Regression Discontinuity

World Bank SIEF / APHRC
Impact Evaluation Training
2015

Owen Ozier
Development Research Group
The World Bank

6 May 2015
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:

- **Academic test scores:** scholarships or prizes, higher education admission, certificates of merit
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:

- **Academic test scores**: scholarships or prizes, higher education admission, certificates of merit
- **Poverty scores**: (proxy-)means-tested anti-poverty programs
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:

- **Academic test scores**: scholarships or prizes, higher education admission, certificates of merit
- **Poverty scores**: (proxy-)means-tested anti-poverty programs
- **Land area**: fertilizer program or debt relief initiative for owners of plots below a certain area
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:

- **Academic test scores:** scholarships or prizes, higher education admission, certificates of merit
- **Poverty scores:** (proxy-)means-tested anti-poverty programs
- **Land area:** fertilizer program or debt relief initiative for owners of plots below a certain area
- **Date:** age cutoffs for pensions; dates of birth for starting school with different cohorts; date of loan to determine eligibility for debt relief
Regression discontinuity - basic idea

A precise rule based on a continuous characteristic determines participation in a program.

When do we see such rules? Here are five categories, but surely there are more:

- **Academic test scores**: scholarships or prizes, higher education admission, certificates of merit
- **Poverty scores**: (proxy-)means-tested anti-poverty programs
- **Land area**: fertilizer program or debt relief initiative for owners of plots below a certain area
- **Date**: age cutoffs for pensions; dates of birth for starting school with different cohorts; date of loan to determine eligibility for debt relief
- **Elections**: fraction that voted for a candidate of a particular party
Regression discontinuity - basic idea ("sharp")

Regression discontinuity - basic idea ("sharp")

Regression discontinuity - basic idea ("sharp")


Note: Local Average Treatment Effect
Regression discontinuity - basic idea ("sharp")
Regression discontinuity - basic idea

Y axis: perhaps log earnings; X axis: perhaps qualification for labor market program
Regression discontinuity - basic idea ("fuzzy")

Note: “Always-takers,” “Nevertakers,” “Compliers,” and the LATE
Several themes stand out in the half century of RDD’s history. One is its repeated independent discovery. …

Campbell (1960; psychology / education) first named the design regression-discontinuity;

Goldberger (1972; economics) referred to it as deterministic selection on the covariate;

Sacks and Spiegelman (1977,78,80; statistics) studiously avoided naming it;

Rubin (1977; statistics) first wrote about it as part of a larger discussion of treatment assignment based on the covariate;

Finkelstein et al (1996; biostatistics) called it the risk-allocation design;

and Trochim (1980; statistics) finished up calling it the cutoff-based design.”
“Several themes stand out in the half century of RDD’s history. One is its repeated independent discovery. ..." 

Campbell (1960; psychology / education) first named the design regression-discontinuity; 

Goldberger (1972; economics) referred to it as deterministic selection on the covariate; 

Sacks and Spiegelman (1977,78,80; statistics) studiously avoided naming it; 

Rubin (1977; statistics) first wrote about it as part of a larger discussion of treatment assignment based on the covariate; 

Finkelstein et al (1996; biostatistics) called it the risk-allocation design; 

and Trochim (1980; statistics) finished up calling it the cutoff-based design."

But starting in the late 1990s, a large amount of research has appeared in economics. Some papers use the technique to find program impacts; others formalize details of the methodology. 

See Journal of Econometrics, 2008, Volume 142, Number 2 - special issue on RD.
Regression-discontinuity analysis:
An alternative to the ex post facto experiment

Donald L. Thistlethwaite and Donald T. Campbell
National Merit Scholarship Corporation
Northwestern University
Time travel: back to 1960

Observation: academically advanced scholarship winners have “intellectual” attitudes.
Observation: academically advanced scholarship winners have “intellectual” attitudes. Are their attitudes changed by receiving the scholarship? (Is it a causal link?)
Observation: academically advanced scholarship winners have “intellectual” attitudes. Are their attitudes changed by receiving the scholarship? (Is it a causal link?)

Outcome: scholarships

Fig. 2. Regression of success in winning scholarships on exposure determinants.
Time travel: back to 1960

Observation: academically advanced scholarship winners have “intellectual” attitudes. Are their attitudes changed by receiving the scholarship? (Is it a causal link?)

Outcome: scholarships

Outcome: attitudes

Fig. 2. Regression of success in winning scholarships on exposure determine

Fig. 4. Regression of attitudes toward intellectualism on exposure determine
RD, a little more formally

*Angrist and Pischke, Chapter 6, pp. 251-267*
RD, a little more formally

Angrist and Pischke, Chapter 6, pp. 251-267

Treatment rule in sharp RD:

\[ D_i = \begin{cases} 
1 & \text{if } x_i \geq x_0 \\ 
0 & \text{if } x_i < x_0 
\end{cases} \]  \hspace{1cm} (1)
RD, a little more formally

*Angrist and Pishke, Chapter 6, pp. 251-267*

Treatment rule in sharp RD:

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

Data generating process:

\[
E[Y_{0i}|x_i] = \alpha + \beta x_i \\
Y_{1i} = Y_{0i} + \rho
\]
RD, a little more formally

*Angrist and Pishke, Chapter 6, pp. 251-267*

Treatment rule in sharp RD:

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

Data generating process:

\[ E[Y_{0i}|x_i] = \alpha + \beta x_i \tag{2} \]
\[ Y_{1i} = Y_{0i} + \rho \tag{3} \]

Regression equation:

\[ Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \tag{4} \]
RD, a little more formally

*Angrist and Pishke, Chapter 6, pp. 251-267*

Treatment rule in sharp RD:

\[
D_i = \begin{cases} 
1 & \text{if } x_i \geq x_0 \\
0 & \text{if } x_i < x_0 
\end{cases}
\]  \hspace{1cm} (1)

Data generating process:

\[
E[Y_{0i}|x_i] = \alpha + \beta x_i \]  \hspace{1cm} (2)

\[
Y_{1i} = Y_{0i} + \rho \]  \hspace{1cm} (3)

Regression equation:

\[
Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \]  \hspace{1cm} (4)

More general (possibly nonlinear) scenario:

\[
Y_i = f(x_i) + \rho D_i + \eta_i \]  \hspace{1cm} (5)
RD, a little more formally

*Angrist and Pishke, Chapter 6, pp. 251-267*

Treatment rule in sharp RD:

\[ D_i = \begin{cases} 
1 & \text{if } x_i \geq x_0 \\
0 & \text{if } x_i < x_0 
\end{cases} \quad (1) \]

Data generating process:

\[ E[Y_{0i}|x_i] = \alpha + \beta x_i \quad (2) \]
\[ Y_{1i} = Y_{0i} + \rho \quad (3) \]

Regression equation:

\[ Y_i = \alpha + \beta x_i + \rho D_i + \eta_i \quad (4) \]

More general (possibly nonlinear) scenario:

\[ Y_i = f(x_i) + \rho D_i + \eta_i \quad (5) \]

What do we do here?
RD, a little more formally

We can (locally) approximate any smooth function:
RD, a little more formally

We can (locally) approximate any smooth function:

\[ Y_i = f(x_i) + \rho D_i + \eta_i \]  

(6)

Substitute:

\[ f(x_i) \approx \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p \]  

(7)

But because the smooth function may behave differently on either side of the cutoff, we will expand on this. First, transform \( x_i \) notationally (and for ease of regression). Let

\[ \tilde{x}_i = x_i - x_0 \]  

(9)

Angrist and Pishke, Chapter 6, pp. 251-267
RD, a little more formally

We can (locally) approximate any smooth function:

\[ Y_i = f(x_i) + \rho D_i + \eta_i \]  \hfill (6)

Substitute:

\[ f(x_i) \approx \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p \]  \hfill (7)

And thus:

\[ Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \rho D_i + \eta_i \]  \hfill (8)
RD, a little more formally

We can (locally) approximate any smooth function:

\[ Y_i = f(x_i) + \rho D_i + \eta_i \]  

(6)

Substitute:

\[ f(x_i) \approx \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p \]  

(7)

And thus:

\[ Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \rho D_i + \eta_i \]  

(8)

But because the smooth function may behave differently on either side of the cutoff, we will expand on this. First, transform \( x_i \) notationally (and for ease of regression). Let

\[ \tilde{x}_i = x_i - x_0 \]  

(9)
RD, a little more formally

We can (locally) approximate any smooth function:

\[ Y_i = f(x_i) + \rho D_i + \eta_i \]  

(6)

Substitute:

\[ f(x_i) \approx \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p \]  

(7)

And thus:

\[ Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \rho D_i + \eta_i \]  

(8)

But because the smooth function may behave differently on either side of the cutoff, we will expand on this. First, transform \( x_i \) notationally (and for ease of regression). Let

\[ \tilde{x}_i = x_i - x_0 \]  

(9)

*Angrist and Pishke, Chapter 6, pp. 251-267*
RD, a little more formally
*Angrist and Pishke, Chapter 6, pp. 251-267*

Then, allowing different trends (and indeed, completely different polynomials) on either side of the cutoff (with and without the program), we can write the conditional expectation functions:

\[
E[Y_{0i}] = f_0(x_i) = \alpha + \beta_{01} \tilde{x}_i + \beta_{02} \tilde{x}_i^2 + \ldots + \beta_{0p} \tilde{x}_i^p \tag{10}
\]

\[
E[Y_{1i}] = f_1(x_i) = \alpha + \rho + \beta_{11} \tilde{x}_i + \beta_{12} \tilde{x}_i^2 + \ldots + \beta_{1p} \tilde{x}_i^p \tag{11}
\]
RD, a little more formally

Angrist and Pishke, Chapter 6, pp. 251-267

Then, allowing different trends (and indeed, completely different polynomials) on either side of the cutoff (with and without the program), we can write the conditional expectation functions:

\[
E[Y_{0i}] = f_0(x_i) = \alpha + \beta_{01}\bar{x}_i + \beta_{02}\bar{x}_i^2 + \ldots + \beta_{0p}\bar{x}_i^p \\
E[Y_{1i}] = f_1(x_i) = \alpha + \rho + \beta_{11}\bar{x}_i + \beta_{12}\bar{x}_i^2 + \ldots + \beta_{1p}\bar{x}_i^p
\]  

(10)  

(11)

And because \(D_i\) is a deterministic function of \(x_i\) (this is important for writing the conditional expectation):

\[
E[Y_i|X_i] = E[Y_{0i}] + (E[Y_{1i}] - E[Y_{0i}])D_i
\]  

(12)
RD, a little more formally

Angrist and Pishke, Chapter 6, pp. 251-267

Then, allowing different trends (and indeed, completely different polynomials) on either side of the cutoff (with and without the program), we can write the conditional expectation functions:

\[ E[Y_{0i}] = f_0(x_i) = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + ... + \beta_{0p}\tilde{x}_i^p \quad (10) \]

\[ E[Y_{1i}] = f_1(x_i) = \alpha + \rho + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 + ... + \beta_{1p}\tilde{x}_i^p \quad (11) \]

And because \( D_i \) is a deterministic function of \( x_i \) (this is important for writing the conditional expectation):

\[ E[Y_{i|X_i}] = E[Y_{0i}] + (E[Y_{1i}] - E[Y_{0i}])D_i \quad (12) \]

So, substituting in for the regression equation, we can define \( \beta^* = \beta_1 - \beta_0 \), and write:

\[ Y_i = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + ... + \beta_{0p}\tilde{x}_i^p + \]

\[ \rho D_i + \beta_1^* D_i\tilde{x}_i + \beta_2^* D_i\tilde{x}_i^2 + ... + \beta_p^* \tilde{x}_i^p + \eta_i \quad (13) \]
RD, a little more formally

Angrist and Pishke, Chapter 6, pp. 251-267

Then, allowing different trends (and indeed, completely different polynomials) on either side of the cutoff (with and without the program), we can write the conditional expectation functions:

\[ E[Y_{0i}] = f_0(x_i) = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p \]  \hspace{1cm} (10)

\[ E[Y_{1i}] = f_1(x_i) = \alpha + \rho + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 + \ldots + \beta_{1p}\tilde{x}_i^p \]  \hspace{1cm} (11)

And because \( D_i \) is a deterministic function of \( x_i \) (this is important for writing the conditional expectation):

\[ E[Y_i|X_i] = E[Y_{0i}] + (E[Y_{1i}] - E[Y_{0i}])D_i \]  \hspace{1cm} (12)

So, substituting in for the regression equation, we can define \( \beta^* = \beta_1 - \beta_0 \), and write:

\[ Y_i = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \ldots + \beta_{0p}\tilde{x}_i^p + \rho D_i + \beta_1^* D_i\tilde{x}_i + \beta_2^* D_i\tilde{x}_i^2 + \ldots + \beta_p^* \tilde{x}_i^p + \eta_i \]  \hspace{1cm} (13)
RD, a little more formally

But this can all really be simplified in many practical cases. For **small** values of $\Delta$:

\[
E[Y_i | x_0 - \Delta < x_i < x_0] \approx E[Y_0 i | x_i = x_0] \quad (15)
\]

\[
E[Y_i | x_0 \leq x_i < x_0 + \Delta] \approx E[Y_1 i | x_i = x_0] \quad (16)
\]

and then, in the most extreme case:

\[
\lim_{\Delta \to 0} E[Y_i | x_0 \leq x_i < x_0 + \Delta] - E[Y_i | x_0 - \Delta < x_i < x_0] = E[Y_1 i - Y_0 i | x_i = x_0] \quad (17)
\]

So the difference in means in an extremely (vanishingly!) narrow band on each side of the cutoff might be enough to estimate the effect of the program, $\rho$. In practice, usually include linear terms and use a narrow region around the cutoff.
RD, a little more formally

But this can all really be simplified in many practical cases. For small values of $\Delta$:

$$E[Y_i|x_0 - \Delta < x_i < x_0] \approx E[Y_{0i}|x_i = x_0]$$

(15)

$$E[Y_i|x_0 \leq x_i < x_0 + \Delta] \approx E[Y_{1i}|x_i = x_0]$$

(16)
RD, a little more formally

But this can all really be simplified in many practical cases. For small values of $\Delta$:

$$E[Y_i|x_0 - \Delta < x_i < x_0] \approx E[Y_{0i}|x_i = x_0]$$

(15)

$$E[Y_i|x_0 \leq x_i < x_0 + \Delta] \approx E[Y_{1i}|x_i = x_0]$$

(16)

and then, in the most extreme case:

$$\lim_{\Delta \to 0} E[Y_i|x_0 \leq x_i < x_0 + \Delta] - E[Y_i|x_0 - \Delta < x_i < x_0] = E[Y_{1i} - Y_{0i}|x_i = x_0]$$

(17)
RD, a little more formally

But this can all really be simplified in many practical cases. For small values of $\Delta$:

\[
E[Y_i|x_0 - \Delta < x_i < x_0] \approx E[Y_{0i}|x_i = x_0] \\
E[Y_i|x_0 \leq x_i < x_0 + \Delta] \approx E[Y_{1i}|x_i = x_0]
\] (15)

and then, in the most extreme case:

\[
\lim_{\Delta \to 0} E[Y_i|x_0 \leq x_i < x_0 + \Delta] - E[Y_i|x_0 - \Delta < x_i < x_0] = E[Y_{1i} - Y_{0i}|x_i = x_0]
\] (17)

So the difference in means in an extremely (vanishingly!) narrow band on each side of the cutoff might be enough to estimate the effect of the program, $\rho$.

In practice, usually include linear terms and use a narrow region around the cutoff.
RD, a little more formally

But this can all really be simplified in many practical cases. For small values of $\Delta$:

$$E[Y_i|x_0 - \Delta < x_i < x_0] \approx E[Y_0_i|x_i = x_0] \quad (15)$$

$$E[Y_i|x_0 \leq x_i < x_0 + \Delta] \approx E[Y_1_i|x_i = x_0] \quad (16)$$

and then, in the most extreme case:

$$\lim_{\Delta \to 0} E[Y_i|x_0 \leq x_i < x_0 + \Delta] - E[Y_i|x_0 - \Delta < x_i < x_0] = E[Y_{1i} - Y_{0i}|x_i = x_0] \quad (17)$$

So the difference in means in an extremely (vanishingly!) narrow band on each side of the cutoff might be enough to estimate the effect of the program, $\rho$.

In practice, usually include linear terms and use a narrow region around the cutoff.

*Angrist and Pishke, Chapter 6, pp. 251-267*
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[ \text{Prob}(D_i = 1|x_i) = \begin{cases} \frac{g_1(x_i)}{g_0(x_i)} & \text{if } x_i \geq x_0 \\ \frac{g_1(x_i)}{g_0(x_i)} & \text{if } x_i < x_0 \end{cases}, \text{where } g_1(x_0) \neq g_0(x_0) \]  

(18)
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[
Prob(D_i = 1|x_i) = \begin{cases} 
    g_1(x_i) & \text{if } x_i \geq x_0 \\
    g_0(x_i) & \text{if } x_i < x_0 
\end{cases}, \text{where } g_1(x_0) \neq g_0(x_0) \tag{18}
\]

We need a different notation for being on the left or the right of the cutoff, now that \( D_i \) doesn’t jump from zero to one. Let \( T_i = \mathbb{1}(x_i \geq x_0) \).
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[
\Pr(D_i = 1|x_i) = \begin{cases} 
  g_1(x_i) & \text{if } x_i \geq x_0 \\
  g_0(x_i) & \text{if } x_i < x_0
\end{cases}, \text{ where } g_1(x_0) \neq g_0(x_0) \tag{18}
\]

We need a different notation for being on the left or the right of the cutoff, now that \(D_i\) doesn’t jump from zero to one. Let \(T_i = 1(x_i \geq x_0)\).

Now, following the equations in the text, we arrive at two (piecewise) polynomial approximations:

\[
Y_i = \mu + \kappa_1 x_i + \kappa_2 x_i^2 + \ldots + \kappa_p x_i^p + \pi \rho T_i + \zeta_{2i} \tag{19}
\]
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[
\text{Prob}(D_i = 1|x_i) = \begin{cases} 
  g_1(x_i) & \text{if } x_i \geq x_0 \\
  g_0(x_i) & \text{if } x_i < x_0
\end{cases}, \text{where } g_1(x_0) \neq g_0(x_0) \tag{18}
\]

We need a different notation for being on the left or the right of the cutoff, now that \( D_i \) doesn’t jump from zero to one. Let \( T_i = 1(x_i \geq x_0) \).

Now, following the equations in the text, we arrive at two (piecewise) polynomial approximations:

\[
Y_i = \mu + \kappa_1 x_i + \kappa_2 x_i^2 + \ldots + \kappa_p x_i^p + \pi \rho T_i + \zeta_{2i} \tag{19}
\]

\[
D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \ldots + \gamma_p x_i^p + \pi T_i + \zeta_{1i} \tag{20}
\]
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[
\text{Prob}(D_i = 1|x_i) = \begin{cases} 
  g_1(x_i) & \text{if } x_i \geq x_0 \\
  g_0(x_i) & \text{if } x_i < x_0
\end{cases}, \text{where } g_1(x_0) \neq g_0(x_0) \tag{18}
\]

We need a different notation for being on the left or the right of the cutoff, now that \(D_i\) doesn’t jump from zero to one. Let \(T_i = 1(x_i \geq x_0)\).

Now, following the equations in the text, we arrive at two (piecewise) polynomial approximations:

\[
Y_i = \mu + \kappa_1 x_i + \kappa_2 x_i^2 + \ldots + \kappa_p x_i^p + \pi \rho T_i + \zeta_2 i \tag{19}
\]

\[
D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \ldots + \gamma_p x_i^p + \pi T_i + \zeta_1 i \tag{20}
\]

So to estimate \(\rho\), we use instrumental variables, and in essence divide the coefficient estimate on \(T_i\) in the “first stage” regression (variations on Equation 20) by the coefficient estimation on \(T_i\) in the “reduced form” regression (variations on Equation 19).
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[
\text{Prob}(D_i = 1|x_i) = \begin{cases} 
    g_1(x_i) & \text{if } x_i \geq x_0 \\
    g_0(x_i) & \text{if } x_i < x_0 
\end{cases}, \text{ where } g_1(x_0) \neq g_0(x_0) \tag{18}
\]

We need a different notation for being on the left or the right of the cutoff, now that \(D_i\) doesn't jump from zero to one. Let \(T_i = 1(x_i \geq x_0)\).

Now, following the equations in the text, we arrive at two (piecewise) polynomial approximations:

\[
Y_i = \mu + \kappa_1 x_i + \kappa_2 x_i^2 + ... + \kappa_p x_i^p + \pi \rho T_i + \zeta_2i \tag{19}
\]

\[
D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + ... + \gamma_p x_i^p + \pi T_i + \zeta_1i \tag{20}
\]

So to estimate \(\rho\), we use instrumental variables, and in essence divide the coefficient estimate on \(T_i\) in the “first stage” regression (variations on Equation 20) by the coefficient estimate on \(T_i\) in the “reduced form” regression (variations on Equation 19).

Again, as in IV: Exclusion restriction, standard errors
RD, a little more formally

What if the assignment rule is discontinuous, but does not completely determine treatment status?

\[ \text{Prob}(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}, \text{ where } g_1(x_0) \neq g_0(x_0) \] (18)

We need a different notation for being on the left or the right of the cutoff, now that \( D_i \) doesn’t jump from zero to one. Let \( T_i = 1(x_i \geq x_0) \).

Now, following the equations in the text, we arrive at two (piecewise) polynomial approximations:

\[ Y_i = \mu + \kappa_1 x_i + \kappa_2 x_i^2 + \ldots + \kappa_p x_i^p + \pi \rho T_i + \zeta_{2i} \] (19)

\[ D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \ldots + \gamma_p x_i^p + \pi T_i + \zeta_{1i} \] (20)

So to estimate \( \rho \), we use instrumental variables, and in essence divide the coefficient estimate on \( T_i \) in the “first stage” regression (variations on Equation 20) by the coefficient estimage on \( T_i \) in the “reduced form” regression (variations on Equation 19).

Again, as in IV: **Exclusion restriction, standard errors**

*Angrist and Pishke, Chapter 6, pp. 251-267*
Practical considerations

Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:
Practical considerations

Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:

- Visualization
Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:

- Visualization
- Specification: polynomial order, “kernel”
Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:

- Visualization
- Specification: polynomial order, “kernel”
- Bandwidth
Practical considerations

Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:

- Visualization
- Specification: polynomial order, “kernel”
- Bandwidth
- Standard errors (confidence interval)
Practical considerations

Five basic issues are highlighted by Guido Imbens and Thomas Lemieux in their paper, *Regression discontinuity designs: A guide to practice*:

- Visualization
- Specification: polynomial order, “kernel”
- Bandwidth
- Standard errors (confidence interval)
- Specification tests: density, covariates, other jumps
Manipulation of the running variable

What if the population of potential program participants is able to precisely influence the running variable, and knows the program assignment rule?
Manipulation of the running variable

What if the population of potential program participants is able to precisely influence the running variable, and knows the program assignment rule?

Example from Camacho and Conover (2011) in Colombia: program rule became known in 1997; watch what happens.
Poverty score distribution - Camacho and Conover (2011) in Colombia

1994

Percent

Poverty index score

Notes:

Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Notes:
Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Notes:
Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.

<table>
<thead>
<tr>
<th>Year</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>6</td>
</tr>
<tr>
<td>1995</td>
<td>5</td>
</tr>
<tr>
<td>1996</td>
<td>4</td>
</tr>
<tr>
<td>1997</td>
<td>3</td>
</tr>
<tr>
<td>1998</td>
<td>2</td>
</tr>
<tr>
<td>1999</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td>2001</td>
<td>0</td>
</tr>
<tr>
<td>2002</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>0</td>
</tr>
</tbody>
</table>

Poverty index score:

0 7 14 21 28 35 42 49 56 63 70 77 84 91 98
Poverty score distribution - Camacho and Conover (2011) in Colombia
Poverty score distribution - Camacho and Conover (2011) in Colombia

Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Notes:
Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Notes:
Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Notes:
Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
Poverty score distribution - Camacho and Conover (2011) in Colombia

Each figure corresponds to the interviews conducted in a given year, restricting the sample to urban households living in strata levels below four. The vertical line indicates the eligibility threshold of 47 for many social programs.
An example