

# 政策评估与因果推断: Stata 应用概述

王群勇 (经济学教授、博士生导师)

南开大学 数量经济研究所

2018年8月19-20日, 广东·顺德

# Contents

## 1 Rubin causal model

- Rubin causal model
- regression and inverse probability weighting
- matching method
- Applications using Stata

## 2 Regression Discontinuity

- sharp regression discontinuity
- fuzzy regression discontinuity
- kink regression discontinuity
- supplementary analysis
- Applications using Stata

## 3 Synthetic control method

- difference in difference
- synthetic control approach for case study
- Applications using Stata

# Contents

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- Given treatment  $W$ , the potential outcome  $Y_i(W)$  can be written

$$Y_i = Y_{0i} + W_i(Y_{1i} - Y_{0i}).$$

Rubin causal model:

$$\tau_i = Y_{1i} - Y_{0i}$$

Counterfactual: we never observe  $Y_{1i}, Y_{0i}$  together (“fundamental problem of causal inference”). So, we focus on the average treatment effect for the population or subpopulation.

$$\tau_{ATE} = E(Y_1 - Y_0)$$

$$\tau_{ATE1} = E(Y_1 - Y_0 | W = 1)$$

$$\tau_{ATE0} = E(Y_1 - Y_0 | W = 0)$$

- Conditional on covariates  $X$ , define

$$\begin{aligned}\mu_1(X) &= E(Y_1|X) = E(Y|X, W = 1) \\ \mu_0(X) &= E(Y_0|X) = E(Y|X, W = 0)\end{aligned}$$

The conditional treatment effect

$$\begin{aligned}\tau_{ATE}(X) &= E(Y_1 - Y_0|X) \\ \tau_{ATE_T}(X) &= E(Y_1 - Y_0|X, W = 1) \\ \tau_{ATE_C}(X) &= E(Y_1 - Y_0|X, W = 0)\end{aligned}$$

From the law of iterated expectations,

$$\tau = E[\tau(X)] = E[\mu_1(X) - \mu_0(X)].$$

- Hernando de Soto (2000): granting de jure property titles to poor land squatters augments their access to credit markets by allowing them to use their property to collateralize debt, fostering broad socioeconomic development.
- compare poor squatters who possess titles to those who don't?
- Problems of confounding factors.
  - (1) observed and unobserved confounders.
  - (2) how to control the observed confounders.

- How to solve the confounding factor problem? Randomized controlled experiment, conditional independence assumption.
- three hallmarks of Randomized controlled experiment (gold standard for drawing inference).
  - (1) The response of experimental subjects assigned to receive a treatment is compared to the response of subjects assigned to a control group.
  - (2) The assignment of subjects to treatment and control groups is done at random, through a randomizing device such as coin flip.
  - (3) The manipulation of the treatment (intervention) is under the control of an experimental researcher.

- Note:
  - (1) Random assignment establishes ex ante symmetry between treatment and control groups and therefor obviates confounding. It ensures any differences in outcomes between the groups are due either to chance error or to the causal effect.
  - (2) Experimental manipulation of treatment establishes further evidence for a causal relationship.



- Difficult or impossible to implement randomized controlled experiment in social studies
  - (1) effect of education on labor market
  - (2) effect of minimum wage on employment
- typical observational studies/data:
  - (1) self-selection into treatment and control groups is the norm.
  - (2) no experimental manipulation.
- natural experiments share attribute (1), and at least partially share attribute (2), but not attribute (3). Natural experiment is observational studies, and it is neither “natural” nor “experiment”.

- Assumption 1: unconfoundness (also called ignorability, conditional independence). Conditional on  $X$ ,  $W$  and  $(Y_0, Y_1)$  are independent.
- mean version of unconfoundness (conditional mean independence)

$$E(Y_0|X, W) = E(Y_0|X), \quad E(Y_1|X, W) = E(Y_1|X).$$

- implications:
  - (1) The assignment mechanism doesn't depend on potential outcome (conditional on  $X$ ), so self-selection is excluded.
  - (2) all confounding factors (i.e., factors correlated with both potential outcomes and with the assignment to the treatment) are observed.
  - (3) conditional on observed confounders, the treatment is as good as randomly assigned.

- identification. Write  $Y = Y_0 + W(Y_1 - Y_0)$ ,

$$\begin{aligned}E(Y|X, W) &= E(Y_0|X, W) + W[E(Y_1|X, W) - E(Y_0|X, W)] \\ &= E(Y_0|X) + W[E(Y_1|X) - E(Y_0|X)] \\ &= \mu_0(X) + W(\mu_1(X) - \mu_0(X))\end{aligned}$$

- 

$$\tau_{ATE}(X) = E(Y_1 - Y_0|X) = E(Y|X, W = 1) - E(Y_0|X, W = 0)$$

- Method to estimate  $\mu_1(X), \mu_0(X)$ :
  - (1)  $Y$  is continuous or limited.
  - (2) parametric or nonparametric.

- sample

|       | bweight | mbsmoke | mage | y0   | y1   |
|-------|---------|---------|------|------|------|
| 3776. | 4026    | 0       | 24   | 4026 | .    |
| 3777. | 4366    | 0       | 27   | 4366 | .    |
| 3778. | 3544    | 0       | 31   | 3544 | .    |
| 3779. | 3500    | 1       | 24   | .    | 3500 |
| 3780. | 3289    | 1       | 23   | .    | 3289 |
| 3781. | 3430    | 1       | 31   | .    | 3430 |
| 3782. | 3147    | 1       | 28   | .    | 3147 |

# Contents

## 1 Rubin causal model

- Rubin causal model
- **regression and inverse probability weighting**
- matching method
- Applications using Stata

## 2 Regression Discontinuity

- sharp regression discontinuity
- fuzzy regression discontinuity
- kink regression discontinuity
- supplementary analysis
- Applications using Stata

## 3 Synthetic control method

- difference in difference
- synthetic control approach for case study
- Applications using Stata

- Second way to establish identification: inverse probability weighting. Note that  $WY = WY_1$ ,

$$\begin{aligned}
 E\left[\frac{WY}{p(X)} \mid X\right] &= E\left[\frac{WY_1}{p(X)} \mid X\right] = E\left[E\left(\frac{WY_1}{p(X)} \mid X, W\right) \mid X\right] \\
 &= E\left[\frac{WE(Y_1 \mid X, W)}{p(X)} \mid X\right] = E\left[\frac{WE(Y_1 \mid X)}{p(X)} \mid X\right] \\
 &= E\left[\frac{W}{p(X)} \mid X\right] E(Y_1 \mid X) = E(Y_1 \mid X).
 \end{aligned}$$

Similarly,

$$E\left[\frac{(1-W)Y}{1-p(X)} \mid X\right] = \mu_0(X)$$

- So, ATE is

$$E \left[ \frac{(W - p(X))Y}{p(X)(1 - p(X))} | X \right] = \mu_1(X) - \mu_0(X) = \tau_{ATE}(X).$$

- ATET is

$$\tau_{ATET} = E \left[ \frac{W - p(X)Y}{(1 - p(X))N_1 / (N_0 + N_1)} \right]$$

- overlap (common support) assumption:  $0 < P(W_i = 1 | X_i) < 1$ .
- Lack of complete overlap creates problems because it means that there are treatment observations for which we have no counterfactuals
- i.i.d. (SUTVA, Stable Unit Treatment Value Assumption).

- potential-outcome linear model

$$Y = (1 - W)Y_0 + WY_1$$

$$Y_0 = X\beta_0 + u_0$$

$$Y_1 = X\beta_1 + u_1$$

and

$$W = 1(Z\gamma + v > 0)$$



# Contents

## 1 Rubin causal model

- Rubin causal model
- regression and inverse probability weighting
- **matching method**
- Applications using Stata

## 2 Regression Discontinuity

- sharp regression discontinuity
- fuzzy regression discontinuity
- kink regression discontinuity
- supplementary analysis
- Applications using Stata

## 3 Synthetic control method

- difference in difference
- synthetic control approach for case study
- Applications using Stata

- idea of matching: search the subjects in the control group which are close enough to the individuals in the treated group, so that we get a balanced sample.
- Two types: covariate matching, propensity score matching.
- let  $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,p})'$ , distance between  $X_i$  and  $X_j$  is

$$\|X_i - X_j\| = [(X_i - X_j)'S^{-1}(X_i - X_j)]^{1/2}$$

$S$  is determined by distance type:

euclidean:  $S = I$ .

invariance:  $S$  is the diagonal matrix of covariance (standardized Euclidean distance)

mahalanobis:  $S$  is the covariance matrix of covariates

- nearest neighbor matching based on covariates.
- propensity score matching based on propensity score

- propensity score

$$P(X_i) = E(W_i|X_i) = P(W_i = 1|X_i)$$

- The propensity score is a balance index, which means  $W_i \perp X_i | P(X_i)$ .

Proof: we need to prove

$$P(W_i = 1|X_i, P(X_i)) = P(W = 1|P(X_i)).$$

$$P(W_i = 1|X_i, P(X_i)) = E[W_i|X_i, P(X_i)] = E(W_i|X_i) = P(X_i).$$

- Theorem: If  $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp W_i | X_i$ , then  $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp W_i | P(X_i)$ .  
Proof:

$$\begin{aligned}
 P(W_i = 1 | Y_{0i}, Y_{1i}, P(X_i)) &= E(W_i | Y_{0i}, Y_{1i}, P(X_i)) \\
 &= E[E(W_i | Y_{0i}, Y_{1i}, X_i, P(X_i)) | Y_{0i}, Y_{1i}, P(X_i)] \\
 &= E[E(W_i | X_i, P(X_i)) | Y_{0i}, Y_{1i}, P(X_i)] \\
 &= E[E(W_i | P(X_i)) | Y_{0i}, Y_{1i}, P(X_i)] \\
 &= E(W_i | P(X_i)).
 \end{aligned}$$

- The unconfoundedness holds conditional on  $X$  or  $P(X)$ .

- matching based on linearized propensity score (log odds ratio)

$$lor_i = \ln \left( \frac{P(X_i)}{1 - P(X_i)} \right)$$

- In logit model,

$$P(X_i) = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}$$

so,

$$lor_i = X_i\beta.$$

- Note: the main purpose of the propensity score estimation is not to predict selection into treatment as good as possible but to balance all covariates (Augurzky and Schmidt, 2000).

- matching quality: Broadly speaking, any differences across groups can be referred to as lack of balance across groups.
- standardized difference in averages

$$sdiffave = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{(s_1^2 + s_0^2)/2}}$$

where  $s_1^2, s_0^2$  are sample variance of the treated and control groups.

- Don't use  $t$ -test to test the balance between groups to avoid the effect of sample size (Imbens and Rubin, 2015)
- One possible problem with the standardised bias approach is that we do not have a clear indication for the success of the matching procedure, even though in most empirical studies a bias reduction below 3% or 5% is seen as sufficient.
- test:

1 t-test: no significant differences should be found.

- what to do if matching quality is not satisfactory?
  - 1 Mis-specification of propensity score equation. Take a step back, include e.g. interaction or higher-order terms in the score estimation and test the quality once again.
  - 2 If after re-specification the quality indicators are still not satisfactory, it may indicate a failure of the CIA (Smith and Todd, 2005) and alternative evaluation approaches should be consider.

- log ratio of standard deviations:

$$sdifsd = \ln(s_1) - \ln(s_0)$$

- linearized propensity score

$$difflor = \frac{\bar{l}or_1 - \bar{l}or_0}{\sqrt{(s_{1,lor}^2 + s_{0,lor}^2)/2}}$$



- Unobserved Heterogeneity: Rosenbaum (2002) Bounds.

$$P(x_i) = P(W_i = 1|x_i) = F(x_i\beta + \gamma u_i)$$

where  $u_i$  is the unobserved variable.

- the odd ratio

$$\frac{P(x_i)/(1 - P(x_i))}{P(x_j)/(1 - P(x_j))} = \frac{P_i(1 - P_j)}{P_j(1 - P_i)} = \frac{\exp(x_i\beta + \gamma u_i)}{\exp(x_j\beta + \gamma u_j)}$$

If  $(x_i, x_j)$  is balanced,

$$\frac{P(x_i)/(1 - P(x_i))}{P(x_j)/(1 - P(x_j))} = \exp(\gamma(u_i - u_j))$$

If there is no unobserved variables,  $\gamma = 0$ .  $e^\gamma$  is a measure of the degree of departure from a study that is free of hidden bias.

- Rosenbaum (2002) shows

$$\frac{1}{e^\gamma} \leq \frac{P(x_i)/(1 - P(x_i))}{P(x_j)/(1 - P(x_j))} \leq e^\gamma$$

- Aakvik (2001) suggests using the Mantel and Haenszel (MH, 1959) test statistic.

$$Q_{MH} = \frac{|Y_1 - \sum_{s=1}^S E(Y_{1s})| - 0.5}{\sqrt{\sum_{s=1}^S \text{Var}(Y_{1s})}}$$

$$Q_{MH}^+ = \frac{|Y_1 - \sum_{s=1}^S E(Y_{1s})^+| - 0.5}{\sqrt{\sum_{s=1}^S \text{Var}(Y_{1s})^+}}$$

$$Q_{MH}^- = \frac{|Y_1 - \sum_{s=1}^S E(Y_{1s})^-| - 0.5}{\sqrt{\sum_{s=1}^S \text{Var}(Y_{1s})^-}}$$

## summarization of steps of matching study

- ① Which one to use PSM or covariate matching?
- ② choose covariates: economic theory, empirical studies, data-driven (significance, cross-validation etc.).
- ③ choose matching algorithm: nearest neighbor, kernel.
- ④ check over-lap (common support):
  - ① method: Minima and Maxima comparison. Assume the propensity score lies within the interval  $[0.07, 0.94]$  in the treatment group and within  $[0.04, 0.89]$  in the control group. Hence, with the 'minima and maxima criterion', the common support is given by  $[0.07, 0.89]$ . Lechner (2002) suggests to check the sensitivity of the results when the minima and maxima are replaced by the 10th smallest and 10th largest observation.
  - ② Bryson, Dorsett, and Purdon (2002) note that when the proportion of lost individuals is small, this poses few problems. However, if the number is too large, there may be concerns whether the estimated effect on the remaining individuals can be

## steps in empirical studies using matching

- 1 check matching quality: balance the distribution of the relevant variables in both the control and treatment group.
  - 1 kernel density plot of propensity score of two groups (unmatched and matched)
  - 2 box plot of propensity score of two groups (unmatched and matched)
  - 3 Q-Q plot
  - 4 After conditioning on  $P(W = 1|X)$ , additional conditioning on  $X$  should not provide new information about the treatment decision.
- 2 sensitivity analysis.

## steps in empirical studies using matching

- 1 check matching quality: balance the distribution of the relevant variables in both the control and treatment group.
  - 1 kernel density plot of propensity score of two groups (unmatched and matched)
  - 2 box plot of propensity score of two groups (unmatched and matched)
  - 3 Q-Q plot
  - 4 After conditioning on  $P(W = 1|X)$ , additional conditioning on  $X$  should not provide new information about the treatment decision.
- 2 sensitivity analysis.

# Contents

## 1 Rubin causal model

- Rubin causal model
- regression and inverse probability weighting
- matching method
- **Applications using Stata**

## 2 Regression Discontinuity

- sharp regression discontinuity
- fuzzy regression discontinuity
- kink regression discontinuity
- supplementary analysis
- Applications using Stata

## 3 Synthetic control method

- difference in difference
- synthetic control approach for case study
- Applications using Stata

- syntax

```
teffects ra (ovar omvarlist , omodel) (tvar)
teffects ipw (ovar) (tvar tmvarlist, tmodel)
teffects aipw (ovar omvarlist, omodel) (tvar tmvarlist, tmodel)
teffects ipwra (ovar omvarlist, omodel) (tvar tmvarlist, tmodel)
```

- *omodel* includes: linear, logit, probit, hetprobit, flogit, fprobit, fhetprobit, poisson.
- *tmodel* includes: logit, probit, hetprobit

- syntax of matching

```
teffects psmatch (ovar) (tvar tmvarlist), options  
teffects nnmatch (ovar omvarlist) (tvar), options
```

- some options:

`nneighbor(k)`: Each individual is matched with at least the specified number of individuals from the other treatment level.

`ematch(varlist)`: exact matching

`gen(newvar)`: observation number of matched individuals

`caliper(c)`: Rosenbaum and Rubin (1985) suggested using  $0.25s(\textit{lor})$  as caliper, where  $s(\textit{lor})$  is the standard deviation of linearized propensity score.



- syntax of matching

```
predict newvars, stats
```

where *stats* includes: *te*, *po*, *distance*.

- syntax of sensitivity analysis

```
. rbounds diff [if] , gamma(numlist)  
. mhbounds outcome [if] , gamma(numlist) treated(varname)  
support(varname)
```

- example:

```
teffects nnmatch re78 ..., gen(mobs)  
gen diff=re78-re78[mobs1]  
rbounds diff, gamma(1(0.2)2)
```

- example:

```
use http://www.stata-press.com/data/r15/cattaneo2
global xlist "prenatal1 mmarried mage fbaby"
global tlist "mmarried c.mage##c.mage fbaby medu"
teffects ra (bweight $xlist)
teffects ipw (bweight) (mbsmoke $tlist)
teffects aipw (bweight $xlist) (mbsmoke $tlist)
teffects ipwra (bweight $xlist) (mbsmoke $tlist)
```

- example:

```
use jtrain2, clear
global xlist "age educ black hispanic married re1974 re1975"
teffects ra (re78 $xlist)
teffects ipw (re78) (train $xlist)
teffects aipw (re78 $xlist) (train $xlist)
teffects ipwra (re78 $xlist) (train $xlist)
```

- nearest neighbor matching (“STEP China.dta”):

```
qui gen sp = 1 if tedu==1
local n=1
while `n'>0 {
capture drop osp
capture teffects nnmatch (lnwage age tenure) (over) if sp==1, /
ematch(female informal occat) osample(osp)
qui count if osp==1
local n=r(N)
if `n'>0 {
dis "`n'"
qui replace sp = 0 if osp==1
}
}
teffects nnmatch (lnwage age tenure) (over) if sp==1, ///
ematch(female informal occat cog2 tech2 noncog2) gen(mobs) nn(1)
tebalance summ
```

- balance test (“STEP China.dta”):

```
gen obs = _n if !mi(mobs1)
gen over2 = over[mobs1]
global v "tenure"
gen x0 = $v[obs] if over==0
replace x0 = $v[mobs1] if over==1
gen x1 = $v[mobs1] if over==0
replace x1 = $v[obs] if over==1
qqplot x0 x1
ttest x0 = x1
sdtest x0 = x1
sktest x0 x1
```

```
gen z0 = $v if over==0
gen z1 = $v if over==1
qqplot z0 z1
ttest $v, by(over)
sdtest $v, by(over)
```

- illustration of balance test

|       | tenure | over | obs  | mobs1 | overCopy | x0   | x1   |
|-------|--------|------|------|-------|----------|------|------|
| 104.  | 3.36   | 0    | 1699 | 1215  | 1        | 3.36 | 2.64 |
| 105.  | 0.96   | 0    | 1723 | 499   | 1        | .96  | .6   |
| 106.  | 1.08   | 0    | 1735 | 17    | 1        | 1.08 | 1.56 |
| 107.  | 1.44   | 0    | 1757 | 496   | 1        | 1.44 | 1.2  |
| 108.  | 2.04   | 0    | 1982 | 1555  | 1        | 2.04 | 2.28 |
| 1571. | 1.56   | 1    | 17   | 43    | 0        | 1.32 | 1.56 |
| 1572. | 0.72   | 1    | 33   | 855   | 0        | .96  | .72  |
| 1573. | 3.12   | 1    | 35   | 505   | 0        | 3.24 | 3.12 |
| 1574. | 0.18   | 1    | 100  | 1192  | 0        | .72  | .18  |
| 1575. | 0.84   | 1    | 103  | 508   | 0        | .72  | .84  |

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata



- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - **sharp regression discontinuity**
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- discontinuities in incentives or ability to receive a treatment.  
examples: school district boundaries, birthdate cutoffs, eligibility threshold.
  - 1 anti-poverty program: households below a given poverty index
  - 2 pension program: targeted to population above a certain age.
  - 3 scholarship: targeted to students with high scores on standardized test.
- forcing (or assignment, running) variable (address, birthday, income, age, etc).

- Angrist and Lavy look at the effects of school class size on kid' s outcomes.
- Maimonides (a twelfth century Rabbinic scholar) rule:  
Twenty-five children may be put it charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed.

- Chay and Greenstone (2005): Willingness to pay for clean air.
- They solve the identification problem by making use of the Clean Air Act Amendments of 1970. A county violates federal standards if:
  - Annual geometric mean of TSP exceeds 75  $\mu\text{g}/\text{m}^3$
  - Second highest daily measure exceeds 260  $\mu\text{g}/\text{m}^3$If you fail the test (nonattainment) the county needs to derive a plan to clean something else.

- Jacob and Lefgren (2004): causal effect of attending summer school.
- (1) rule: third-graders who scored below a threshold (2.75) on either reading or mathematics were required to attend summer school.
- (2) outcome: math score (normalized) after summer school

- when to use RD?
  - ① The beneficiaries/non-beneficiaries can be ordered along a quantifiable dimension.
  - ② This dimension can be used to compute a well-defined index.
  - ③ The index has a cut-off point for eligibility.
  - ④ The index value is what drives the assignment of a potential beneficiary to the treatment (or to non-treatment.)
- The basic idea behind the RD design is that assignment to the treatment is determined, either completely or partly, by the value of a forcing variable being on either side of a fixed threshold. This predictor may itself be associated with the potential outcomes, but this association is assumed to be smooth, and so any discontinuity of the conditional distribution of the outcome as a function of this covariate at the cutoff value is interpreted as evidence of a causal effect of the treatment.

- illustration plot

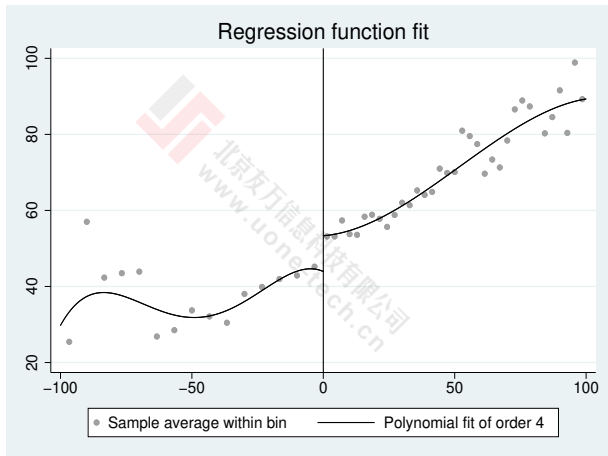


Figure: Sharp RD design

- design: individuals close to the threshold but on different side on otherwise comparable, so any difference in average outcomes between individuals just to one side or the other can be attributed to the treatment.
- two types:  
sharp RD:  $P(W = 1|X \geq c) = 1, P(W = 0|X < c) = 0$ .  
fuzzy RD:  $0 < P(W = 1|X \geq c) < 1, 0 < P(W = 0|X < c) < 1$



- illustration plot

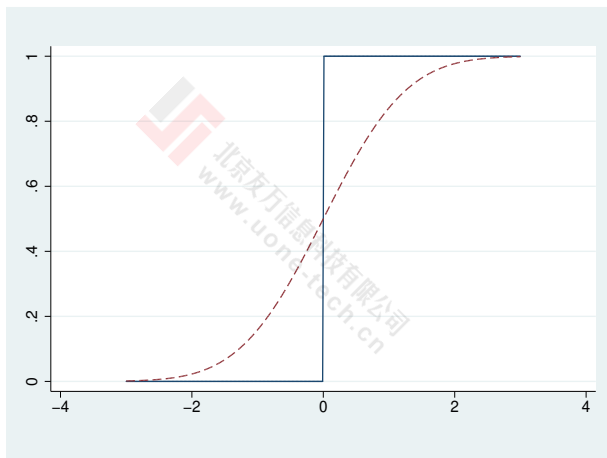


Figure: Sharp RD design

- Assumptions 1: unconfoundedness:  $(Y_{1i}, Y_{0i}) \perp W_i | X_i$ . Because  $W$  is a deterministic function of  $X$ , ignorability necessarily holds.

$$E(Y_g | X, W) = E(Y_g | X), \quad g = 0, 1.$$

- overlap is absolutely violated since  $p(W_i = 1 | X_i < c) = 0$ . So, there is an unavoidable need for extrapolation. We focus on the ATE at  $X = c$  to avoid non-trivial.

- RD estimates the ATE at the discontinuity point, defined as

$$\tau = E(Y_1 - Y_0 | X = c) = \mu_1(c) - \mu_0(c).$$

$$Y = (1 - W)Y_0 + WY_1 = 1(X < c)Y_0 + 1(X \geq c)Y_1$$

So,

$$\begin{aligned} E(Y|X) &= 1(X < c)E(Y_0|X) + 1(X \geq c)E(Y_1|X) \\ &= 1(X < c)\mu_0(X) + 1(X \geq c)\mu_1(X) \end{aligned}$$

- Assumption 2 :  $E[Y_{1i}|X_i], E[Y_{0i}|X_i]$  are continuous.

$$E(Y_0|X = c) = E_{x \uparrow c}(Y|X = x)$$

$$E(Y_1|X = c) = E_{x \downarrow c}(Y|X = x)$$

- identification:

$$\begin{aligned}\tau_{SRD} &= E[Y_1|X = c] - E[Y_0|X = c] \\ &= E_{x \downarrow c}E(Y|X = x) - E_{x \uparrow c}E(Y|X = x)\end{aligned}$$

- Two characteristics of RD: discontinuity at a cut-point, local randomization.

- local randomization

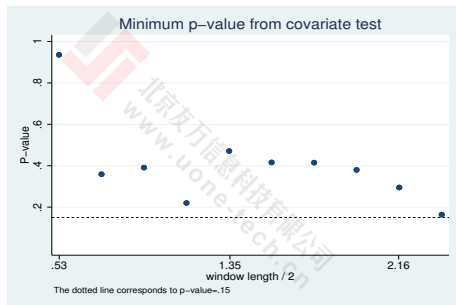


Figure: bandwidth selection for local randomization in RD

- local randomization

| Window length / 2 | Bal. test<br>p-value | Var. name<br>(min p-value) | Bin. test<br>p-value | Obs<c | Obs>=c |
|-------------------|----------------------|----------------------------|----------------------|-------|--------|
| 0.529             | 0.901                | demvoteshlag1              | 0.327                | 10    | 16     |
| 0.733             | 0.311                | demvoteshlag1              | 0.200                | 15    | 24     |
| 0.937             | 0.338                | demvoteshlag1              | 0.126                | 16    | 27     |
| 1.141             | 0.163                | demvoteshlag1              | 0.161                | 20    | 31     |
| 1.346             | 0.325                | population                 | 0.382                | 28    | 36     |
| 1.550             | 0.380                | demvoteshlag1              | 0.644                | 35    | 40     |
| 1.754             | 0.402                | demvoteshlag1              | 0.916                | 44    | 46     |
| 1.958             | 0.370                | demvoteshlag1              | 0.760                | 46    | 50     |
| 2.163             | 0.282                | demvoteshlag1              | 0.621                | 48    | 54     |
| 2.367             | 0.160                | demvoteshlag1              | 0.781                | 56    | 60     |

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 **Regression Discontinuity**
  - sharp regression discontinuity
  - **fuzzy regression discontinuity**
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- FRD

$$P(W = 1|X) = F(X)$$

- From  $Y = Y_0 + W(Y_1 - Y_0)$ ,

$$\begin{aligned} E(Y|X) &= E(Y_0|X) + E(W|X)E(Y_1 - Y_0|X) \\ &= \mu_0(X) + E(W|X)\tau(X) \end{aligned}$$

So,

$$\tau_{FRD} = \frac{E_{X \downarrow c}(Y|X = x) - E_{X \uparrow c}(Y|X = x)}{E_{X \downarrow c}(W|X = x) - E_{X \uparrow c}(W|X = x)}$$



- Assumption 3:

$$E_{x \downarrow c}(W|X = x) \neq E_{x \uparrow c}(W|X = x)$$

- Assumption 4: Local randomization

$$(Y_{0i}, Y_{1i}) \perp W_i | X_i \in (c - \delta, c + \delta)$$

where  $\delta$  is an arbitrary small positive value. The units closest to the cutoff are viewed as being part of a local randomized experiment.

- local linear regression, local polynomial regression, etc.
- Imbens and Lemieux (2008) recommends using local linear methods for estimation process, rather than local constant.
- bandwidth selection: cross-validation.

# Contents

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 **Regression Discontinuity**
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - **kink regression discontinuity**
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 **Regression Discontinuity**
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - **supplementary analysis**
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- test the validity of RD design: individuals could not manipulate systematically the forcing variable around the cutoffs.
- method: test that the densities of the forcing variable are smooth and, in particular, do not jump at the cutoffs (Lee and Lemieux, 2010). If the density of the forcing variable is discontinuous at the threshold, which would suggest that the forcing variable is being manipulated.
- density discontinuity test: McCrary (2008), Otsu, Xu, and Matsushita (2013), Cattaneo, Jansson, and Ma (2016).
- placebo test for assumption: all covariates should be uncorrelated with the treatment when the forcing variable is close to the threshold. (Lee, 2008).

- test for discontinuities of covariates at the threshold. Use each covariate as pseudo-outcome, make RD analysis. The treatment effect shouldn't be significant.
- use other threshold to make RD analysis (pseudo-cutoff point). For example, use the median of the left sample and right sample. The treatment effect should not be significant.
- sensitivity analysis of bandwidth. Try different bandwidth to check the robustness of discontinuity.

- Plot the probability of receiving treatment as a function of the rating variable (judge whether the design is sharp or fuzzy).
- plot the relationship between  $Y$  and  $X$  (to visualize the impact of treatment).
- plot the relationship between covariate and rating variable (to check the internal validity of the design).
- plot the density of the rating variable (to check whether there is any manipulation of  $X$  around the cutpoint).

- Steps to plot relationship between  $Y$  and  $X$ : construct a series of interval  $(b_k, b_{k+1})$ ,  $k = 1, 2, \dots, K$  ( $K = K_0 + K_1$ ). For bandwidth  $h$ ,

$$b_k = c - (K_0 - k + 1)h$$

Plot the average of  $k$ th interval

$$\bar{Y}_k = N_k^{-1} \sum_{i=1}^N Y_i 1(b_k < X_i \leq b_{k+1}).$$

- Use bandwidth selection method.



- the design never allow the researchers to estimate the overall average effect of the treatment. So, the design has fundamentally only a limited degree of external validity. The external validity can be assessed by the credibility of extrapolations to other subpopulations.

# Contents

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - **Applications using Stata**
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- nonparametric local polynomial regression

```
. rdplot dep run, c(cutoff) h(min max) binselect(method) nbin(lnum rnum)  
p(order) kernel(type)
```

- `h(min max)`: support bandwidth (full support by default)
- `kernel(type)`: may be `tri`, `epan`, or `unif`
- `nbins(lnum rnum)`: number of bins
- `binselect(method)`: `es`, the integrated mean squared error (IMSE)-optimal evenly spaced method using spacing estimators; `espr`, IMSE-optimal evenly spaced method using polynomial regression; `esmv`, the mimicking-variance evenly spaced method using spacing estimators; `esmvpr`, mimicking-variance evenly spaced method using polynomial regression; `qs`, IMSE-optimal quantile-spaced method using spacing estimators; `qspr`, IMSE-optimal quantile-spaced method using polynomial regression; `qsmv`, mimicking-variance quantile-spaced method using spacing estimators; `qsmvpr`, mimicking-variance

- nonparametric local polynomial regression

```
. rdrobust dep run, c(cutoff) h(min max) bwselect(method)  
p(order) kernel(type) q(order) covs(varlist) all
```

- nonparametric local polynomial regression

```
. rdbwselect dep run, [ c(cutoff) bwselect(method) q(order)  
p(order) kernel(type) covs(varlist) all ]
```

- RD command for local randomization

```
. rdrandinf dep run, options wl(num) wu(num) cov(varlist)  
. rdwinselect run covariates, options plot
```

where *options* include

```
c(cutoff) kernel(type) stat(statistics) approx  
minobs(num)|wmin(value) obsstep(num)|wstep(value) nwindow(num)
```

- $wl(value)$ : left limit of window
- $wu(value)$ : right limit of window
- $stat(method)$ : may be `ttest`, `ksmirnov`, `ranksum`, or `all`

- RD command for local randomization

```
. rdsensitivity dep run, c(num) wlist(numlist) tlist(numlist) covs(varlist)  
. rdrbounds dep run, prob(varname) gammalist(numlist)
```

- `wlist(numlist)`: list of window lengths to be evaluated. By default, 10 windows around the cutoff, the first one including 10 treated and control observations and then adding 5 observations to each group in subsequent windows.
- `tlist(numlist)`: list of values of the treatment effect under the null to be evaluated. By default, the program uses 10 evenly spaced points within the asymptotic confidence interval for a constant treatment effect in the smallest window to be used.
- `stat(method)`: may be `ttest`, `ksmirnov`, `ranksum`, or `all`

- example (rdrobust\_senate.dta)

```
global cov = "population dopen"  
rdplot demvoteshfor2 demmv, p(2) kernel(tri)  
rdrobust demvoteshfor2 demmv, covs($cov) all  
rdbwselect demvoteshfor2 demmv, all  
  
rdwinselect demmv $cov, c(0) approx wmin(.5) wstep(.125) nwin(20) approx plot  
rdrandinf demvoteshfor2 demmv, wl(-.75) wr(.75) stat(all)  
rdrandinf demvoteshfor2 demmv, wl(-.75) wr(.75) stat(all) covariate($cov)  
rdsensitivity demvoteshfor2 demmv, wlist(.75(.25)2) tlist(0(1)20) reps(1000)
```



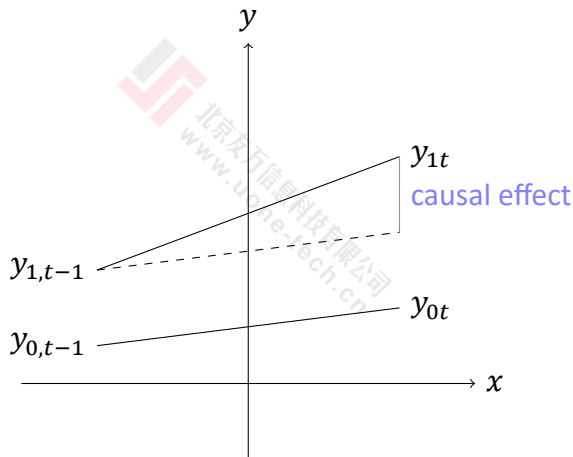
- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

# Contents

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- difference-in-differences (Card, 1990; Peri and Yasenov, 2015).
- Card's question: the effect of the Mariel boatlift, which brought low-skilled Cuban workers to Miami. How the boatlift affected the Miami labor market, and specifically the wages of low-skilled workers?
- Solution of DID: He compares the change in the outcome of interest for the treatment city (Miami) to the corresponding change in a control city. He considers various possible control cities, including Houston, Petersburg, and Atlanta.

- causal effect by DID



- 对于只有两年的面板数据，共同趋势假设是无法直接验证的。在多年的面板数据下，有两种方式可以用以关注 CT：画图和回归。
- (1) 假设考察某一政策冲击对企业生产率的影响，政策发生在 2001 年，样本期间为 1995-2006 年。画出 1995-2001 年间实验组和对照组的年度生产率（年度生产率均值）趋势图，如果两条线的走势完全一致或基本一致，说明 CT 假设是满足的。
- (2) 回归模型：

$$y_{it} = \alpha_0 + \alpha_1 du_i + \sum_{j=1995}^{2000} \alpha_j dt_j + \sum_{j=1995}^{2000} \beta_j du_i \times dt_j + \epsilon_{it}$$

交互项的系数反映的便是，对于政策实施前的某一年，实验组和对照组的差异。如果回归得到的所有交互项都不显著，说明政策实施前实验组和对照组不存在明显的差别；从而 CT

- 安慰剂检验: a) 选取政策实施之前的年份进行处理, 比如原来的政策发生在 2008 年, 研究区间为 2007-2009 年。可以假定政策实施年份为 2006 年, 并将研究区间前移至 2005-2007 年, 然后进行回归;
- b) 选取已知的并不受政策实施影响的群组作为处理组进行回归。如果不同虚构方式下的 DID 估计量的回归结果依然显著, 说明原来的估计结果很有可能出现了偏误。
- 利用不同的对照组进行回归, 看研究结论是否依然一致。
- 选取一个完全不受政策干预影响的因素作为被解释变量进行回归, 如果 DID 估计量的回归结果依然显著, 说明原来的估计结果很有可能出现了偏误。如果回归结果显著, 说明原结果是一定有问题的, 而如果回归结果不显著, 并不一定能表明原结果没问题。

- 预期效应：

$$y_{it} = \alpha_0 + \alpha_1 du_i + \sum_{j=1995}^{2000} \alpha_j dt_j + \sum_{j=1995}^{2000} \beta_j du_i \times dt_j + \epsilon_{it}$$

政策实施年份：2001。检验 2000、1999、... 与 du 交叉项的显著性。如果显著，则说明可能存在预期效应。

- 滞后效应：加入政策之后的时间虚拟变量与 du 的交叉积（比如到 2003 年）。

$$y_{it} = \alpha_0 + \alpha_1 du_i + \sum_{j=1995}^{2003} \alpha_j dt_j + \sum_{j=1995}^{2003} \beta_j du_i \times dt_j + \epsilon_{it}$$

政策实施年份：2001。检验 2001、2002、2003 与 du 交叉项的显著性。如果显著，则说明可能存在滞后效应。

- causal effect by DID

$$\begin{aligned}\tau_{ATT} &= E[Y_{1it} - Y_{0it} | W_i = 1, X_{it}] \\ &= E[Y_{it} | W_i = 1, X_{it}] - E[Y_{i,t-1} | W_i = 1, X_{it}] \\ &\quad - [E[Y_{it} | W_i = 0, X_{it}] - E[Y_{i,t-1} | W_i = 0, X_{it}]] \\ &= E[Y_{it} | W_i = 1, X_{it}] - E[Y_{it} | W_i = 0, X_{it}] \\ &\quad - [E[Y_{i,t-1} | W_i = 1, X_{it}] - E[Y_{i,t-1} | W_i = 0, X_{it}]]\end{aligned}$$



- Propensity score matching DID: 先用 PSM 在原始样本中挑选出基本特征都比较相似的新的实验组和对照组，然后再基于匹配的实验组和对照组进行 DID 回归，这种情况下 CT 假设容易满足。
- 截面数据做 DID? 参考：Chen and Zhou (2007) 研究大饥荒对健康的影响 (CHNS)。
- 连续型政策变量做 DID? 参考：Nancy Qian。

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- synthetic control approach developed by Abadie, Diamond, and Hainm- ueller (2010, 2014) and Abadie and Gardeazabal (2003).
- Solution of synthetic control approach to Card's question: choose weights for each of the three cities so that the weighted average is more similar to Miami than any single city would be.
- choice of weight:
  - (1) minimum distance approach
  - (2) LASSO (Least Absolute Shrinkage and Selection Operator)
  - (3) elastic nets

- treated group: one individual  $i = 1$ .  
control group: many individuals  $i = 2, \dots, N + 1$ .
- treatment effect for  $i$

$$\tau_{it} = Y_{it}^{(1)} - Y_{it}^{(0)}$$

the treatment effect of interest:

$$\tau_{1t} = Y_{1t} - Y_{1t}^{(0)}$$

- synthetic control method:

$$Y_{1t}^{(0)} = \sum_{j=2}^{N+1} w_j Y_{jt}$$

the weight  $w_j$  is chosen to balance  $Y_{1,t-1}$  and  $\sum_{j=1}^{N+1} Y_{j,t-1}$ .

- placebo analysis: replicates the primary analysis with the outcome replaced by a pseudo- outcome that is known not to be affected by the treatment. Thus, the true value of the estimand for this pseudo-outcome is zero, and the goal of the supplementary analysis is to assess whether the adjustment methods employed in the primary analysis, when applied to the pseudo-outcome, lead to estimates that are close to zero. These are not standard specification tests that suggest alternative specifications when the null hypothesis is rejected. The implication of rejection here is that it is possible the original analysis was not credible at all.

# Contents

- 1 Rubin causal model
  - Rubin causal model
  - regression and inverse probability weighting
  - matching method
  - Applications using Stata
- 2 Regression Discontinuity
  - sharp regression discontinuity
  - fuzzy regression discontinuity
  - kink regression discontinuity
  - supplementary analysis
  - Applications using Stata
- 3 Synthetic control method
  - difference in difference
  - synthetic control approach for case study
  - Applications using Stata

- example: Effect of minimum wage on employment (Card and Krueger, 1994)

```
. use cardkrueger1994, clear
. diff fte, t(treated) p(t)
. diff fte, t(treated) p(t) cov(bk kfc roys)
. diff fte, t(treated) p(t) cov(bk kfc roys) qdid(0.5)
. diff fte, t(treated) p(t) cov(bk kfc roys) kernel
```

- example: Effect of California's tobacco control program (Abadie et al., 2010)

```
. use smoking, clear
. egen cigave = mean(cigsale) if state!=3, by(year)
. twoway (line cigsale year if state==3) ///
    line cigave year if state==1, lp(dash))
. synth cigsale beer(1984/1988) lncincome retprice age15to24 ///
cigsale(1988) cigsale(1980) cigsale(1975), ///
trunit(3) trperiod(1988) xperiod(1980(1)1988) ///
resultperiod(1970(1)2000) fig
```



- Which variables will be averaged?
  - (1) By default, all predictor variables are averaged over the entire pre-intervention period (missing values are ignored).
  - (2) particular predictor the user can specify the period over which the variable will be averaged. Examples:

```
. synth Y X1(1980) X2(1982&1986&1988)  
X3(1980(1)1990) X4
```
  - (3) lagged dependent variable can also be used as predictor.