

# RDD

James J. Heckman  
University of Chicago

Econ 350, Spring 2023  
This draft, March 30, 2023 12:37 Noon

# Regression Discontinuity Estimators and LATE

- Campbell (1969) developed the regression discontinuity design estimator.
- Hahn, Todd, and Van der Klaauw (2001) present an exposition of the regression discontinuity estimator within a LATE framework.

- Standard IV assumptions hold, except that we relax independence assumption (A-1) to assume that  $(Y_1 - Y_0, U_D)$  is independent of  $Z$  conditional on  $X$ .
- *We do not impose the condition that  $Y_0$  is independent of  $Z$  conditional on  $X$ .*

- Using  $Y = Y_0 + D(Y_1 - Y_0)$ , we obtain:

$$\begin{aligned}
 E(Y|X = x, Z = z) &= E(Y_0|X = x, Z = z) \\
 &\quad + E(D(Y_1 - Y_0)|X = x, Z = z) \\
 &= E(Y_0|X = x, Z = z) \\
 &\quad + \int_0^{P(z)} E(Y_1 - Y_0|X = x, U_D = u) du.
 \end{aligned}$$

So

$$\begin{aligned}
 \frac{\frac{\partial}{\partial z} E(Y|X = x, Z = z)}{\frac{\partial}{\partial z} P(z)} &= \frac{\frac{\partial}{\partial z} E(Y_0|X = x, Z = z)}{\frac{\partial}{\partial z} P(z)} \\
 &\quad + E(Y_1 - Y_0|X = x, U_D = P(z))
 \end{aligned}$$

- Likewise, for discrete changes of IV:

$$\begin{aligned}
 & \frac{E(Y|X = x, Z = z) - E(Y|X = x, Z = z')}{P(z) - P(z')} \\
 = & \underbrace{\frac{E(Y_0|X = x, Z = z) - E(Y_0|X = x, Z = z')}{P(z) - P(z')}}_{\text{Bias for LATE}} \\
 & + \underbrace{E(Y_1 - Y_0|X = x, P(z) > U_D > P(z'))}_{\text{LATE}}
 \end{aligned}$$

Recover LATE *plus* a bias term.

- A regression discontinuity design allows analysts to recover a LATE parameter at **a particular value of Z**.
- If  $E(Y_0|X = x, Z = z)$  is continuous in  $z$ , while  $P(z)$  is discontinuous in  $z$  at a particular point, then it will be possible to use a regression discontinuity design to recover a LATE parameter.

- While the regression discontinuity design does have the advantage of allowing  $Y_0$  to depend on  $Z$  conditional on  $X$ , it only recovers a LATE parameter at a particular value of  $Z$  and cannot in general be used to recover either other treatment parameters such as the average treatment effect or the answers to policy questions such as the PRTE.



- The following discussion is motivated by the analysis of Hahn et al. (2001).
- For simplicity, assume that  $Z$  is scalar.
- First, consider LIV while relaxing independence assumption (A-1) to assume that  $(Y_1 - Y_0, U_D)$  is independent of  $Z$  conditional on  $X$  but without imposing that  $Y_0$  is independent of  $Z$  conditional on  $X$ .
- In order to make the comparison with the regression discontinuity design easier, we will condition on  $Z$  instead of  $P(Z)$ .

- Using  $Y = Y_0 + D(Y_1 - Y_0)$ , we obtain:

$$\begin{aligned}
 E(Y|X = x, Z = z) &= E(Y_0|X = x, Z = z) \\
 &\quad + E(D(Y_1 - Y_0)|X = x, Z = z) \\
 &= E(Y_0|X = x, Z = z) \\
 &\quad + \int_0^{P(z)} E(Y_1 - Y_0|X = x, U_D = u) du.
 \end{aligned}$$

So

$$\begin{aligned}
 \frac{\frac{\partial}{\partial z} E(Y|X = x, Z = z)}{\frac{\partial}{\partial z} P(z)} &= \frac{\frac{\partial}{\partial z} E(Y_0|X = x, Z = z)}{\frac{\partial}{\partial z} P(z)} \\
 &\quad + E(Y_1 - Y_0|X = x, U_D = P(z))
 \end{aligned}$$

- We have assumed that  $\frac{\partial}{\partial z}P(z) \neq 0$ .
- We have also assumed that  $E(Y_0|X = x, Z = z)$  is differentiable in  $z$ .
- Notice that under our stronger independence condition (A-1),  $\frac{\partial}{\partial z}E(Y_0|X = x, Z = z) = 0$  so that we identify MTE as before.
- With  $Y_0$  possibly dependent on  $Z$  conditional on  $X$ , we now get MTE plus the bias term that depends on  $\frac{\partial}{\partial z}E(Y_0|X = x, Z = z)$ .

- Likewise, if we consider the discrete change form of IV:

$$\begin{aligned}
 & \frac{E(Y|X = x, Z = z) - E(Y|X = x, Z = z')}{P(z) - P(z')} \\
 = & \underbrace{\frac{E(Y_0|X = x, Z = z) - E(Y_0|X = x, Z = z')}{P(z) - P(z')}}_{\text{Bias for LATE}} \\
 & + \underbrace{E(Y_1 - Y_0|X = x, P(z) > U_D > P(z'))}_{\text{LATE}}
 \end{aligned}$$

so that we now recover LATE plus a bias term.

- Consider a regression discontinuity design.
- Suppose that there exists an evaluation point  $z_0$   $Z$  such that  $P(\cdot)$  is discontinuous at  $z_0$ ,
- Suppose that  $E(Y_0|X = x, Z = z)$  is continuous at  $z_0$ .
- Suppose that  $P(\cdot)$  is increasing in a neighborhood of  $z_0$ .
- Let

$$P(z_0-) = \lim_{\epsilon \downarrow 0} P(z_0 - \epsilon),$$

$$P(z_0+) = \lim_{\epsilon \downarrow 0} P(z_0 + \epsilon).$$

- Note that the conditions that  $P(\cdot)$  is increasing in a neighborhood of  $z_0$  and discontinuous at  $z_0$  imply that  $P(z_0+) > P(z_0-)$ .

- Let

$$\mu(x, z_0-) = \lim_{\epsilon \downarrow 0} E(Y|X = x, Z = z_0 - \epsilon),$$

$$\mu(x, z_0+) = \lim_{\epsilon \downarrow 0} E(Y|X = x, Z = z_0 + \epsilon).$$

- Note that

$$\mu(x, z_0-) = E(Y_0|X = x, Z = z_0) + \int_0^{P(z_0-)} E(Y_1 - Y_0|U_D = u_D) du_D$$

and

$$\begin{aligned} \mu(x, z_0+) &= E(Y_0|X = x, Z = z_0) \\ &+ \int_0^{P(z_0+)} E(Y_1 - Y_0|X = x, U_D = u_D) du_D. \end{aligned}$$

- We used the fact that  $E(Y_0|X = x, Z = z)$  is continuous at  $z_0$ .

$$\mu(x, z_0+) - \mu(x, z_0-) = \int_{P(z_0-)}^{P(z_0+)} E(Y_1 - Y_0 | X = x, U_D = u_D) du_D$$

$$\Rightarrow \frac{\mu(x, z_0+) - \mu(x, z_0-)}{P(z_0+) - P(z_0-)} = E(Y_1 - Y_0 | X = x, P(z_0+) \geq U_D > P(z_0-))$$

- We have recovered a LATE parameter for a particular point of evaluation.

- Note that if  $P(z)$  is only discontinuous at  $z_0$ , then we only identify  $E(Y_1 - Y_0|X = x, P(z_0+) \geq U_D > (z_0-))$  and not any LATE or MTE at any other evaluation points.
- While this discussion assumes that  $Z$  is a scalar, it is straightforward to generalize the discussion to allow for  $Z$  to be a vector.
- For more discussion of the regression discontinuity design estimator and an example, see Hahn et al. (2001), *Econometrica*.