项目评估计量经济学方法与Stata应用

张川川　浙江大学

2021年Stata洞察数据科学大会

内容概要

- 社会科学经验研究的核心：因果推断
  - OLS
  - Rubin Model

- 因果推断方法及STATA实现
  - RCT
  - IV
  - RD
  - DID

因果推断

- 因果推断的基本框架
  - OLS
  - Rubin Model

OLS

- OLS估计：最小化残差平方和

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

OLS

- OLS估计的假设
  - The Gauss-Markov Assumptions
  - Zero Conditional Mean

$$E(u|x) = 0$$

  - Homoskedasticity

$$Var(u|x) = \sigma^2$$

OLS

- OLS估计量的性质：无偏性

$$y_i - \bar{y} = (\beta_0 + \beta_1 x_i + u_i) - (\beta_0 + \beta_1 \bar{x}) = \beta_1 (x_i - \bar{x}) + u_i,$$

$$\begin{aligned}
\Rightarrow E(\widehat{\beta}_1) &= E\left(\frac{\sum_{i=1}^n \beta_1 (x_i - \bar{x})^2 + (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
&= \beta_1 + E_x\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) E(u_i|X)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
&= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} E(u_i|X) \\
&= \beta_1
\end{aligned}$$

OLS

- OLS估计的有效性（efficiency）

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\begin{aligned}
\Rightarrow \sigma^2_{\widehat{\beta}_1} &= Var(\widehat{\beta}_1) = E[(\widehat{\beta}_1 - E(\widehat{\beta}_1))^2] = E[(\widehat{\beta}_1 - \beta_1)^2] \\
&= E[(\frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2})^2] \\
&= \ldots \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}
\end{aligned}$$

Rubin Model

- 因果效应的定义：潜在结果的比较
  - *comparison of so-called potential outcomes:* ***pairs of outcomes defined for the same unit*** *given different levels of exposure to the treatment.*

Rubin Model

- 模型设定
  - A pair of **potential outcomes:** Yi(0) and Yi(1).

$$Y_i = Y_i(W_i) = Y_i(0) \cdot (1 - W_i) + Y_i(1) \cdot W_i = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases}$$

- 因果效应(unit level): Yi(1)-Yi(0)

Rubin Model v.s. OLS

- As a comparison, researchers **conventionally** write down a regression function: $Yi=a+bWi+ui$, with b as the causal effect (**OLS**).
  - 有哪些不同？体会微小的差异

- The **counterfactual approach** has only opened up a new perspective on traditional estimation.

因果推断

- 识别平均处理效应（ATE）：
  - ATE= E[Yi(1)-Yi(0)]
    =E[Yi(1)-Yi(0)|Wi=1]*Pr(Wi=1)+E[Yi(1)-Yi(0)|Wi=0]*Pr(Wi=0)
    =[E[Yi(1)|Wi=1]-E[Yi(0)|Wi=1]]*Pr(Wi=1)+[E[Yi(1)|Wi=0]-E[Yi(0)|Wi=0]]*[1-Pr(Wi=1)]
    = ATT*Pr(Wi=1)+ATUT*Pr(Wi=0)

- 我们可以基于数据识别的部分包括:
  - E[Yi(1)|Wi=1]; E[Yi(0)|Wi=0]; Pr(Wi=1)

- 无法观察到两个反事实状态:
  - E[Yi(0)|Wi=1]; E[Yi(1)|Wi=0]
  - Knowledge of these would be sufficient to identify the two parameters.

因果推断

- 重新理解chance 和choice

- Assignment mechanisms:
  - 1. Randomized experiments
  - 2. The assignment probabilities do not depend on the potential outcomes (given covariates):

$$W_i \perp \Big(Y_i(0), Y_i(1)\Big) \Big| X_i,$$

    - Unconfoundedness (Rubin, 1990), selection on observables, exogeneity, and conditional independence.
  - 3. Assignments with some dependence on potential outcomes.

因果推断方法

- RCT
- OLS/Matching
- IV/RD/DID

因果推断方法

- RCT
  - A Benchmark

- If we randomly force some people to receive the treatment and others not to receive the treatments then **Wi is random and uncorrelated with everything** so:
  - $E[Yi(1)|Wi=1]=E[Yi(1)|Wi=0]$
  - $E[Yi(0)|Wi=1]=E[Yi(0)|Wi=0]$

Nonrandom Assignment

- The nonrandomness is inherent to a policy for two distinct reasons:
  - The **self-selection** into the program operated by individuals.
  - The selection mechanism of the agency managing the program.

- **Choice or Chance**

Nonrandom Assignment

- **DIM（or OLS）估计的选择偏误**

- The **DIM**:
  - ✓ E(Yi|Wi=1)-E(Yi|Wi=0)={E[Yi(1)|Wi=1]-E[Yi(0)|Wi=1]}
    +{E[Yi(0)|Wi=1]-E[Yi(0)|Wi=0]}
    =ATT+ {E[Yi(0)|Wi=1]-E[Yi(0)|Wi=0}

- We thus have a **selection bias** equals to {E[Yi(0)|Wi=1]-E[Yi(0)|Wi=0}.

- Note that the selection bias is unobservable since we cannot observe: E[Yi(0)|Wi=1].

Selection on observables

- OLS/Matching
  - 关键假设 I：Conditional Independence Assumption (CIA)

$$W_i \perp\!\!\!\perp \Big(Y_i(0), Y_i(1)\Big) \Big| \; X_i.$$

  - The CIA assumption, or the unconfoundedness assumption means that beyond the observed covariates, Xi, there are no unobserved factors associated both with the potential outcomes and the treatment.

  - Assuming **homogeneous treatment effect**, this assumption is equivalent to independence of εi and of Wi, with the following regression model specification:

$$Y_i = \alpha + \tau \cdot W_i + \beta' X_i + \varepsilon_i,$$

Selection on observables

- **关键假设II: Overlap**

$$0 < \mathrm{pr}(W_i = 1 | X_i = x) < 1, \qquad \text{for all } x.$$

  - The combination of unconfoundedness and overlap was referred to by Rosenbaum and Rubin (1983) as **strong ignorability**.
  - Pr(W=1 | X) is the so-called propensity score.

- With the strong ignorability, for each x, the conditional ATE can be identified:

$$\begin{aligned}\tau(x) &= \mathbb{E}[Y_i(1)|X_i = x] - \mathbb{E}[Y_i(0)|X_i = x] \\ &= \mathbb{E}[Y_i(1)|W_i = 1, X_i = x] - \mathbb{E}[Y_i(0)|W_i = 0, X_i = x] \\ &= \mathbb{E}[Y_i|W_i = 1, X_i = x] - \mathbb{E}[Y_i|W_i = 0, X_i = x],\end{aligned}$$

- The ATE can be identified by taking expected values across the population distribution of the covariates (given the overlap assumption satisfied).

Matching

- Matching
  - 方法介绍（Rubin，1973，Biometrics；Rubin，1979，JASA；Rosenbaum & Rubin，1983，Biometrika; Heckman et al., 1998, RES；Imbens，2000，Biometrika；Smith & Todd，2001，AER; Abadie & Imbens, 2004, NBER wp）
  - 经典应用（Heckman et al., 1997, RES; Dehejia & Wahba; 1999, JASA）
  - **STATA：psmatch2**

Selection on Unobservables

- **DIM:**

  $=E[Y|X, W=1]-E[Y|X,W=0]$

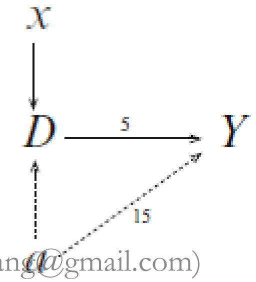
  $=E[Y(1)|X,W=1]-E[Y(0)|X,X=1]+ E[Y(0)|X,W=1]-E[Y(0)|X,W=0]$

  $=ATT+ E[Y(0)|X,W=1]-E[Y(0)|X,W=0]$

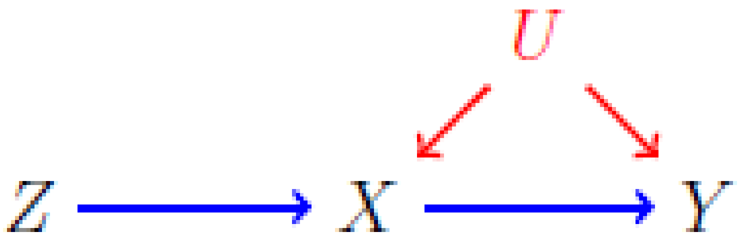
  The DIM produces a biased estimation of the causal effect of W on Y:

  - **$E[Y(0)|X,W=1]-E[Y(0)|X,W=0]$**

- Suppose that a treatment, D, is affected by two factors, one observable x and one unobservable a. Suppose a determines not only D but also the outcome Y in a direct way. A change in D produces a change in Y of 20, but the actual effect is only 5:



IV



传统IV

- 经典线性回归模型中的IV估计：

$$cov(z,y)=\beta_1 cov(z,x)+cov(z,u)$$

$$\beta_1=\frac{Cov(z,y)}{Cov(z,x)}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})}{\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})}$$

传统IV

- 两阶段最小二乘：

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$$

LATE

- 识别假设

Random Assignment:

$$Z_i \perp (Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), W_i(0), W_i(1))$$

Exclusion Restriction: $Y_i(z,w) = Y_i(z',w)$, for all z, z', w

First-stage: $E[W_i(1) - W_i(0)] \neq 0$, and $0 < P[z=1] < 1$

Monotonicity: $W_i(1) \geq W_i(0)$

LATE

- With the above assumptions, Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) show that the average causal effect for a subpopulation , so-called compliers, can be identified.

Table 1: COMPLIANCE TYPES

		$W_i(0)$	
		0	1
$W_i(1)$	0	never-taker	defier
	1	complier	always-taker

LATE

Table 2: COMPLIANCE TYPE BY TREATMENT AND INSTRUMENT

		Z_i = 0	Z_i = 1
W_i	0	complier/never-taker	never-taker/defier
	1	always-taker/defier	complier/always-taker

LATE

Table 3: Compliance Type by Treatment and Instrument given Monotonicity

		Z_i = 0	Z_i = 1
W_i	0	complier/never-taker	never-taker
	1	always-taker	complier/always-taker

LATE

- We can infer the average outcome by treatment status for compliers, and thus the average effect for compliers:

$$\mathbb{E}[Y(1) - Y_i(0)|\text{complier}] = \mathbb{E}[Y_i(1)|\text{complier}] - \mathbb{E}[Y_i(0)|\text{complier}]$$

$$\beta_{IV} = E[Y_i(1) - Y_i(0)|complier] = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[W_i|Z_i=1]-E[W_i|Z_i=0]}$$

IV文献

- 工具变量（IV）方法
  - 方法介绍（Angrist, Imbens & Rubin， 1996， JASA； Imbens & Angrist， 1994， Econometrica）
  - 经典应用（Angrist， 1990， AER; Angirst & Krueger， 1991， QJE;）
  - **STATA: ivreg2**

RD

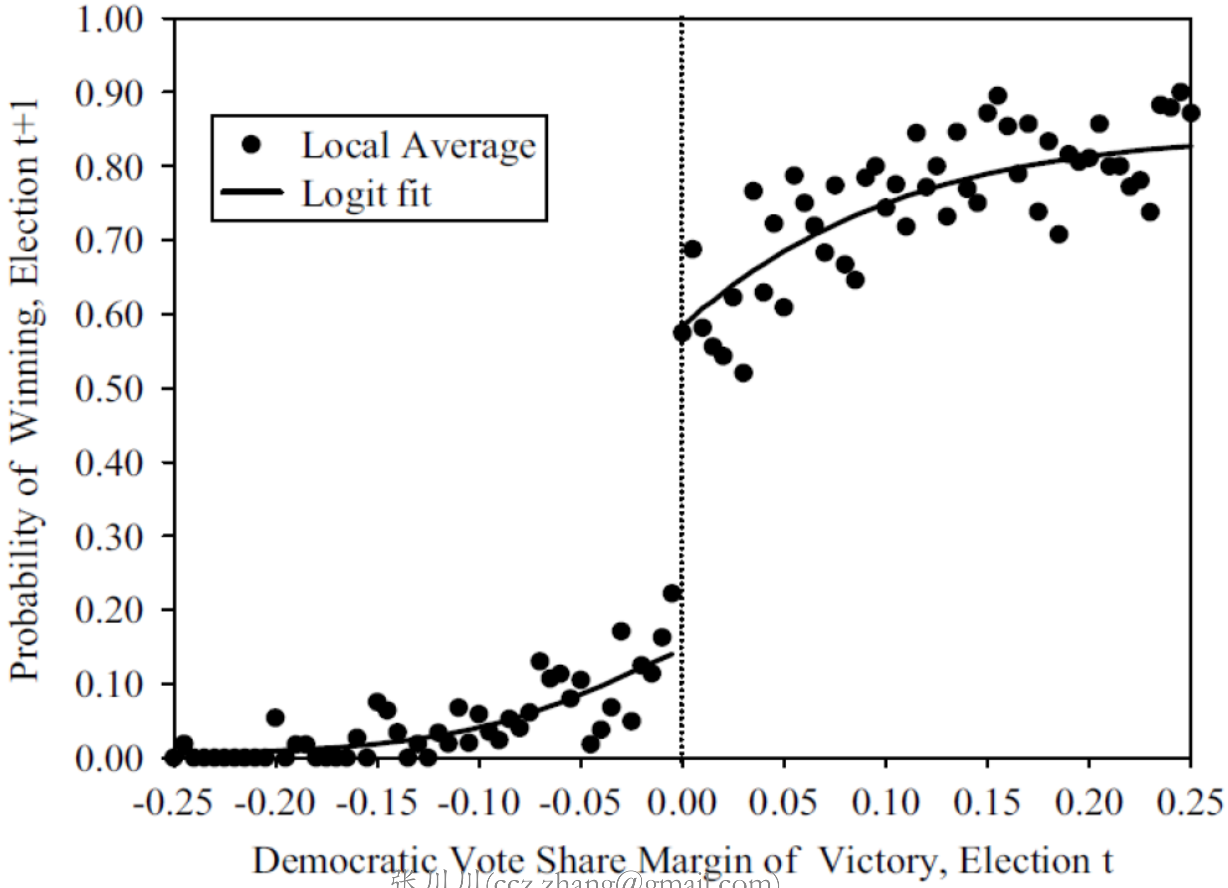
- 断点回归（regression discontinuity）
  - RD 利用由制度或政策或地理因素等引起的处理状态（treatment status）的非连续性跳跃
    - 这种非连续性跳跃被认为是相对外生的
  - RD designs require seemingly mild assumptions compared to those needed for other nonexperimental approaches (Hahn, Todd, and van der Klaauw, 2001)

RD方法的应用：最初的应用

- Donald L. Thistlethwaite and Donald T. Campbell (1960)： 奖学金如何影响学生的学术表现

RD方法的应用：政党在位优势

- Lee，2008，JOE



RD估计量

- Sharp RD

$$B - A = \lim_{\varepsilon \downarrow 0} E[Y_i | X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i | X_i = c + \varepsilon],$$

which would equal

$$E[Y_i(1) - Y_i(0) | X = c].$$

RD估计量

- Fuzzy RD

$$\tau_F = \frac{\lim_{\varepsilon \downarrow 0} E[Y \mid X = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y \mid X = c + \varepsilon]}{\lim_{\varepsilon \downarrow 0} E[D \mid X = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[D \mid X = c + \varepsilon]}$$

RD估计的有效性

- RD估计的有效性
  - 不能精准控制驱动变量(Imprecise control over the assignment variable, X).
    - Density test (Justin McCrary, 2008).
  - 所有其他因素连续**(All other factors** evolving "smoothly" with respect to X).
    - Fundamentally untestable.
    - Only "**observable factors**".

- Test the manipulation of assignment variable

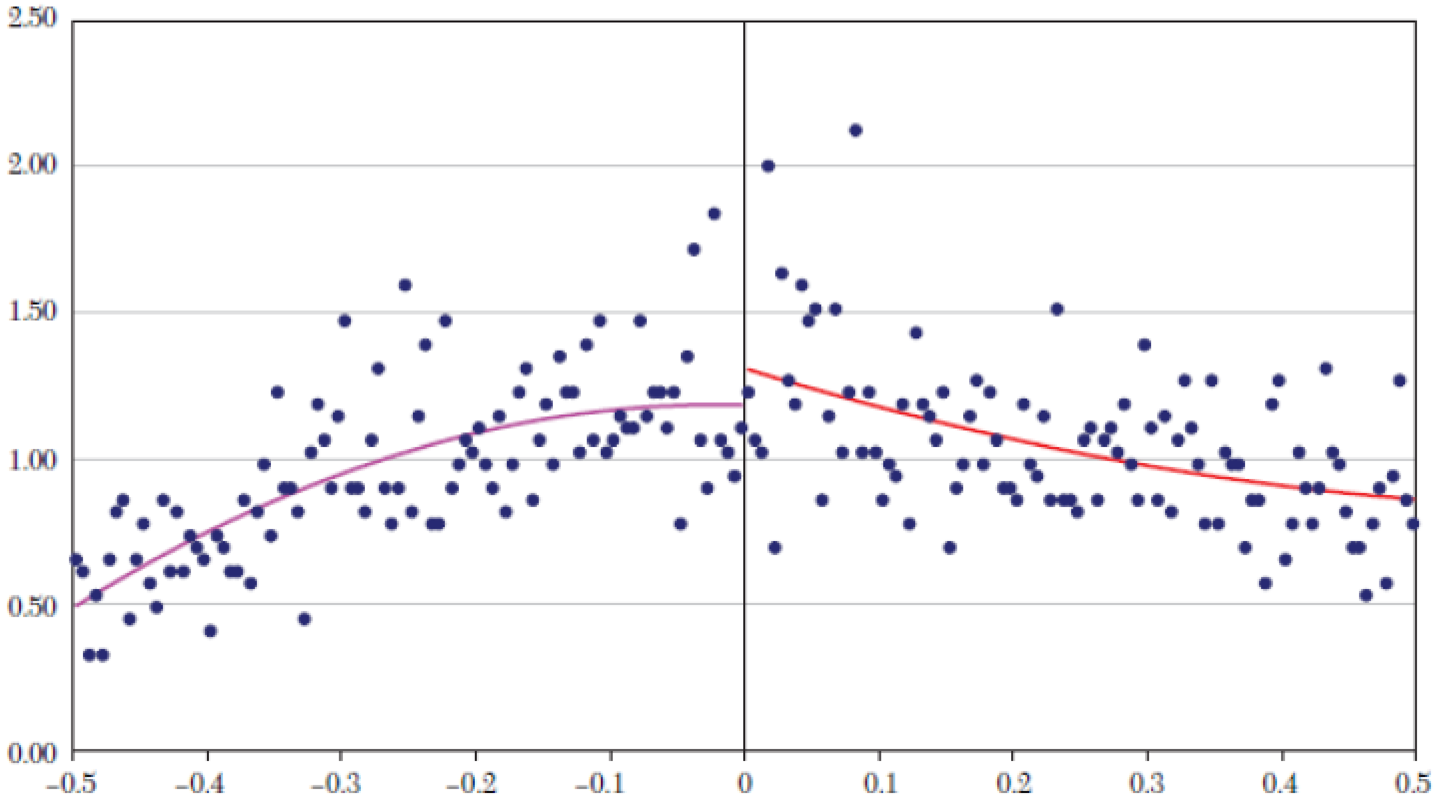


Figure 16. Density of the Forcing Variable (Vote Share in Previous Election)

RD文献

- RD
  - 方法介绍（Hahn et al., 2001, Econometrica; JOE，2008，Vol. 142 专刊）
  - 经典应用（Angrist & Lavy, 1999, QJE; Klaauw, 2002, IER）
  - 一篇重要的综述：Lee & Lemieux，2010，JLE
  - **STATA: rdrobust**

DID

- 双重差分（Difference-in-differences）
  - 广泛用于政策评估
  - 经典应用（Card & Krueger, 1994, AER; Meyer et al., 1995, AER; Duflo， 2001， AER）
  - STATA

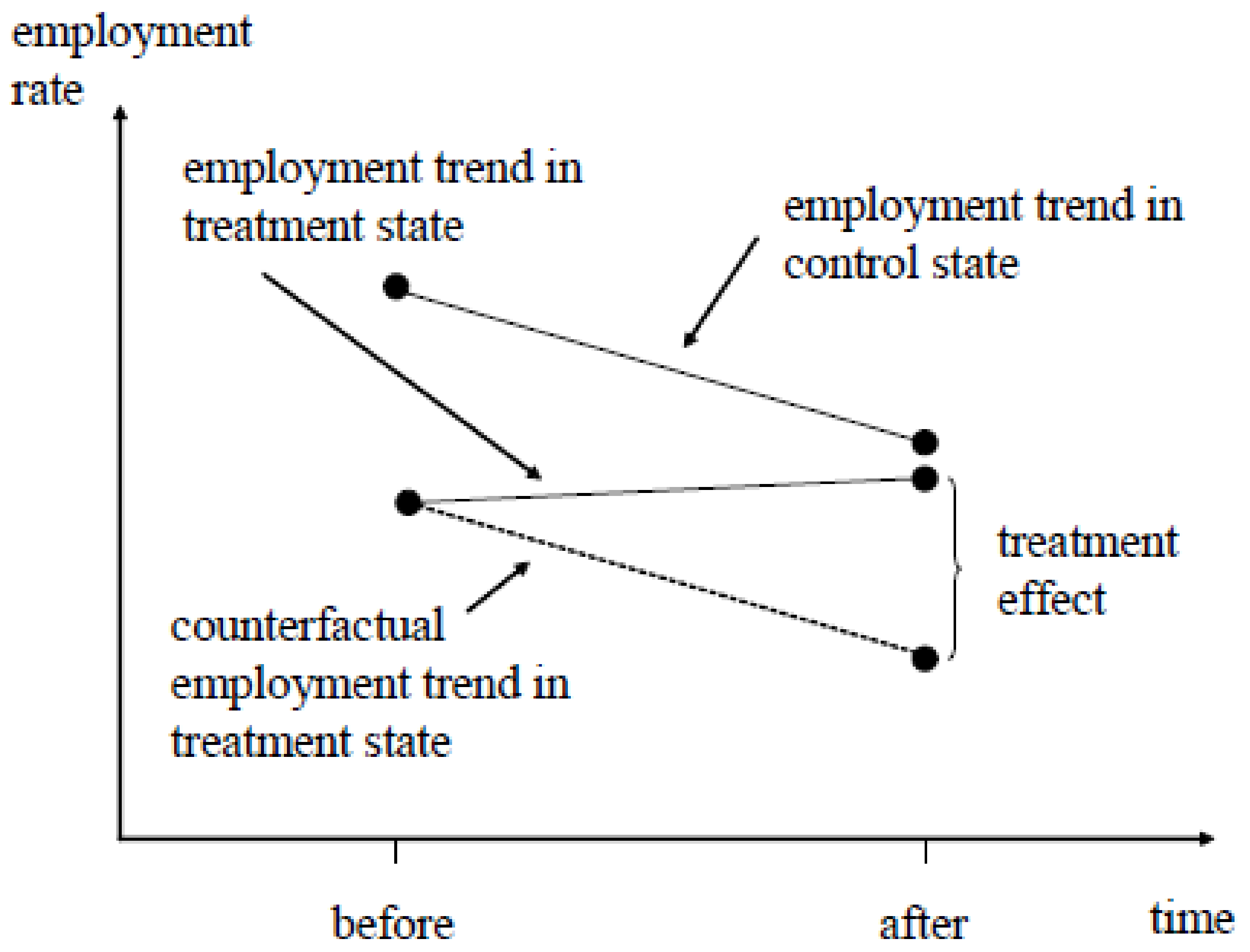


Figure 5.2.1: Causal effects in the differences-in-differences model

DID方法的应用：最低工资法案

- Card & Krueger, 1994, AER

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \beta(NJ_s \cdot d_t) + \varepsilon_{ist}$$

$$Y_i = \alpha + \beta \cdot T_i + \gamma \cdot G_i + \tau \cdot I_i + \varepsilon_i$$

DID方法的应用：建学校与教育

- Duflo，2001，AER

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

- Duflo（2001）中的证伪检验：

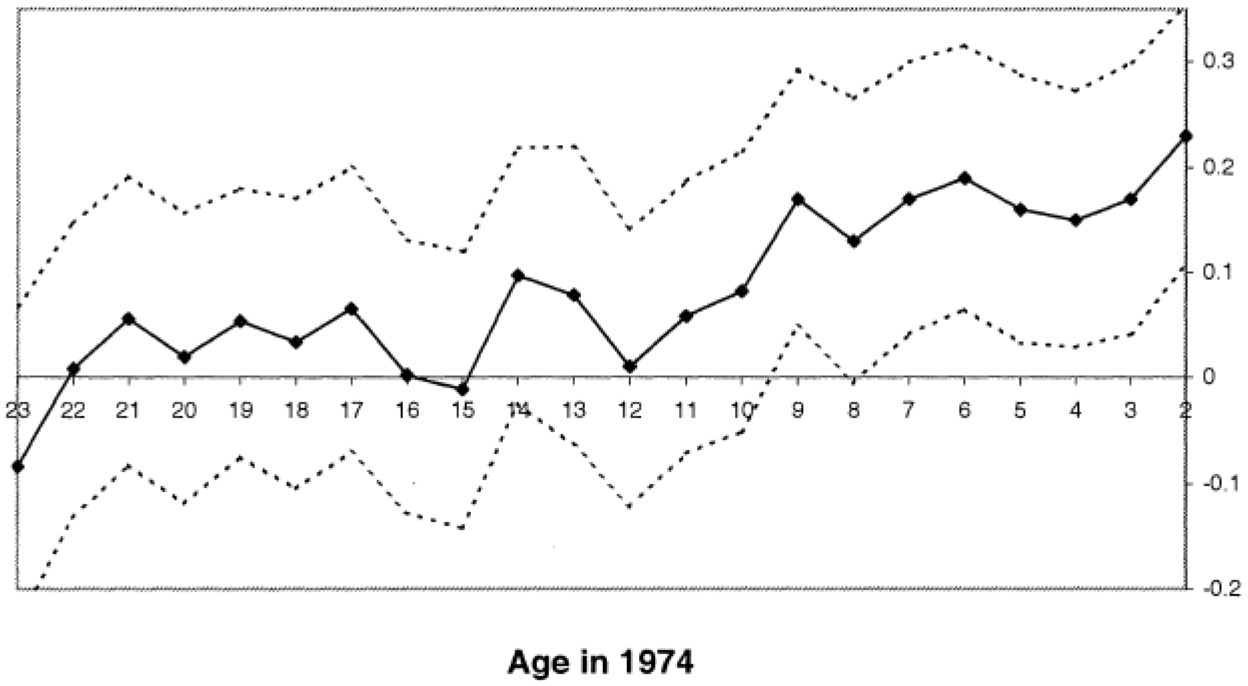


FIGURE 1. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE EDUCATION EQUATION

Thank You!