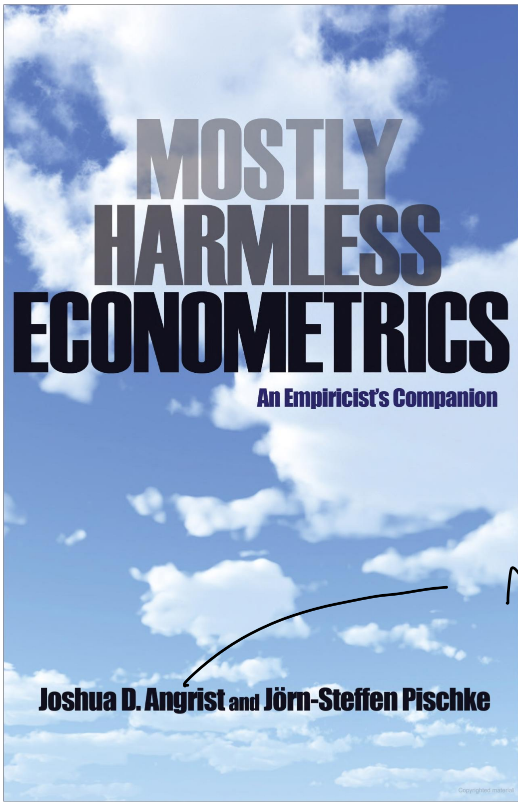


一些残余的
计量经济学的
话题

Nobel Prize

Contents

RCT : OLS

OLS 做因果推断
观测性研究

IV : econometrics

TSLS

LATE

? 面板数据

双重差分

Chapter 20

chapter 24

今天先讲这两章

Chapter 20

$0 < e(x) < 1$ ~ 围性 ： 阅读

一个极端 ： $e(x) = 0$ or 1

$$Z = \mathbb{1}(X \geq x_0)$$

Z：处理 二值

X：协变量 running/forcing variable

x_0：固定常数

$$e(X) = Pr(Z=1 \mid X) = \begin{cases} 1 & 若 X \geq x_0 \\ 0 & 若 X < x_0 \end{cases}$$

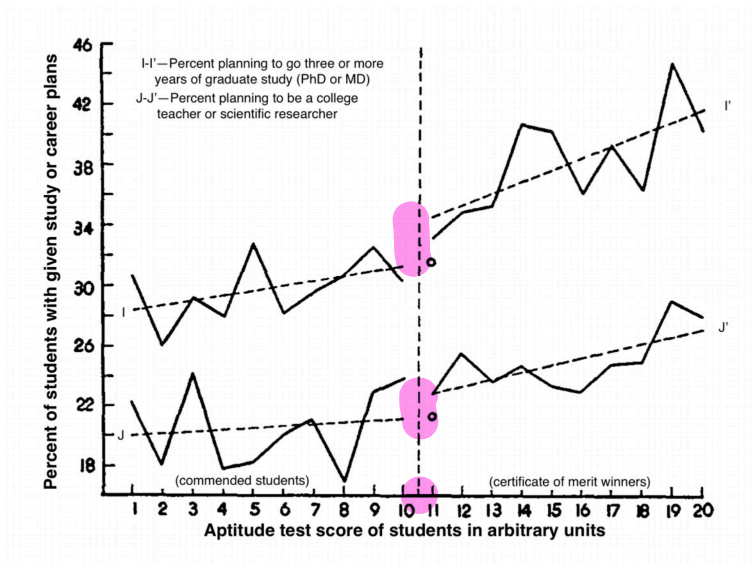


FIGURE 20.1: A graph from Thistlethwaite and Campbell (1960) with minor modifications of the unclear text in the original paper

RD: regression discontinuity

RD 第一篇文章

Carpenter & Dobkin (2009)

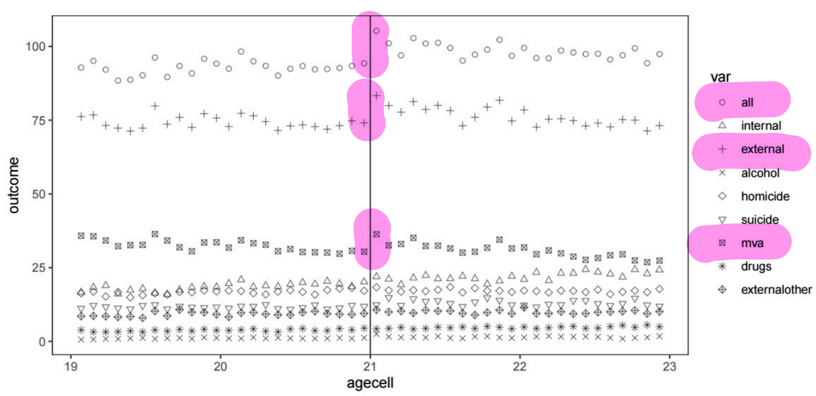


FIGURE 20.2: Minimum legal drinking age example

如何数学化 RD?

$$Z \in \{0, 1\}$$

$$Y(0), \quad Y(1)$$

$$\tau(x_0) = E\{Y(1) - Y(0) \mid X = x_0\}$$

Local Average Treatment Effect

LATE

可识别的，因为

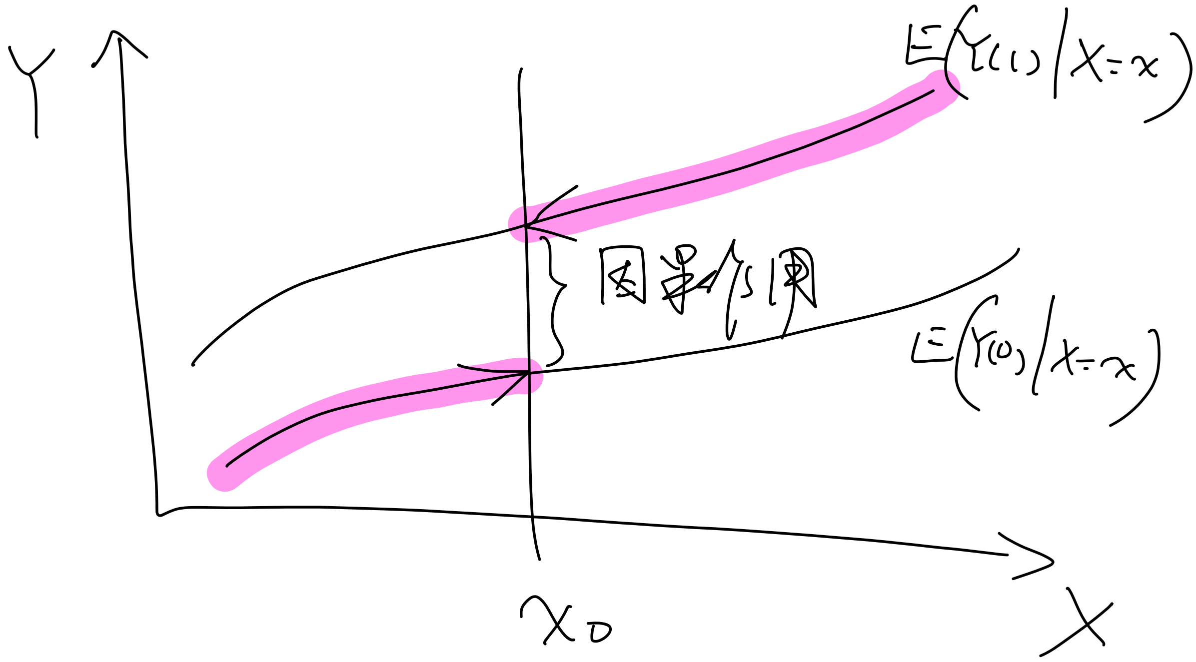
$$E\{Y(1) \mid X = x_0\} = \lim_{\varepsilon \to 0+} E\{Y(1) \mid X = x_0 + \varepsilon\}$$

假定连续性

$$= \lim_{\varepsilon \to 0+} E\{Y(\cancel{1}) \mid Z = 1, X = x_0 + \varepsilon\}$$

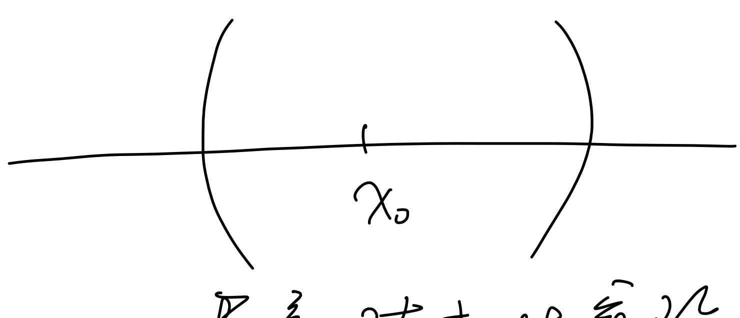
$$= \lim_{\varepsilon \to 0+} E(Y \mid Z = 1, X = x_0 + \varepsilon)$$

$$E\{Y(0) \mid X = x_0\} = \lim_{\varepsilon \to 0+} E(Y \mid Z = 0, X = x_0 - \varepsilon)$$



这种观点：基于 $E(Y(z)|X=x)$ 的连续性

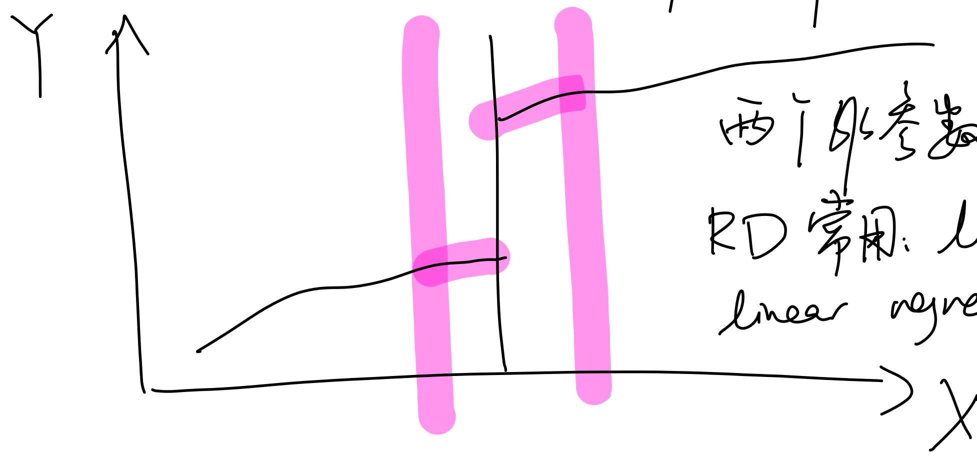
另一种观点：可能更直观



局部随机化实验

统计推断：两种观点类似

——如何选择局部点



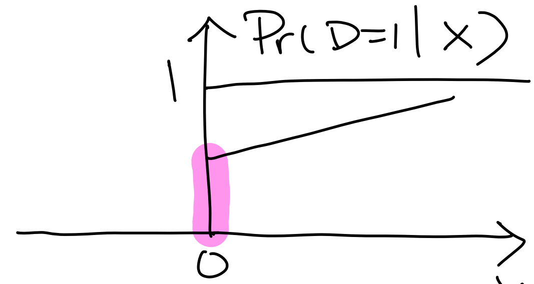
两个非参数回归

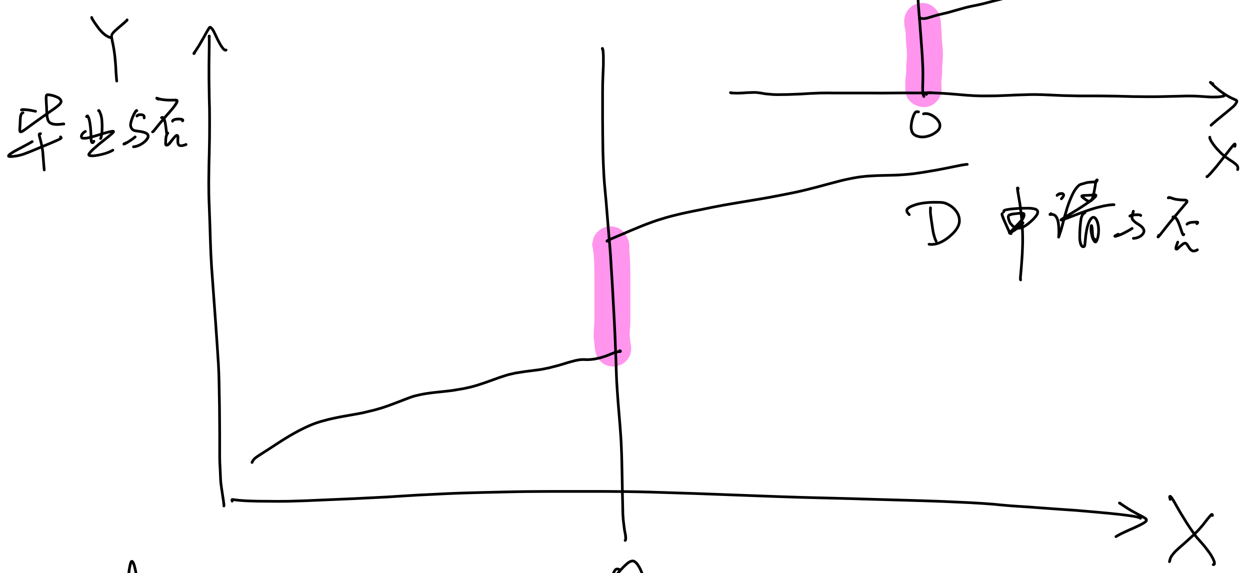
RD常用：local linear regression

chapter 24 RD + IV

= fuzzy RD

Li et al (2015)





$Z = \mathbb{1}(X \geq 0)$

|| 15000 euro — 家庭年收入

Z: 处理分配 $Z = \mathbb{1}(X \geq x_0)$

D: 接受的处理

Y: 结果

$$\tau_D(x_0) = E\left(D(1) - D(0) \mid X = x_0\right)$$

$$= \lim_{\varepsilon \to 0+} E\left(D \mid Z=1, X = x_0 + \varepsilon\right) - \lim_{\varepsilon \to 0+} E\left(D \mid Z=0, X = x_0 - \varepsilon\right)$$

$$\tau_Y(x_0) = E\left(Y(Z=1) - Y(Z=0) \mid X = x_0\right)$$

$$= \lim_{\varepsilon \to 0+} E\left(Y \mid Z=1, X = x_0 + \varepsilon\right) - \lim_{\varepsilon \to 0+} E\left(Y \mid Z=0, X = x_0 - \varepsilon\right)$$

定理 24.1 HW

假设 $\begin{cases} D(1) \geq D(0) \\ D(1) = D(0) \Rightarrow Y(1) = Y(0) \end{cases}$

\+ 连续性

那么 $\tau_C(x_0) \overset{\text{定义}}{=} E\left\{Y(1) - Y(0) \mid D(1) > D(0), X = x_0\right\}$

等于 $\tau_C(x_0) = \dfrac{E\{Y(1) - Y(0) \mid X = x_0\}}{E\{D(1) - D(0) \mid X = x_0\}}$

可以被识别:

$$\tau_c(x_0) = \frac{\tau_Y(x_0)}{\tau_D(x_0)}$$

公式

见上面.

计量经济推荐 h^*

例子

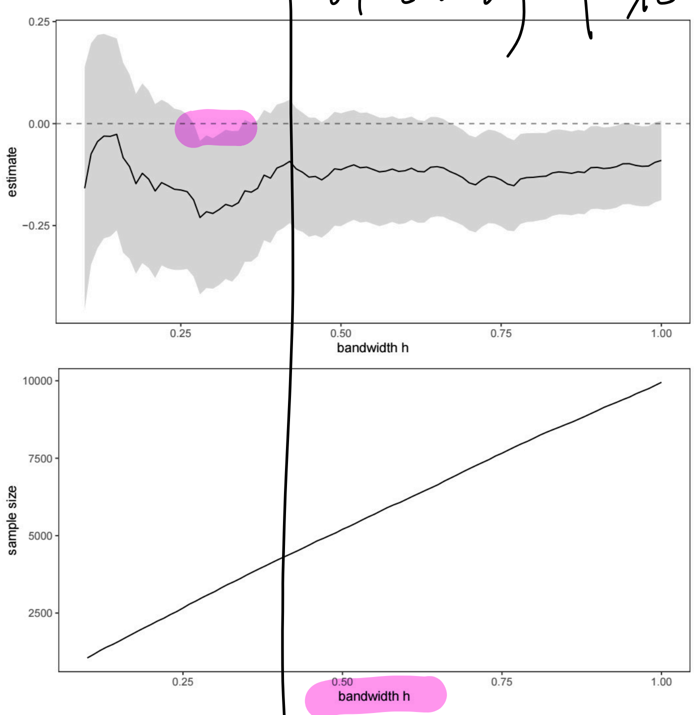


FIGURE 24.3: Re-analyzing Li et al. (2015)'s data, with point estimates and standard errors from TSLS.

Difference-in-differences

(DID)

双重差分

1994

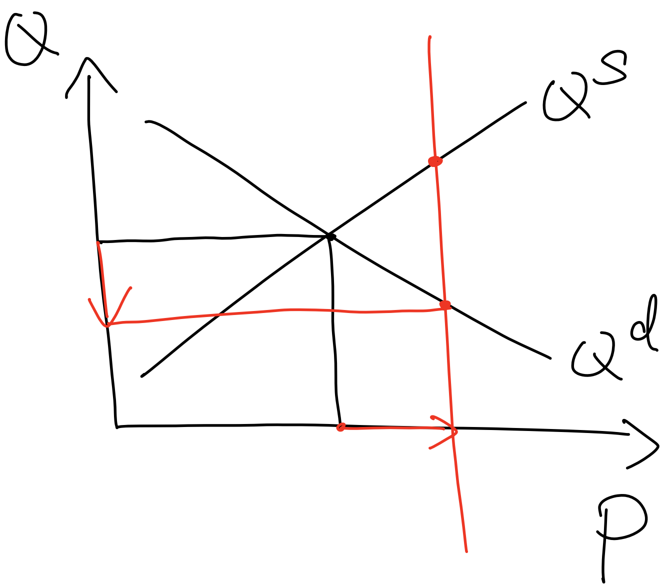
AER

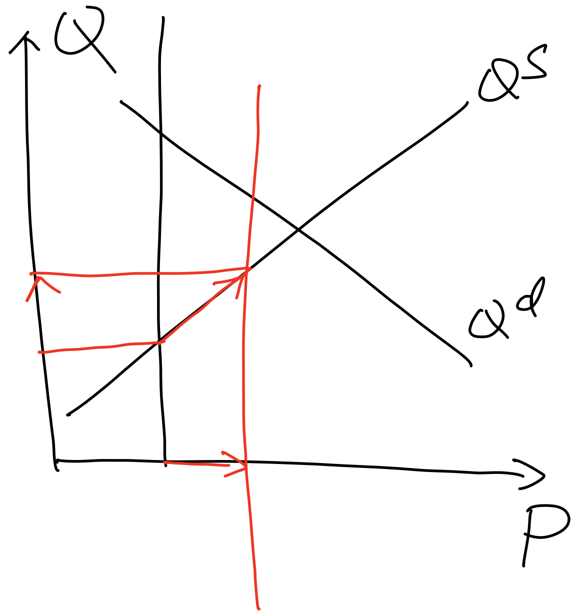
Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

By David Card and Alan B. Krueger*

On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above $5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (*JEL* J30, J23)

$$\hat{C} = \left(\bar{Y}_{NJ,\,after} - \bar{Y}_{NJ,\,before}\right) - \left(\bar{Y}_{PA,\,after} - \bar{Y}_{PA,\,before}\right)$$





DID 因果推断基础

$G_i = \begin{cases} 0 & 两组 \\ 1 & \end{cases}$ PA NJ

$T_i = \begin{cases} t & 前 \\ t+1 & 后 \end{cases}$ 只有两期

$D_{it} = 0$ 无处理

$D_{i,t+1} = G_i$，完全记号

$$Y_{it}, \quad Y_{it+1}$$
$$\parallel \qquad \parallel$$
$$Y_{it}(0) \qquad G_i Y_{i,t+1}(1) + (1-G_i) Y_{i,t+1}(0)$$

潜在结果：D = 1 or 0

记号系统

A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment

Peng Ding[1] **and Fan Li**[2]

[1] *Department of Statistics, University of California, 425 Evans Hall, Berkeley, CA 94720, USA. Email: pengdingpku@berkeley.edu*
[2] *Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708, USA. Email: fl35@duke.edu*

读书笔记

Abstract

Difference-in-differences is a widely used evaluation strategy that draws causal inference from observational panel data. Its causal identification relies on the assumption of parallel trends, which is scale-dependent and may be questionable in some applications. A common alternative is a regression model that adjusts for the lagged dependent variable, which rests on the assumption of ignorability conditional on past outcomes. In the context of linear models, Angrist and Pischke (2009) show that the difference-in-differences and lagged-dependent-variable regression estimates have a bracketing relationship. Namely, for a true positive effect, if ignorability is correct, then mistakenly assuming parallel trends will overestimate the effect; in contrast, if the parallel trends assumption is correct, then mistakenly assuming ignorability will underestimate the effect. We show that the same bracketing relationship holds in general nonparametric (model-free) settings. We also extend the result to semiparametric estimation based on inverse probability weighting. We provide three examples to illustrate the theoretical results with replication files in Ding and Li (2019).

参数: ATT

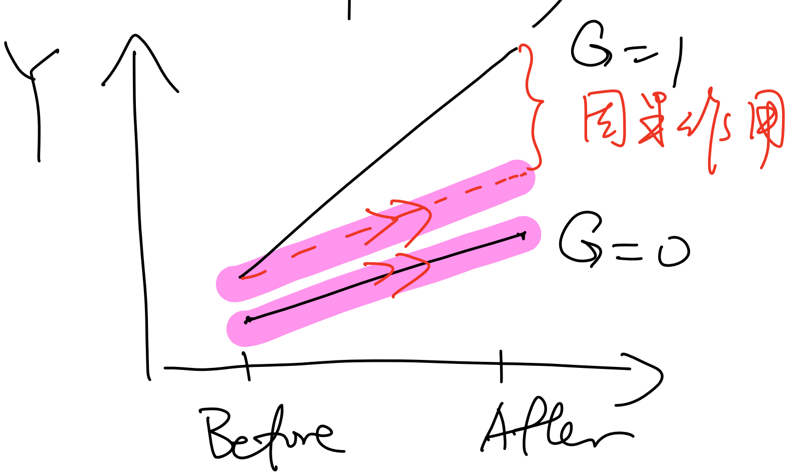
$$\tau_{ATT} = E\{Y_{i,t+1}(1) - Y_{i,t+1}(0) \mid G_i = 1\}$$

$$= \underbrace{E\{Y_{i,t+1}(1) \mid G_i = 1\}}_{\text{容易}} - \underbrace{E\{Y_{i,t+1}(0) \mid G_i = 1\}}_{\text{困难}}$$

可观测的　　　　反事实

假设1 (Parallel trends)

平行趋势



$$E\{Y_{i,t+1}(0)-Y_{i,t}(0) \mid G_i=1\} = E\{Y_{i,t+1}(0)-Y_{i,t}(0) \mid G_i=0\}$$

也可加协变量

X_i X_i

反事实

可观测

$$\mathbb{E}\left\{ Y_{i,t+1}(0) \mid G_i = 1 \right\}$$ 反事实？

$$= \mathbb{E}\left[\mathbb{E}\left\{ Y_{i,t}(\cancel{0}) \mid X_i, G_i = 1 \right\} + \mathbb{E}\left\{ Y_{i,t+1}(\cancel{0}) - Y_{i,t}(\cancel{0}) \mid X_i, G_i = 0 \right\} \,\middle|\, G_i = 1 \right]$$

$$= \mathbb{E}\left(Y_{it} \mid G_i = 1 \right) + \int \mathbb{E}\left(Y_{i,t+1} - Y_{it} \mid X_i = x, G_i = 0 \right) \cdot f(x \mid G_i = 1)\, dx$$

扔掉 x:

$$\mathbb{E}\left\{ Y_{i,t+1}(0) \mid G_i = 1 \right\}$$ 反事实

$$= \mathbb{E}\left(Y_{it} \mid G_i = 1 \right) + \mathbb{E}\left(Y_{i,t+1} - Y_{it} \mid G_i = 0 \right)$$

$$\Rightarrow \tau_{ATT} = \left[\mathbb{E}\left(Y_{i,t+1} \mid G_i = 1 \right) - \mathbb{E}\left(Y_{it} \mid G_i = 1 \right) \right] - \left[\mathbb{E}\left(Y_{i,t+1} \mid G_i = 0 \right) - \mathbb{E}\left(Y_{it} \mid G_i = 0 \right) \right]$$

"DID" 名字由来

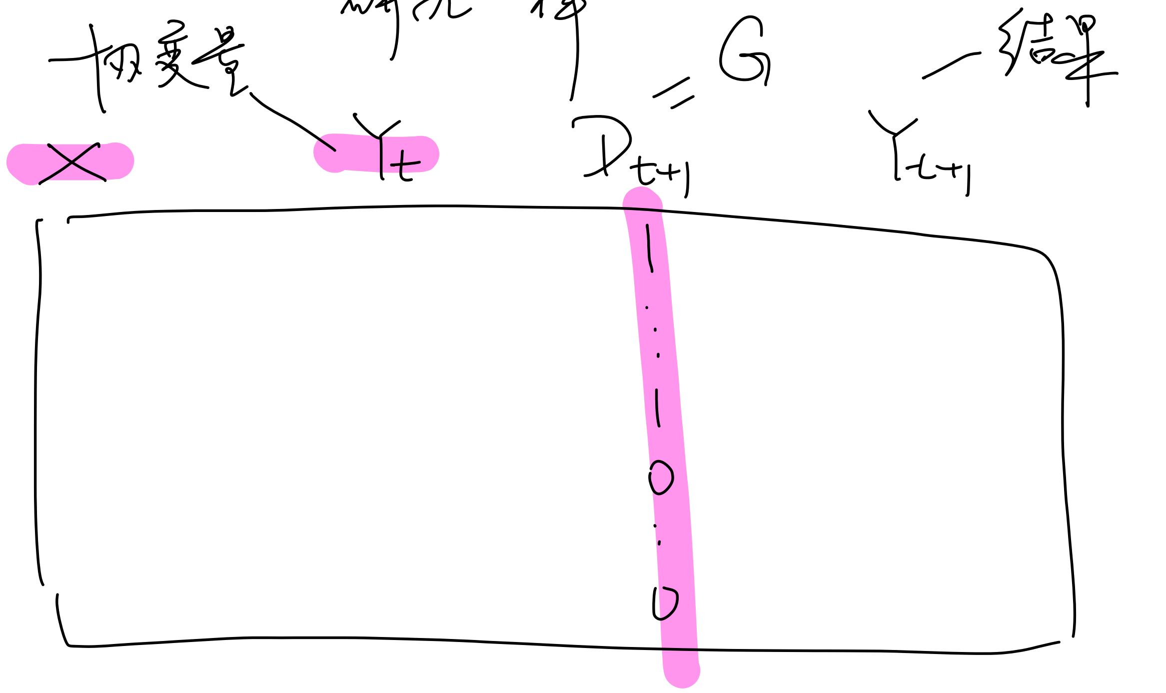
批评：为何要DID？

└ parallel trends？太强

scale dependent

└ 数据结构和标准观察性

研究一样



LDV : lagged-dependent-variable adjustment

(X, Y_t) 为协变量

$\Rightarrow$ 标准问题

假设 $Y_{i,t+1}(0) \perp\!\!\!\perp D_{i,t+1} \mid (X_i, Y_{it})$

Not scale dependent

不可验证或证伪的假定

$$E\{Y_{i,t+1}(0) \mid G_i=1\}$$

$$= \iint E(Y_{i,t+1} \mid G_i=0, Y_{it}=y, X_i=x)\, f(Y_{it}=y, X_i=x \mid G_i=1)\, dy\, dx$$

ATT 标准公式

回顾 Angrist & Pischke (2009) 教科书中的结果：

忽略 X.

$$\hat{\tau}_{DID} = \left(\bar{Y}_{1,t+1} - \bar{Y}_{1,t}\right) - \left(\bar{Y}_{0,t+1} - \bar{Y}_{0,t}\right)$$

可以从如下的回归得到：

时间 T 的固定效应

$$E\left(Y_{iT} \mid D_{iT}\right) = \alpha_i + \lambda_T + \tau D_{iT}$$

个体 i 的固定效应

fixed effect

"作用"

Wooldridge 书：Econometric Analysis of Cross-Section and Panel Data

LDV: 用 OLS

$$E(Y_{t+1} \mid G=0, Y_t = y) = \hat{\alpha} + \hat{\beta} y$$

用 $G=0$ 数据

$\Rightarrow$ 预测 $G=1$ 的潜在结果均值

$$\hat{\alpha} + \hat{\beta}\, \overline{Y}_{1,t}$$

$$\Rightarrow \hat{\tau}_{LDV} = \overline{Y}_{1,t+1} - (\hat{\alpha} + \hat{\beta}\, \overline{Y}_{1,t})$$

$$\hat{\alpha} \overset{OLS}{=} \overline{Y}_{0,t+1} - \hat{\beta}\, \overline{Y}_{0,t}$$

$$= (\overline{Y}_{1,t+1} - \overline{Y}_{0,t+1}) - \hat{\beta}(\overline{Y}_{1,t} - \overline{Y}_{0,t})$$

$\hat{\beta}$ 一般 <1 才有平稳的时间序列

我们的结果：非参数意义下也对

$$\tau_{ATT} = E(Y_{i,t+1} \mid G_i=1) - \underbrace{E\{Y_{i,t+1}(0) \mid G_i=1\}}_{\text{核心}}$$

Parallel trends

$$\tilde{\mu}_{0,DID} = E(Y_{it} \mid G_i=1) + E(Y_{i,t+1} \mid G_i=0) - E(Y_{it} \mid G_i=0)$$

ignorability

$$\tilde{\mu}_{0,LDV} = \int E(Y_{i,t+1} \mid G_i=0, Y_{it}=y)\, f(Y_{it}=y \mid G_i=1)\, dy$$

比较

$$\tilde{\mu}_{0,LDV} - \tilde{\mu}_{0,DID}$$

$$= \int E(Y_{i,t+1} \mid G_i=0, Y_{it}=y)\, f(Y_{it}=y \mid G_i=1)\, dy - \int y\, f(Y_{it}=y \mid G_i=1)\, dy$$

$$- \int E(Y_{i,t+1} \mid G_i=0, Y_{it}=y)\, f(Y_{it}=y \mid G_i=0)\, dy + \int y\, f(Y_{it}=y \mid G_i=0)\, dy$$

定义 $\Delta(y) = E(Y_{i,t+1} \mid G_i=0, Y_{it}=y) - y$

$$\tilde{\mu}_{0,LDV} - \tilde{\mu}_{0,DID} = \int \Delta(y) f(Y_{it}=y \mid G_i=1)dy - \int \Delta(y) f(Y_{it}=y \mid G_i=0)dy$$

假设 $\begin{cases} \dfrac{\partial \Delta(y)}{\partial y} < 0 \quad \forall y \\ F_{Y_t}(y \mid G=1) \geq F_{Y_t}(y \mid G=0) \\ \leq \end{cases}$

$\Rightarrow \tilde{\mu}_{0,LDV} - \tilde{\mu}_{0,DID}$ 能定

引理: $\Pr(A \leq x) \geq \Pr(B \leq x)$

$\Longleftrightarrow E(u(A)) \geq E(u(B))$

对所有 $u(\cdot)$ ↓ 成立

批评这个结果：

DID, LDV 两个都有偏

$\Rightarrow$ 无法得到其参数

正确的解读：

DID, LDV 的大小被估计的

要提前确定

Panel data + causal interence

$$\left(Y_{it},\ X_{it} \ : \ \begin{matrix} i = 1 \cdots n \\ t = 1 \cdots T \end{matrix} \right)$$

面板数据

回归: $Y_{it} = X_{it}'\beta + \varepsilon_{it}$

计量特色: $Y_{it} = \alpha_i + X_{it}'\beta + \varepsilon_{it}$

固定效应

$$Y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + \varepsilon_{it}$$

DID: 特别 $T = 2$

$$Y_{it} = \alpha_i + \lambda_t + \tau D_{it} + \varepsilon_{it}$$

参数

$\Downarrow$

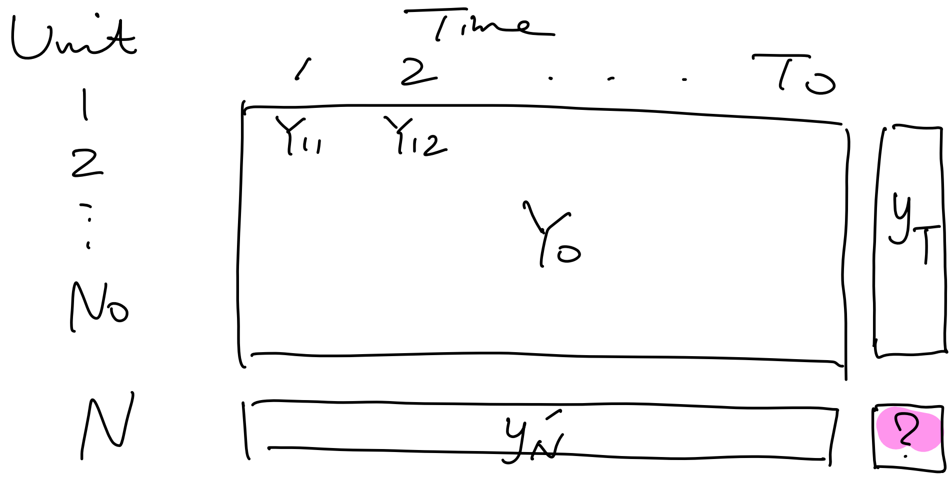
OLS $\Rightarrow$ $\hat{\tau}_{DID}$

incidental parameter problem

最后一个话题: SC

synthetic control

合成对照

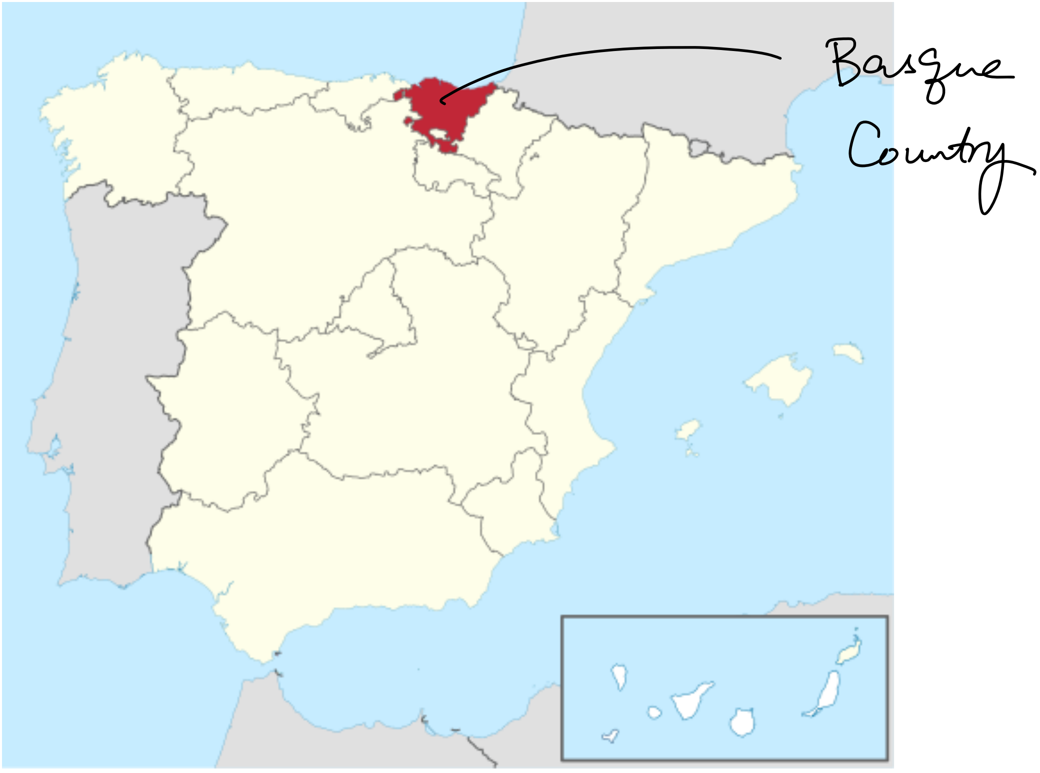


AER(2003)

The Economic Costs of Conflict: A Case Study of the Basque Country

By ALBERTO ABADIE AND JAVIER GARDEAZABAL*

This article investigates the economic effects of conflict, using the terrorist conflict in the Basque Country as a case study. We find that, after the outbreak of terrorism in the late 1960's, per capita GDP in the Basque Country declined about 10 percentage points relative to a synthetic control region without terrorism. In addition, we use the 1998–1999 truce as a natural experiment. We find that stocks of firms with a significant part of their business in the Basque Country showed a positive relative performance when truce became credible, and a negative relative performance at the end of the cease-fire. (*JEL* D74, G14, P16)



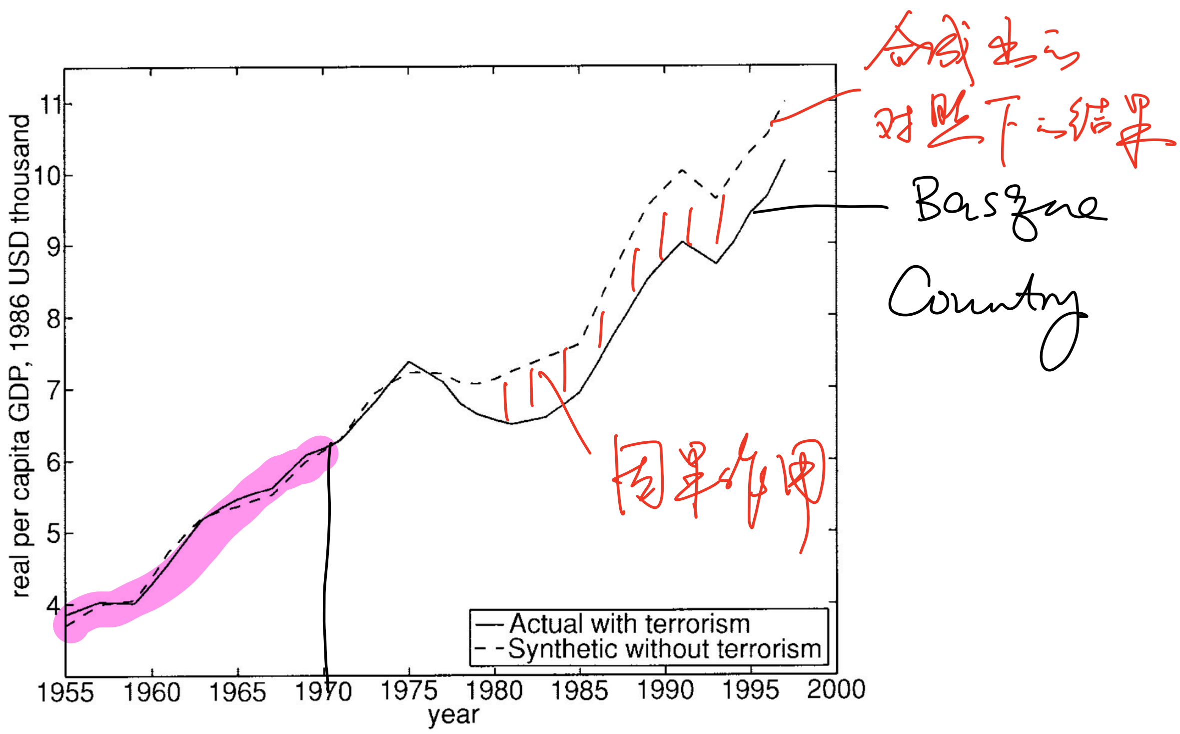


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 3–32

The State of Applied Econometrics: Causality and Policy Evaluation

Susan Athey and Guido W. Imbens

Here we discuss two recent developments to the difference-in-differences approach: the synthetic control approach and the nonlinear changes-in-changes method. The synthetic control approach developed by Abadie, Diamond, and Hainmueller (2010, 2014) and Abadie and Gardeazabal (2003) is arguably the most important innovation in the policy evaluation literature in the last 15 years. This method builds on difference-in-differences estimation, but uses systematically more attractive comparisons. To gain some intuition about these methods, consider the classic difference-in-differences study by Card (1990; see also Peri and Yasenov 2015). Card is interested in the effect of the Mariel boatlift, which brought low-skilled Cuban workers to Miami. The question is how the boatlift affected the Miami labor market, and specifically the wages of low-skilled workers. 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.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1716–1730: Theory and Methods Special Section on Synthetic Control Methods
https://doi.org/10.1080/01621459.2021.1891924

Taylor & Francis
Taylor & Francis Group

Check for updates

Matrix Completion Methods for Causal Panel Data Models

Susan Athey[a], Mohsen Bayati[b], Nikolay Doudchenko[b], Guido Imbens[c], and Khashayar Khosravi[d]

[a]Graduate School of Business, Stanford University, SIEPR, and NBER, Stanford, CA; [b]Graduate School of Business, Stanford University, Stanford, CA; [c]Graduate School of Business, and Department of Economics, Stanford University, SIEPR, and NBER, Stanford, CA; [d]Department of Electrical Engineering, Stanford University, Stanford, CA

ABSTRACT
In this article, we study methods for estimating causal effects in settings with panel data, where some units are exposed to a treatment during some periods and the goal is estimating counterfactual (untreated) outcomes for the treated unit/period combinations. We propose a class of matrix completion estimators that uses the observed elements of the matrix of control outcomes corresponding to untreated unit/periods to impute the "missing" elements of the control outcome matrix, corresponding to treated units/periods. This leads to a matrix that well-approximates the original (incomplete) matrix, but has lower complexity according to the nuclear norm for matrices. We generalize results from the matrix completion literature by allowing the patterns of missing data to have a time series dependency structure that is common in social science applications. We present novel insights concerning the connections between the matrix completion literature, the literature on interactive fixed effects models and the literatures on program evaluation under unconfoundedness and synthetic control methods. We show that all these estimators can be viewed as focusing on the same objective function. They differ solely in the way they deal with identification, in some cases solely through regularization (our proposed nuclear norm matrix completion estimator) and in other cases primarily through imposing hard restrictions (the unconfoundedness and synthetic control approaches). The proposed method outperforms unconfoundedness-based or synthetic control estimators in simulations based on real data.

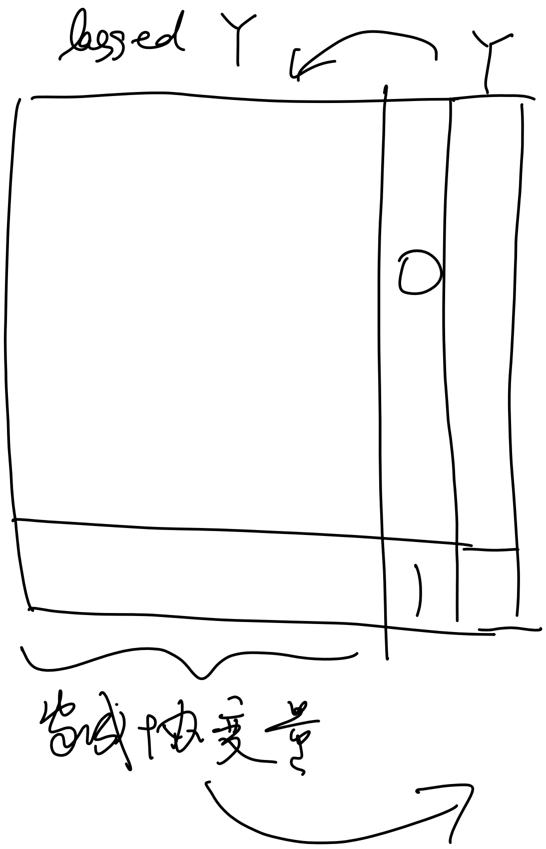
ARTICLE HISTORY
Received October 2018
Accepted February 2021

KEYWORDS
Causality; Interactive fixed effects; Low-rank matrix estimation; Synthetic controls; Unconfoundedness

3.3.1. Horizontal Regression and the Unconfoundedness Literature

The unconfoundedness literature (Rosenbaum and Rubin 1983; Rubin 2006; Imbens and Wooldridge 2009; Abadie and Cattaneo 2018) focuses primarily on the single-treated-period block structure with a thin matrix ($N \gg T$), a substantial number of treated and control units, and imputes the missing potential outcomes in the last period using control units with similar lagged outcomes:

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & ? \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & ? \end{pmatrix}.$$



A simple version of the unconfoundedness approach is to regress the last period outcome on the lagged outcomes and use the estimated regression to predict the missing potential outcomes. That is, for the units with $(i, T) \in \mathcal{M}$, the predicted outcome is

$$\hat{Y}_{iT} = \hat{\beta}_0 + \sum_{s=1}^{T-1} \hat{\beta}_s Y_{is}, \quad \text{where}$$

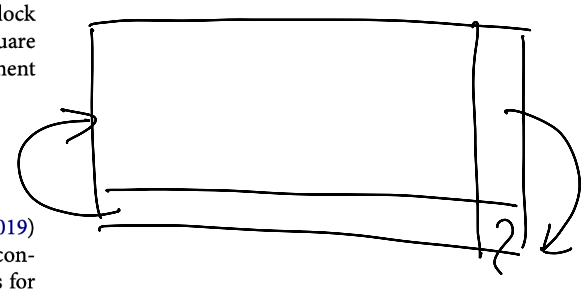
$$\hat{\beta} = \arg\min_{\beta} \sum_{i:(i,T)\in\mathcal{O}} \left(Y_{iT} - \beta_0 - \sum_{s=1}^{T-1} \beta_s Y_{is} \right)^2. \qquad (2)$$

We refer to this as a *horizontal* regression, where the rows of the $\mathbf{Y}$ matrix form the units of observation. A more flexible, nonparametric, version of this estimator would correspond to matching where we find for each treated unit i a corresponding control unit j with Y_{jt} approximately equal to Y_{it} for all pretreatment periods $t = 1, \ldots, T-1$.

3.3.2. Vertical Regression and the Synthetic Control Literature

The synthetic control literature (Abadie, Diamond, and Hainmueller 2010) focuses primarily on the single-treated-unit block structure with a relatively fat ($T \gg N$) or approximately square matrix ($T \approx N$), and a substantial number of pretreatment periods:

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \ldots & \checkmark & \checkmark & \ldots & \checkmark \\ \checkmark & \checkmark & \ldots & \checkmark & \checkmark & \ldots & \checkmark \\ \checkmark & \checkmark & \ldots & \checkmark & ? & \ldots & ? \end{pmatrix}.$$



Doudchenko and Imbens (2016) and Ferman and Pinto (2019) showed how the Abadie–Diamond–Hainmueller synthetic control method can be interpreted as regressing the outcomes for the treated unit prior to the treatment on the outcomes for the control units in the same periods. That is, for the treated unit in period t, for $t = T_0, \ldots, T$, the predicted outcome is

$$\hat{Y}_{Nt} = \hat{\gamma}_0 + \sum_{i=1}^{N-1} \hat{\gamma}_i Y_{it}, \quad \text{where}$$

$$\hat{\gamma} = \arg\min_{\gamma} \sum_{t:(N,t)\in\mathcal{O}} \left(Y_{Nt} - \gamma_0 - \sum_{i=1}^{N-1} \gamma_i Y_{it} \right)^2. \qquad (3)$$

Submitted to *Econometrica*

在很多重要场景.

SAME ROOT DIFFERENT LEAVES: TIME SERIES AND CROSS-SECTIONAL METHODS IN PANEL DATA

$VT = HZ$

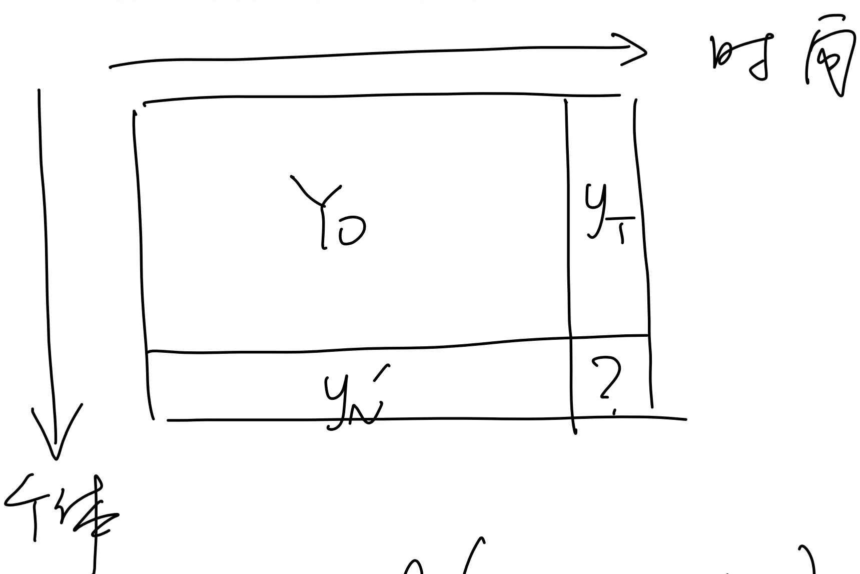
DENNIS SHEN
Simons Institute for the Theory of Computing, University of California, Berkeley

PENG DING
Department of Statistics, University of California, Berkeley

JASJEET SEKHON
Departments of Statistics & Data Science and Political Science, Yale University

BIN YU
Departments of Statistics and EECS, University of California, Berkeley

One dominant approach to evaluate the causal effect of a treatment is through panel data analysis, whereby the behaviors of multiple units are observed over time. The information across time and units motivates two general approaches: (i) horizontal regression (i.e., unconfoundedness), which exploits time series patterns, and (ii) vertical regression (e.g., synthetic controls), which exploits cross-sectional patterns. Conventional wisdom often considers the two approaches to be different. We establish this position to be partly false for estimation but generally true for inference. In the absence of any assumptions, we show that both approaches yield algebraically equivalent point estimates for several standard estimators. However, the source of randomness assumed by each approach leads to a distinct estimand and quantification of uncertainty even for the same point estimate. This emphasizes that researchers should carefully consider where the randomness stems from in their data as it has direct implications for the accuracy of inference.



$$HZ: \ lm(y_T \sim Y_0) \Rightarrow \hat{\alpha}$$

$$\Rightarrow \ ? = \langle y_N, \hat{\alpha} \rangle$$

$$VT: \ lm(y_N \sim Y_0') \Rightarrow \hat{\beta}$$

$$\Rightarrow \ ? = \langle y_T, \hat{\beta} \rangle$$

结论: OLS + minimum ℓ_2-norm estimation

$$VT = HZ$$

$$\hat{\alpha} = (Y_0' Y_0)^{-1} Y_0' y_T$$

$$= Y_0^{+} y_T$$

pseudo inverse

伪逆

$$\Rightarrow \ ? = \langle y_N, \hat{\alpha} \rangle$$

$$\hat{\beta} = (Y_0')^{+} y_N$$

$$\Rightarrow \ ? = \langle y_T, \hat{\beta} \rangle$$