Lecture 6: Regression Discontinuity
Part I

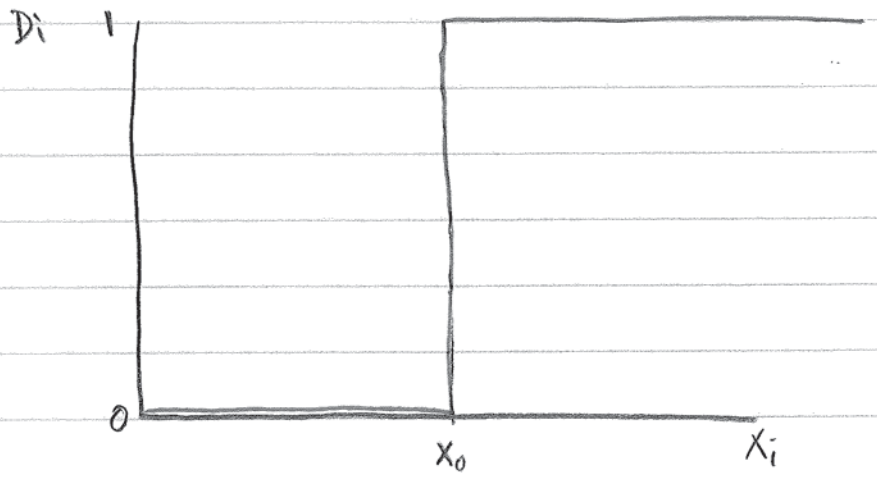
Propensity Score: No precise knowledge of why individuals are treated. We try to estimate a "rule" (the propensity score) for why people are "assigned" to treatment.

Regression Discontinuity: A precise rule for assignment to treatment exists. We don't need to try to guess what this rule is. We just need to use it!

→ Fuzzy Discontinuity: The rule for assignment exists, but it's not stringently enforced ⇒ some who are not assigned to treatment get it anyway; some who are assigned don't get it. ↛ Binding on one side of discontinuity could be problematic!

RD used when treatment is a deterministic and discontinuous function of a covariate, x_i:

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \\ 0 & \text{if } x_i < x_0 \end{cases}$$



Note: differently from p-score matching, there is no value of X_i for which we have both treatment and control observations!
⇒ we need to be willing to extrapolate across covariate values.

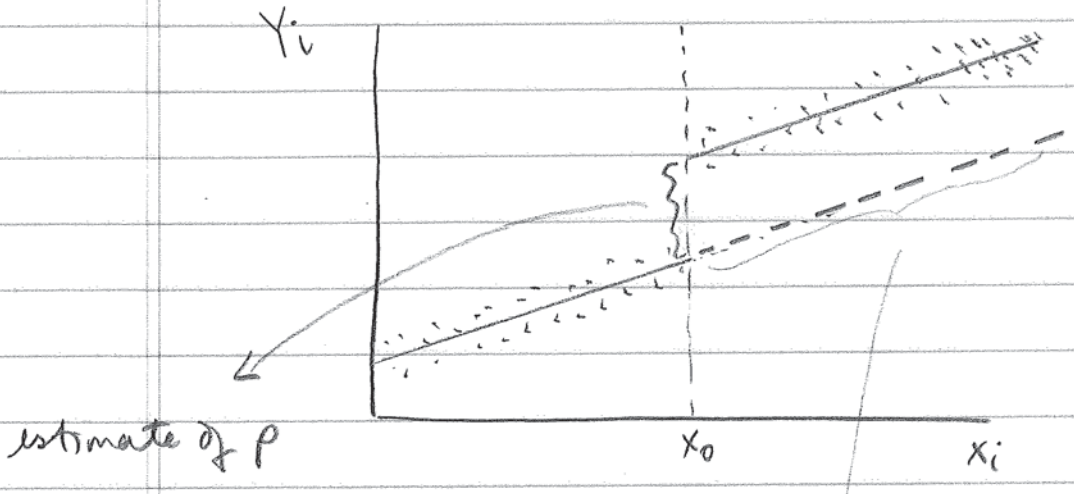
→ Suppose a linear, constant effects model:

$$E[Y_{0i}|X_i] = \alpha + \beta X_i$$

$$Y_{1i} = Y_{0i} + \rho$$

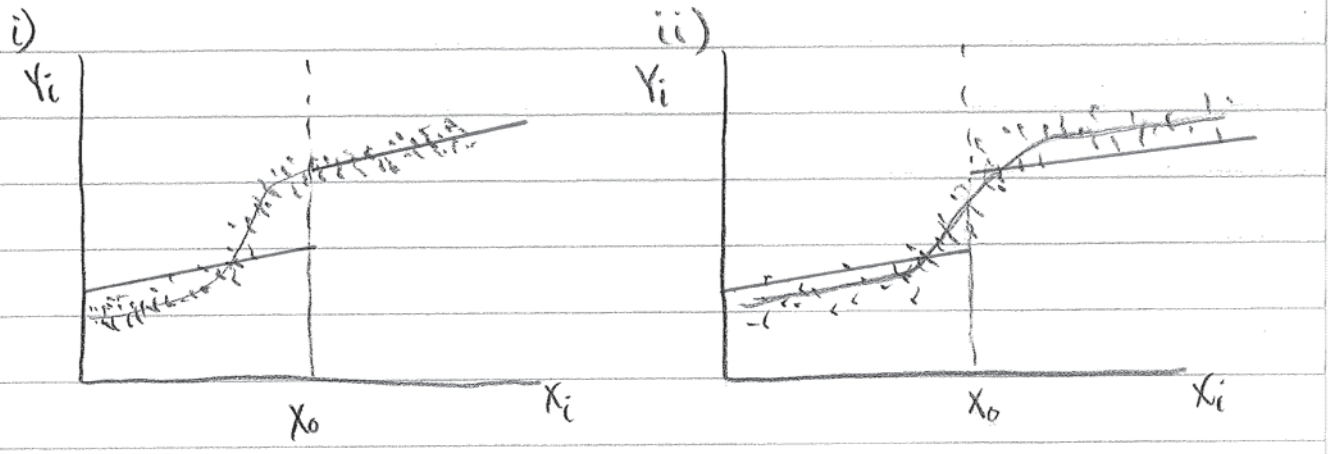
Regression version:

$$Y_i = \alpha + \beta X_i + \rho D_i + \eta_i \qquad (1)$$



We are effectively just extrapolating the line from $X_i < X_0$ and assuming that Y_{1i} would be Y_{0i} at $X_i \geq X_0$ if not for the intervention $D_i = 1$.

→ What if I suppose linear, but $E[Y_{0i} | X_i]$ is really non-linear in X_i?



In case (i), non-linear increase occurs at some $x_i < x_0$. At $x_0 - \Delta$ and $x_0 + \Delta$, there's really no discontinuity. But running a regression like

$$Y_i = \alpha + \beta X_i + \rho D_i + \eta_i \qquad (1)$$

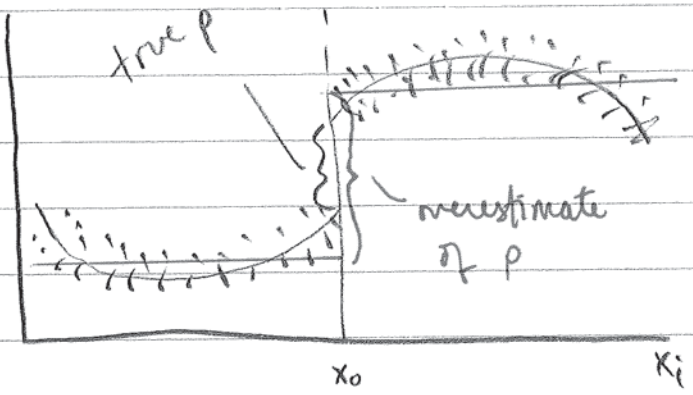
would find a large value of ρ associated with $D_i = 1$. Note, however, that $\rho > 0$ has nothing to do with $D_i = 1$. We've simply misspecified eqn (1). Same in case (ii) except now the non-linearity occurs right at x_0.

→ Can we specify a non-linear CEF?

$$E[Y_{0i} | X_i] = f(x_i)$$

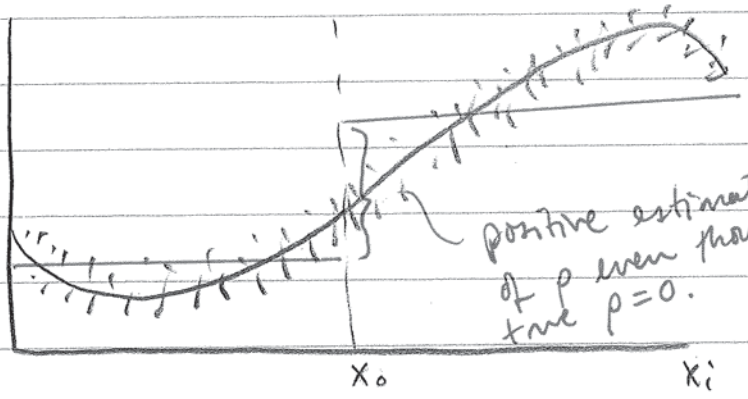
$$Y_i = f(x_i) + \rho D_i + \eta_i$$

i)



Discontinuity in treatment leads to discontinuity in outcome.

ii)



Discontinuity in treatment leads to <u>no</u> discontinuity in outcome.

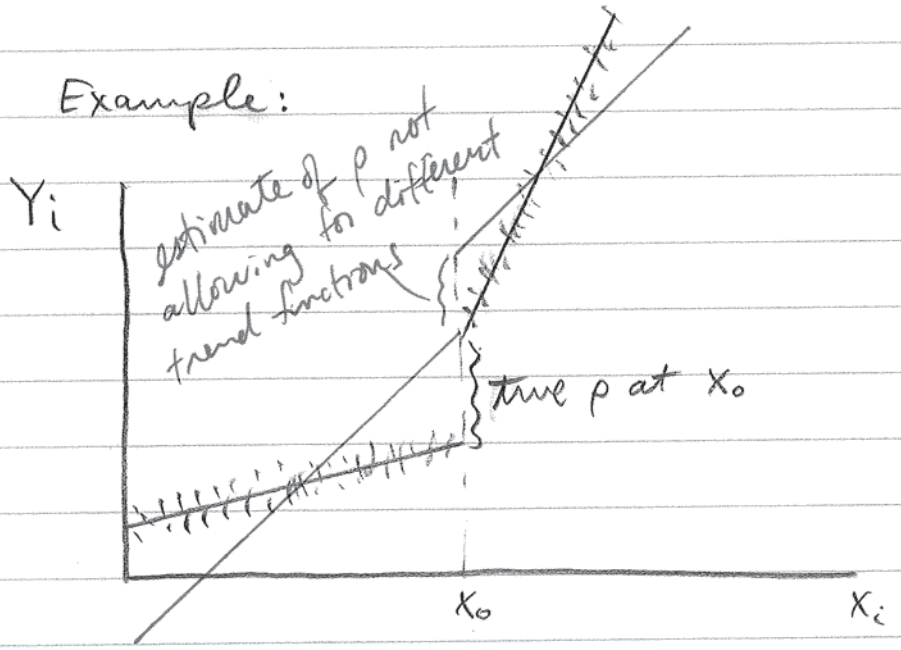
Linear estimation of the CEF would lead to overestimate of ρ in (i) and to positive estimate of ρ in (ii) even though true ρ is zero.

Estimate:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_p X_i^p + \rho D_i + \eta_i \quad (2)$$

→ What if we want to allow different trend functions for $E[Y_{0i}|X_i]$ and $E[Y_{1i}|X_i]$?

Example:



$$E[Y_{0i}|X_i] = f_0(X_i) = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \ldots + \beta_{0p}\tilde{X}_i^p \quad (3)$$
$$E[Y_{1i}|X_i] = f_1(X_i) = \alpha + \rho + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + \ldots + \beta_{1p}\tilde{X}_i^p$$

where $\tilde{X}_i \equiv X_i - X_0$

Why do this? Ensures that at $X_i = X_0$, treatment effect is simply ρ.

Estimate:

$$Y_i = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \ldots + \beta_{0p}\tilde{X}_i^p \quad (4)$$
$$+ \rho D_i + \beta_1^*\tilde{X}_i D_i + \beta_2^*\tilde{X}_i^2 D_i + \ldots + \beta_p^*\tilde{X}_i^p D_i + \eta_i$$

$\beta_1^* = \beta_{11} - \beta_{01}$, $\beta_2^* = \beta_{12} - \beta_{02}$, etc. from (3).

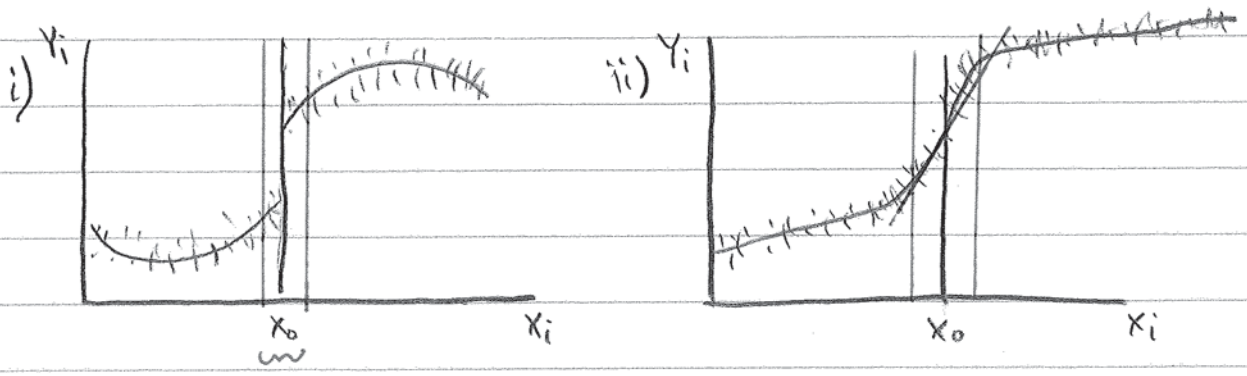
Gives us how the trend changes when D_i changes.

In linear case with trend changes:

$$Y_i = \alpha + \beta_{01}\tilde{X}_i + \rho D_i + \beta_1 \tilde{X}_i D_i + \eta_i$$

At $X_i = X_0$, treatment effect is simply ρ.
At $X_i > X_0$, treatment effect is $\rho + \beta_1 \tilde{X}_i$

→ How much are we willing to extrapolate?
Even a non-linear function is "locally" linear.



Observations within a small window (Δ) of X_0 might be fairly similar. A lot of extrapolation to say that observations far from X_0 are similar to each other.

Overall, fn is non-linear. But within a small window (Δ) of X_0, fn is "locally" linear.

We can look only at data in the neighborhood around a discontinuity, i.e., the interval $[X_0 - \Delta, X_0 + \Delta]$.

Then we have:

$$E[Y_i \mid X_0 - \Delta < X_i < X_0] \simeq E[Y_{0i} \mid X_i = X_0]$$

$$E[Y_i \mid X_0 \leq X_i < X_0 + \Delta] \simeq E[Y_{1i} \mid X_i = X_0]$$

so that:

$$\lim_{\Delta \to 0} E[Y_i \mid X_0 \leq X_i < X_0 + \Delta] - E[Y_i \mid X_0 - \Delta < X_i < X_0] =$$

$$E[Y_{1i} - Y_{0i} \mid X_i = X_0].$$

This is the average causal effect: ρ.

In effect, estimate:

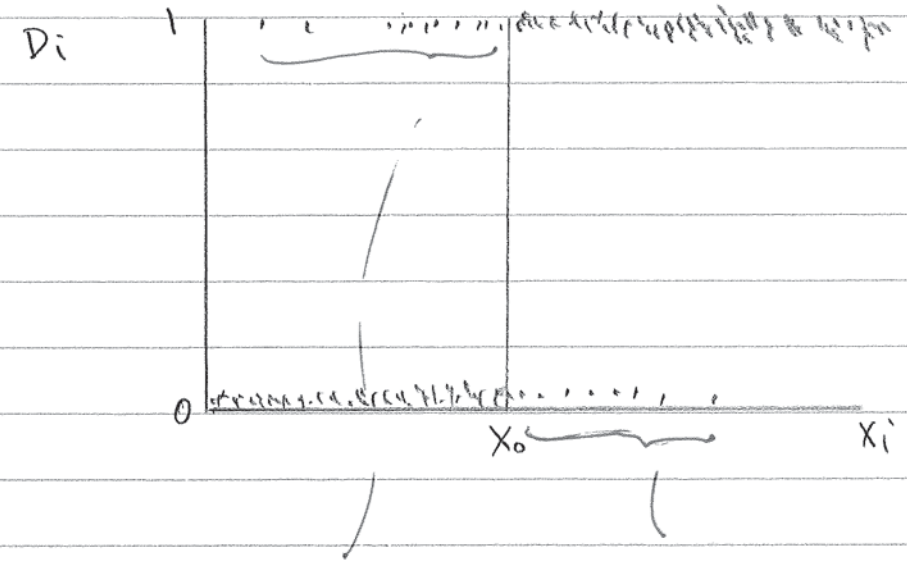
$Y_i = \alpha + \beta X_i + \rho D_i + \eta_i$ where you have dropped all data <u>not</u> in the interval $[X_0 - \Delta, X_0 + \Delta]$.

Problems with this:

1) How to choose Δ? It's arbitrary; choose several and hopefully your estimates of ρ remain stable for different choices of Δ.
2) Reduction in sample size! As $N \downarrow$, standard errors will increase. More difficult to achieve significance.

→ So choose several Δ's. Hopefully coefficient estimate on ρ does not change much even though standard errors $\uparrow$ as $N \downarrow$.

→ What if regression discontinuity is "fuzzy"?



some observations with $x_i < x_0$ get treatment ($D_i = 1$)

some observations with $x_i \geq x_0$ do not get treatment ($D_i = 0$)

D_i is no longer deterministically related to $x_i \geq x_0$. But there is a discrete, discontinuous jump in the probability of being treated at $x_i \geq x_0$.

$$P(D_i = 1 \mid x_i) = \begin{cases} g_0(x_i) & \text{if } x_i < x_0 \\ g_1(x_i) & \text{if } x_i \geq x_0 \end{cases}$$

where $g_1(x_0) \neq g_0(x_0)$.

$$E[D_i|X_i] = P[D_i=1|X_i] = g_0(X_i) + [g_1(X_i) - g_0(X_i)]T_i$$

where $T_i = \begin{cases} 1 \text{ if } X_i \geq X_0 \\ 0 \text{ if } X_i < X_0 \end{cases}$

Now, as we did in eqn (4), we can estimate:

$$D_i = \gamma + \delta_{01}\tilde{X}_i + \delta_{02}\tilde{X}_i^2 + \dots + \delta_{0p}\tilde{X}_i^p + \pi T_i + \delta_1^*\tilde{X}_i T_i + \delta_2^*\tilde{X}_i^2 T_i + \dots + \delta_p^*\tilde{X}_i^p T_i + \varepsilon_i \quad (5)$$

This is our <u>first stage</u>: how <u>assignment to treatment</u> (T_i) affects whether you receive treatment (D_i).

Our structural equation is:

$$Y_i = \alpha + \sum_p \beta_{0p}\tilde{X}_i^p + \rho D_i + \sum_p \beta_p^*\tilde{X}_i^p D_i + \eta_i \quad (6)$$

Our reduced form is:

$$Y_i = \tilde{\alpha} + \sum_p \tilde{\beta}_{0p}\tilde{X}_i^p + \tilde{\rho}T_i + \sum_p \tilde{\beta}_p^*\tilde{X}_i^p T_i + \tilde{\eta}_i \quad (7)$$

Our second stage in 2SLS is:

$$Y_i = \alpha + \sum_p \beta_{0p} \tilde{X}_i^p + \rho \hat{D}_i + \sum_p \beta_p^* \tilde{X}_i^p \hat{D}_i + \eta_i \quad (8)$$

where $\hat{D}_i$ is the predicted, not actual, receipt of treatment based on your assignment to treatment (T_i).

↳ You can also instrument for the probability of receiving treatment around a narrow window of x_0: in the interval $[x_0 - \Delta, x_0 + \Delta]$.

Simple Example:

$$D_i = \gamma + \delta x_i + \pi T_i + \varepsilon_i \qquad \text{1st stage}$$

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \qquad \text{structural}$$

$$Y_i = \tilde{\alpha} + \tilde{\beta} x_i + \tilde{\rho} T_i + \tilde{\eta}_i \qquad \text{reduced form}$$

$$Y_i = \alpha + \beta x_i + \rho \hat{D}_i + \eta_i \qquad \text{2nd stage}$$

We want an estimate of ρ:

$$\pi = \frac{\Delta D_i}{\Delta T_i}$$

$$\tilde{\rho} = \frac{\Delta Y_i}{\Delta T_i}$$

$$\rho = \frac{\Delta Y_i}{\Delta D_i} = \frac{\Delta Y_i}{\cancel{\Delta T_i}} \cdot \frac{\cancel{\Delta T_i}}{\Delta D_i} = \frac{\tilde{\rho}}{\pi} = \frac{\text{reduced form est.}}{\text{1st stage est.}}$$

This is all that IV-2SLS is doing!

→ As with all treatment and control experiments, check that covariates (other X's) balance around discontinuity (X_0).