examples:

- 25-percent or less of patients hospitalized with heart failure would be eligible for three of the major clinical trials of heart failure (Masoudi et al., 2003).
- 12% of the target population of US people infected with HIV would be eligible for seminal AIDS Clinical Trial Group (Cole and Stuart, 2010).
- Medical RCTs often have arbitrary upper age limits and other poorly justified exclusion criteria (Cherubini 2011).

Randomized Experiments (cont'd)

Even when groups not explicitly excluded, may not be present in trial because of how data is screened or implication of other exclusion criteria.

- For example, Heiat et al (2002) show that many heart failure trials do not contain any elderly patients even when there is no explicit age exclusion criteria. In contrast, most patients hospitalized with heart failure are elderly.

Randomized Experiments (cont'd)

Large biostat literature using reweighting RCT data so that weighted sample is representative of target population in terms of observed covariates
(see, e.g., Hartman et al 2015).

- Closely related to matching, will discuss next.
- Assumptions/strategies at odds with complete exclusion of groups. Cannot weight trial data without elderly to have same age distribution as target population with elderly.

Randomized Experiments (cont'd)

Same worry with unobserved characteristics:

- treatment effect may vary with unobserved characteristics that have a different distribution in the RCT sample versus the target population.

Individual's decision to to enroll in RCT might be directly related to their own treatment effect.

- See, e.g., Malani, 2008 and Basu 2014 for economic models of selection into RCTs.

RCT: Connection between Selection and Treatment Effects

With typical implementation, RCTs contain no information on connection between selection and treatment effects.

Limits policy relevance of RCTs, cannot address such questions as

- Is treatment currently targeted at those who benefit from the treatment?
- Should treatment be encouraged or discouraged relative to the status quo?
- Would those currently not taking the treatment benefit on average from taking the treatment?

Related to points made by Heckman and Smith (1995), Heckman, Smith with Clements (1997), Basu (2014).

RCT: Connection between Selection and Treatment Effects

There do exist strategies to identify connection between selection and treatment effects by appropriate design of RCT, e.g.

- elicitation of preferences for treatment as part of trial (Berry et al, 2011, Chassang et al, 2012),
- Experimental variation in incentives combined with MTE framework (see, e.g., Basu 2014).

We will discuss these approaches further after covering the MTE framework.

Randomized Experiments (cont'd)

Other issues of running experiments

- Ethical concerns (clinical equipoise)
- Feasibility
- Expense.
- Length of followup, attrition.
- Answers very limited question.

Above discussion following Heckman and Smith (1995), see their work for further analysis of randomized experiments. See also Burtless (1995) for an alternative perspective.

Randomized Experiments (cont'd)

Other uses of randomized experiments:

- To validate nonexperimental treatment-effect approaches, by checking whether nonexperimental estimates agree with experimental estimates, see e.g.
  - Lalonde (1986)
  - Heckman, Ichimura and Todd (1997)

  How to interpret validation exercise with treatment effect heterogeneity?

- Combine with structural estimation, to validate structural model, to help identify structural model, and/or to provide answers to policy questions not directly answered by the experiment, see e.g.
  - Todd and Wolpin (2003, 2006)
  - Attanasio, Meghir and Santiago (2010)

Matching/Conditional Independence Assumption

Suppose cross-section data on treated and untreated.
(can also do differences-in-differences version of matching, see e.g. Heckman, Ichimura and Todd, 1997, and Abadie, 2005).

Matching/CIA Assumptions:

MATCH-1 (CIA)

Conditional Independence Assumption: $Y_1, Y_0 \perp\!\!\!\perp D | X$

MATCH-2 (CSA)

Common Support Assumption: $0 < \Pr[D = 1|X] < 1$

Matching/Conditional Independence Assumption (cont'd)

CIA assumption, $Y_1, Y_0 \perp\!\!\!\perp D|X$,
D is exogenous (conditional on X).

- How plausible? Behavioral restriction?
- Can weaken CIA assumption to mean independence, $E(Y_1|D, X) = E(Y_1|X)$, $E(Y_0|D, X) = E(Y_0|X)$,
- Can further weaken CIA assumption to only assume $E(Y_0|D, X) = E(Y_0|X)$ if only need TT parameter.

Matching/Conditional Independence Assumption (cont'd)

CSA assumption, $0 < \Pr[D = 1|X] < 1$

- Requires some treated and some nontreated for all values of X.
- How plausible? Behavioral restriction?
- For estimation theory, require stronger assumption that

$$0 < c_0 \leq \Pr[D = 1|X] \leq c_1 < 1$$

- Can weaken CSA assumption to $\Pr[D = 1|X] < 1$
  (or $\Pr[D = 1|X] \leq c_1 < 1$ for estimation)
  if only need TT parameter.
- In practice, often use trimming, only use observations where $\hat{f}(X_i|D = 1) > q$, $\hat{f}(X_i|D = 0) > q$, for some $q > 0$.

Matching/Conditional Independence Assumption (cont'd)

CIA: $Y_1, Y_0 \perp\!\!\!\perp D|X$
CS: $0 < \Pr[D = 1|X] < 1$

Then:

$$\begin{aligned} & E\left[E(Y|D=1,X) - E(Y|D=0,X)\right] \\ &\quad = E\left[E(Y_1|D=1,X) - E(Y_0|D=0,X)\right] \\ &\quad = E\left[E(Y_1|X) - E(Y_0|X)\right] \\ &\quad = E\left[E(Y_1 - Y_0|X)\right] \\ &\quad = E(Y_1 - Y_0) \end{aligned}$$

outer expectation is expectation over X.

Matching/Conditional Independence Assumption (cont'd)

$$E\left[E(Y|D=1,X)-E(Y|D=0,X)\right]=E(Y_1-Y_0)$$

where outer expectation is expectation over unconditional distribution of X.

Likewise

$$E\left[E(Y|D=1,X)-E(Y|D=0,X)|D=1\right] = E(Y_1-Y_0|D=1)$$

where outer expectation is expectation over distribution of X conditional on $D=1$.

Matching/Conditional Independence Assumption (cont'd)

Alternative way to express above results, more useful for estimation:

- For ATE:

$$E(Y_1 - Y_0) = E\big(D\,[Y - E(Y|D=0,X)] + (1-D)\,[E(Y|D=1,X) - Y]\big)$$

- For TT:

$$\begin{aligned} E(Y_1 - Y_0|D=1) &= E(Y - E(Y|D=0,X)|D=1) \\ &= E\,(D\,[Y - E(Y|D=0,X)])\,/\,\Pr[D=1] \end{aligned}$$

Matching/Conditional Independence Assumption (cont'd)

Suggests Two Stage Estimation Strategy for TT
(estimation strategy for ATE is parallel)

(1) Nonparametric regression of Y on (D, X) to recover $\hat{E}(Y|D, X)$.

- For example, estimate $\hat{E}(Y|D, X)$ by kernel regression,

$$\hat{E}(Y|D = d, X = x) = \sum_{i:D_i=1} Y_i K((X_i - x)/h) / \sum_{i:D_i=1} K((X_i - x)/h)$$

  with k the kernel and h the bandwidth.
- For example, estimate $\hat{E}(Y|D, X)$ by local linear regression.
- For example, estimate $\hat{E}(Y|D, X)$ by series regression.

Matching/Conditional Independence Assumption (cont'd)

(2) Estimate TT by

$$\frac{1}{N_1}\sum D_i\left[Y_i - \hat{E}(Y|D=0, X_i)\right]$$

where $N_1 =$ no. of observations with $D_i = 1$.

Note role of $\Pr[D=1|X]$ bounded away from one

Propensity Score Matching

Define $P(X)$ as the "propensity score," $P(X) = \Pr[D = 1|X]$.

Important Result (Rosenbaum and Rubin, 1983):

- $Y_0, Y_1 \perp\!\!\!\perp D|X$ if and only if $Y_0, Y_1 \perp\!\!\!\perp D|P(X)$.

Thus, if sufficient to condition on random vector X to obtain conditional independence, than it is also sufficient to condition on scalar $P(X)$.

Parallel result for mean independence.

Propensity Score Matching

Rosenbaum and Rubin result suggests 3 stage approach:

1. Regress D on X to recover $\hat{P}(X)$.
2. Regress Y on $(D, \hat{P}(X))$ to recover $\hat{E}(Y|D, P(X))$.
3. Estimate TT by

$$\frac{1}{N_1} \sum D_i \left[Y_i - \hat{E}(Y|D=0, \hat{P}(X_i)) \right]$$

Estimation Theory for Regression-Based Matching

- Above estimation strategies referred to as regression based matching
- See, e.g., Heckman, Ichimura and Todd for estimation theory for regression-based matching, including for use of kernel or LLR estimation of $E(Y|D, X)$.
- Under regularity conditions, obtain $\sqrt{N}$-consistent, asymptotically normal estimators of ATE or TT.
- Typically use bootstrap for inference.

Estimation Theory for Regression-Based Matching (cont'd)

- Curse of dimensionality?
- Possible sizable finite sample bias from large dimensional X, asymptotics approximate can be poor if X is high dimensional.
- Does use of propensity score help?
- Typically, in practice, do propensity score matching with parametric model for $\Pr[D = 1|X]$, e.g. estimate logit or probit model.

Estimation Theory for Regression-Based Matching (cont'd)

Additional Issue: Trimming

If density of X conditional on D is not bounded away from zero on the same set of value, then need to trim to common support over region where both conditional densities are bounded away from zero.

Estimation Theory for Regression-Based Matching (cont'd)

- Let $\mathcal{S} = \{x : f(x|D=1) > q, f(x|D=0) > q\}$.
- Let $\hat{\mathcal{S}} = \{x : \hat{f}(x|D=1) > q, \hat{f}(x|D=0) > q\}$, where $\hat{f}$ is kernel density estimator.
- Estimator of TT conditional on $X \in \mathcal{S}$:

$$\frac{\frac{1}{N_1}\sum D_i\left[Y_i - \hat{E}(Y|D=0,\hat{P}(X_i))\right]\mathbf{1}[X_i \in \hat{\mathcal{S}}]}{\frac{1}{N_1}\sum D_i \mathbf{1}[X_i \in \hat{\mathcal{S}}]}$$

- Estimator is now consistently estimating $E(Y_1 - Y_0|D=1, X \in \mathcal{S})$. Sometimes called "feasible" version of parameter. Changes definition of parameter.
- Typically need to trim, but problems with trimming too much.

Nearest Neighbor Matching

Alternative way to implement matching, e.g., M-nearest neighbor matching.

Consider k-nearest neighbor matching of TT (estimation of ATE by M-nearest neighbor is parallel)

- Let M denote some positive integer (most often $M = 1$)
- Let $d(\cdot,\cdot)$ denote a distance measure, for example:
  - Euclidean metric:

$$d(U, V) = ||U - V||$$

  - Mahalanobis metric (Rosenbaum and Rubin):

$$d(U, V) = (U - V)\Sigma_U^{-1}(U - V)$$

Nearest Neighbor Matching (cont'd)

For each treated i (i s.t. $D_i = 1$),

1. For $m = 1, ..., M$, let $l_m(i)$ denote the index l of the control individual ($D_l = 0$) whose X_l is the mth closest to X_i among all control individuals, i.e.

$$\sum_{j:D_j=0} \mathbf{1}\left[d(X_j, X_i) \leq d(X_l, X_i)\right] = m$$

2. Let $\mathcal{J}_M(i)$ denote the indices of the M closest control individuals,

$$\mathcal{J}_M(i) = \{l_1(i), ..., l_M(i)\}$$

   This is the set of control individuals "matched" to treated individual i.

Nearest Neighbor Matching (cont'd)

Define

$$\hat{Y}_i(0) = \frac{1}{M} \sum_{m \in \mathcal{J}_M(i)} Y_m$$

$\hat{Y}_i(0)$ is imputed value of $Y_i(0)$ using M matched control observations.

Estimate TT by

$$\frac{1}{N_1} \sum D_i \left[Y_i - \hat{Y}_i(0) \right]$$

Nearest Neighbor Matching (cont'd)

- Typically examine "balance" in matched sample, (standardized) sample moments of X in treated sample compared to matched control sample. Standardized sample moments being too different is a cause for concern.

Nearest Neighbor Matching (cont'd)

Advantage of estimator:

1. Very simple to implement.
2. Only need to pick number of nearest neighbors (M), usually single nearest neighbor. Do not need to pick kernel or bandwidth.

Estimation theory developed by Abadie and Imbens (2006, 2008). Estimator has some disadvantages.

Nearest Neighbor Matching (cont'd)

Abadie and Imbens (2006, 2008) show the following results for nearest neighbor matching estimators:

- Not an efficient estimator
- Estimator is consistent under regularity conditions
- When number of continuous covariates $k \geq 2$, estimator is generally not $\sqrt{N}-$consistent
  - bias generally does not disappear at $\sqrt{N}$-rate when $k = 2$
  - bias generally dominates at $\sqrt{N}$-rate when $k \geq 3$.
- Best result with one continuous covariate, in which case estimator is $\sqrt{N}$-consistent and asymptotically normal.
- Even with only one continuous covariate, bootstrap is not valid, though can use analytical s.e. or can use subsampling for inference.

Nearest Neighbor Matching (cont'd)

- Often nearest neighbor matching used with matching on propensity score:
  - If propensity score known, than at most one continuous covariate, and estimator is $\sqrt{N}$-consistent and asymptotically normal.
  - Theoretical estimation results when propensity score is estimated developed by Abadie and Imbens, "Matching on the Estimated Propensity Score" (2015). Need to adjust inference for first stage estimation of the propensity score.
- Generally also have to trim, make sure closest control is close enough. (caliper)
  - Shifts definition of parameter. Typically need to trim, but disadvantages to trimming too much.
- Sometimes create matched sample, and then perform other estimation approach on matched sample (e.g., run linear regression of Y on D and X in matched sample).

Other Matching Approaches

Another popular method for matching:
inverse propensity score weighting.

- Sample average, weighted by inverse estimated propensity score.
- Has good theoretical properties.
- See, e.g., Hirano, Imbens and Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score."

Other Matching Approaches

Another method, recently becoming more popular:
Doubly Robust Estimation, developed by James Robins.

- See, e.g., Bang, H. and Robins, J. M. (2005). "Doubly robust estimation in missing data and causal inference models."

Another method recently popular:
matching for high dimensional problems exploiting sparsity

- See, e.g., Belloni, Chernozhukov, and Hansen, "Inference on Treatment Effects After Selection Among High Dimensional Controls."

Instrumental Variables

$$
\begin{aligned}
Y &= Y_0 + D(Y_1 - Y_0) \\
&= E(Y_0) + DE(Y_1 - Y_0) + \{\varepsilon + \eta D\}
\end{aligned}
$$

where $\epsilon = [Y_0 - E(Y_0)]$, $\eta = (Y_1 - Y_0) - E(Y_1 - Y_0)$.

Suppose that D is potentially endogenous

Suppose have instrument, in particular, suppose $\text{Cov}(D, Z) \neq 0$, $Z \perp\!\!\!\perp Y_0, Y_1$.

For example, in randomized experiment with imperfect compliance, Z is assignment to treatment and D is receipt of treatment.

Instrumental Variables (cont'd)

Probability limit of IV:

$$\text{plim} IV = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)}$$

Special Case:

- If $Z = 0, 1$, we have the Wald expression for IV,

$$\begin{aligned}\text{plim} IV &= \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} \\ &= \frac{E(Y|Z = 1) - E(Y|Z = 0)}{\Pr[D = 1|Z = 1] - \Pr[D = 1|Z = 0]}\end{aligned}$$

Does linear IV recover a parameter of interest?

Instrumental Variables

Does IV recover a parameter of interest?

- If Δ is a constant, homogeneous treatment effect, classical IV results hold, IV recovers the effect of treatment.
- If Δ is heterogeneous, and in particular if essential heterogeneity, then classical IV results do not hold, in general do not recover interpretable parameter.
- If Δ is heterogeneous, and if impose selection model (equivalently, Imbens-Angrist LATE conditions), IV recovers interpretable parameter which may or may not be of interest.

Δ is a constant

- Suppose Δ is a constant $\Rightarrow \eta = 0$.
- If there is an instrument Z, with the property that

$$\begin{aligned} \text{Cov}(Z, D) &\neq 0 \\ \text{Cov}(Z, Y_0) = 0 \Rightarrow \text{Cov}(Z, \epsilon) &= 0, \end{aligned}$$

then

$$\text{plim IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, D)} = E(Y_1 - Y_0).$$

- If other instruments exist, each identifies the same parameter.

Heterogeneous response case: Δ is a variable even conditioning on X

Suppose $\text{Cov}(Z, D) \neq 0$, $Z \perp\!\!\!\perp Y_0, Y_1$, but Δ varies across people.

Can we identify $E(Y_1 - Y_0)$ using IV?

- In general we cannot (Heckman and Robb, 1985).

$Y = E(Y_0) + DE(Y_1 - Y_0) + \{\epsilon + \eta D\}$
where $\epsilon = [Y_0 - E(Y_0)]$, $\eta = (Y_1 - Y_0) - E(Y_1 - Y_0)$.

- Need Z to be uncorrelated with $[\varepsilon + \eta D]$ to use IV identify $E(Y_1 - Y_0)$.
- This condition will be satisfied if policy adoption is made without knowledge of $\eta = (Y_1 - Y_0) - E(Y_1 - Y_0)$.
- If decisions about D are made with partial or full knowledge of η, IV does not identify $E(Y_1 - Y_0)$.

- $E(\eta D \mid Z) = E(\eta \mid D = 1, Z)\Pr(D = 1 \mid Z)$.
- Even if $\eta \perp\!\!\!\perp Z$, $\eta \not\perp\!\!\!\perp Z \mid D = 1$.
- The IV condition is

$$E\left[\varepsilon + \eta D \mid Z\right] = 0.$$

- $E\left(\varepsilon \mid Z\right) = 0, \quad E\left(\eta \mid Z\right) = 0.$
- But $E\left(\eta \mid Z, D = 1\right) \neq 0$, in general, if agents have some information about the gains.

- Linear IV does not identify ATE or any standard treatment parameters.
- Without more conditions, IV does not identify any interpretable parameter.
- With more conditions (under selection model), IV does identify interpretable parameter.

- Imbens and Angrist (1994) establish that IV can identify an interpretable parameter in the model with essential heterogeneity.
- Their parameter is a discrete approximation to the marginal gain parameter of Björklund and Moffitt (1987).
- This parameter can be interpreted as the marginal gain to outcomes induced from a marginal change in the costs of participating in treatment (Björklund-Moffitt).

Imbens Angrist conditions (1994)

IV-1 (Independence)

$$Z \perp\!\!\!\perp \left(Y_1, Y_0, \{D(z)\}_{z \in \mathcal{Z}}\right).$$

IV-2 (Rank)

$\Pr(D = 1 \mid Z)$ *depends on* Z.

IV-3 (Monotonicity)

For all $z, z' \in \mathcal{Z}$, *either* $D_i(z) \geq D_i(z')$ *for all* i,
or $D_i(z) \leq D_i(z')$ *for all* i.

Imbens Angrist conditions (1994)

- *Uniformity* of responses *across* persons.
- Uniformity is satisfied when, for $z < z'$, $D_i(z) \leq D_i(z')$ for all i, while for $z'' > z'$, $D_i(z'') \leq D_i(z')$ for all i.

Vytlacil (2002) Equivalence

Let $\mathbf{1}[\cdot]$ denote the logical indicator function.

Vytlacil (2002) shows Imbens-Angrist conditions are equivalent to the nonparametric selection model:

SELECTION-1 (Selection Model)

$D_i = \mathbf{1}[\mu(Z_i) \geq U_i]$, $Z_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}, U_i)$, and $\mu(\cdot)$ is a nontrivial function of Z_i.

Imbens Angrist conditions (1994)

- These conditions imply the following LATE parameter.

$$
\begin{aligned}
&E\left(Y \mid Z=z\right)-E\left(Y \mid Z=z'\right)\\
&\quad =E(Y_0+D(Y_1-Y_0)\mid Z=z)-E(Y_0+D(Y_1-Y_0)\mid Z=z')\\
&\quad =E(D(Y_1-Y_0)\mid Z=z)-E(D(Y_1-Y_0)\mid Z=z')\\
&\quad =E((D(z)-D(z'))(Y_1-Y_0))\ \text{(Independence)}
\end{aligned}
$$

Imbens Angrist conditions (1994)

Using iterated expectations,

$$
\begin{aligned}
& E\left(Y \mid Z=z\right)-E\left(Y \mid Z=z'\right) \\
& \quad = E((D(z)-D(z'))(Y_1-Y_0)) \\
& \quad = \left(\begin{array}{c} E\left(Y_1-Y_0 \mid D\left(z\right)-D\left(z'\right)=1\right) \\ \cdot \Pr\left(D\left(z\right)-D\left(z'\right)=1\right) \end{array}\right) \\
& \qquad\quad - \left(\begin{array}{c} E\left(Y_1-Y_0 \mid D\left(z\right)-D\left(z'\right)=-1\right) \\ \cdot \Pr\left(D\left(z\right)-D\left(z'\right)=-1\right) \end{array}\right) \\
& \quad = \left(\begin{array}{c} E\left(Y_1-Y_0 \mid D\left(z\right)=1, D\left(z'\right)=0\right) \\ \cdot \Pr\left(D\left(z\right)=1, D\left(z'\right)=0\right) \end{array}\right) \\
& \qquad\quad - \left(\begin{array}{c} E\left(Y_1-Y_0 \mid D\left(z\right)=0, D\left(z'\right)=1\right) \\ \cdot \Pr\left(D\left(z\right)=0, D\left(z'\right)=1\right) \end{array}\right).
\end{aligned}
$$

Imbens Angrist conditions (1994) (cont'd)

- Using iterated expectations,

$$
\begin{aligned}
& E\left(Y \mid Z = z\right) - E\left(Y \mid Z = z'\right) \\
& \quad = \begin{pmatrix} E\left(Y_1 - Y_0 \mid D\left(z\right) = 1, D\left(z'\right) = 0\right) \\ \cdot \Pr\left(D\left(z\right) = 1, D\left(z'\right) = 0\right) \end{pmatrix} \\
& \qquad - \begin{pmatrix} E\left(Y_1 - Y_0 \mid D\left(z\right) = 0, D\left(z'\right) = 1\right) \\ \cdot \Pr\left(D\left(z\right) = 0, D\left(z'\right) = 1\right) \end{pmatrix}.
\end{aligned}
$$

- Invoke monotonicity condition,
- Suppose that $\Pr(D\left(z\right) = 0, D\left(z'\right) = 1) = 0$. Then,

$$
\begin{aligned}
& E\left(Y \mid Z = z\right) - E\left(Y \mid Z = z'\right) = \\
& E\left(Y_1 - Y_0 \mid D\left(z\right) = 1, D\left(z'\right) = 0\right) \Pr\left(D\left(z\right) = 1, D\left(z'\right) = 0\right).
\end{aligned}
$$

Imbens Angrist conditions (1994)

$$
\begin{aligned}
LATE &= \frac{E\left(Y \mid Z = z\right) - E\left(Y \mid Z = z'\right)}{\Pr(D = 1 \mid Z = z) - \Pr(D = 1 \mid Z = z')} \\
&= E\left(Y_1 - Y_0 \mid D\left(z\right) = 1, D\left(z'\right) = 0\right)
\end{aligned}
$$

- The mean gain to those induced to switch from "0" to "1" by a change in Z from z' to z.
- Observe LATE = ATE if

$$\Pr\left(D = 1 \mid Z = z\right) = 1 \quad \text{while} \quad \Pr\left(D = 1 \mid Z = z'\right) = 0.$$

- "Identification at infinity" plays a crucial role throughout the entire literature on policy evaluation.

IV with Random Coefficient, Imbens-Angrist

In case of RCT with imperfect compliance

- Can use random assignment as Z.
- LATE assumptions particularly plausible in RCT context.
- LATE parameter is average effect of treatment on those who would receive treatment if randomized into treatment but not if randomized out.
- Implications for comparing evidence from RCT to nonexperimental estimates?

Imbens Angrist conditions (1994)

- In general, LATE $\neq E(Y_1 - Y_0)$.
- Not treatment on the treated: $E(Y_1 - Y_0 \mid D = 1)$.
- Different instruments define different parameters.
- Having a wealth of different strong instruments does not improve the precision of the estimate of any particular parameter.
- When there are more than two distinct values of Z, Imbens and Angrist use Yitzhaki (1989)weights.

Application: Flu Vaccination

Consider effect of flu vaccination on incidence of the flu.

- Questions of interest:
  - Does receiving flu shots cause people to be less likely to contract the flu?
  - If so, by how much?
- Based on Hirano, Imbens, Rubin, and Zhou (1999), "Assessing the Effect of Influenza Vaccine in an Encouragement Design."
  (I am simplifying their analysis).

Application: Flu Vaccination

Let Y denote dummy variable for being hospitalized with flu.
Let D denote dummy variable for being vaccinated.

Hirano et al estimate

$$\widehat{\Pr}[Y = 1 \mid D = 1] = 0.084$$

$$\widehat{\Pr}[Y = 1 \mid D = 0] = 0.085.$$

$$\Rightarrow \widehat{\Pr}[Y = 1 \mid D = 1] - \widehat{\Pr}[Y = 1 \mid D = 0] = -.001.$$

Mean differences/OLS suggest that the flu vaccine has little effect on incidence of the flu.

Application: Flu Vaccination

Potential problem with OLS estimation:

- Individuals who are at a higher risk for the flu are more likely to receive flu shots.
- May worry that difference would remain even conditioning on X variables that we observe.

RCT with random assignment of vaccination?

- Unethical, would violate clinical equipose.

Application: Flu Vaccination

Solution used by Hirano et al (1999):
Exploit "encouragement design" experiment:

- Use data from an experiment in which doctors were randomly sent letters encouraging them to give flu shots.
- Instrument Z is dummy variable for whether subject's doctor was sent the letter.

Application: Flu Vaccination

Z is dummy variable for whether subject's doctor was sent encouragement letter.

In this context, how to interpret D_0, D_1?

In this context, how to interpret LATE assumptions:

- IV Exogeneity: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$.
- IV Relevance: $\Pr[D = 1|Z = 1] \neq \Pr[D = 1|Z = 0]$.
- IV Monotonicity: $D_1 \geq D_0$.

Are these restrictions plausible in this context?

Application: Flu Vaccination

Z is dummy variable for whether subject's doctor was sent encouragement letter.

- In this context, how to interpret, how to interpret LATE, $E(Y_1 - Y_0 | D_1 = 1, D_0 = 0)$?
- In this context, is LATE of policy interest?

Application: Flu Vaccination

They find:

$$\begin{aligned} \widehat{\Pr}(D=1|Z=1) &= .307 \\ \widehat{\Pr}(D=1|Z=0) &= .190 \\ \Rightarrow \widehat{\Pr}(D=1|Z=1) - \widehat{\Pr}(D=1|Z=0) &= .117. \end{aligned}$$

t-value on difference is 7.3.

Instrument has strong first stage.

Application: Flu Vaccination

$$\begin{aligned}
\widehat{\Pr}[Y=1 \mid Z=1] &= 0.078 \\
\widehat{\Pr}[Y=1 \mid Z=0] &= 0.092 \\
\widehat{\Pr}(D=1|Z=1) &= .307 \\
\widehat{\Pr}(D=1|Z=0) &= .190
\end{aligned}$$

"Intention to Treat":

$$\widehat{\Pr}[Y=1 \mid Z=1] - \widehat{\Pr}[Y=1 \mid Z=0] = -0.014$$

Wald-IV estimate of LATE:

$$\frac{0.078-0.092}{0.307-0.190} = -.12$$

Application: Angrist and Evans, Fertility and Labor Supply

Angrist and Evans (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size"

- Consider effect of fertility on female labor supply.
- Women with more children work less, but how much is this relationship causal versus the result of selection?
- Focus on effect of going from two children to having three or more children.

Application: Angrist and Evans, Fertility and Labor Supply

- Y is measure of labor supply of mother
  - Alternatively consider: Work/not work; weeks worked; hours worked per week; labor income.
- $D = 1$ if have three or more chidren, $D = 0$ if only two. (in sample with 2 or more children)
- Instruments Z :
  1. First two children same sex.
     - In some specifications, split to first two children male and first two children female as separate instruments.
  2. Second birth resulted in twins.

Application: Angrist and Evans, Fertility and Labor Supply

- Y is measure of labor supply of mother
- $D = 1$ if have three or more chidren, $D = 0$ if only two. (in sample with 2 or more children)
- Z is whether first two children same sex, or alternatively second birth resulted in twins.

In this context, what is

- Y_0, Y_1?
- D_0, D_1?

Application: Angrist and Evans, Fertility and Labor Supply

- Y is measure of labor supply of mother
- $D = 1$ if have three or more chidren, $D = 0$ if only two. (in sample with 2 or more children)
- Z is whether first two children same sex, or alternatively second birth resulted in twins.

In this context, how to interpret:

- IV Exogeneity: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$.
- IV Relevance: $\Pr[D = 1|Z = 1] \neq \Pr[D = 1|Z = 0]$.
- IV Monotonicity: $D_1 \geq D_0$.

Are these restrictions plausible?

Application: Angrist and Evans, Fertility and Labor Supply

- Y is measure of labor supply of mother
- $D = 1$ if have three or more chidren, $D = 0$ if only two. (in sample with 2 or more children)
- Z is whether first two children same sex, or alternatively second birth resulted in twins.

In this context, how to interpret LATE, $E(Y_1 - Y_0 | D_1 = 1, D_0 = 0)$?

In this context, is LATE of policy interest?

In this context, how to interpret differences in LATE from same-sex vs twin birth instruments?

Application: Angrist and Evans, Fertility and Labor Supply

Data:

- From 1980 and 1990 United States Census Public Use Micro Sample (PUMS)
- Restrict sample to women aged 21-35 with 2 or more children.
- Results in sample of 398,835 women in 1980, and 380,007 women in 1990.

Application: Angrist and Evans, Fertility and Labor Supply

Sample statistics for 1980,
women aged 21-35 with 2 or more children.

- Fraction of women
    - with 3 or more children: .40.
    - first two births same sex: .51.
    - second birth resulted in twins: 0.009.
    - work for pay: .57.
- Average
    - Weeks worked: 21.
    - Hours per week: 19.
    - Labor income: 7,160.

Application: Angrist and Evans, Fertility and Labor Supply

Fraction of Families that Had Another Child by Parity and Sex of Chidlren

Sex of first two children in families with two or more children	1980 PUMS (394,835 observations): Fraction of sample	1980 PUMS (394,835 observations): Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)
two girls	0.242	0.441 (0.002)
two boys	0.264	0.423 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)
(2) both same sex	0.506	0.432 (0.001)
difference (2) − (1)	—	0.060 (0.002)

Application: Angrist and Evans, Fertility and Labor Supply

Wald/LATE Estimates

1980 PUMS

Variable	Mean difference by *Same sex*	Wald estimate using as covariate: *More than 2 children*	Wald estimate using as covariate: *Number of children*
More than 2 children	0.0600 (0.0016)	—	—
Number of children	0.0765 (0.0026)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)

Application: Angrist and Evans, Fertility and Labor Supply

OLS and TSLS, 1980 PUMS

Estimation method	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:			
Worked for pay	−0.176 (0.002)	−0.120 (0.025)	−0.113 (0.025) [0.013]
Weeks worked	−8.97 (0.07)	−5.66 (1.11)	−5.37 (1.10) [0.017]
Hours/week	−6.66 (0.06)	−4.59 (0.95)	−4.37 (0.94) [0.030]
Labor income	−3768.2 (35.4)	−1960.5 (541.5)	−1870.4 (538.5) [0.126]
ln(*Family income*)	−0.126 (0.004)	−0.038 (0.064)	−0.045 (0.064) [0.319]

Application: Angrist and Evans, Fertility and Labor Supply

Instruments: Same-Sex vs Twins, 1980 PUMS

Instrument for *More than 2 children*	*Same sex*	*Twins-2*
Dependent variable:		
Worked for pay	−0.125	−0.079
	(0.026)	(0.013)
Weeks worked	−5.82	−3.64
	(1.15)	(0.60)
Hours/week	−4.76	−3.33
	(0.98)	(0.51)
Labor income	−1961.7	−1262.2
	(560.5)	(292.8)
ln(*Family income*)	−0.021	−0.071
	(0.067)	(0.035)

Selection Models

Heckman, Vytlacil and co-authors

- Impose Nonparametric Selection Model
  - By Vytlacil (2002), is equivalent to Imbens and Angrist (1994) assumptions
- Goals:
  - To understand relationship between selection and treatment effect heterogeneity.
  - Unify literature with a common set of underlying parameters interpretable across studies.
  - To understand how to connect the results of various disparate IV estimands within a unified framework.
  - Consider strategies other than linear IV.

We will extensively cover the Heckman-Vytlacil analysis in future lectures.

Regression Discontinuity: Sharp Design

Approach originated in Thistlethwaite and Campbell (1960), relatively recent in economics.

Suppose $D = 1$ if $Z \geq z_0$, and $D = 0$ otherwise

$$\Rightarrow \begin{cases} E(Y|Z = z) = E(Y_0|Z = z) & \text{for} \quad z < z_0 \\ E(Y|Z = z) = E(Y_1|Z = z) & \text{for} \quad z \geq z_0 \end{cases}$$

Assume $E(Y_1|Z = z)$, $E(Y_0|Z = z)$ are continuous in z.

$$\Rightarrow \begin{cases} \lim\limits_{\epsilon > 0, \epsilon \downarrow 0} E(Y_0|Z = z_0 - \epsilon) = E(Y_0|Z = z_0) \\ \lim\limits_{\epsilon > 0, \epsilon \downarrow 0} E(Y_1|Z = z_0 + \epsilon) = E(Y_1|Z = z_0) \end{cases}$$

Z sometimes called "forcing variable" or "assignment variable."

Regression Discontinuity: Sharp Design

Key Assumption:
$E(Y_1|Z=z)$, $E(Y_0|Z=z)$ are continuous in z.

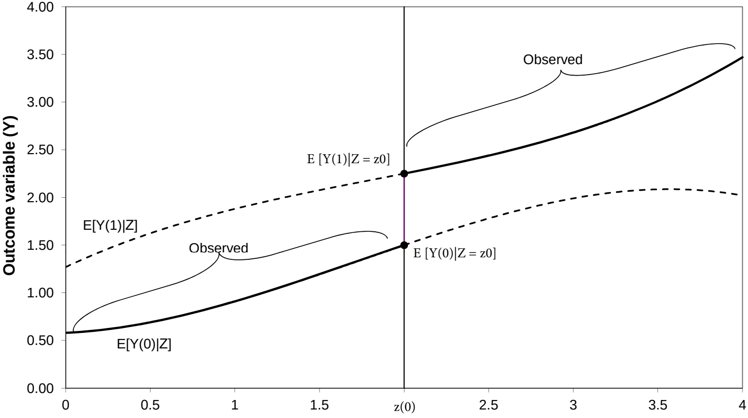
Potential threats to validity of assumption:

1. Manipulation of Z, see e.g. McCrary (2008).
2. Other treatments with same threshold.

Require Z to be continuous, precisely measured.

- See Lee and Card (2008) for analysis of RD with rounding error/discreteness of Z.

Regression Discontinuity (from Lee and Lemieux, 2013)



Source: Lee and Lemieux, 2013, "Regression Discontinuity Designs in Social Sciences."

Regression Discontinuity: Sharp Design (cont'd)

Using that $D = 1$ iff $Z \geq z_0$, and that $E(Y_1|Z = z)$, $E(Y_0|Z = z)$ are continuous in z, we have:

$$\lim_{\epsilon>0,\epsilon\downarrow 0} E(Y|Z = z_0 + \epsilon) - \lim_{\epsilon>0,\epsilon\downarrow 0} E(Y|Z = z_0 - \epsilon)$$

$$\begin{aligned}
&= \lim_{\epsilon>0,\epsilon\downarrow 0} E(Y_1|Z = z_0 + \epsilon) - \lim_{\epsilon>0,\epsilon\downarrow 0} E(Y_0|Z = z_0 - \epsilon) \\
&= E(Y_1|Z = z_0) - E(Y_0|Z = z_0) \\
&= E(Y_1 - Y_0|Z = z_0).
\end{aligned}$$

Regression Discontinuity: Sharp Design (cont'd)

Suggests estimating $\theta = E(Y_1 - Y_0 | Z = z_0)$ by

$$\hat{\theta} \equiv \hat{\mu}_1(z_0) - \hat{\mu}_0(z_0),$$

where

- $\hat{\mu}_1(z_0)$ is the sample average of Y_i among observations with $Z_i \in [z_0, z_0 + h_N]$,
- $\hat{\mu}_0(z_0)$ is the sample average of Y_i among observations with $Z_i \in [z_0 - h_N, z_0)$,
- h_N a bandwidth that decreases to zero slowly as $N \to \infty$.

Estimator is a kernel regression with rectangular kernel on both sides of threshold, see Imbens and Lemieux (2008).

Regression Discontinuity: Sharp Design (cont'd)

This estimator is equivalent to OLS regression on model

$$Y_i = \beta_0 + \theta D_i + \epsilon_i,$$

only using observations with $Z_i \in [z_0 - h_N, z_0 + h_N]$.

Alternatively could run regression on model

$$Y_i = \beta_0 + \theta D_i + \beta_1 (Z_i - z_0) + (\beta_2 - \beta_1) D_i (Z_i - z_0) + \epsilon_i,$$

with observations with $Z_i \in [z_0 - h_N, z_0 + h_N]$. (local linear regression with rectangular kernel on both sides of threshold, has theoretical advantage over local kernel at boundary).

Could also use higher order local polynomial regression. Could use alternative kernels (e.g., triangular kernel).

Regression Discontinuity: Sharp Design (cont'd)

Issues with choice of h_N window:

- Bias vs variance tradeoff in choice of h_N.
- Potential for data mining.
- Choice of h_N for estimation vs for inference.
    - Choices of h_N that are optimal for estimation (minimize MSE) and standard data-driven choices of h_N (e.g., cross-validation) lead to h_N that is too-large for bias of estimator to be negligible, resulting in confidence intervals that are not properly centered and with empirical coverage substantially below nominal coverage. Bootstrapping does not address problem.

See Imbens and Lemieux (2008), Lee and Lemieux (2010), Imbens and Kalyanaraman (2012), and Calonico, Cattaneo and Titiunik (2014).

Regression Discontinuity: Sharp Design (cont'd)

Also common for estimation to fit global polynomial using all observations, not just those within a neighborhood of the threshold. For example:

$$
\begin{aligned}
Y_i = \beta_0 + \theta D_i + \beta_{11}(Z - z_0) + \beta_{12}(Z - z_0)^2 + \beta_{13}(Z - z_0)^3 \\
+ (\beta_{21} - \beta_{11}) D_i (Z - z_0) + (\beta_{22} - \beta_{12}) D_i (Z - z_0)^2 + (\beta_{23} - \beta_{13}) D_i (Z - z_0)^3 \\
+ \epsilon_i .
\end{aligned}
$$

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Application: Effect of Incumbency in U.S. House Elections

- Lee, 2008, "Randomized experiments from non-random selection in U.S. House Elections," *Journal of Econometrics*.
- Political party that won last election is more likely to win next election.
- Is this due (in part) to a causal effect of incumbency?
- Difficult to evaluate: Same underlying factors that caused the political party to win the last election might drive the same political party to win the next election.

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Electoral Success of Incumbents (Lee, 2008)

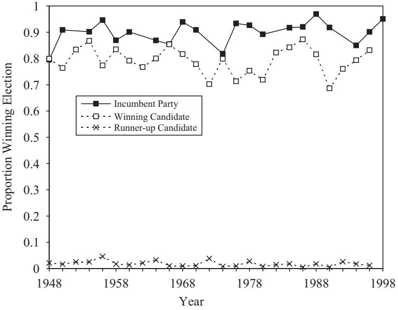


Fig. 1. Electoral success of U.S. House incumbents: 1948–1998. Note: Calculated from ICPSR study 7757 (ICPSR, 1995). Details in Appendix A. Incumbent party is the party that won the election in the preceding election in that congressional district. Due to re-districting on years that end with "2", there are no points on those years. Other series are the fraction of individual candidates in that year, who win an election in the following period, for both winners and runner-up candidates of that year.

Source: Lee (2008) "Randomized experiments from non-random selection in U.S. House Elections."

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Lee (2008) uses RD design, with:

- Outcome Y dummy variable for Democratic candidate winning next election.
- Treatment variable D dummy variable for Democratic candidate winning last election (i.e., being incumbent party).
- Forcing variable Z is fraction of votes awarded to Democratic candidate last election.
- Threshold $z_0 = 1/2$, Democratic candidate won last election if $Z > 1/2$.
- Uses data from 6,558 U.S. House elections over 1946-98.

Effect of Incumbency in US Congressional Elections (Lee, 2008)

In this example, how to interpret:

- Y_0, Y_1?
- Assumption that $D = 1$ if and only if $Z \geq 1/2$?
- Assumption that $E(Y_0|Z = z)$ and $E(Y_1|Z = z)$ are continuous in z?
- Parameter estimated, $E(Y_1 - Y_0|Z = .5)$?

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Graphical presentation:

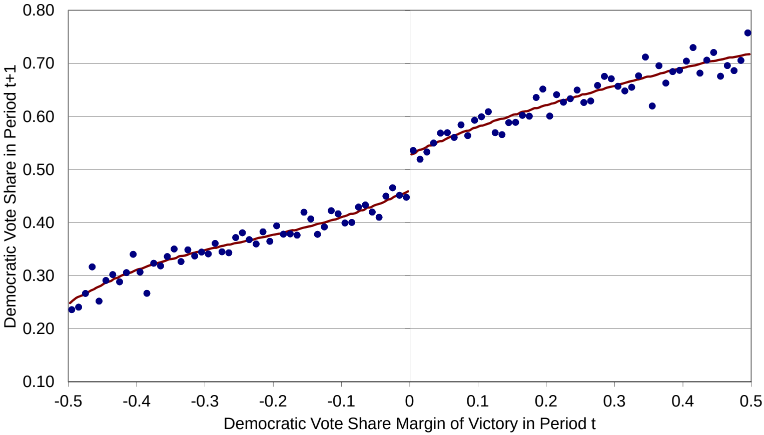
1. Plot fitted values from regression of Y on polynomial (quartic) of Z, fit separately on each side of discontinuity.
2. On same figure, include non-overlapping, binned local averages, using width of 0.01 for the bins.

Using figure:

1. Inspect functional form,
2. Inspect magnitude of jump at threshold.
3. Look for existence of jumps at other points.

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Share of vote in next election, bandwidth of 0.01 (100 bins)



Source: Lee and Lemieux, 2013, "Regression Discontinuity Designs in Social Sciences."

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Local polynomial analysis to estimate treatment effect.

- For local linear regression, cross-validation suggests choosing $h_N = 0.28$ (Lee and Lemieux, 2010).
- Try range of h_N, and range of orders for the polynomial.
- Find strong evidence of causal effect of incumbency, with estimated treatment effects in the range of 0.05 to 0.10.
- Is it plausible that this effect is similar at values of Z far from $1/2$?

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Table 1: RD estimates of the effect of winning the previous election on the share of votes in the next election

Bandwidth:	1.00	0.50	0.25	0.15	0.10	0.05	0.04	0.03	0.02	0.01
Polynomial of order:										
Zero	0.347	0.257	0.179	0.143	0.125	0.096	0.080	0.073	0.077	0.088
	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)	(0.009)	(0.011)	(0.012)	(0.014)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.047]	[0.778]	[0.821]	[0.687]	
One	0.118	0.090	0.082	0.077	0.061	0.049	0.067	0.079	0.098	0.096
	(0.006)	(0.007)	(0.008)	(0.011)	(0.013)	(0.019)	(0.022)	(0.026)	(0.029)	(0.028)
	[0.000]	[0.332]	[0.423]	[0.216]	[0.543]	[0.168]	[0.436]	[0.254]	[0.935]	
Two	0.052	0.082	0.069	0.050	0.057	0.100	0.101	0.119	0.088	0.098
	(0.008)	(0.010)	(0.013)	(0.016)	(0.020)	(0.029)	(0.033)	(0.038)	(0.044)	(0.045)
	[0.000]	[0.335]	[0.371]	[0.385]	[0.458]	[0.650]	[0.682]	[0.272]	[0.943]	
Three	0.111	0.068	0.057	0.061	0.072	0.112	0.119	0.092	0.108	0.082
	(0.011)	(0.013)	(0.017)	(0.022)	(0.028)	(0.037)	(0.043)	(0.052)	(0.062)	(0.063)
	[0.001]	[0.335]	[0.524]	[0.421]	[0.354]	[0.603]	[0.453]	[0.324]	[0.915]	
Four	0.077	0.066	0.048	0.074	0.103	0.106	0.088	0.049	0.055	0.077
	(0.013)	(0.017)	(0.022)	(0.027)	(0.033)	(0.048)	(0.056)	(0.067)	(0.079)	(0.063)
	[0.014]	[0.325]	[0.385]	[0.425]	[0.327]	[0.560]	[0.497]	[0.044]	[0.947]	
Optimal order of the polynomial	6	3	1	2	1	2	0	0	0	0
Observations	6558	4900	2763	1765	1209	610	483	355	231	106

Notes: Standard errors in parentheses. P-values from the goodness-of-fit test in square brackets. The goodness-of-fit test is obtained by jointly testing the significance of a set of bin dummies included as additional regressors in the model. The bin width used to construct the bin dummies is .01. The optimal order of the polynomial is chosen using Akaike's criterion (penalized cross-validation)

Source: Lee and Lemieux, 2013, "Regression Discontinuity Designs in Social Sciences."

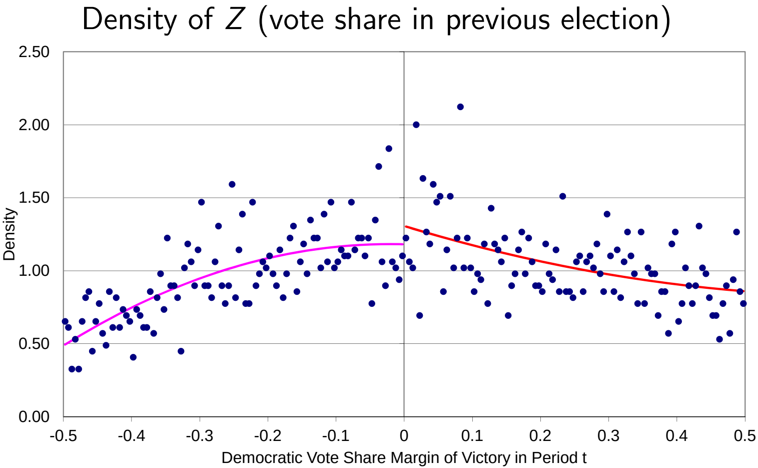
Effect of Incumbency in US Congressional Elections (Lee, 2008)

Visually examine evidence of manipulation of Z by inspecting histogram of its empirical distribution.

- Plots non-overlapping, binned local frequencies using widths of 0.005 for each bin.
- On same figure, fitted values from smooth second order polynomial, fit separately on each side of threshold.

Inspect whether there is evidence of bunching on one side of threshold

Effect of Incumbency in US Congressional Elections (Lee, 2008)



Source: Lee and Lemieux, 2013, "Regression Discontinuity Designs in Social Sciences."

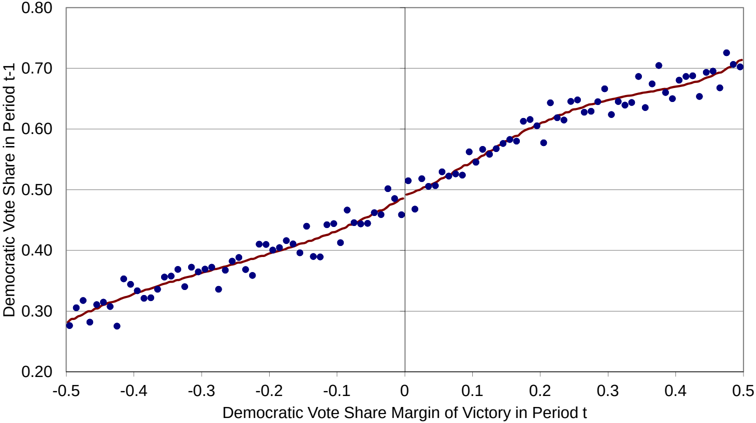
Effect of Incumbency in US Congressional Elections (Lee, 2008)

Falsification analysis:

- Repeat analysis, but now replacing Y with variable known not to be effected by the treatment.

Effect of Incumbency in US Congressional Elections (Lee, 2008)

Discontinuity in baseline covariate (share of vote in prior election)



Source: Lee and Lemieux, 2013, "Regression Discontinuity Designs in Social Sciences."

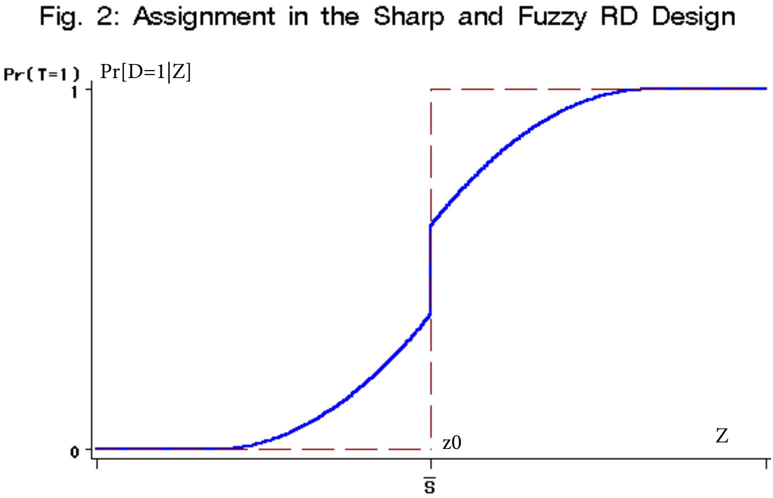
Regression Discontinuity: Fuzzy Design

Can also do "Fuzzy Design"

Suppose $\Pr[D = 1|Z = z]$ discontinuous at z_0
$\Rightarrow \lim_{\epsilon>0,\epsilon\downarrow 0} \Pr(D = 1|Z = z_0 + \epsilon) \neq \lim_{\epsilon>0,\epsilon\downarrow 0} \Pr(D = 1|Z = z_0 - \epsilon)$

Suppose $E(Y_1|Z = z)$, $E(Y_0|Z = z)$ are continuous in z.

Regression Discontinuity, $\Pr[D = 1|Z]$, (from van der Klaauw, 2002)



Regression Discontinuity: Fuzzy Design (cont'd)

Estimator based on:

$$\frac{\lim_{\epsilon>0,\epsilon\downarrow 0} E(Y|Z=z_0+\epsilon)-\lim_{\epsilon>0,\epsilon\downarrow 0} E(Y|Z=z_0-\epsilon)}{\lim_{\epsilon>0,\epsilon\downarrow 0} \Pr(D=1|Z=z_0+\epsilon)-\lim_{\epsilon>0,\epsilon\downarrow 0} \Pr(D=1|Z=z_0-\epsilon)}$$

Connection to IV, in particular Wald form of IV?

Regression Discontinuity: Fuzzy Design (cont'd)

Suppose $E(Y_1 - Y_0|D, Z) = E(Y_1 - Y_0|Z)$.

- Includes special case of constant treatment effects conditional on Z
- Is NOT imposing exogeneity of D.
- Relationship to IV assumptions?

Define:

- $\Delta = Y_1 - Y_0$
- $\Delta(Z) = E(Y_1 - Y_0|Z)$

Regression Discontinuity: Fuzzy Design (cont'd)

$$\begin{aligned} E(Y|Z = z_0 + \epsilon) &= E(Y_0 + \Delta D|Z = z_0 + \epsilon) \\ &= E(Y_0|Z = z_0 + \epsilon) + \Delta(z_0 + \epsilon)P(z_0 + \epsilon) \end{aligned}$$

$$\begin{aligned} \Rightarrow E(Y|Z = z_0 + \epsilon) &- E(Y|Z = z_0 - \epsilon) \\ &= \{E(Y_0|Z = z_0 + \epsilon) - E(Y_0|Z = z_0 - \epsilon)\} \\ &\quad + \{\Delta(z_0 + \epsilon)P(z_0 + \epsilon) - \Delta(z_0 - \epsilon)P(z_0 - \epsilon)\} \end{aligned}$$

$$\begin{aligned} \Rightarrow \lim_{\epsilon > 0, \epsilon \downarrow 0} \{E(Y|Z = z_0 + \epsilon) - E(Y|Z = z_0 - \epsilon)\} \\ = \Delta(z_0)\{P(z_0+) - P(z_0-)\} \end{aligned}$$

Regression Discontinuity: Fuzzy Design (cont'd)

$$\lim_{\epsilon>0,\epsilon\downarrow 0}\left\{E(Y|Z=z_0+\epsilon)-E(Y|Z=z_0-\epsilon)\right\}$$
$$=\Delta(z_0)\left\{P(z_0+)-P(z_0-)\right\}$$

$$\Rightarrow \frac{\lim_{\epsilon>0,\epsilon\downarrow 0}E(Y|Z=z_0+\epsilon)-\lim_{\epsilon>0,\epsilon\downarrow 0}E(Y|Z=z_0-\epsilon)}{\lim_{\epsilon>0,\epsilon\downarrow 0}\Pr(D=1|Z=z_0+\epsilon)-\lim_{\epsilon>0,\epsilon\downarrow 0}\Pr(D=1|Z=z_0-\epsilon)}$$
$$=\frac{\Delta(z_0)\left\{P(z_0+)-P(z_0-)\right\}}{\left\{P(z_0+)-P(z_0-)\right\}}=\Delta(z_0)=E(Y_1-Y_0|Z=z_0)$$

Regression Discontinuity: Fuzzy Design (cont'd)

More generally, with heterogeneous treatment effect, analysis similar to instrumental variables with heterogeneous effect but with Z as a valid instrument only locally to the discontinuity.

- Identify average effect of treatment at $Z = z_0$ for those who would receive treatment if $Z > z_0$ and would not receive treatment if $Z < z_0$.

See Hahn, Todd, and Van der klaauw (2001) for estimation theory, further analysis.

In practice, estimation often done with parametric TSLS, or series based TSLS.

Effect of Financial Aid on Enrollment (van der Klaauw, 2002)

Application: Effect of Financial Aid on Enrollment

- van der Klaauw (2002), "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach" (International Economic Review)
- Examine effect of financial aid offer on enrollment in particular college.
- Difficult to evaluate:
    - Same characteristics that determine financial aid offer also determine financial aid offers from other schools, outside options.
    - Do not observe all relevant characteristics.

Effect of Financial Aid on Enrollment (van der Klaauw, 2002)

van der Klaauw has data on particular U.S. college.

- Let
  $S = \phi_0 \times (\text{first three digits of total SAT score}) + \phi_1 \times \text{GPA}$
- Financial aid offer depends on S as well as other factors.
- College placed students in categories based on threshold values of S, financial aid offers discontinuous at those threshold values.
- Financial aid offer

$$F_i = E(F_i|S_i) + \epsilon_i$$
$$= f(s_i) + \gamma_1 \mathbf{1}[S_i \geq \bar{S}_1] + \gamma_2 \mathbf{1}[S_i \geq \bar{S}_2] + \gamma_3 \mathbf{1}[S_i \geq \bar{S}_3] + \epsilon_i.$$

where $f(\cdot)$ continuous function.

Effect of Financial Aid on Enrollment (van der Klaauw, 2002)

Key idea:

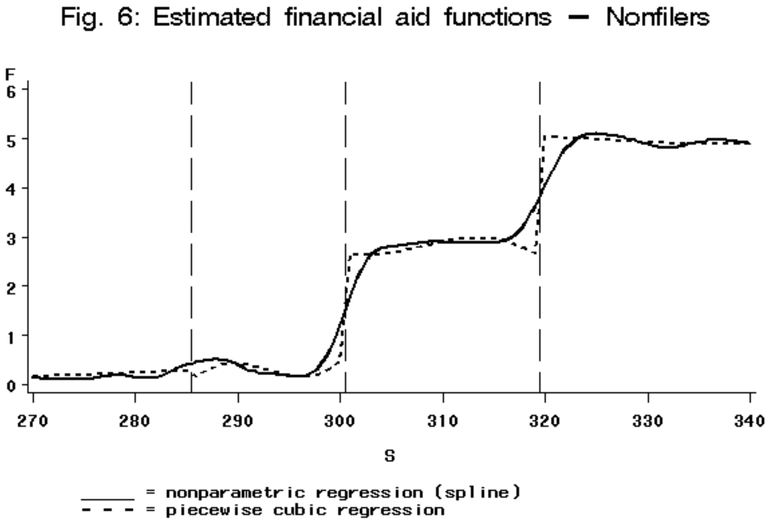
- Outside options likely to depend on GPA and SAT scores (and thus on S) as well as other factors that determine F_i.
- But financial aid offers discontinuous at $S_i =$ each of the three thresholds, outside options unlikely to be discontinuous at exact same values.
- Motivates RD strategy.

Effect of Financial Aid on Enrollment (van der Klaauw, 2002)

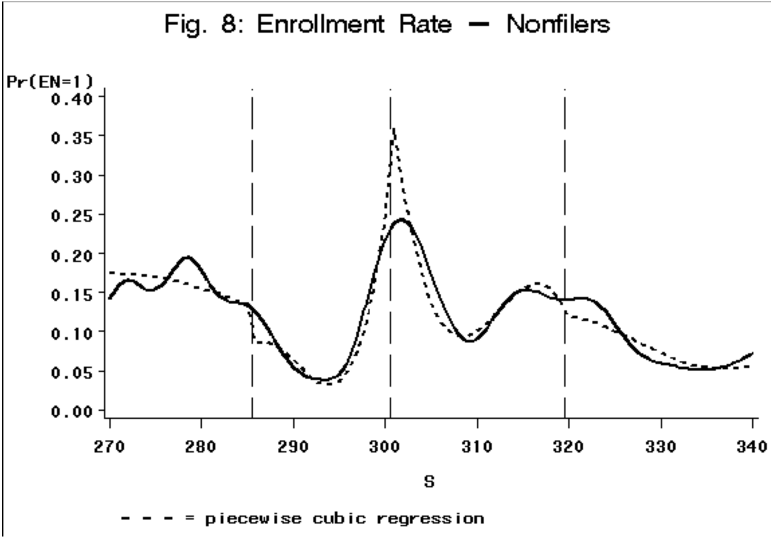
TABLE 1: SAMPLE MEANS

	FILERS				NONFILERS			
Variable	Total	Enrol	Not Enrol	% Enrol	Total	Enrol	Not Enrol	% Enrol
GPA	3.40	3.34	3.43		3.26	3.19	3.28	
SAT	1160	1133	1171		1179	1159	1182	
F	5080	6018	4673		1012	777	1052	
$\%F = 0$	0.12	0.07	0.14		0.73	0.77	0.72	

Effect of Financial Aid on Enrollment (van der Klaauw, 2002)



Effect of Financial Aid on Enrollment (van der Klaauw, 2002)



Effect of Financial Aid on Enrollment (van der Klaauw, 2002)

Table 7: Comparison to OLS estimates

OLS Estimates				RD Estimate
no covariates	covariates	quadratic in S	covariates+ quadratic in S	
FILERS (2225 observations)				
0.030	0.014	0.049	0.013	0.051
(0.003)	(0.003)	(0.003)	(0.003)	(0.015)
NONFILERS (1150 observations)				
-0.011	0.013	0.005	0.013	0.019
(0.005)	(0.010)	(0.012)	(0.014)	(0.011)

The OLS estimates were obtained by regressing EN_i on a constant, the actual financial aid offer F_i, as well as the additional covariates listed. For filers 15 variables were included: GPA, SAT, the individual's age, gender, US citizenship, 2 binary indicators for race, 6 indicators for the applicant's state of residence, a quadratic in parental income and a quadratic in transferable federal and state aid. For nonfilers all variables except the parental income and federal/state aid variables were included. Heteroskedasticity-consistent standard errors in parentheses.

Bounds/Set Identification

If Y is bounded, than without any further assumptions we can bound/set-identitify the treatment parameters (see, e.g., Manski, 1989).

For example, suppose $Y = 0, 1$.

Consider TT.

$$E(Y_1 - Y_0|D = 1) = \Pr[Y_1 = 1|D = 1] - \Pr[Y_0 = 1|D = 1]$$

- We identify $\Pr[Y_1 = 1 \mid D = 1] = \Pr[Y = 1|D = 1]$.
- We do not identify $\Pr[Y_0 = 1 \mid D = 1]$, but we know that:

$$0 \leq \Pr[Y_0 = 1 \mid D = 1] \leq 1$$

Bounds/Set Identification (cont'd)

$$E(Y_1 - Y_0|D = 1) = \Pr[Y_1 = 1|D = 1] - \Pr[Y_0 = 1|D = 1]$$

where:

- $\Pr[Y_1 = 1 \mid D = 1] = \Pr[Y = 1|D = 1]$
- $0 \leq \Pr[Y_0 = 1 \mid D = 1] \leq 1$

and thus

$$Pr[Y = 1|D = 1] - 1 \leq E(Y_1 - Y_0|D = 1) \leq \Pr[Y = 1|D = 1]$$

i.e., we *set-identify* (bound) TT.

Bounds/Set Identification (cont'd)

- Key trade-off: strength of restrictions versus size of bounds.
- Important: obtaining tight bounds.
- Some examples of bounding/identification analysis for treatment effects:
    - Manski (1989), Manski (1990), Heckman, Smith with Clements (1997), Heckman and Vytlacil (2001), Manski and Pepper (2002), Blundell, Gosling, Ichimura and Meghir (2007), Shaikh and Vytlacil (2010), Fan and Wu (2010), Fan and Park (2010), Chiburis (2010),...

Bounds/Set Identification (cont'd)

There are non-trivial issues with constructing confidence intervals for set-identified parameters.

- Confidence interval that contains the set with the correct nominal probability, or that contain any value in the set with the correct nominal probability?
- Sometimes use explicit form of bounds for estimation/inference, alternatively often represent bounds as set of solutions to a minimization problem.
- Often issues of uniformity for confidence intervals/inference.
- Often naive bootstrap invalid, often nonstandard distribution theory.

Bounds/Set Identification (cont'd)

Examples of important papers on estimation/inference for set-identified parameters include:

- Manski and Tamer (2002), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Andrews and Guggenberger (2009), Romano and Shaikh (2008, 2010),...