EC203: Applied Econometrics
Regression Discontinuity Designs

Dr. Tom Martin

University of Warwick

Illustrative reading:

- Mostly Harmless Econometrics: Chapter 6

Regression Discontinuity Designs

RDDs are examples of quasi-experimental designs to estimate causal effects among individuals, firms, countries, ... etc. The basic principle behind RDD is to use an arbitrary rule in an attempt to satisfy the CIA: $E[\epsilon|X] = 0$. For example,

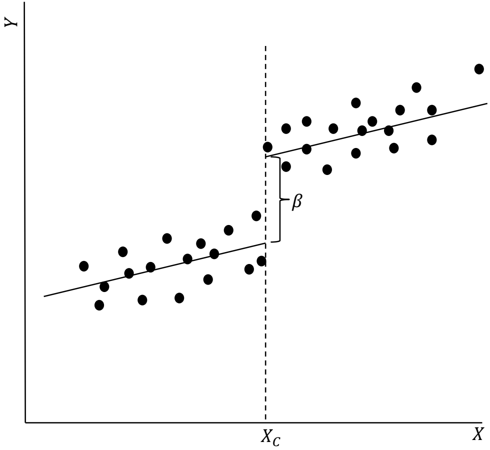
- Question: What is the impact of achieving a merit in a math test (during the year) on a child's final math exam score (at the end of the year)?
- A merit is awarded for a mark on or above 70. The final exam score is measured between $Y \in (0, 100)$.
- Intuitively: comparing the final exam scores, between children just above and below the 70 boundary, will give the causal impact of achieving a merit. Under the assumption that children above and below 70 are similar in all other respects.

Regression Discontinuity Designs

The general set-up is the following:

- We are interesting in estimating the impact of some treatment (D) on some outcome variable (Y).
- You observe a continuous variable, X say, for a group of individuals.
- There is a cut-off along X, X_C say, where if:
  - if $X \geq X_C$ the individual receives the treatment, such that $D = 1$.
  - if $X < X_C$ the individual does not receive the treatment, such that $D = 0$.

Regression Discontinuity Designs



Regression Discontinuity Designs

The causal parameter β can be estimated using a regression of the form:

$$Y_i = \alpha + \delta X_i + \beta D_i + \epsilon_i$$

Where:

- Y_i is the outcome variable
- X_i is known as the assignment variable (or forcing variable or running variable)
- X_C is the cut-off along the assignment variable which assigns the treatment
- D_i is a dummy variable equal to one if $X_i \geq X_C$ and zero if $X_i < X_C$
- ϵ_i is the error term
- The model can be estimated using OLS and β will give the causal estimate if $E[\epsilon|D] = 0$.

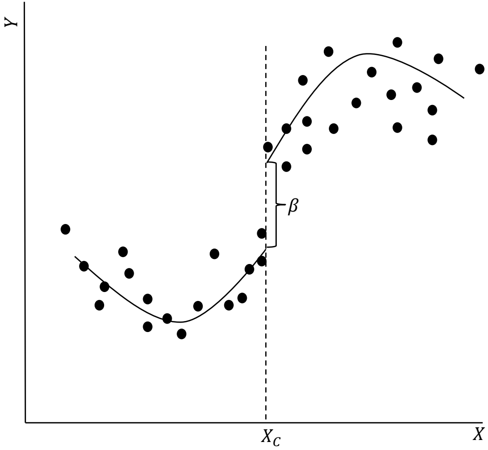
Regression Discontinuity Designs

Note, the closer the observations are to the cut-off the greater our trust in the causal estimate is likely to be, however, we are often restricted in terms of sample size. In the above diagram it is clear that:

- You do not observe treated and control individuals at the same level of X.[1]
- As such the causal (treatment) effect is based on extending (extrapolating) the regression function.
- For this reason it is very important to try different specifications of the regression function.
  - The above graph is an example of a linear assignment mechanism.
  - Another well used mechanisms tested are higher order polynomial forms.

[1] Unlike, for example, observing both male are female wages at different education levels in the regression: $ln(w_i) = \alpha + \beta_1 educ_i + \beta_2 male_i + \epsilon_i$.

Regression Discontinuity Designs



Regression Discontinuity Designs

The causal parameter β can be estimated using a regression of the form:

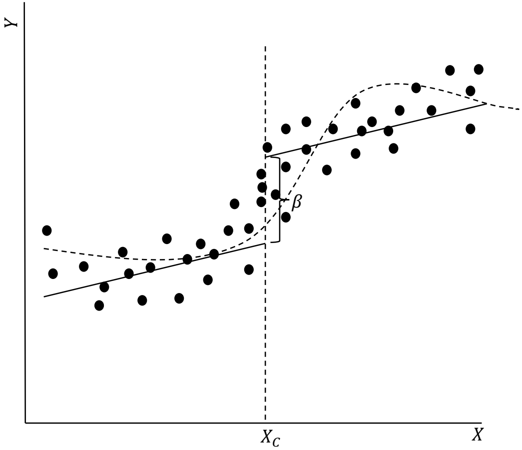
$$Y_i = \alpha + \delta_1 X_i + \delta_2 X_i^2 + \delta_3 X_i^3 + \beta D_i + \epsilon_i$$

Where:

- Y_i is the outcome variable
- The assignment variable is now a polynomial of degree 3 (capture by X_i, X_i^2, X_i^3)
- X_C is the cut-off along the assignment variable which assigns the treatment
- D_i is a dummy variable equal to one if $X_i \geq X_C$ and zero if $X_i < X_C$.
- ϵ_i is the error term
- Note, there is no need for control variables as D_i is random; however, controls may improve precision.

Regression Discontinuity Designs

Mistaking a linear trend (solid line) for a non-linear tread (dashed line). Note: rdplot is a nice command for plotting RDD graphics in Stata.



Regression Discontinuity Designs

In any of the above designs, for β to capture the causal effect we require $E[\epsilon|D] = 0$. As we have seen in other cases, this is an impossible assumption to test. However, as is common practice in randomised control trials, we can check the balancing of covariates:

- Observable characteristics, prior to assignment, such as:
  - gender, age, past test scores, ... etc
- If the treatment is 'as good as randomised' all characteristics should all balance: there are no statistical differences in observables between those individuals just below and just above the cut-off point. You can use t-tests for this purpose.

Regression Discontinuity Designs

Regression discontinuity designs will fail ($E[\epsilon|D] \neq 0$) if individuals can precisely manipulate the assignment variable.

- For example, if students could precisely **choose** their test score, through effort for example. Those who choose a score X_c or just above, will be systematically different from a student who chooses just below the cut-off.

This may or may not be the case, however, the important point is that:

- the existence of a treatment being a discontinuous function of an assignment variable is not sufficient to justify the validity of an RDD.
- If anything, discontinuous rules may generate incentives, causing behavior that would invalidate the RDD approach.