Problem Set 2 Solution

Juan C. Yamin
Brown University

December 10, 2024

Roadmap of Talk

Problem Set Solution - Question 1

Problem Set Solution - Question 2

Regression Discontinuity Design

- Key components:
  1. Running Variable X_i: margin of victory of PAN mayoral candidate
  2. Treatment $D_i = \mathbb{1}\{X_i > 0\}$: PAN mayor electoral victory
  3. Cutoff: $\mathbb{P}(D_i = 1 \mid X_i = x)$ changes discontinuously at the cutoff
- The observed outcome is

$$Y_i = \begin{cases} Y_i(0) & \text{if } X_i \leq 0 \\ Y_i(1) & \text{if } X_i > 0 \end{cases}$$

$\Rightarrow$ The conditional expectation function is given by:

$$\mathbb{E}[Y_i \mid X_i] = \begin{cases} \mathbb{E}[Y_i(0) \mid X_i] & \text{if } X_i < 0 \\ \mathbb{E}[Y_i(1) \mid X_i] & \text{if } X_i \geq 0 \end{cases}$$

Regression Discontinuity Design

- For any value x, the average treatment effect is $\mathbb{E}[Y_i(1) \mid X_i = x] - \mathbb{E}[Y_i(0) \mid X_i = x]$ (*i.e.*, vertical distance between the curves at x)

- Problem: This distance cannot be directly estimated because we never observe both curves for the same value of x

- Solution: Extrapolation - at the cutoff, we *almost* observe both curves!
  - Consider units with $X_i = 0$ (treatment) and $X_i = -\epsilon$ (control)
  - Assume mean potential outcomes at 0 are similar to the mean potential outcomes at points near 0
  - The units with $X_i = 0$ and $X_i = -\epsilon$ are very similar except for their treatment statust!

- RD treatment effect is thus $\tau \equiv \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = 0]$

Regression Discontinuity Design

- The main assumption is continuity: As x gets closer to the cutoff, the mean potential outcomes $\mathbb{E}[Y_i(1) \mid X_i = x]$ gets closer and closer to its value at the cutoff, $\mathbb{E}[Y_i(1) \mid X_i = 0]$ (and analogously for the untreated)

- Because of continuity, we can focus on observations above and below the cutoff in a very small neighborhood around it:

$$\begin{aligned}
\lim_{x \to 0^+} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i \mid X_i = x] &= \lim_{x \to 0^+} \mathbb{E}[Y_i(1) \mid X_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i(0) \mid X_i = x] \\
&= \mathbb{E}[Y_i(1) \mid X_i = 0] - \mathbb{E}[Y_i(0) \mid X_i = 0] \\
&= \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = 0]
\end{aligned}$$

$\Rightarrow$ observations just below the cutoff approximate the average outcome that units just above the cutoff would have had if they had received the control condition instead of the treatment

Parametric RD

- If we knew $\mu_1(x) = \mathbb{E}[Y_i(1) \mid X_i = x]$ and $\mu_0(x) = \mathbb{E}[Y_i(0) \mid X_i = x]$ were in some parametric class (*i.e.*, $\mu_d(x) = p(x)'\gamma_d$ where $p(x)$ is a vector of polynomials in x):

$$\begin{aligned}\mathbb{E}[Y_i \mid X_i] &= \mathbb{E}[Y_i(0) \mid X_i = x] + D_i\Big(\mathbb{E}[Y_i(1) \mid X_i] - \mathbb{E}[Y_i(0) \mid X_i]\Big) \\ &= p(X_i)\gamma_0 + p(X_i)D_i(\gamma_1 - \gamma_0)\end{aligned}$$

- Estimate the effect by regression:

$$Y_i = \beta_0 + \beta_1 D_i + \gamma_1 X_i + \gamma_2 X_i^2 + \gamma_3 D_i X_i + \gamma_4 D_i X_i^2 + \epsilon_i$$

- Note that under the parametric assumption, we can identify effects everywhere, not just at the threshold

- Problem: Functional form assumptions are really strong...

Non-Parametric RD

- Previous approach fits a polynomial approximation globally
    - For sufficiently high-order polynomials, it delivers a good approximation overall
    - But... a poor approximation at boundary points (which is what we care about)
    - Why? The point estimator is heavily influenced by observations far from the boundary

- `rdrobust`: employs local polynomial methods, which focus on approximating the regression functions only near the cutoff
    - localizes the polynomial fit to near the cutoff (omits far-away observations)
    - employs a low-order polynomial approximation (less overfitting)
    - puts higher weight on observations closer to the cutof

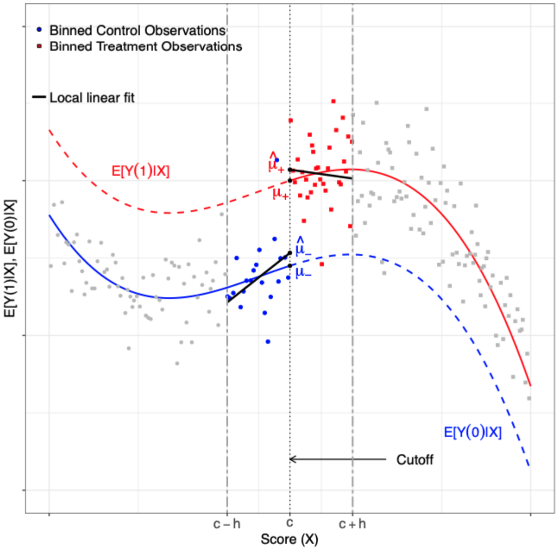
Crash course on local polynomial estimation

1. Choose a polynomial order p, a bandwidth h, a kernel function $K(X_i / h)$
2. For $X_i \geq 0$, estimate by WLS inside h:

$$Y_i = \mu_+ + \beta_1 X_i + \beta_2 X_i^2 + \ldots + X_i^p$$

   $\Rightarrow \hat{\mu}_+$ in an estimate of $\mathbb{E}[Y_i(1) \mid X_i = 0]$
3. Same procedure for $X_i < 0$ to get $\hat{\mu}_-$
4. Compute the estimated effect: $\hat{\tau} = \hat{\mu}_+ - \hat{\mu}_-$

Non-Parametric RD



Crash course on `rdrobust`

- Estimates a local polynomial
- But... there is bias in the local polynomial estimation:
  - Local polynomials imperfectly approximate the true conditional expectation
  - Variance-bias tradeoff in choosing optimal h: larger h, more precision but more bias
- `rdrobust` estimates the bias and corrects the estimates:
  1. *Conventional*: standard local linear estimates + SEs
  2. *Bias-corrected*: adjusted coefficient, standard SEs
  3. *Robust*: adjusted coefficient, SEs adjusted for bias estimation

Crash course on `rdrobust`

```
. rdrobust Y margin_victory,all  vce(hc3) kernel(tri) p(1)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 0 | Left of c  Right of c            Number of obs =        152
-------------------+----------------------            BW type       =      mserd
     Number of obs |        82          70            Kernel        = Triangular
Eff. Number of obs |        22          20            VCE method    =        HC3
    Order est. (p) |         1           1
    Order bias (q) |         2           2
       BW est. (h) |     0.012       0.012
      BW bias (b) |     0.017       0.017
        rho (h/b) |     0.676       0.676

Outcome: Y. Running variable: margin_victory.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  78.697     36.305   2.1676   0.030    7.54003       149.854
    Bias-corrected |  88.248     36.305   2.4307   0.015    17.0912       159.405
            Robust |  88.248     48.626   1.8149   0.070   -7.05618       183.553
--------------------------------------------------------------------------------
```

RD Validation and Sensitivity

- RD continuity fails when:
    1. The cutoff is set systematically: confounding factors change discontinuously
    2. The running variable is manipulable

- Validation:
    - Balance in terms of predetermined covariates
    - No jump on the density of running variable
    - Context-specific placebos

- Sensitivity:
    - Bandwidth choice: sensitivity as units are added or removed at the end points of the neighborhood
    - Donut hole: If systematic manipulation of X_i has occurred, the units closest to the cutoff are those most likely to have engaged in manipulation

- Caution with the inclusion of covariates:
    - ✓ Intended to increase the precision of the RD treatment effect estimator (as in RCTs)
    - ✗ Increase the plausibility of the design $\Rightarrow$ generally requires additional parametric assumptions!

Roadmap of Talk

Problem Set Solution - Question 1

Problem Set Solution - Question 2

IV with controls

Setup

- Many instrumental variables are only valid after conditioning on additional covariates:

$$\Big(Y_i(1), Y_i(0), D_i(1), D_i(0)\Big) \perp\!\!\!\perp Z_i \mid X_i$$

- Under fully saturated models with discrete covariates, Angrist and Imbens (1995):
  - The 2SLS estimand is a convex combination of conditional LATEs:

$$\beta_{2sls} = \mathbb{E}\left[\frac{V\{\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i\}}{\mathbb{E}[V\{\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i\}]}\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i, G = \mathrm{CP}]\right].$$

  - The weights are proportional to the average conditional variance of the population first-stage fitted value, $\mathbb{E}[D_i \mid X_i, Z_i]$, at each value of X_i
  - More weight to covariate values where the instrument creates more variation in fitted values

IV with controls

Problem

- Problem: Most empirical papers don't use a saturated specification... What happens in those cases?

- In general, the LATE interpretation will not apply!

- That is, *saturation* is both sufficient and necessary for 2SLS to have a LATE interpretation

IV with controls

Wald and 2SLS

- The IV estimand is given by

$$\beta_{iv} = \frac{\mathbb{E}[Y\tilde{Z}]}{\mathbb{E}[D\tilde{Z}]}, \text{ where } \tilde{Z} \equiv Z - \mathbb{L}(Z \mid X),$$

  where $\mathbb{L}(Z \mid X) = X'\mathbb{E}[X'X]^{-1}\mathbb{E}[X'Z]$ are the population fitted values from regressing (linearly projecting) Z onto X.

- Intuition: Frisch–Waugh–Lovell (FWL) theorem

IV with controls

IV Decomposition

Proposition (Blandhol, Bonney, Mogstad & Torgovitsky (2022))

Suppose that $\mathbb{E}[Y(d)|X] = \eta_d' X$ *for some (unknown) parameters* η_d, $d = 0, 1$. *Let*

$$\Delta(cp, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = cp, X = x]$$
$$\Delta(at, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = at, X = x]$$

denote the conditional average treatment effects for the compliers and always-takers. Then:

$$\beta_{iv} = \mathbb{E}[\omega(cp, X)\Delta(cp, X)] + \mathbb{E}[\omega(at, X)\Delta(at, X)],$$

where

$$\omega(cp, X) \equiv \mathbb{E}[Z|X](1 - \mathbb{E}[Z|X])\mathbb{P}[G = cp|X]\mathbb{E}[\tilde{Z}D]^{-1},$$
$$\omega(at, X) \equiv \mathbb{E}[\tilde{Z}|X]\mathbb{P}[G = at|X]\mathbb{E}[\tilde{Z}D]^{-1}.$$

IV with controls

Blandhol, Bonney, Mogstad & Torgovitsky (2022)

- If $\mathbb{E}[\tilde{Z}D] > 0$ (*i.e.*, first stage coefficient is positive), then $\omega(\text{cp}, X)$ are negative if and only if $\mathbb{E}[Z|X] > 1$
- In general, $\omega(\text{at}, X)$ are strictly negative with positive probability
- Negative weights $\Rightarrow$ β_{iv} can be negative (positive) even when both the compliers and the always-takers have positive (negative) treatment effects.
- Key intuition:

$$\mathbb{E}[Y\tilde{Z}] = \mathbb{E}\left[\mathbb{E}\left[Y\tilde{Z}|X\right]\right] = \mathbb{E}\left[\underbrace{Cov[Y, \tilde{Z}|X]}_{\text{only contains complier treatment effects}}\right] + \mathbb{E}\left[\underbrace{\mathbb{E}[Y|X]}_{\text{contains all three groups}} \quad \mathbb{E}[\tilde{Z}|X]\right]$$

$\Rightarrow$ If not fully saturated, $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X] \neq 0$, and $\mathbb{E}[Y\tilde{Z}]$ will depend on $\mathbb{E}[Y|X]$

IV with controls

Conclusion

- As long as we fully saturate, $\mathbb{E}[Z|X] = \mathbb{L}[Z|X]$, and Blandhol et al. result is equivalent to Angrist and Imbens
- This can be viewed as a misspecification problem: if we assume $\mathbb{E}[Z|X]$ it is linear, but it is not, we run into issues
- Even if the negative weights seem like a very serious issue, it is still not clear how important these problems are empirically!
  - In our simulation, excluding the interaction did not matter much...
- Recent work on applied econometrics shows many conventional estimators may suffer from related issues (*e.g.*, negative weights $+$ contamination bias):
  - OLS with selection-on-observables with multiple treatments
  - Difference-in-Difference with staggered adoption
  - IV with multiple instruments