定量社会科学的因果推断

Causal Inference in Quantitative Social Sciences

江　艇
中国人民大学经济学院

Last updated: May 8, 2021

Lecture 5　面板数据的参数和非参数方法

5.1　面板数据固定效应模型

- 我们已经看到，控制固定效应就是控制相应类别的不可观测的选择性。类别划分得越细致，控制就越细致，剩余的不可观测的潜在选择性就越少，所需要的识别假设就越弱，识别就越干净。但对于截面数据而言，类别无法划分到个体层面。
- 面板数据可以用来处理一类特殊的假定**LS.1**/**LS.2**被违背的情形——不随时间变化的不可观测的选择性。其基本思想在于，通过拓展数据的时间维度，得以控制最细致的截面固定效应——个体固定效应。

$$Y_{it}=\beta D_{it}+u_i+\varepsilon_{it},\ i=1\ldots,n,\ t=1\ldots,T$$

其中 D_{it} 为核心解释变量，u_i 为不可观测的个体固定效应。

- 处理固定效应的方式：组内去平均变换 + OLS

$$\bar{Y}_i = \beta \bar{D}_i + u_i + \bar{\varepsilon}_i$$

其中 $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\bar{D}_i = \frac{1}{T}\sum_{t=1}^{T} D_{it}$, $\bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- 固定效应模型的关键假设：

$$\mathbb{E}\left(D_{is}\varepsilon_{it}\right) = 0,\ \ s, t = 1, \ldots, T$$

这一假设允许 D_{it} 和 u_i 之间具有任意形式的相关性。

- 与混合横截面模型及随机效应模型的关键假设的比较：

$$\begin{cases} \mathbb{E}(D_{it}u_i) = 0 \\ \mathbb{E}(D_{it}\varepsilon_{it}) = 0 \end{cases}$$

- 如果 $\mathbb{E}(D_{it}u_i) = 0$ 成立，那么采用随机效应估计量可以改进估计效率 (efficiency)。但渐进效率的改进是有限的，特别是在有限样本下。更重要的是，不能把对解决内生性的指望寄托在 Hausman 检验的功效上。所以，**不要使用随机效应模型，也无需做 Hausman 检验。**

- 实际应用时，结构模型中还会包含控制变量，

$$Y_{it} = \beta_1 D_{it} + \beta_2 X_{it} + u_i + \varepsilon_{it}$$

此时的识别假设仍然是适当形式的“条件不相关”假设。

$$\text{Cov}(D_{is}, \varepsilon_{it} | X_{i1}, X_{i1}, \ldots, X_{iT}; u_i) = 0,\ s, t = 1, \ldots, T$$

- 处理固定效应的等价方式：把固定效应看作参数而非随机变量，从而进行虚拟变量回归。

$$Y_{it} = \beta D_{it} + \sum_{j=1}^{n} \delta_j F_{it}^j + \varepsilon_{it},\ F_{it}^j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

由这种方式可以得到固定效应的估计量，

$$\hat{u}_i = \hat{\delta}_i = \bar{Y}_i - \hat{\beta} \bar{D}_i$$

但这一估计量是不一致的，因为我们处理的是短面板 (T 固定，n 充分大)。

- Stata 会报告截距项，其定义为

$$\beta_0 \triangleq \mathbb{E}(u_i)$$

其一致估计量为

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i$$

- 实践中更经常采用的模型是双向乃至多向模型。

$$Y_{it} = \beta D_{it} + u_i + \eta_t + \varepsilon_{it}$$

- 当数据层级较多时，有更丰富的控制固定效应的方式。例如，因城市而异的时间趋势 `i.city#c.year`；因城市而异的时间固定效应 `i.city#i.year`

			个体固定效应						城市固定效应				时间固定效应			时间趋势	因城市而异的时间趋势				因城市而异的时间固定效应											
firm	city	year	D1	D2	D3	D4	D5	D6	C1	C2	C3	C4	T1	T2	T3	T	C1T	C2T	C3T	C4T	C1T1	C1T2	C1T3	C2T1	C2T2	C2T3	C3T1	C3T2	C3T3	C4T1	C4T2	C4T3
1	1	1	1						1				1			1	1				1											
1	1	2	1						1					1		2	2					1										
1	1	3	1						1						1	3	3						1									
2	1	1		1					1				1			1	1				1											
2	1	2		1					1					1		2	2					1										
2	1	3		1					1						1	3	3						1									
3	2	1			1					1			1			1		1						1								
3	2	2			1					1				1		2		2							1							
3	2	3			1					1					1	3		3								1						
4	3	1				1					1		1			1			1								1					
4	3	2				1					1			1		2			2									1				
4	3	3				1					1				1	3			3										1			
5	3	1					1				1		1			1			1								1					
5	3	2					1				1			1		2			2									1				
5	3	3					1				1				1	3			3										1			
6	4	1						1				1	1			1				1										1		
6	4	2						1				1		1		2				2											1	
6	4	3						1				1			1	3				3												1

- 固定效应模型的参数识别依赖于同一个体随时间的变化，由于我们实际使用的是双向固定效应模型，因此准确地说，参数识别依赖于同一个体剔除宏观趋势变动之后的随时间变化。换句话说，固定效应模型得不到显著的结果，可能并不是因为 D 不影响 Y，而是 D 的逐期变动中的信息含量较少；极端情况下，当 D 不随时间变化时，无法识别其对 Y 的影响（与个体固定效应完全共线性）；类似地，当 D 仅随时间变化、不随个体变化时，其对 Y 的影响也无法识别。
- 来看两组例子。

- 报告 R^2：可以报告 `xtreg` 的 within R^2, 也可以报告 `reghdfe` 的 R^2, 但不建议报告 `xtreg` 的 overall R^2.

$$\text{within } R^2 = \rho^2(Y_{it} - \bar{Y}_i, \hat{\beta}(D_{it} - \bar{D}_i)$$
$$\text{between } R^2 = \rho^2(\bar{Y}_i, \hat{\beta}\bar{D}_i)$$
$$\text{overall } R^2 = \rho^2(Y_{it}, \hat{\beta}D_{it})$$

- 面板数据的“异方差稳健标准误”没有意义，因为没有理由认为同一个体不同时期的扰动项不相关，因此系数估计标准误至少聚类到个体层面。考虑 ε_{ict}，

$$\begin{array}{cc}
 & \begin{array}{cccccccc} \varepsilon_{111} & \varepsilon_{112} & \varepsilon_{211} & \varepsilon_{212} & \varepsilon_{321} & \varepsilon_{322} & \varepsilon_{421} & \varepsilon_{422} \end{array} \\
\begin{array}{c} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{321} \\ \varepsilon_{322} \\ \varepsilon_{421} \\ \varepsilon_{422} \end{array} &
\begin{pmatrix}
\times & \times & & & & & & \\
\times & \times & & & & & & \\
 & & \times & \times & & & & \\
 & & \times & \times & & & & \\
 & & & & \times & \times & & \\
 & & & & \times & \times & & \\
 & & & & & & \times & \times \\
 & & & & & & \times & \times
\end{pmatrix}
\end{array}$$

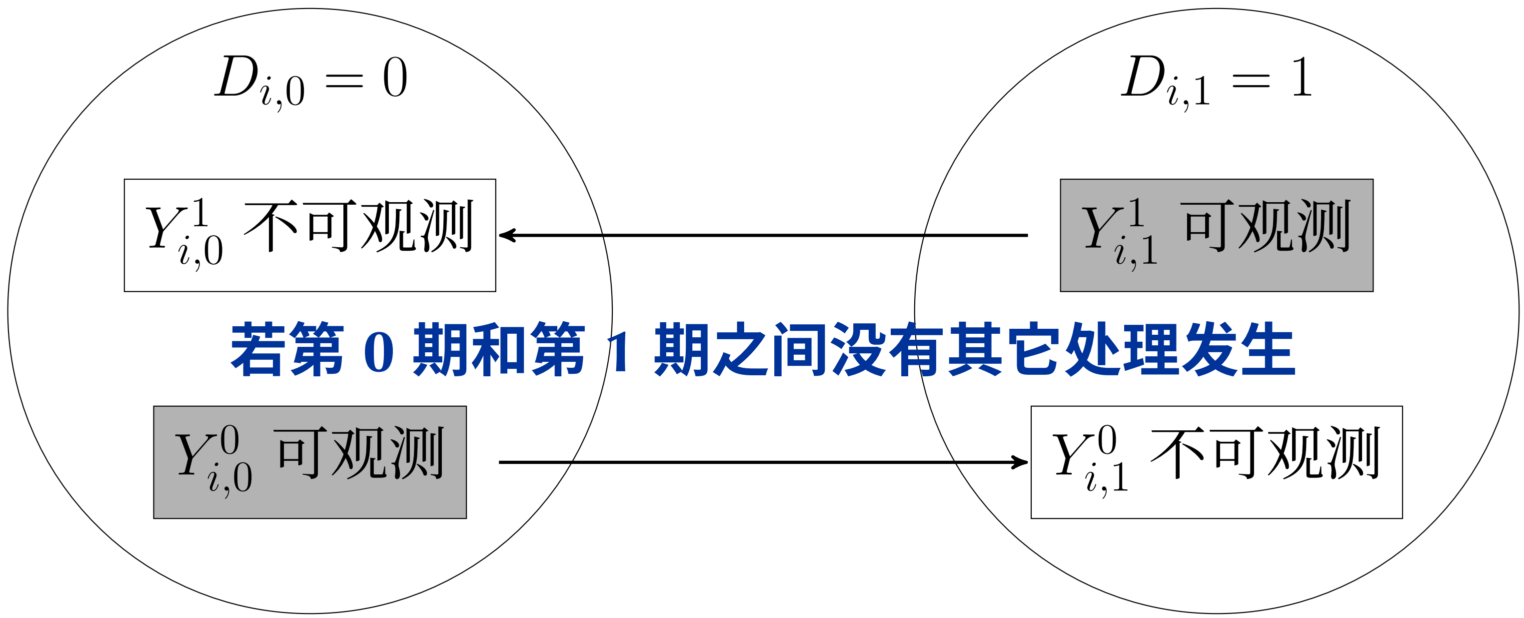
如果有必要，可以聚类到城市层面,

$$
\begin{array}{c|cccccccc}
 & \varepsilon_{111} & \varepsilon_{112} & \varepsilon_{211} & \varepsilon_{212} & \varepsilon_{321} & \varepsilon_{322} & \varepsilon_{421} & \varepsilon_{422} \\
\hline
\varepsilon_{111} & \times & \times & \times & \times & & & & \\
\varepsilon_{112} & \times & \times & \times & \times & & & & \\
\varepsilon_{211} & \times & \times & \times & \times & & & & \\
\varepsilon_{212} & \times & \times & \times & \times & & & & \\
\varepsilon_{321} & & & & & \times & \times & \times & \times \\
\varepsilon_{322} & & & & & \times & \times & \times & \times \\
\varepsilon_{421} & & & & & \times & \times & \times & \times \\
\varepsilon_{422} & & & & & \times & \times & \times & \times
\end{array}
$$

- 不能通过将解释变量替换成其滞后项来解决内生性问题，也不能使用解释变量的滞后项作为其工具变量。
- 对于非线性模型，消除固定效应的方法失效，此时要得到解释变量系数的一致估计需要限制性很强的假设，因此对于结构复杂的面板数据，研究者更多地采用线性概率模型。
- 有时会控制被解释变量的滞后项，主要理由是被解释变量的 persistence？要使用动态面板方法？

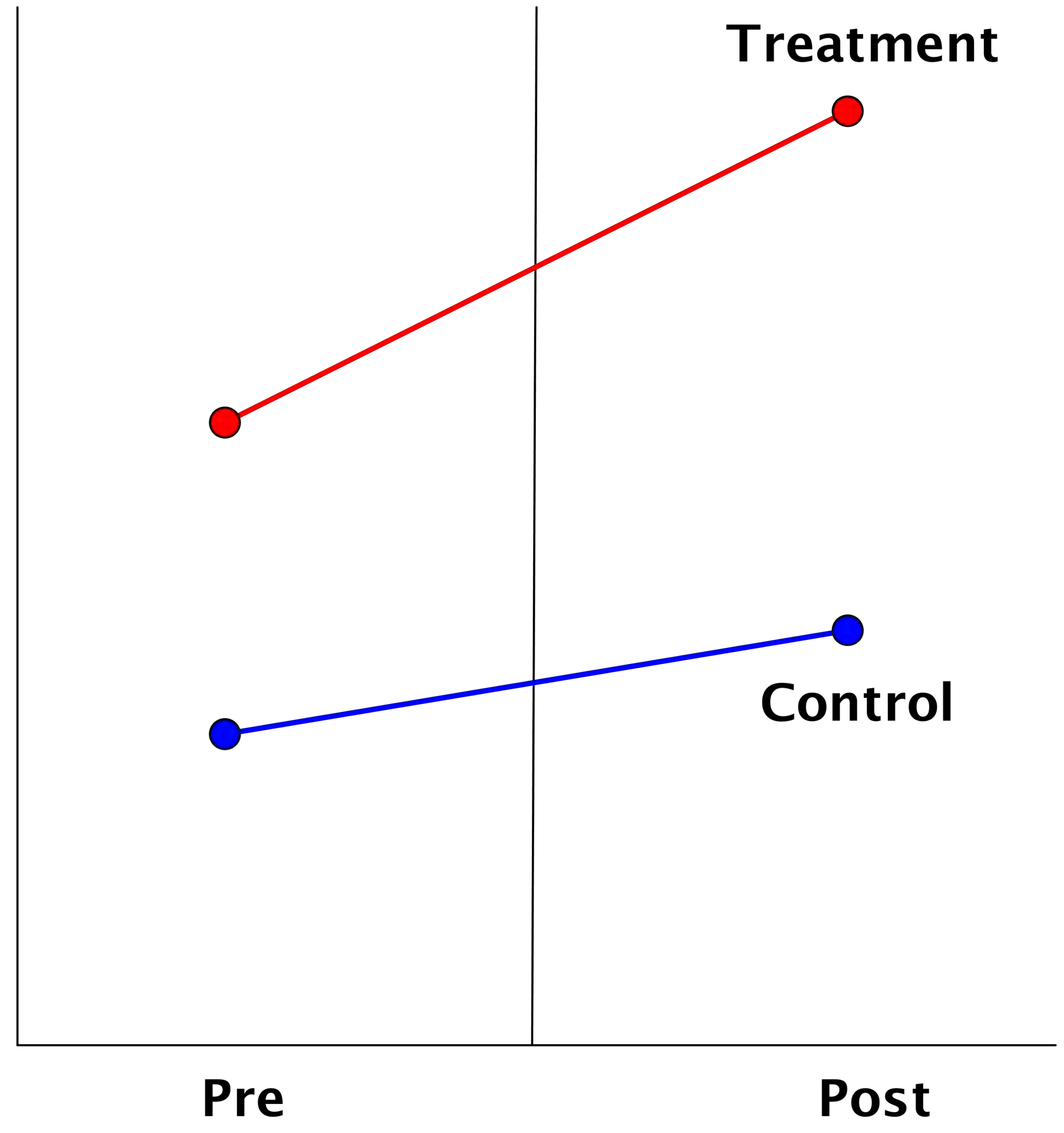
5.2 双重差分

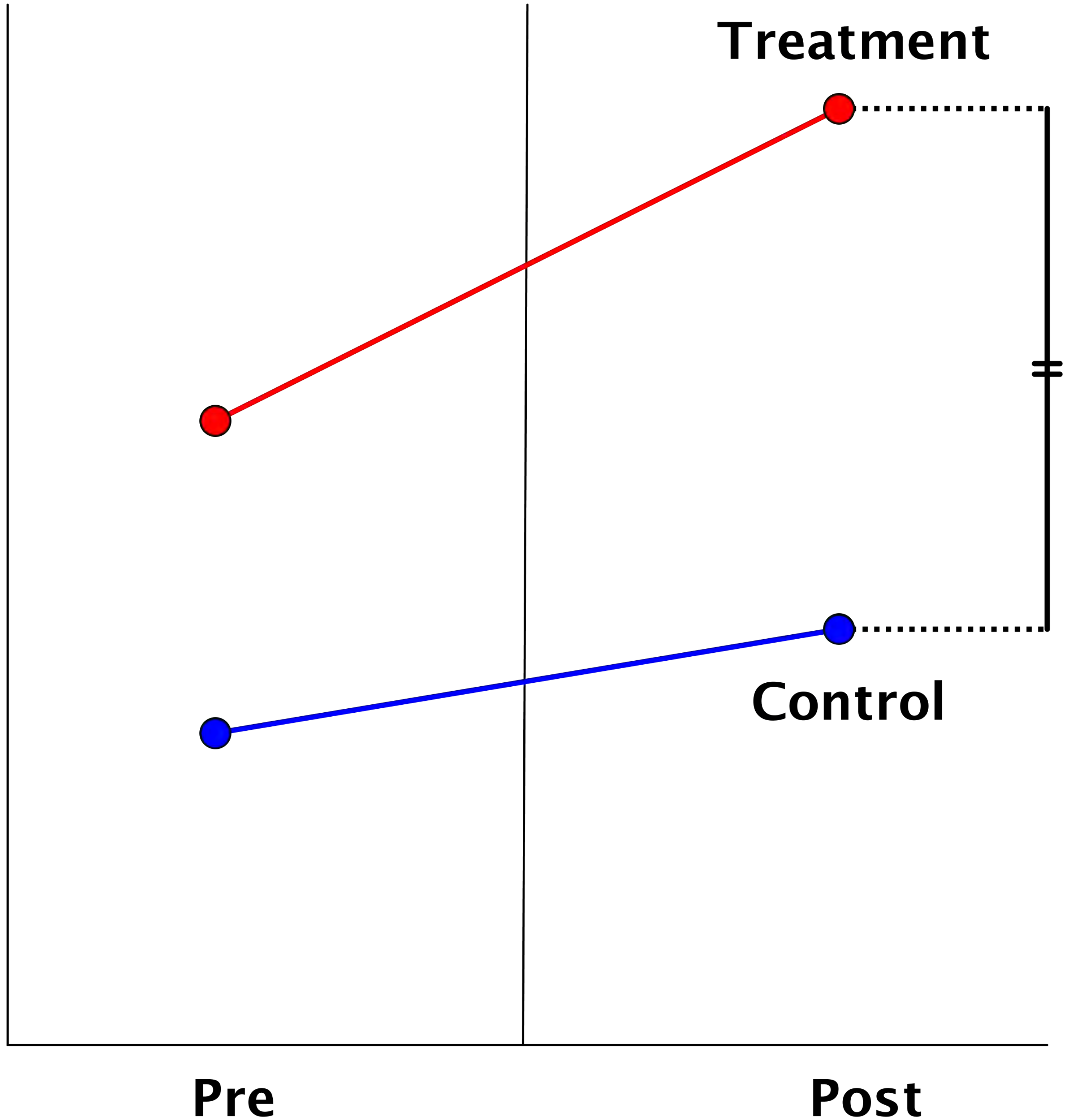
- 当扩充了数据的时间维度，就可以考虑用同一个体不同时期处理状态相反的结果来构造反事实。



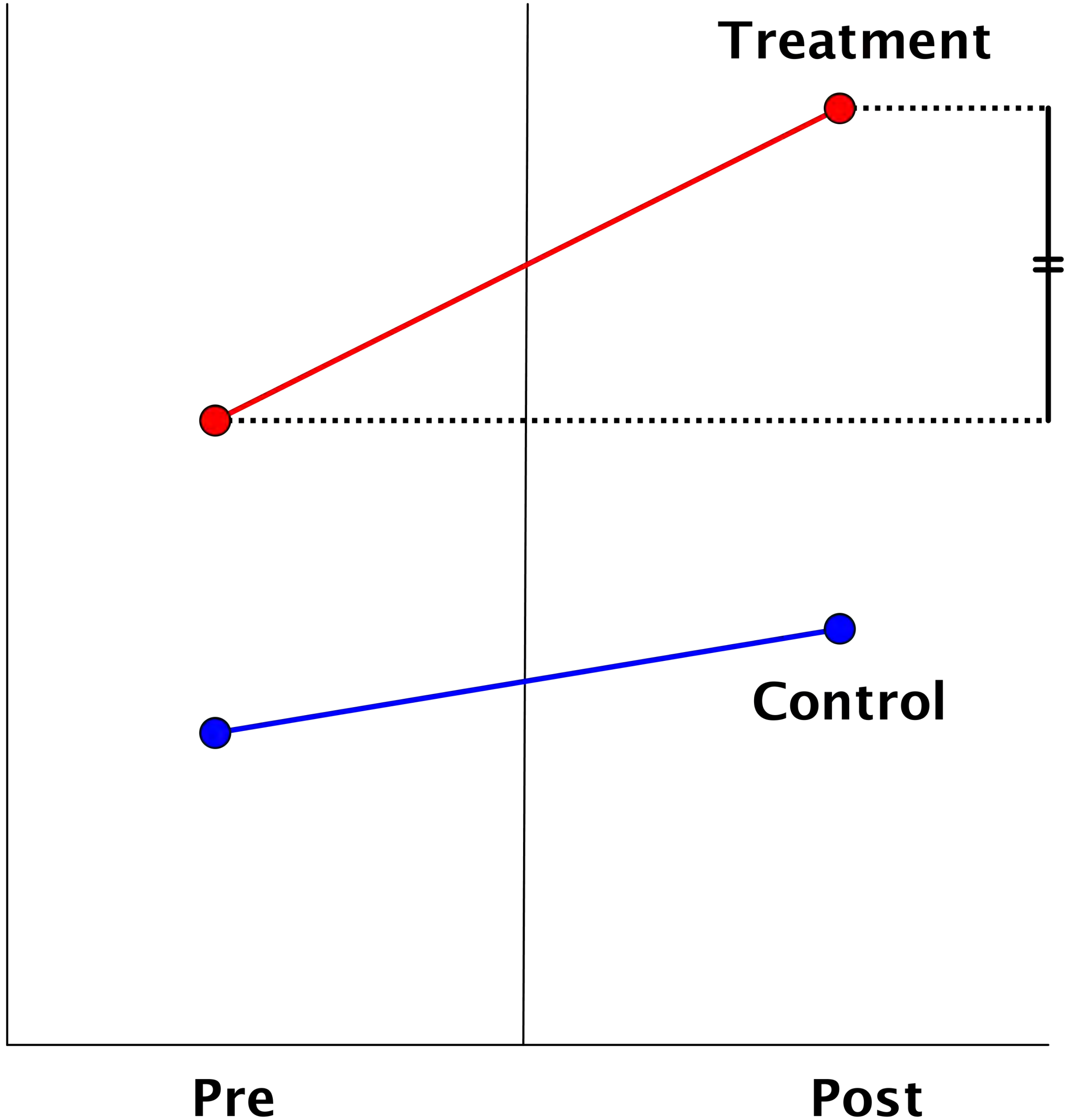
$$
\begin{aligned}
\hat{\tau} = \frac{1}{N}\sum_i \left(Y_{i1} - Y_{i0}\right) \to_p \mathbb{E}(Y_1 - Y_0) \\
&= \mathbb{E}(Y_1^1 - Y_0^0) \\
&= \underbrace{\mathbb{E}(Y_1^1 - Y_1^0)}_{\text{ATE at post}} + \mathbb{E}(Y_1^0 - Y_0^0)
\end{aligned}
$$

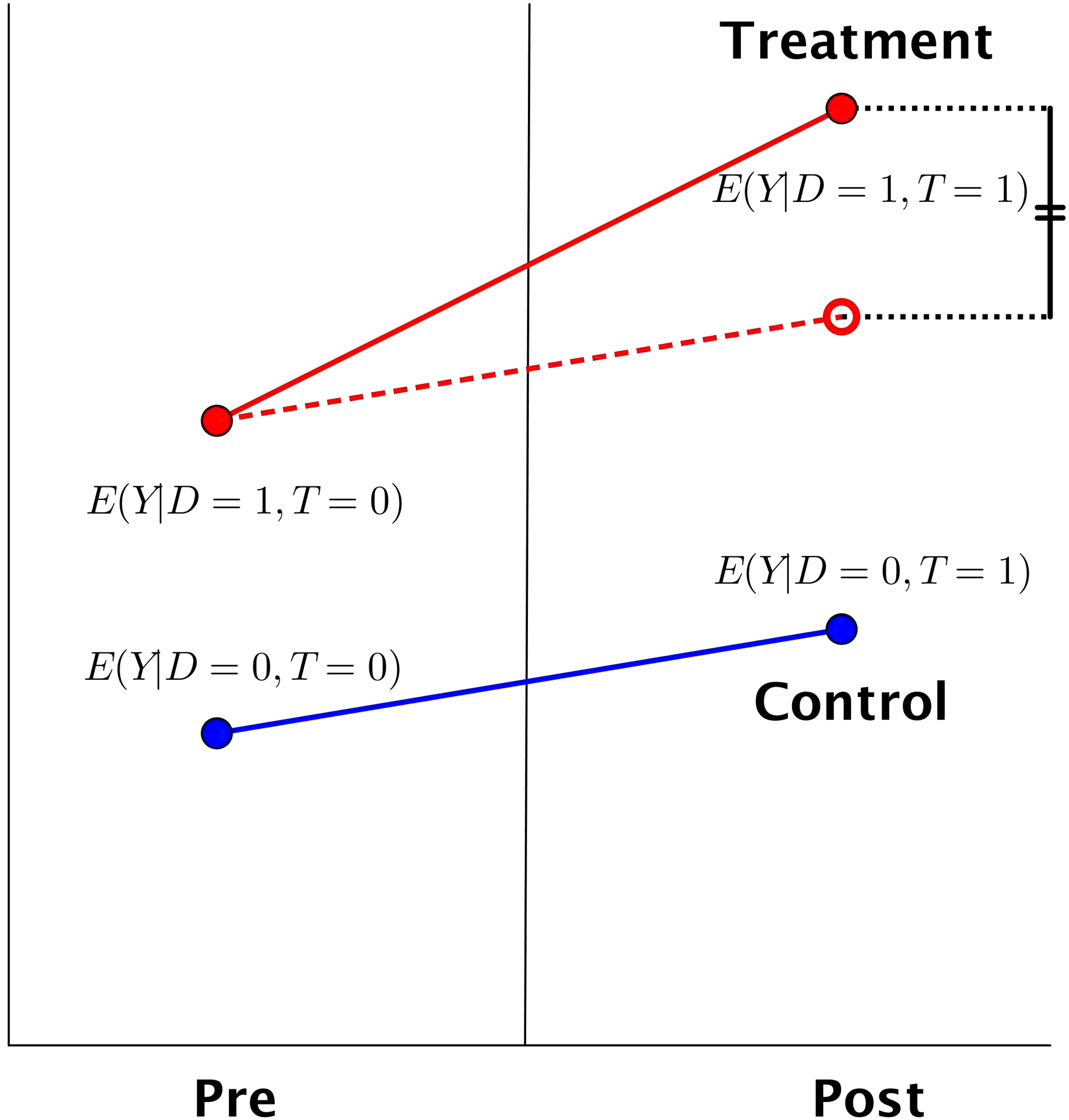
- 事前事后比较的识别假设为 $\mathbb{E}(Y_1^0 - Y_0^0) = 0$，时间间隔越短，该假设越可能成立。
- 但政策评估往往更关注政策的长期效果，因此很难保证第 0 期和第 1 期之间，结果不会因为别的原因发生变化。此时可以考虑用控制组实际未经处理的结果**变化**作为处理组假若未经处理的结果**变化**的反事实。


Treatment
Control
Pre
Post



Treatment
Control
Pre
Post


Treatment
Control
Pre
Post


Treatment
$E(Y|D=1,T=1)$
$E(Y|D=1,T=0)$
$E(Y|D=0,T=1)$
$E(Y|D=0,T=0)$
Control
Pre
Post

DID 的识别

$$
\begin{aligned}
\tau(X) =& \mathbb{E}(Y|X, D=1, T=1) - \mathbb{E}(Y|X, D=1, T=0) \\
& - \left[\mathbb{E}(Y|X, D=0, T=1) - \mathbb{E}(Y|X, D=0, T=0)\right] \\
=& \mathbb{E}(Y^1|X, D=1, T=1) - \mathbb{E}(Y^0|X, D=1, T=0) \\
& - \left[\mathbb{E}(Y^0|X, D=0, T=1) - \mathbb{E}(Y^0|X, D=0, T=0)\right] \\
=& \left[\mathbb{E}(Y^1|X, D=1, T=1) - \mathbb{E}(Y^0|X, D=1, T=1)\right] \\
& + \left[\mathbb{E}(Y^0|X, D=1, T=1) - \mathbb{E}(Y^0|X, D=1, T=0)\right] \\
& - \left[\mathbb{E}(Y^0|X, D=0, T=1) - \mathbb{E}(Y^0|X, D=0, T=0)\right]
\end{aligned}
$$

- 第三类识别假设：

$$
\begin{aligned}
& \mathbb{E}(Y^0|X, D=1, T=1) - \mathbb{E}(Y^0|X, D=1, T=0) \\
=& \mathbb{E}(Y^0|X, D=0, T=1) - \mathbb{E}(Y^0|X, D=0, T=0)
\end{aligned}
$$

- 在此假设下，DID 估计量识别了处理组接受处理后的平均处理效应 (ATT at the post-treatment period).

$$
\tau_{\text{DID}} = \mathbb{E}\left(Y^1 - Y^0 \middle| D=1, T=1\right) = \mathbb{E}_X\left[\tau(X)|D=1, T=1\right]
$$

- 第三类识别假设对分配机制的要求是什么？（DID 要求随机分组么？）这取决于怎么理解随机分组。第三类识别假设可以简写作

  Assumption ID.3: $\mathbb{E}\big(\Delta Y^0|X, D=1\big) = \mathbb{E}\big(\Delta Y^0|X, D=0\big)$

  也就是说给定 X，ΔY^0 均值独立于 D，即关于 ΔY^0 是随机分组的！
- 假设**ID.3**比假设**ID.2**更弱么？

  回忆假设**ID.2**：

  $$\mathbb{E}\big(Y^0|X, D=1\big) = \mathbb{E}\big(Y^0|X, D=0\big)$$

  - 从数学上说，答案显然是否定的。
  - 假设**ID.3**可以（部分）检验。但假设**ID.2**也可以（找到某个 pseudo-outcome，比如 lagged outcome）。
- 当前的现状是双重差分研究泛滥，但对其识别假设是否成立的讨论明显不足。

DID 的估计

- 基本思路

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 T_t + \beta_3 D_i \times T_t + \varepsilon_{it}$$
$$i = 1, 2, \ldots, n;\ t = 0, 1$$

其中

$$D_i = \begin{cases} 1 & \text{个体 } i \text{ 来自处理组} \\ 0 & \text{个体 } i \text{ 来自控制组} \end{cases}$$

$$T_t = \begin{cases} 1 & t = 1\,(\text{处理已实施}) \\ 0 & t = 0\,(\text{处理未实施}) \end{cases}$$

$$\mathbb{E}\left(Y_{it}|\cdot\right) = \begin{cases} \beta_0 + \beta_1 & D_i = 1\ \&\ T_t = 0 \\ \beta_0 + \beta_1 + \beta_2 + \beta_3 & D_i = 1\ \&\ T_t = 1 \\ \beta_0 + & D_i = 0\ \&\ T_t = 0 \\ \beta_0 + \beta_2 & D_i = 0\ \&\ T_t = 1 \end{cases}$$

- 对于面板数据，可以控制双向固定效应和控制变量

$$Y_{it} = \beta D_i \times T_t + \mathbf{X}'_{it}\gamma + u_i + \eta_t + \varepsilon_{it}$$

- 对于两期面板，等价于一阶差分

$$\Delta Y_i = \beta_0 + \beta_1 D_i + \mathbf{X}'_i\gamma + \varepsilon_i$$

- 重复横截面数据

$$Y_{it} = \beta D_i \times T_t + \mathbf{X}'_{it}\gamma + \rho D_i + \eta_t + \varepsilon_{it}$$

  但通常仍会控制更细的横截面固定效应。

- 可以很方便地拓展到多期面板情形。假定处理从第 T^* 期开始实施，则只需重新定义 T_t，

$$T_t = \begin{cases} 1 & t \geq T^* \\ 0 & t < T^* \end{cases}$$

- 对于多期面板，可以采用更灵活的形式

$$Y_{it} = \sum_{l=2}^{T} \beta_l (D_i \times T_t^l) + \mathbf{X}_{it}' \gamma + u_i + \eta_t + \varepsilon_{it}$$

$$T_t^l = \begin{cases} 1 & t = l \\ 0 & t \neq l \end{cases}$$

我们应该期望得到

$$\beta_2 \approx \beta_3 \approx \cdots \beta_{T^*-1} \approx 0;\ \beta_{T^*},\ \beta_{T^*+1},\ \cdots,\ \beta_T \neq 0$$

- 平行趋势本质上是不可检验的——以前平行不见得今后就平行。如果处理不是随机的，而是由某因素 X 所决定，那么这个因素 X 在决定处理与否的同时，很有可能会影响趋势。如果 X 是可观测的，就要尽量控制起来，甚至控制它与时间虚拟变量的交互项。

$$Y_{it} = \beta D_i \times T_t + \mathbf{X}'_{it}\boldsymbol{\gamma} + \sum_{l=2}^{T} (\mathbf{X}'_i \times T_t^l)\boldsymbol{\delta}_l + u_i + \eta_t + \varepsilon_{it}$$

- 检查结果对于个体特定的线性趋势的加入是否敏感。

$$Y_{it} = \tau D_i \times T_t + u_i + \eta_t + u_i \times t + \varepsilon_{it}$$

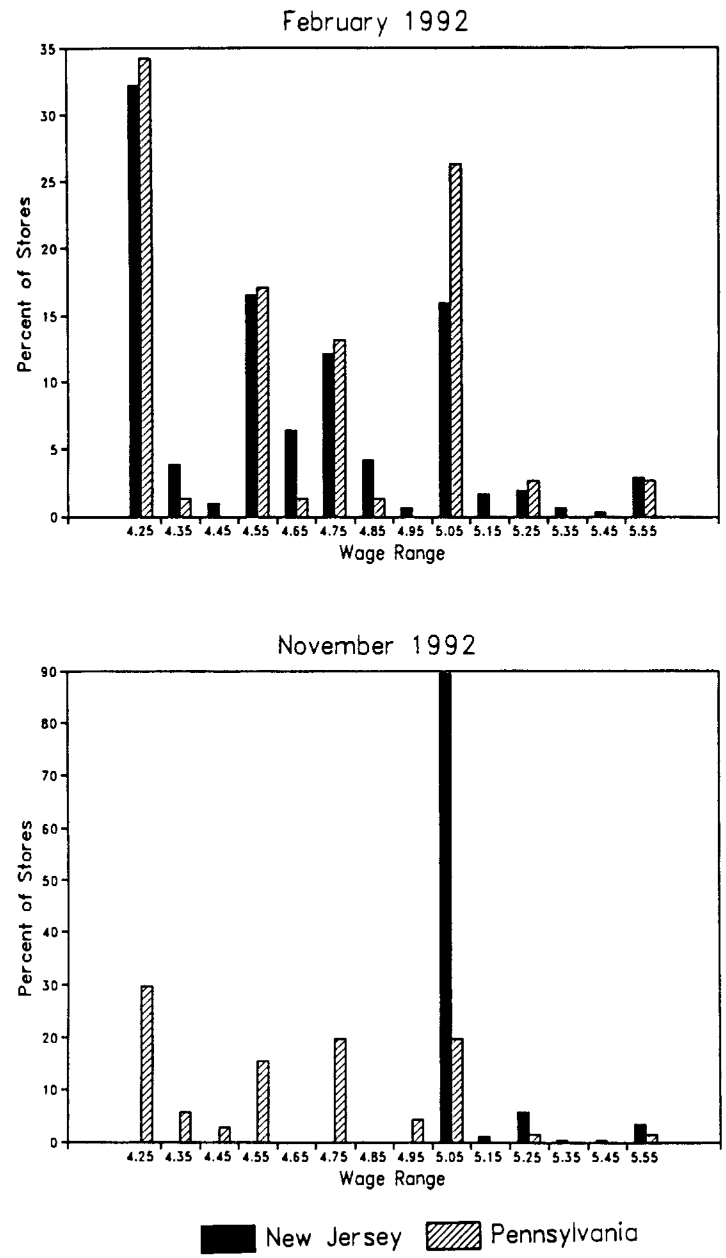
- 尽管处理不是随机的，但仍然希望处理组和控制组越接近越好，可以尝试检验重要控制变量 X 的差异，或者考虑与匹配方法相结合。
- 如果处理组和控制组之间存在明显差异，对于不同的控制组选择，结果需保持稳健。

示例 20. 提高最低工资是否降低就业？(Card and Krueger, 1994, *AER*)

- 政策事件：1992 年新泽西州最低工资标准从 $\$4.25$ 提高到 $\$5.05$。
- 处理组：受政策影响的新泽西州低工资快餐店。
- 控制组：不受政策影响的新泽西州高工资快餐店，以及宾夕法尼亚州东部快餐店。
- 时期：政策事件前后两轮调查。

– 描述性统计：看处理组和控制组之间的可比性

Variable	Stores in: NJ	PA	t^a
1. *Distribution of Store Types (percentages):*			
a. Burger King	41.1	44.3	−0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	−1.1
e. Company-owned	34.1	35.4	−0.2
2. *Means in Wave 1:*			
a. FTE employment	20.4	23.3	−2.0
	(0.51)	(1.35)	
b. Percentage full-time employees	32.8	35.0	−0.7
	(1.3)	(2.7)	
c. Starting wage	4.61	4.63	−0.4
	(0.02)	(0.04)	
d. Wage = $4.25 (percentage)	30.5	32.9	−0.4
	(2.5)	(5.3)	
e. Price of full meal	3.35	3.04	4.0
	(0.04)	(0.07)	
f. Hours open (weekday)	14.4	14.5	−0.3
	(0.2)	(0.3)	
g. Recruiting bonus	23.6	29.1	−1.0
	(2.3)	(5.1)	
3. *Means in Wave 2:*			
a. FTE employment	21.0	21.2	−0.2
	(0.52)	(0.94)	
b. Percentage full-time employees	35.9	30.4	1.8
	(1.4)	(2.8)	
c. Starting wage	5.08	4.62	10.8
	(0.01)	(0.04)	
d. Wage = $4.25 (percentage)	0.0	25.3	—
		(4.9)	
e. Wage = $5.05 (percentage)	85.2	1.3	36.1
	(2.0)	(1.3)	
f. Price of full meal	3.41	3.03	5.0
	(0.04)	(0.07)	
g. Hours open (weekday)	14.4	14.7	−0.8
	(0.2)	(0.3)	
h. Recruiting bonus	20.3	23.4	−0.6
	(2.3)	(4.9)	



– 交叉表/列联表 (cross tabulation)：展示基本结果

Variable	Stores by state: PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Stores in New Jersey[a]: Wage = $4.25 (iv)	Wage = $4.26–$4.99 (v)	Wage ≥ $5.00 (vi)	Differences within NJ[b]: Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)

– 模型 1:

$$\Delta E_i = \beta_0 + \beta_1 \cdot \texttt{NJ}_i + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

– 模型 2:

$$\Delta E_i = \beta_0 + \beta_1 \cdot \texttt{GAP}_i + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

其中

$$\texttt{GAP}_i = \begin{cases} 0 & \text{宾夕法尼亚州快餐店} \\ 0 & \text{新泽西州高工资快餐店} \\ 1 & \text{新泽西州低工资快餐店} \end{cases}$$

– 模型 3：

$$\Delta E_i = \beta_0 + \beta_1 \cdot \texttt{NJ}_i + \beta_2 \cdot \texttt{GAP}_i + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

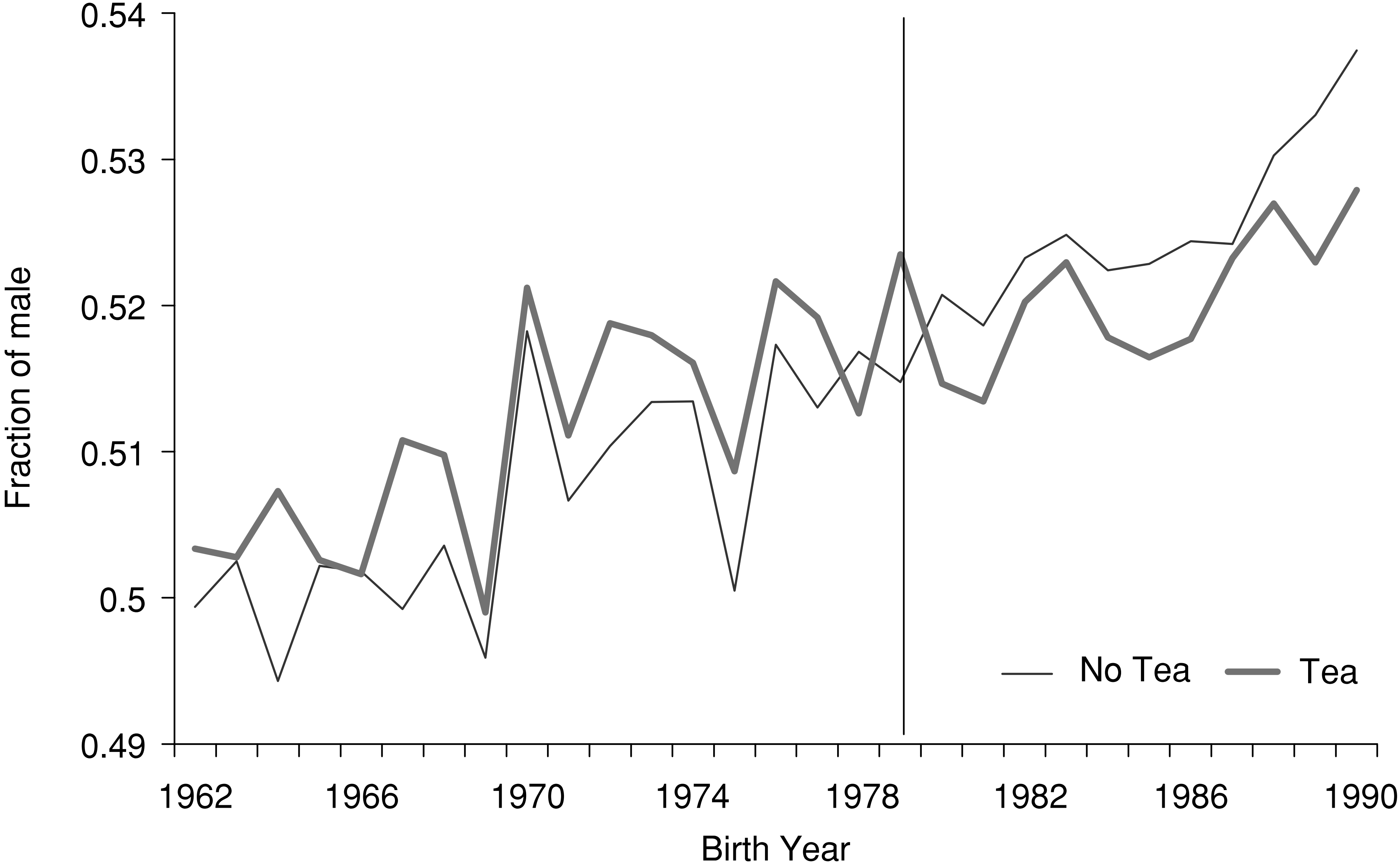
Since GAP_i varies within New Jersey, it is possible to add both GAP_i and NJ_i to the employment model. The estimated coefficient of the New Jersey dummy then provides a test of the Pennsylvania control group. When we estimate these models, the coefficient of the New Jersey dummy is insignificant (with t ratios of 0.3–0.7), implying that inferences about the effect of the minimum wage are similar whether the comparison is made across states or across stores in New Jersey with higher and lower initial wages.

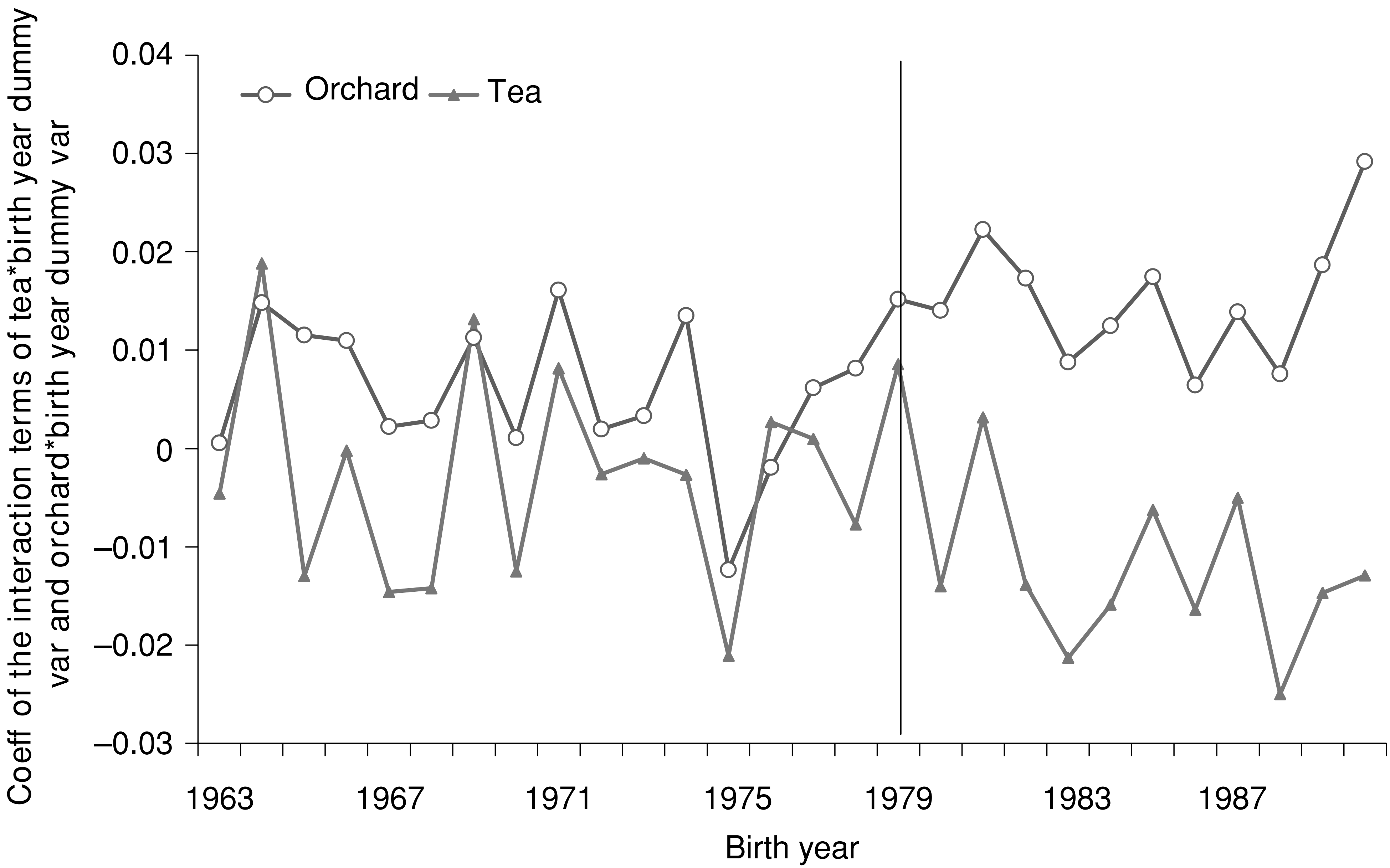
- 考虑政策处理的工具变量。

示例 21. 茶叶的价格与消失的女性 (Qian, 2008, *QJE*).

- 结果变量：各地区每年新生儿中男性占比。
- 政策事件：中央计划经济时代，国家控制主要农作物的价格。1978 年价格体制改革，放开了包括茶叶和果物在内的经济作物价格。
- 处理组：适宜种植茶叶的地区和适宜种植果树的地区。
- 控制组：其余地区。
- 时期：1962-1978 年，改革前；1979-1990 年，改革后。
- 故事：女性种植茶叶有比较优势，男性种植果树有比较优势，因此在放开价格以后，适宜种植茶叶地区，女性对家庭收入增加的贡献较高，男女性别比下降；适宜种植果树的地区，男性对家庭收入增加的贡献较高，男女性别比升高。

$$\texttt{sex}_{it} = \sum_{l=1963}^{1990} \beta_l \cdot \texttt{tea}_i \times D_t^l + \sum_{l=1963}^{1990} \delta_l \cdot \texttt{orchard}_i \times D_t^l + \mathbf{X}_{it}'\gamma + u_i + \eta_t + \varepsilon_{it}$$


0.54
0.53
0.52
0.51
0.5
0.49
Fraction of male
1962
1966
1970
1974
1978
1982
1986
1990
Birth Year
No Tea
Tea


Orchard
Tea
0.04
0.03
0.02
0.01
0
–0.01
–0.02
–0.03
Coeff of the interaction terms of tea*birth year dummy var and orchard*birth year dummy var
1963
1967
1971
1975
1979
1983
1987
Birth year

– 将适宜种植茶叶的地区定义为 1997 年中国农业普查数据中茶叶实际种植面积大于零的地区。

Another problem of the empirical strategy is that if, at the time of the reforms, there is a change in the attitudes that drive sex preference in tea-planting counties, then the estimate of the effect of planting tea will capture both the relative female income effect and the effect of the attitude change. Or, if the increase in the value of tea changed the reason for women to pick tea, then the prereform cohort will not be an adequate control group. Although I cannot resolve the former problem, the latter is addressed by instrumenting for tea planting with time-invariant geographic data.[14]

– 考虑到茶叶必须种植在温暖、半湿润的山顶，能够避开强风和暴雨，因此坡度 (hilliness) 可以作为是否适宜种植茶叶的工具变量。

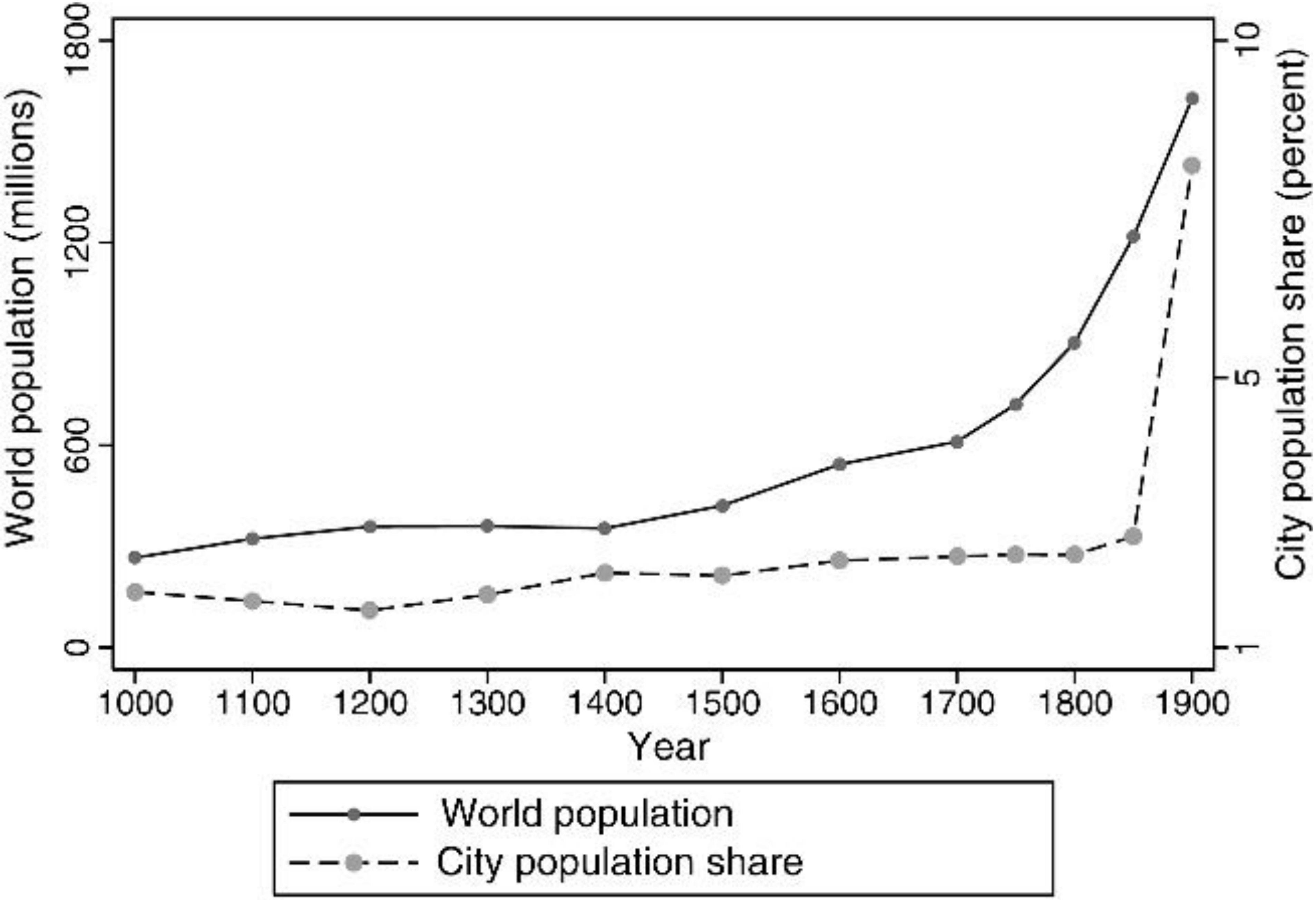
$$\texttt{tea}_i \times \texttt{post}_t = \gamma_0 + \gamma_1 \cdot \texttt{slope}_i \times \texttt{post}_t + \delta_i + \theta_t + u_{it}$$

$$\texttt{sex}_{it} = \beta_0 + \beta_1 \cdot \texttt{tea}_i \times \texttt{post}_t + \mu_i + \eta_t + \varepsilon_{it}$$

- 如果所有个体都在同一时间接受了处理，则 $D_i \times T_t = T_t$ 与时间固定效应共线性，无法识别。此时需要尝试构建处理强度指标 (treatment intensity) 指标，可以称之为连续处理的 DID 或连续型 DID。（事实上**示例 20**中的 $\texttt{GAP}_i$ 变量并不是虚拟变量。）

示例 22. 土豆与人口增长 (Nunn and Qian, 2011, *QJE*).

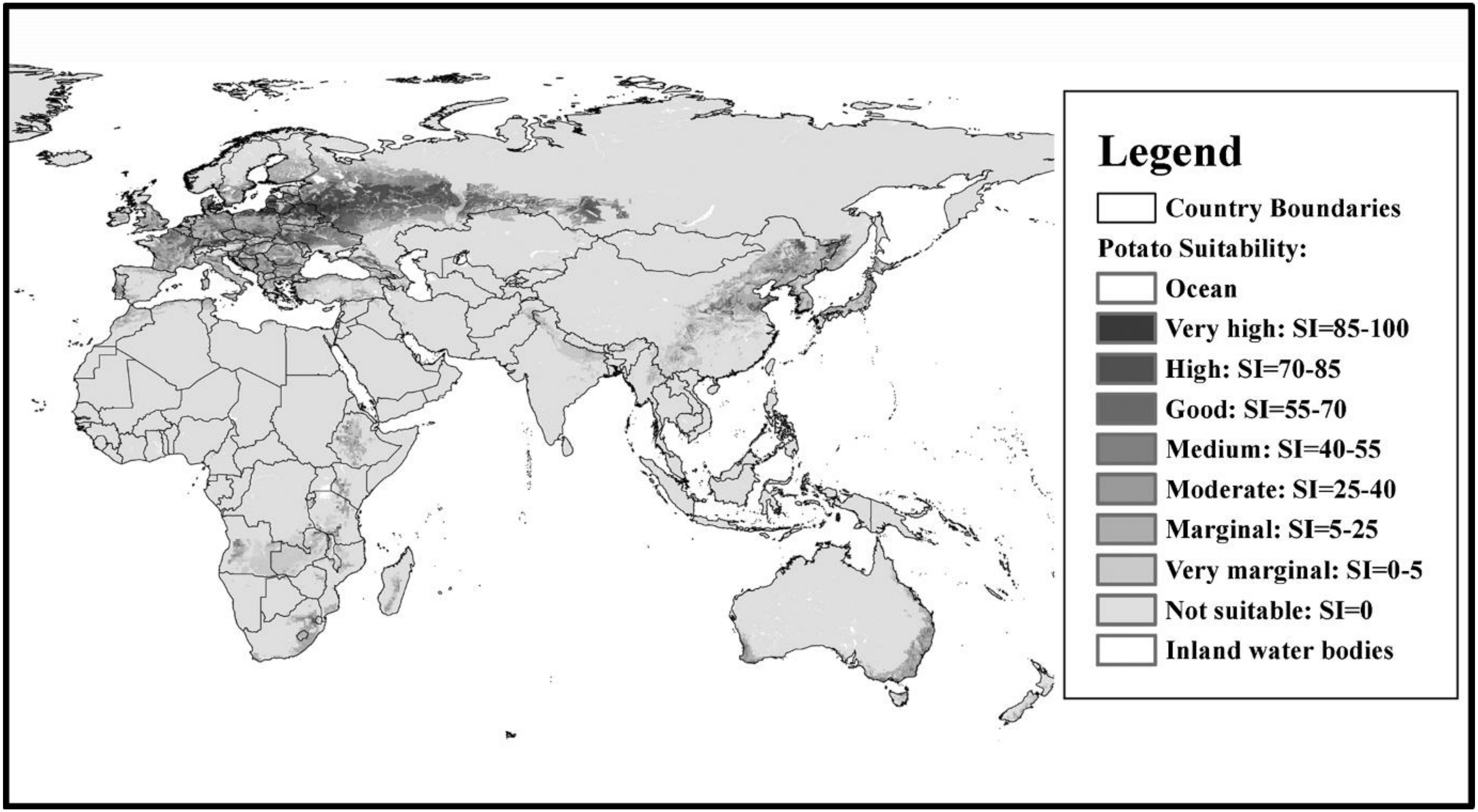
- 跨度为 1000-1900 年的超长期经济史研究。考察土豆从新大陆被引入旧大陆后对世界人口增长和城市化的贡献。
- 与其它粮食作物相比，土豆能提供每亩更多的热量和营养。有许多历史证据表明土豆对人口增长和经济发展的贡献，但定量证据缺乏。
- 学习引言中最重要的几段话的写法：研究这个问题有何难度？本文的研究策略是什么？如何评价这个策略？


World population (millions)
1800
1200
600
0
City population share (percent)
10
5
1
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
Year
World population
City population share

Despite qualitative accounts of the benefits of the potato to the Old World, empirical evidence quantifying the overall impact of the potato is scarce. This is no doubt partly due to the estimation difficulties caused by the endogeneity of potato adoption. First, there is an issue of reverse causality. The adoption of potatoes might have caused population growth, but alternatively, latent population pressure and the associated demand for food might have caused the adoption of potatoes. A number of historians have argued for the latter relationship (e.g., Salaman 1949; Connell 1962; Cullen 1968). The second problem is joint determination. Both population growth and the adoption of new agricultural technologies can be the outcome of a third unobserved factor.

The principal contribution of our study is to provide a rigorous quantitative analysis of the historical role of potatoes in increasing population and urbanization. We expand the scope of Mokyr's analysis by examining the effect of potatoes on the entire Old World during the eighteenth and nineteenth centuries and by examining urbanization, city growth, and adult height in addition to population. Our estimation strategy exploits two sources of variation. The first is time variation arising from the introduction of potatoes as a field crop in the Old World. Potatoes did not exist in the Old World for most of history. They are native to South America and were first discovered by Europeans during the voyages of Columbus. Potatoes were widely adopted as a field crop in Europe towards the end of the seventeenth and beginning of the eighteenth centuries. Their cultivation then spread to the rest of the Old World, mainly through European sailors and missionaries. The second source of variation is cross-sectional and arises from differences in countries' suitability for cultivating potatoes, as determined by time-invariant geoclimatic conditions. Conditional on access to potatoes, regions that are more suitable for potato cultivation will be able to increase food production more. Our identification strategy relies on the interaction of the two sources of variation, and only the interaction can be interpreted as plausibly exogenous. Our strategy, similar in spirit to a differences-in-differences (DD) strategy, compares population and urbanization levels between Old World countries that were more suitable for potato cultivation to regions that were less suitable, before and after potatoes were adopted in the Old World.

Our strategy shares most of the advantages and disadvantages of a standard DD strategy. On the one hand, it allows us to control for both country and time-period fixed effects so that all time-invariant differences across countries—such as geography, food preferences, or institutions (to the extent that they change slowly over time)—and secular changes over time—such as global improvements in health, sanitation, and technological advancements—are controlled for. On the other hand, the strategy relies on there being no other shocks occurring around the same time that potatoes were adopted that are correlated with countries' suitability for potato cultivation. We address this identification concern by directly controlling for time- and country-varying factors that might bias our estimates. In the baseline estimates, we control for potentially important characteristics, each interacted with the full set of time-period indicator variables. This allows the effect of each factor to vary flexibly over time. The characteristics include a country's suitability for cultivating the Old World staple crops wheat and rice, as well as three geographic characteristics that are correlated with potato suitability: terrain ruggedness, elevation, and the presence of a tropical climate. We also control for a number of characteristics that have been identified as being important for historical growth and development.


Legend
Country Boundaries
Potato Suitability:
Ocean
Very high: SI=85-100
High: SI=70-85
Good: SI=55-70
Medium: SI=40-55
Moderate: SI=25-40
Marginal: SI=5-25
Very marginal: SI=0-5
Not suitable: SI=0
Inland water bodies

- 需要正式处理的实证上的威胁：
  - ▷ 有没有可能是诸如玉米、青贮玉米、甜薯、木薯等其它新大陆粮食作物的贡献？
  - ▷ 适宜种植土豆的土地大多集中在欧洲，而西欧 1700 年以后的兴起有其它原因，可能和土豆种植相关。
- 将处理时点设定为同一时点，主要是因为缺乏关于各地区何时开始种植土豆的可信记载，但也由此缓解了内生处理问题。

Because the date of actual adoption in a country could be partly driven by endogenous factors such as latent population demand, we use the same date of initial adoption for all countries. The his-

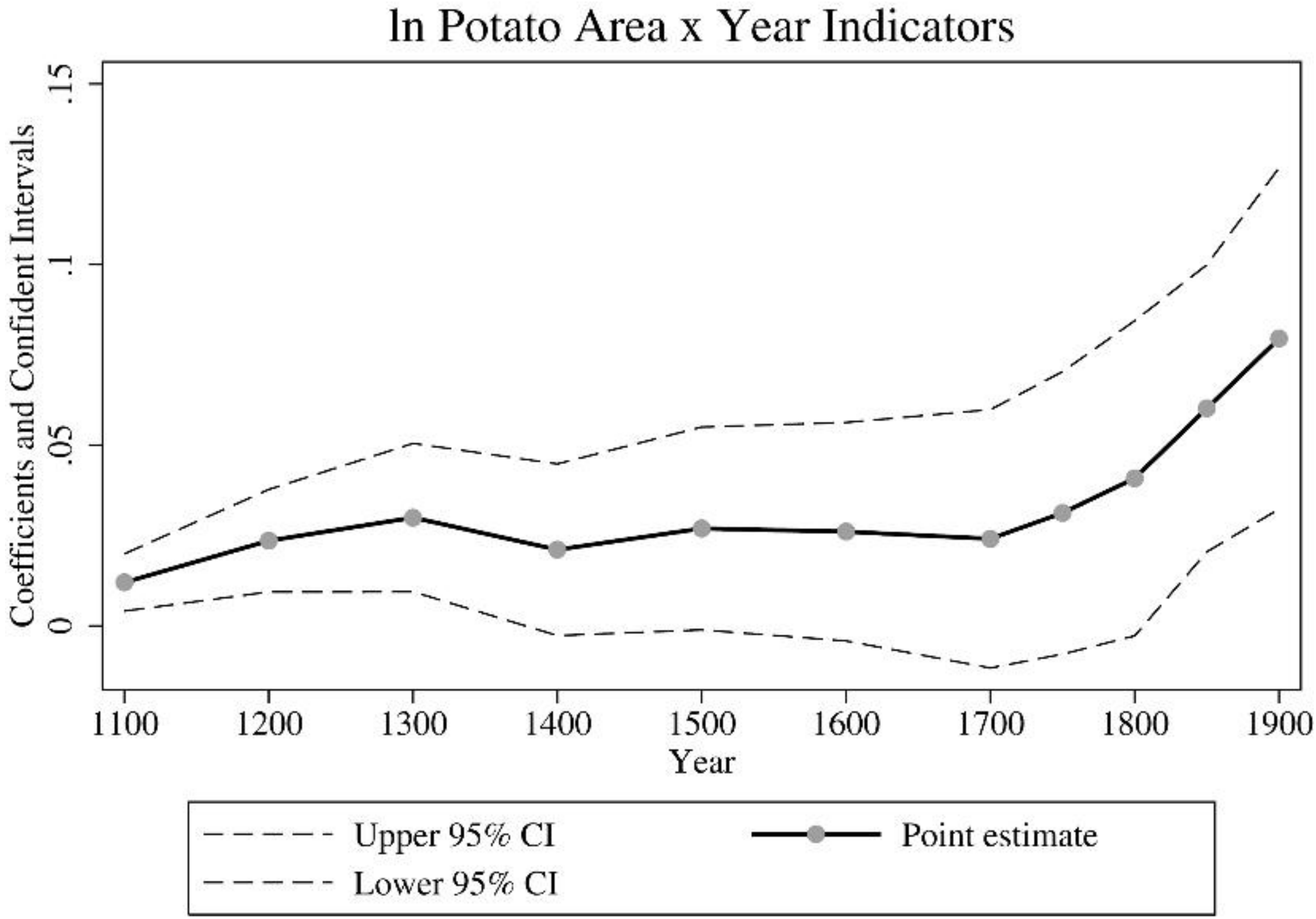
- 历史证据表明，土豆的大规模推广发生在 1750-1800 年间。文章最终将处理前时期设定为 8 期（1000-1700），处理后时期设定为 4 期（1750、1800、1850、1900）。但这需要数据支持，因此文章先进行灵活估计和滚动估计，再进行基准估计。

$$Y_{it} = \beta \ln \texttt{PotatoArea}_i \cdot \mathbb{1}(t > 1700)_t + \sum_{l=1100}^{1900} (\mathbf{X}_i' \times T_t^l)\boldsymbol{\delta}_l + u_i + \eta_t + \varepsilon_{it}$$

$$Y_{it} = \sum_{l=1100}^{1900} \beta_l \ln \texttt{PotatoArea}_i \cdot T_t^l + \sum_{l=1100}^{1900} (\mathbf{X}_i' \times T_t^l)\boldsymbol{\delta}_l + u_i + \eta_t + \varepsilon_{it}$$

FLEXIBLE ESTIMATES: THE RELATIONSHIP BETWEEN POTATO-SUITABLE LAND AREA AND POPULATION OR CITY POPULATION SHARE BY TIME PERIOD

	Dependent Variable					
	ln total population			City population share		
	(1)	(2)	(3)	(4)	(5)	(6)
ln *Potato-Suitable Area* × 1100	0.013	0.011	0.012	–0.0018	–0.0013	–0.0006
	(0.003)	(0.003)	(0.004)	(0.0014)	(0.0009)	(0.0013)
ln *Potato-Suitable Area* × 1200	0.029	0.024	0.024	–0.0011	–0.0013	–0.0012
	(0.005)	(0.005)	(0.007)	(0.0009)	(0.0009)	(0.0012)
ln *Potato-Suitable Area* × 1300	0.039	0.031	0.030	0.0002	–0.0005	0.0014
	(0.007)	(0.007)	(0.010)	(0.0008)	(0.0011)	(0.0014)
ln *Potato-Suitable Area* × 1400	0.019	0.004	0.021	0.0008	0.0002	0.0010
	(0.008)	(0.008)	(0.012)	(0.0012)	(0.0015)	(0.0014)
ln *Potato-Suitable Area* × 1500	0.034	0.014	0.027	0.0003	–0.0002	0.0008
	(0.009)	(0.010)	(0.014)	(0.0009)	(0.0012)	(0.0013)
ln *Potato-Suitable Area* × 1600	0.041	0.021	0.026	0.0002	–0.0010	–0.0000
	(0.009)	(0.011)	(0.015)	(0.0014)	(0.0025)	(0.0029)
ln *Potato-Suitable Area* × 1700	0.043	0.018	0.024	0.0020	0.0017	0.0022
	(0.012)	(0.013)	(0.018)	(0.0010)	(0.0013)	(0.0015)
ln *Potato-Suitable Area* × *1750*	*0.055*	*0.030*	*0.031*	*0.0015*	*0.0011*	*0.0013*
	(0.012)	*(0.014)*	*(0.020)*	*(0.0009)*	*(0.0013)*	*(0.0018)*
ln *Potato-Suitable Area* × *1800*	*0.073*	*0.048*	*0.041*	*0.0020*	*0.0016*	*0.0018*
	(0.014)	*(0.015)*	*(0.022)*	*(0.0009)*	*(0.0013)*	*(0.0017)*
ln *Potato-Suitable Area* × *1850*	*0.095*	*0.069*	*0.060*	*0.0024*	*0.0022*	*0.0031*
	(0.015)	*(0.017)*	*(0.020)*	*(0.0011)*	*(0.0014)*	*(0.0017)*
ln *Potato-Suitable Area* × *1900*	*0.121*	*0.092*	*0.080*	*0.0118*	*0.0123*	*0.0100*
	(0.017)	*(0.021)*	*(0.024)*	*(0.0023)*	*(0.0024)*	*(0.0032)*
Baseline Controls (× Year fixed effects):						
ln *Old World Crops Area*	N	Y	Y	N	Y	Y
ln *Elevation*	N	N	Y	N	N	Y
ln *Ruggedness*	N	N	Y	N	N	Y
ln *Tropical Area*	N	N	Y	N	N	Y
Observations	1552	1552	1552	1552	1552	1552
R-squared	0.99	0.99	0.99	0.42	0.42	0.46
F Stat for Joint Signficance 1750–1900	17.88	13.60	4.20	8.02	8.82	4.89



(a) ln Total Population

TABLE III
THE IMPACT OF THE POTATO WITH ALTERNATIVE CUT-OFFS

	Placebo Treatment Periods											
	1200–1500; Post = 1400, 1500		1300–1600; Post = 1500, 1600		1400–1700; Post = 1600, 1700		1500–1800; Post = 1700, 1800		1600–1900; Post = 1800, 1850, 1900		1600–1900; Post = 1750, 1800, 1850, 1900	
	ln pop.	City share	ln pop.	City share	ln pop.	City share	ln pop.	City share	ln pop.	City share	ln pop.	City share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln *Potato*	–0.002	0.0008	0.001	–0.0008	0.001	0.0002	0.006	0.0014	0.033	0.0038	0.028	0.0030
Area × Post	(0.006)	(0.0012)	(0.005)	(0.0017)	(0.006)	(0.0016)	(0.007)	(0.0017)	(0.011)	(0.0015)	(0.009)	(0.0017)
Observations	516	516	518	518	520	520	650	650	780	780	780	780
R-squared	0.99	0.56	0.99	0.56	0.99	0.59	0.99	0.59	0.99	0.62	0.99	0.62

TABLE IV

THE IMPACT OF THE POTATO: BASELINE ESTIMATES

	Dependent Variable									
	ln total population					City population share				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln *Potato Area* × Post	0.059	0.044	0.032	0.034	0.043	0.0044	0.0046	0.0036	0.0039	0.0039
	(0.009)	(0.011)	(0.012)	(0.011)	(0.014)	(0.0009)	(0.0009)	(0.0012)	(0.0011)	(0.0011)
Baseline Controls (× Year fixed effects):										
ln *Old World Crops Area*	N	Y	Y	N	Y	N	Y	Y	N	Y
ln *Elevation*	N	N	Y	Y	Y	N	N	Y	Y	Y
ln *Ruggedness*	N	N	Y	Y	Y	N	N	Y	Y	Y
ln *Tropical Area*	N	N	Y	Y	Y	N	N	Y	Y	Y
Other Controls (× Year fixed effects):										
ln *All Crops Area*	N	N	N	Y	N	N	N	N	Y	N
ln *Maize Area*	N	N	N	N	Y	N	N	N	N	Y
ln *Silage Maize Area*	N	N	N	N	Y	N	N	N	N	Y
ln *Sweet Potatoes Area*	N	N	N	N	Y	N	N	N	N	Y
ln *Cassava Area*	N	N	N	N	Y	N	N	N	N	Y
Observations	1552	1552	1552	1552	1552	1552	1552	1552	1552	1552
R-squared	0.99	0.99	0.99	0.99	0.99	0.38	0.39	0.44	0.44	0.48

To illustrate the magnitudes of our estimates, we perform a simple calculation that measures how much of the observed increase in log population and urbanization (of the Old World) between 1700 and 1900 can be explained by the introduction of the potato. Based on our data, the natural logarithm of Old World population increased by 0.90, from 20.21 in 1700 to 21.11 in 1900. Using the baseline estimate reported in column (3) of Table IV, we can calculate the counterfactual population in 1900 for each country if potatoes had not been introduced. This is equal to the observed log population in 1900 minus the estimated impact of potatoes, $\hat{\beta}$, multiplied by the natural log of the country's land area suitable for potato cultivation, i.e., $\ln \mathit{Total\,Population}_{i,1900}$— $\hat{\beta} \cdot \ln \mathit{Potato\,Area}_i$. We then aggregate all countries' counterfactual populations and calculate a counterfactual measure of Old World log population. According to the calculations, the counterfactual log population in 1900 would have been 20.87 (rather than 21.11), and the increase would have been 0.66 (rather than 0.90). Therefore, the increase would have only been 74 percent (0.66/0.90) of the observed increase if potatoes had not been introduced. In other words, according to the estimates, the introduction of the potato explains 26 percent of the observed increase in Old World population between 1700 and 1900.

- 当处理组个体接受处理时间不一致时，直接构造表示接受处理的虚拟变量。

$$Y_{it} = \tau D_{it} + \mathbf{X}'_{it}\gamma + u_i + \eta_t + \varepsilon_{it}$$

$$D_{it} = \begin{cases} 1, & \text{个体 } i \text{ 在第 } t \text{ 期接受处理} \\ 0, & \text{其它情形} \end{cases}$$

  - 事件研究法 (event-study specification)

$$Y_{it} = \gamma_{L_d} \cdot D_i \cdot \mathbb{1}(t - T_i^* \leq L_d) + \sum_{l=-L_d+1}^{-2} \gamma_l \cdot D_i \cdot \mathbb{1}(t - T_i^* = l)$$
$$+ \sum_{l=0}^{L_g-1} \beta_l \cdot D_i \cdot \mathbb{1}(t - T_i^* = l) + \beta_{L_g} \cdot D_i \cdot \mathbb{1}(t - T_i^* \geq L_g)$$
$$+ u_i + \eta_t + \varepsilon_{it}$$

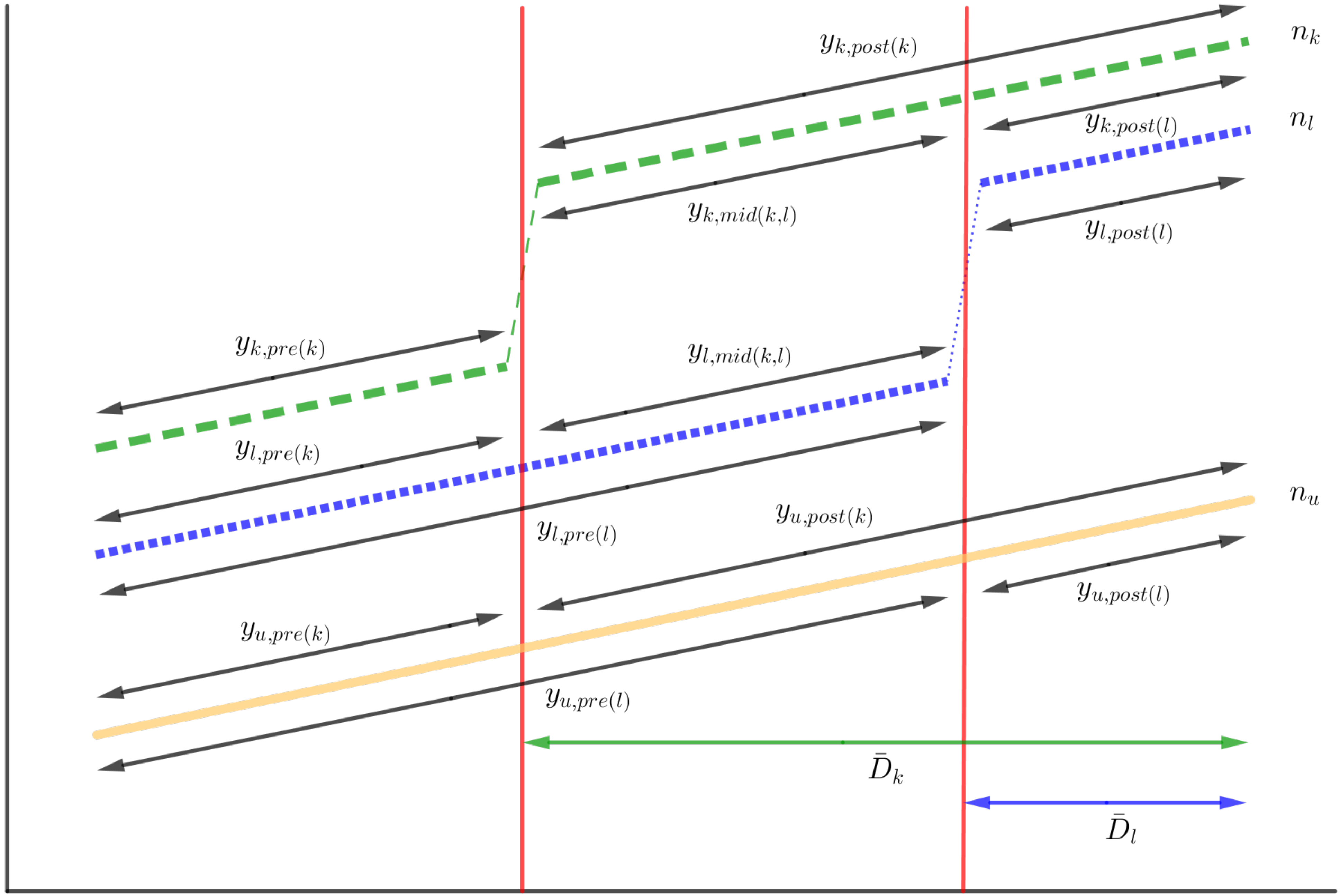
– Stacked DID：假定存在 K 个处理时点组，生成 K 个数据集，在每个数据集内构造 $\texttt{Treated}_{ik}$ 变量，当前开始接受处理的个体取 1，尚未接受处理的个体取 0，然后将这 K 个数据集合并起来（在合并数据集内一些个体重复出现了多次，标准误需要聚类到 K 个数据集层面）。

$$
\begin{aligned}
Y_{ikt} = u_i + \eta_t + \beta_0 \cdot \texttt{Treated}_{ik} + \sum_{l=-L_d}^{l=L_g} \delta_l \cdot T_{kt}^l \\
+ \beta \cdot \left(\texttt{Treated}_{ik} \times \texttt{Post}_{kt}\right) + \varepsilon_{ikt}
\end{aligned}
$$

$$
\begin{aligned}
Y_{ikt} = u_i + \eta_t + \beta_0 \cdot \texttt{Treated}_{ik} + \sum_l \delta_l \cdot T_{kt}^l \\
+ \sum_l \beta_l \cdot \left(\texttt{Treated}_{ik} \times T_{kt}^l\right) + \varepsilon_{ikt}
\end{aligned}
$$

示例 23. 办事处的关闭与残障险的申请 (Deshpande and Li, 2019, *AEJ: Policy*).

- 讨论混淆因素。如果存在可能的混淆事件，将其作为被解释变量进行回归，看是否存在显著的 DID 效应；如果存在可能影响处理进度 (timing) 的前在变量 (preexisting variable)，看它和处理进度之间是否存在显著相关。如是，则要考虑控制起来。
- 所有个体可以分为 $(K+1)$ 组，其中一组未经处理，其余 K 组处理时点各不相同。则 K 组之间可形成 $K(K-1)$ 种基本 DID 比较，其中一半以晚处理组为控制组，另一半以早处理组为控制组；K 组与未经处理组之间可形成 K 种基本 DID 比较。最终的估计量是 K^2 个基本 DID 估计量的加权平均。


$y_{k,post(k)}$
n_k
$y_{k,post(l)}$
n_l
$y_{k,mid(k,l)}$
$y_{l,post(l)}$
$y_{k,pre(k)}$
$y_{l,mid(k,l)}$
$y_{l,pre(k)}$
$y_{l,pre(l)}$
$y_{u,post(k)}$
n_u
$y_{u,post(l)}$
$y_{u,pre(k)}$
$y_{u,pre(l)}$
$\bar{D}_k$
$\bar{D}_l$

- 当处理组和控制组在处理前后存在成分变化 (compositional changes) 时，表明政策的内生性很强，此时平行趋势假定尤其可疑，要慎重使用 DID。
- 因果效应的异质性。

DID 与匹配的结合

- 先差分，然后对差分结果进行非参数匹配估计。

  示例 24. 排污权交易的环境后果 (Fowlie et al., 2012, *AER*).
- 先通过匹配方法构造控制组，然后进行参数估计。
  - $1:1$ 匹配的样本，估计交互项模型（无需再控制 X）。

    示例 25. 医院兼并对竞争的影响 (Schmitt, 2018, *AEJ: Policy*).
  - $1:m$ 匹配的样本，估计用匹配权重加权的交互项模型（无需再控制 X）。

    示例 26. 厂网分离对竞争的影响 (Cicala, 2015, *AER*).

Table 4—Average Treatment Effect using Nearest Neighbors Matching

	Levels	Logs	RECLAIM facilities	Controls
Panel A. Change in NO_x emissions between periods 1 and 4				
OLS	−32.58**	−0.30***	212	1,222
	(13.77)	(0.10)		
Nearest neighbor matching (base specification)	−20.59***	−0.25***	212	1,222
	(7.63)	(0.09)		
Nearest neighbor matching (alternative specification)	−18.12	−0.11	211	1,191
	(11.51)	(0.08)		
Nearest neighbor matching (restricted sample)	−14.16**	−0.20**	199	1,222
	(6.86)	(0.09)		
Panel B. Change in NO_x emissions between periods 2 and 3				
OLS	−6.84	−0.22***	255	1,577
	(6.65)	(0.04)		
Nearest neighbor matching (base specification)	−8.29**	−0.26***	255	1,577
	(3.85)	(0.06)		
Nearest neighbor matching (alternative specification)	−6.18	−0.16***	252	1,493
	(5.06)	(0.06)		
Nearest neighbor matching (unrestricted sample)	−6.37	−0.23***	268	1,577
	(4.57)	(0.06)		

Notes: We define periods as averages of positive emissions in two years: 1990 and 1993 (period 1); 1997–1998 (period 2); 2001–2002 (period 3); and 2004–2005 (period 4). All observations are from historic nonattainment counties. The OLS estimates control for average NO_x emissions during period 1 and four-digit SIC code indicator variables, with standard errors clustered by air basin. For all semiparametric matching, we match on the three closest neighbors with linear bias adjustment in levels and quadratic bias adjustment in logs. The baseline nearest neighbor matching model matches on historic emissions and exactly on four-digit SIC codes. In the alternative specification, industry-specific emissions quartile indicators are added to the exact matching variables; predetermined demographic characteristics (race and income) are added to the matching variables. Panel A's restricted sample omits 13 facilities removed from the program in 2001. Panel B's unrestricted sample includes these facilities. For the log specifications, emissions differences are defined as $\ln(\text{EmitX} + 1) - \ln(\text{Emit1} + 1)$, and all matching is on $\ln(\text{Emit1} + 1)$. Standard errors are reported in parentheses.

示例 25. Schmitt (2018)

- 本文的研究对象是美国的医院和医院系统 (multihospital systems)。
- 在同一个市场上竞争的厂商，经过兼并重组，市场集中度增加，市场价格会升高。但两个厂商如果同时在两个市场上竞争，在市场 2 中的兼并重组，并不会增加市场 1 的集中度，却也会导致市场 1 的价格升高。理论上称之为多市场接触 (multimarket contact, MMC) 的互相宽容 (mutual forebearance) 假说：为了避免单个市场上的竞争引发全面竞争。
- 已有研究往往采用控制市场固定效应的方式，通过同一市场内随时间的变动性 (within-market variation) 来识别多市场接触的因果效应。如下图，两个市场的 MMC 在事后都发生了变化，但市场 2 的变动可能是内生的，本文只使用市场 1 的变动来识别（称之为市场外兼并，out-of-market consolidation）。

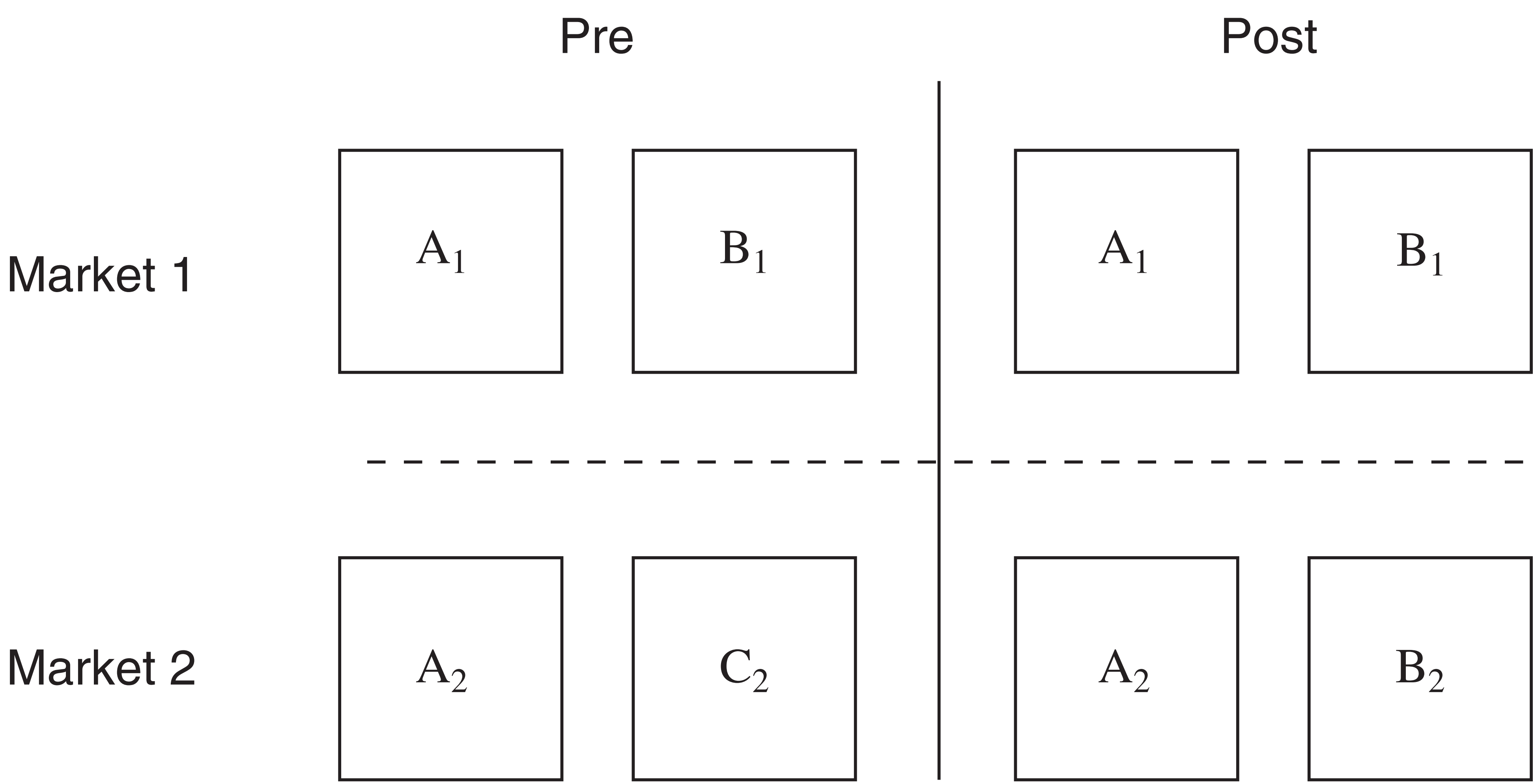


FIGURE 1. TWO MARKET, TWO HOSPITAL EXAMPLE

- 本文使用的市场定义是一家医院方圆 20 英里。
- 样本期为 1996-2014 年，兼并发生在 2000-2010 年。接受多次处理的医院，将处理时点设定为第一次。
- 计量方程：

$$\ln\left(\texttt{price}_{ht}\right)=\alpha_h+\gamma_t+\lambda\cdot\mathbf{1}\left(t\geq\tau_h\right)+\mathbf{X}_{ht}\boldsymbol{\beta}+\varepsilon_{ht}$$

$$\begin{aligned}\ln\left(\texttt{price}_{ht}\right)=\ &\alpha_h+\gamma_t+\lambda_{-4}\cdot\mathbf{1}\left(t\leq\tau_h-4\right)\\&+\lambda_{-3}\cdot\mathbf{1}\left(t=\tau_h-3\right)+\lambda_{-2}\cdot\mathbf{1}\left(t=\tau_h-2\right)\\&+\sum_{k=0}^{3}\lambda_k\cdot\mathbf{1}\left(t=\tau_h+k\right)+\lambda_4\cdot\mathbf{1}\left(t\geq\tau_h+4\right)\\&+\mathbf{X}_{ht}\boldsymbol{\beta}+\varepsilon_{ht}\end{aligned}$$

- 控制变量：病例类型 (case mix index, CMI)、公费医保病人比例、床位数、是否营利性医院、市场集中度（HHI）、同系统医院数量。

– 构造匹配样本

- ▷ 1:1，无放回，准确匹配城市虚拟变量和地区，近似匹配马氏距离，对数优势比 (log odds ratio) 不大于 0.2 倍标准差。
- ▷ 协变量均取 1998 年值。
- ▷ 计算马氏距离的协变量：价格、出院总病人数、CMI、公费医保病人比例、床位数、HHI、同系统医院数量。
- ▷ 计算倾向得分的协变量：上述协变量 + 是否营利性医院。

TABLE 2—DIFFERENCE-IN-DIFFERENCES MMC REGRESSIONS

	All	Control group: All	Control group: Matched	Matched
	(1)	(2)	(3)	(4)
Panel A. Post only (equation (2))				
Post ($t \geq \tau_h$)	0.064	0.070	0.060	0.065
	(0.017)	(0.018)	(0.019)	(0.019)
Control variables		✓		✓
Hospitals	2,950	2,943	694	692
Observations	39,374	39,080	10,645	10,535
R^2	0.766	0.770	0.708	0.713
Panel B. Leads and lags (equation (3))				
$t \leq \tau_h - 4$	−0.015	−0.016	−0.011	−0.013
	(0.028)	(0.028)	(0.029)	(0.029)
$t = \tau_h - 3$	−0.021	−0.026	−0.012	−0.019
	(0.026)	(0.026)	(0.027)	(0.026)
$t = \tau_h - 2$	0.009	0.006	0.011	0.009
	(0.021)	(0.021)	(0.021)	(0.021)
$t = \tau_h - 1$	0	0	0	0
	–	–	–	–
$t = \tau_h$	0.043	0.042	0.044	0.043
	(0.024)	(0.024)	(0.025)	(0.025)
$t = \tau_h + 1$	0.069	0.068	0.067	0.067
	(0.022)	(0.022)	(0.022)	(0.023)
$t = \tau_h + 2$	0.067	0.068	0.067	0.068
	(0.024)	(0.024)	(0.024)	(0.025)
$t = \tau_h + 3$	0.062	0.062	0.062	0.062
	(0.024)	(0.024)	(0.025)	(0.025)
$t \geq \tau_h + 4$	0.050	0.059	0.050	0.056
	(0.025)	(0.026)	(0.027)	(0.027)
Control variables		✓		✓
Hospitals	2,950	2,943	694	692
Observations	39,374	39,080	10,645	10,535
R^2	0.767	0.770	0.709	0.714

- 当存在多个处理时点时，尽量采用不随时间变化的协变量进行匹配。
- 尽管很少见，但理论上可以对每个处理组个体根据时变协变量的处理前取值进行匹配。对于匹配后样本如果要使用事件研究法，鉴于同一个控制组个体可能匹配到不同的处理组个体，可以考虑 expand 多个重复观测值，然后根据匹配到的处理组个体进行相应的时间中心化。

合成控制方法

- 合成控制方法 (synthetic control method, Abadie and Gardeazabal, 2003, *AER*; Abadie et al., 2010, *JASA*; 2015, *AJPS*) 被认为是过去十五年来项目评估 (program evaluation) 领域最重要的创新 (Athey and Imbens, 2017, *JEP*).
- 适用情形：单个处理组个体，较长的处理前时期。
- 基本思想：通过控制组个体的凸组合来构造反事实。
  - $i = 0, 1, \ldots, N$, $i = 0$ 为处理组。
  - $t = 1, \ldots, T_0, T_0 + 1, \ldots, T$, $T_0 + 1$ 之后接受处理。
- 寻找最优权重：用控制组的处理前结果拟合处理组的处理前结果。

$$\mathbf{w} = \arg\min_{\mathbf{w}} \sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{i=1}^{N} w_i \cdot Y_{it} \right)^2$$

- 也可以拟合协变量 X，可能包括处理前个体特征或处理前结果。

$$\mathbf{w}(\Omega)=\underset{\mathbf{w}}{\operatorname{argmin}}\left(X_0-\sum_{i=1}^{N} w_i\cdot X_i\right)'\Omega\left(X_0-\sum_{i=1}^{N} w_i\cdot X_i\right)$$

$$s.t.\ w_i\geq 0\ \forall i,\ \sum_{i} w_i=1$$

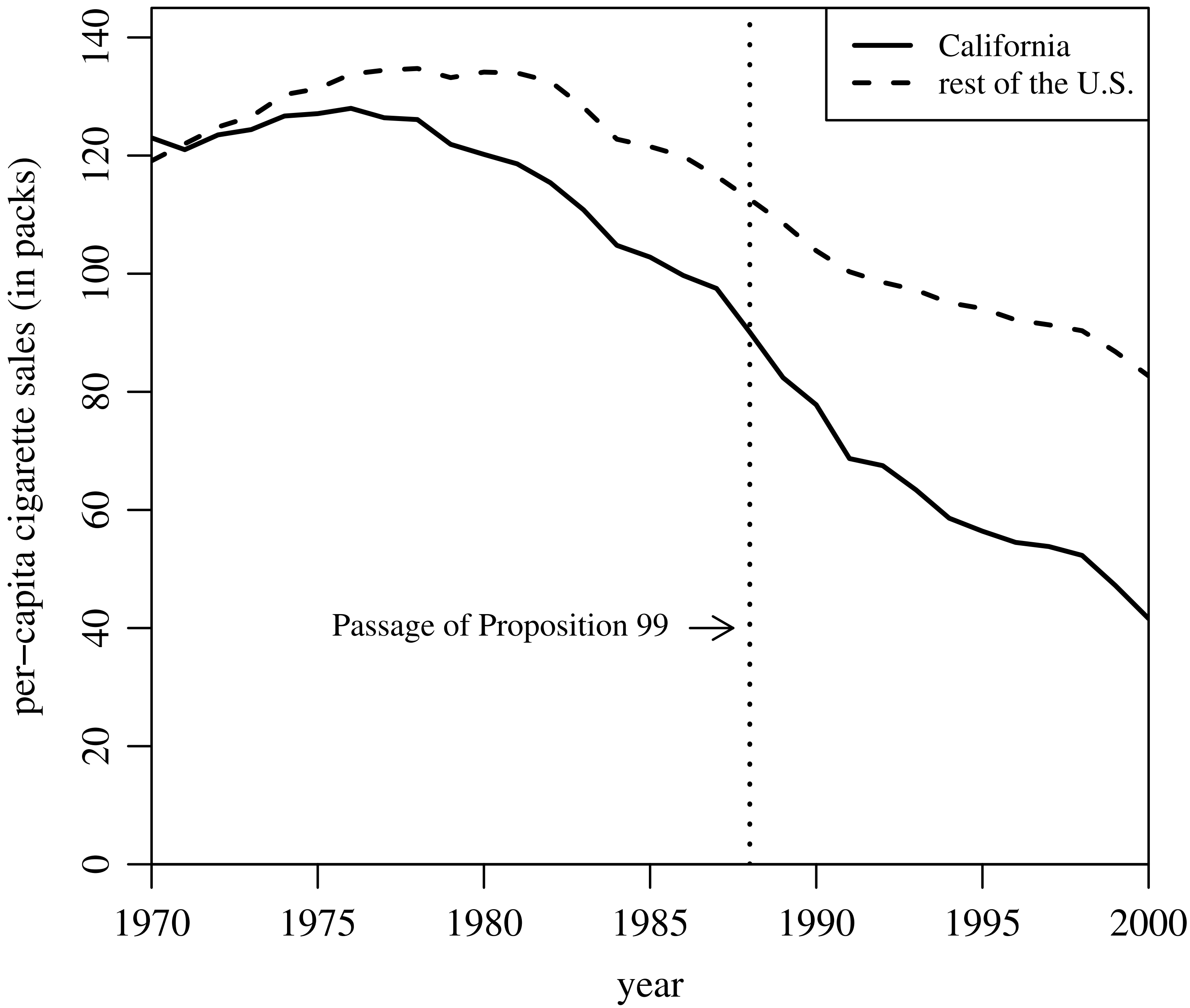
$$\Omega^*=\arg\min_{\Omega}\ \sum_{t=1}^{T_0}\left(Y_{0t}-\sum_{i=1}^{N} w_i(\Omega)\cdot Y_{it}\right)^2$$

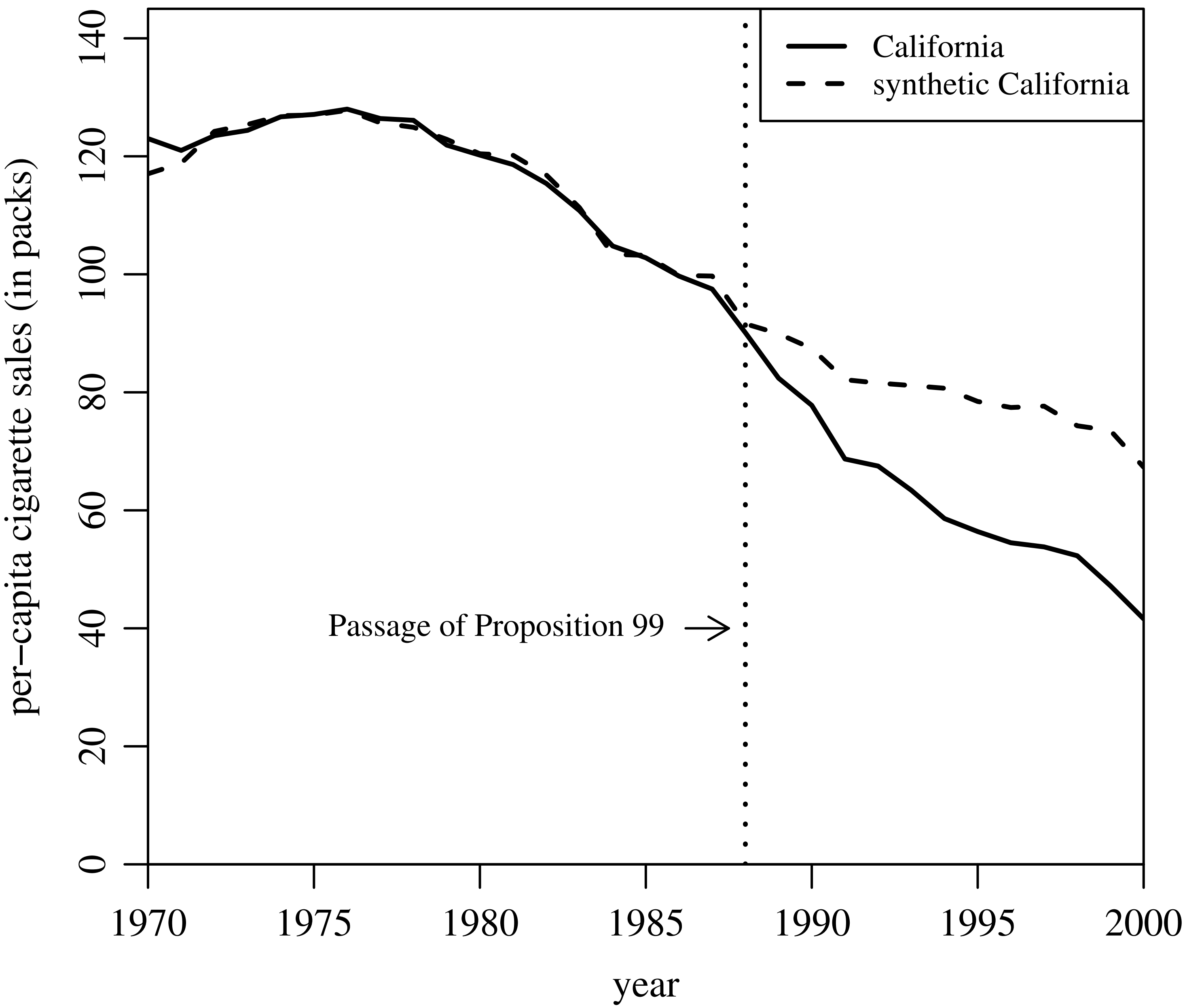
- 处理组的处理效应

$$\hat{\tau}_t^{\text{SCM}}=Y_{0t}-\sum_{i=1}^{N}\hat{w}_i\left(\Omega^*\right)\cdot Y_{it},\ \forall t\geq T_0+1$$

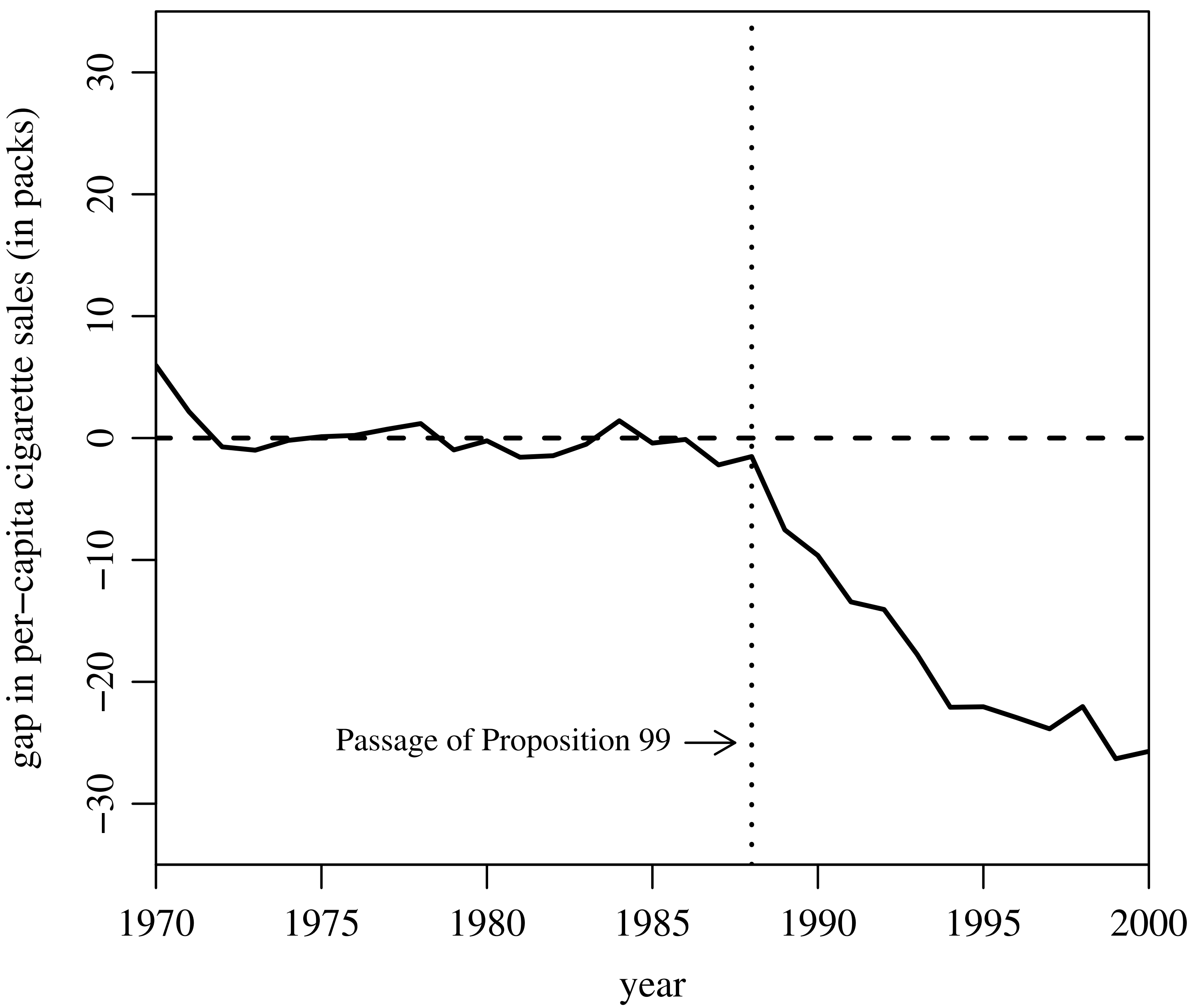
- `STATA` 实现：`synth`, `synth_runner`

示例 27. 加州的控烟条例 (Abadie et al., 2010, *JASA*).



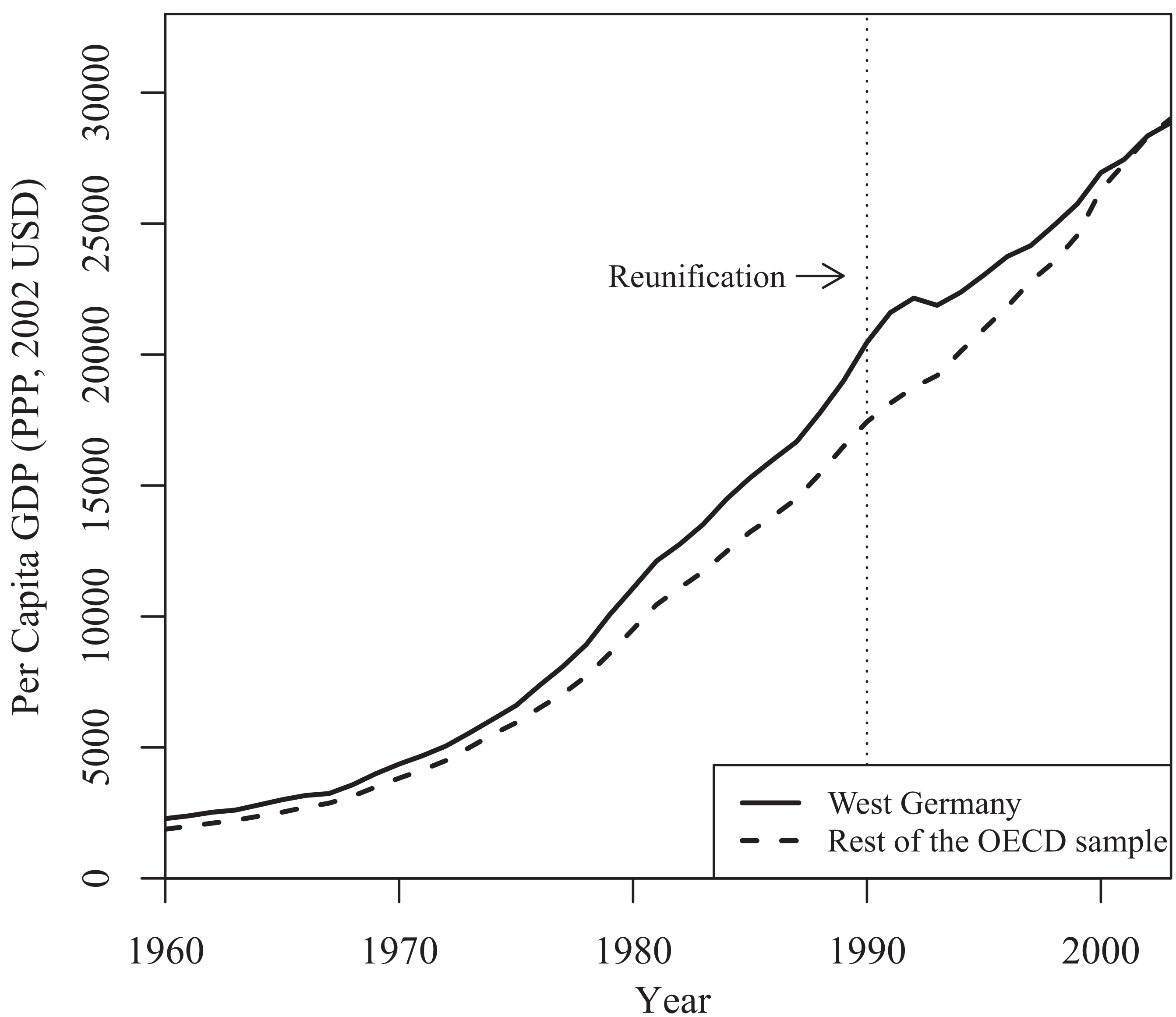


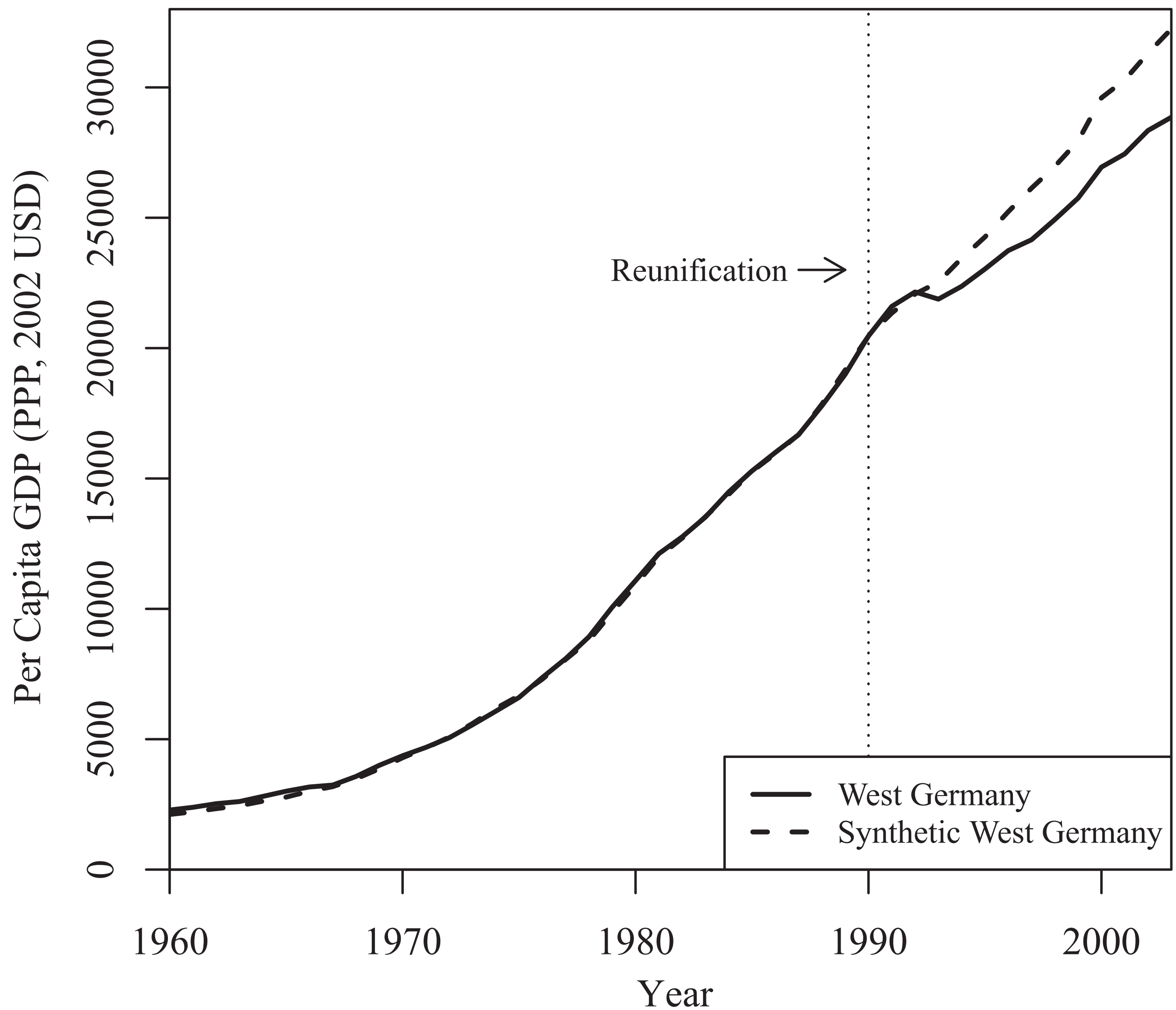
per−capita cigarette sales (in packs)
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000
year
California
synthetic California
Passage of Proposition 99



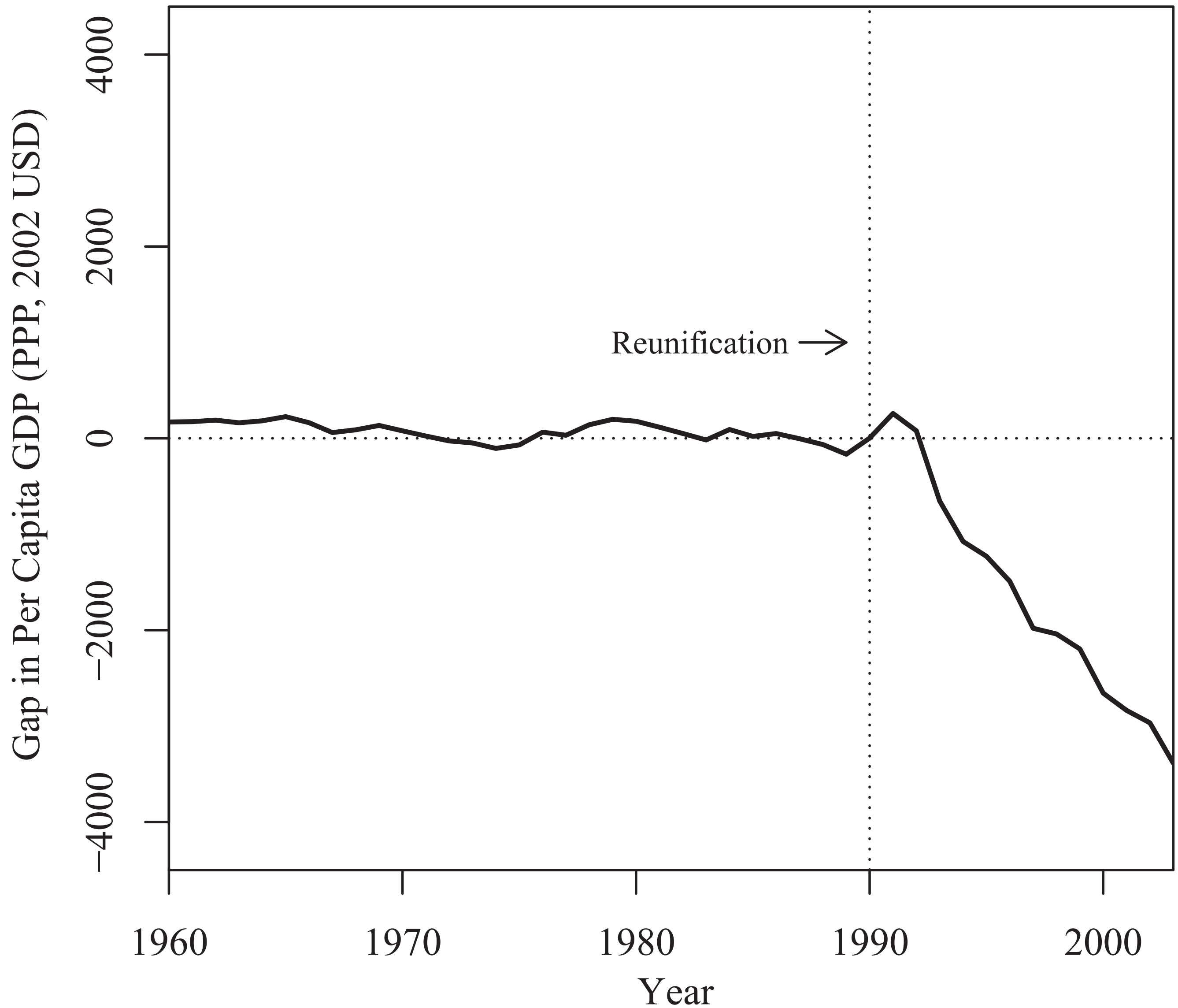
gap in per−capita cigarette sales (in packs)
30
20
10
0
−10
−20
−30
Passage of Proposition 99
1970
1975
1980
1985
1990
1995
2000
year

示例 28. 两德统一的经济后果 (Abadie et al., 2015, *AJPS*).



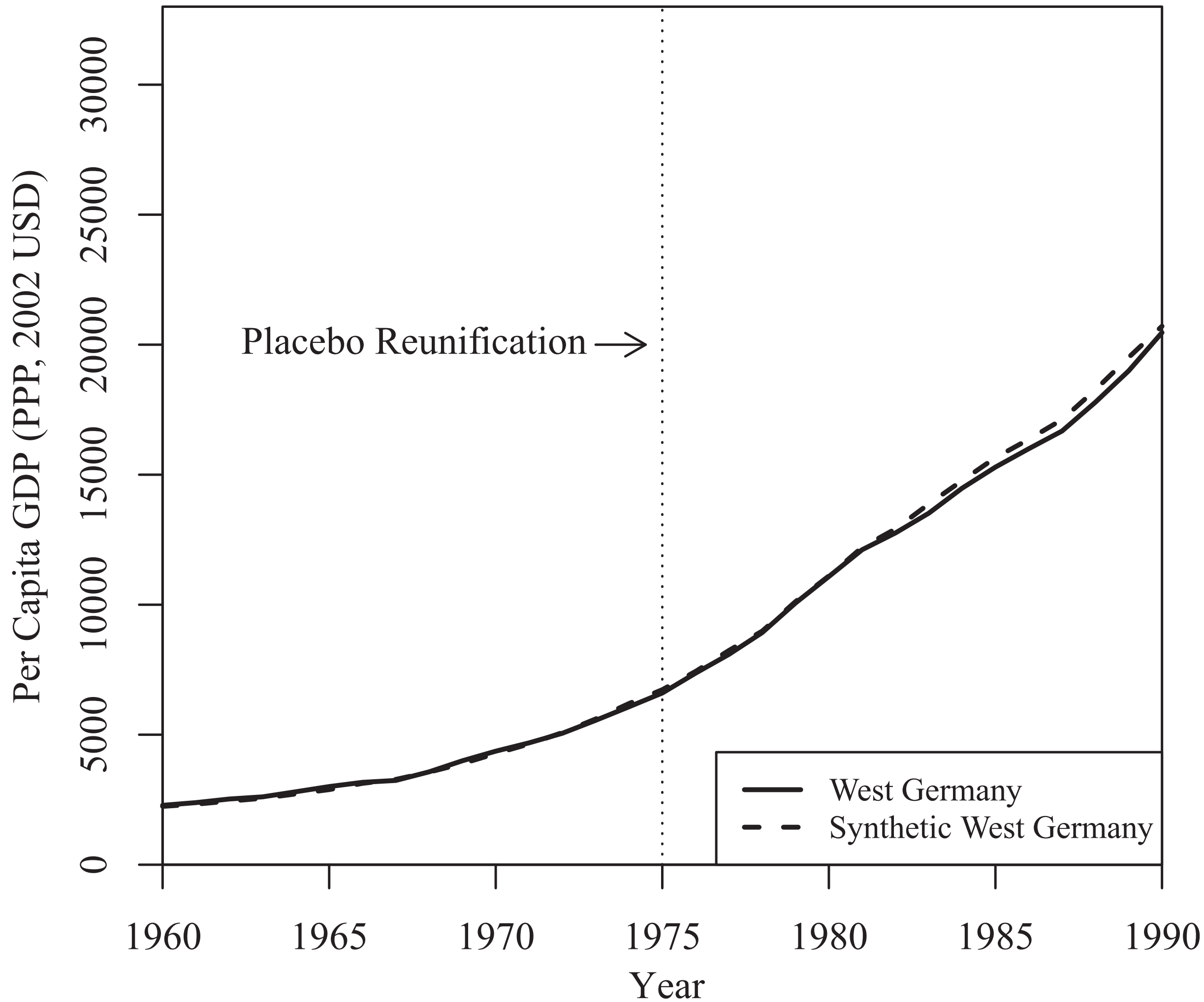


Per Capita GDP (PPP, 2002 USD)
0
5000
10000
15000
20000
25000
30000
Reunification
West Germany
Synthetic West Germany
1960
1970
1980
1990
2000
Year

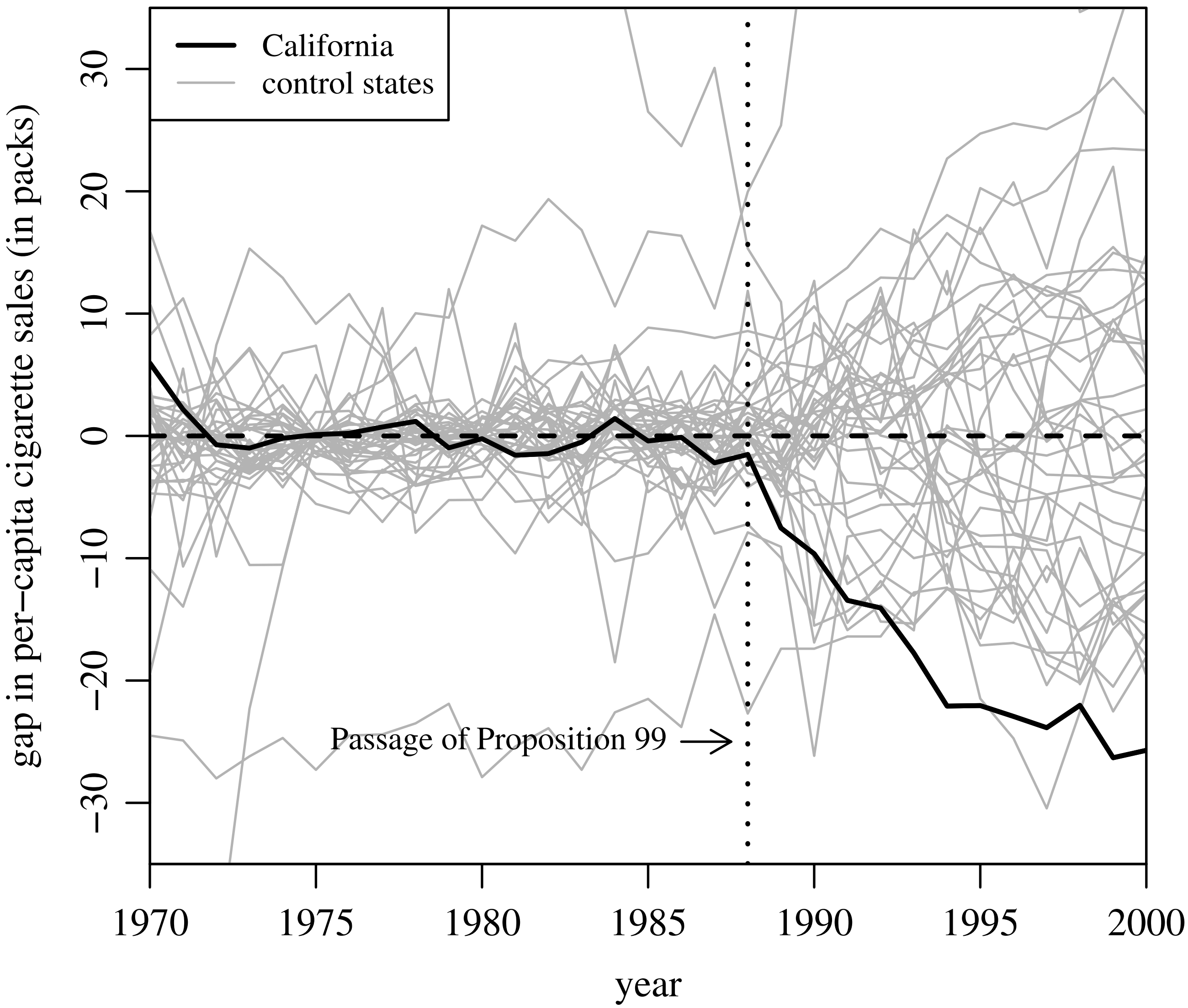


Gap in Per Capita GDP (PPP, 2002 USD)
4000
2000
0
−2000
−4000
Reunification
1960
1970
1980
1990
2000
Year

- 非传统的统计推断：安慰剂检验
  - 对处理前时期进行同样的估计。



– 对控制组个体进行同样的估计，根据安慰剂效应的经验分布计算 p 值。



- 多个处理组个体，以 Acemoglu et al (2016, *JFE*) 为例。

$$\{w_j^{i*}\}_{j\in\text{Control group}}$$

$$= \arg\min \sum_{t\in\text{Estimation window}} \left[R_{it} - \sum_j w_j^i R_{jt}\right]^2, \ \forall i \in \text{Treatment group}$$

$$\hat{R}_{it} = \sum_j w_j^{i*} R_{jt}$$

$$\hat{\phi}(\tau, k) = \frac{\sum_{i\in\text{Treatment group}} \frac{\sum_{t=\tau}^{k} R_{it} - \hat{R}_{it}}{\hat{\sigma}_i}}{\sum_{i\in\text{Treatment group}} \frac{1}{\hat{\sigma}_i}}$$

$$\hat{\sigma}_i = \sqrt{\frac{\sum_{t\in\text{Estimation window}} \left(R_{it} - \hat{R}_{it}\right)^2}{T}}$$

面板数据方法

- Hsiao, Ching, and Wan (2012, *JAE*) 开发了一种类似的方法，用 $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Nt})'$ 来预测处理后的 Y_{0t}^0.

$$Y_{0t} = w_0 + \mathbf{Y}_t'\mathbf{w} + \varepsilon_{0t},\ t = 1, \ldots, T_0$$

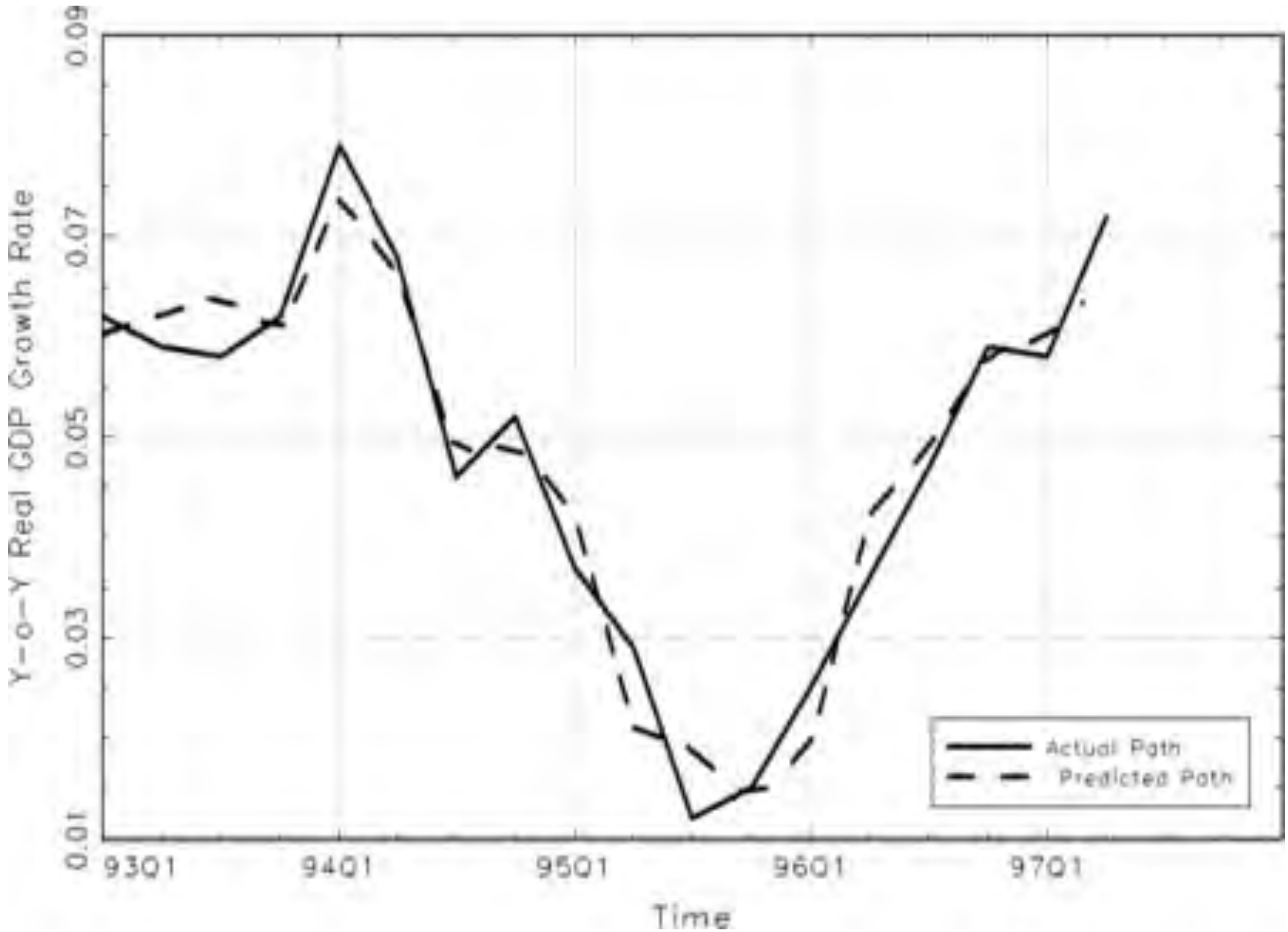
$$\hat{Y}_{0t}^0 = \hat{w}_0 + \mathbf{Y}_t'\hat{\mathbf{w}},\ t = T_0 + 1, \ldots, T$$

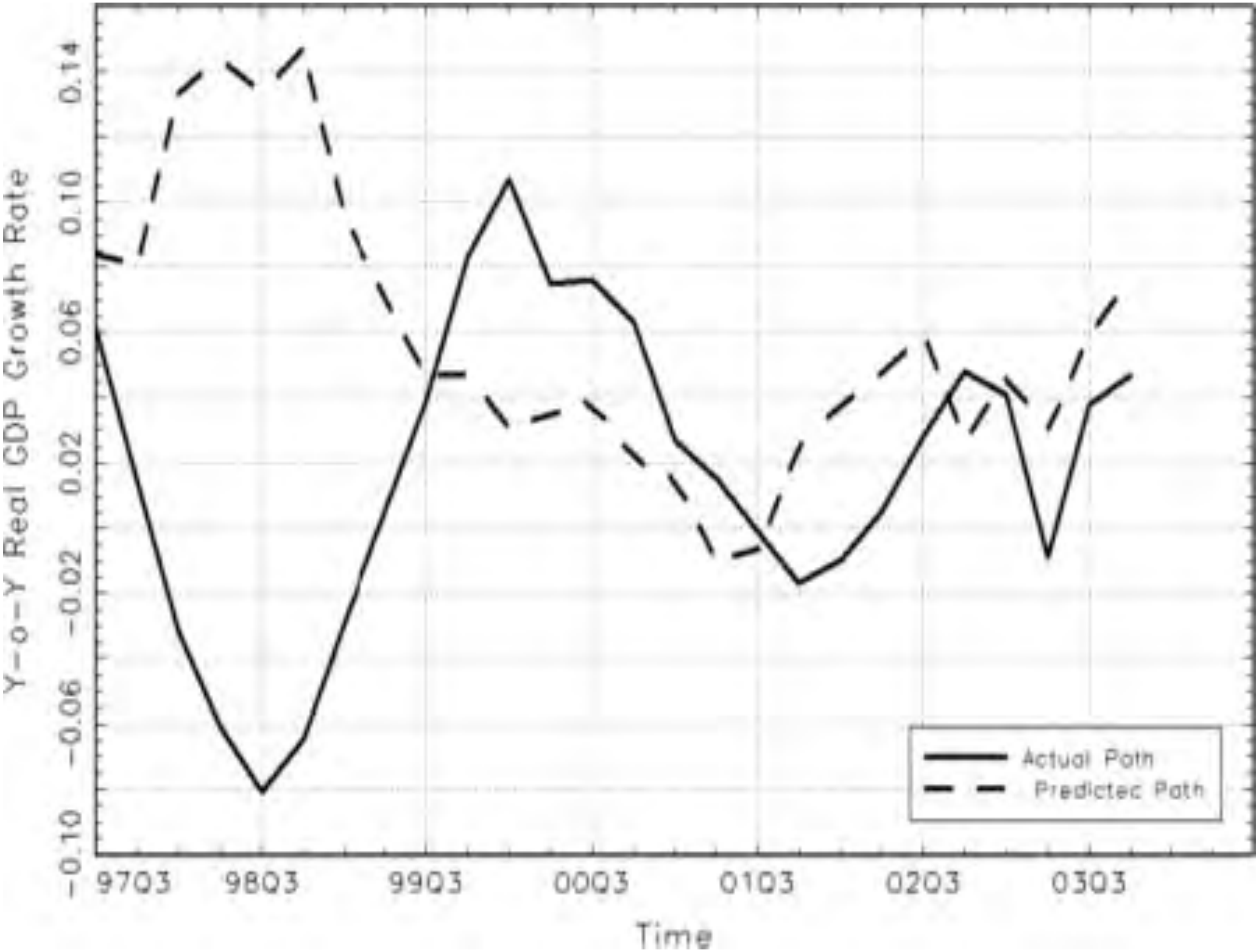
$$\hat{\tau}_t^{\text{PDA}} = Y_{0t} - \hat{Y}_{0t}^0,\ \forall t \geq T_0 + 1$$

- 控制组个体选择策略
  - HCW(2012) 建议，给定控制组个体数目，根据拟合优度大小来确定使用哪几个控制组个体；然后根据 AIC 或 BIC 等模型选择标准 (model selection criterion) 来确定最优的控制组个体数目。
  - Li and Bell (2017, *JOE*) 建议使用 LASSO 方法，

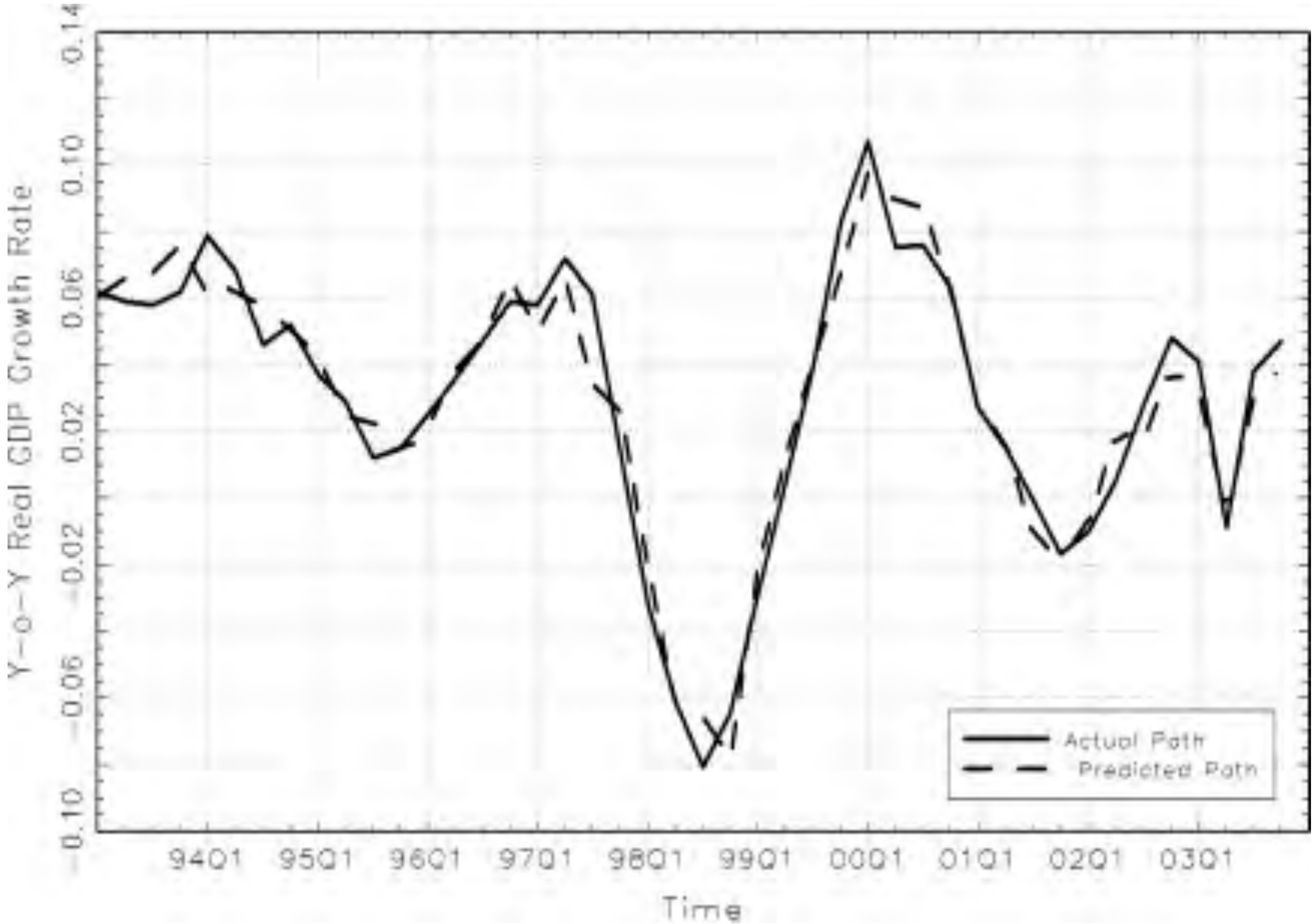
$$(w_0, \mathbf{w}')' = \arg\min \sum_{t=1}^{T_0} \left(Y_{0t} - w_0 - \mathbf{Y}_t'\mathbf{w}\right)^2 + \lambda \sum_{j=0}^{N} w_j$$

示例 29. 香港与大陆的政治融合和经济融合 (Hsiao et al., 2012, *JAE*).

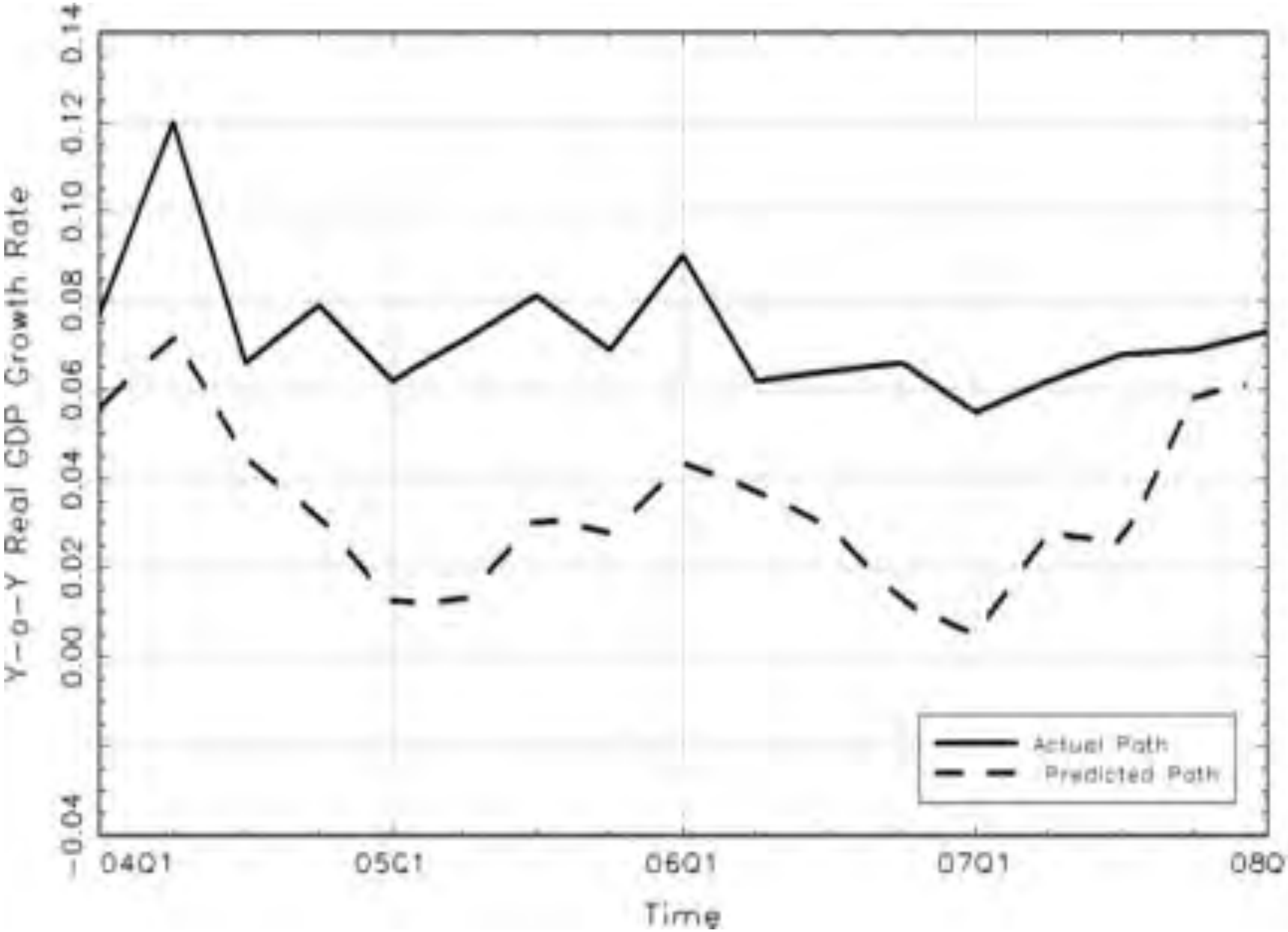




Y-o-Y Real GDP Growth Rate
0.14
0.10
0.06
0.02
-0.02
-0.06
-0.10
97Q3
98Q3
99Q3
00Q3
01Q3
02Q3
03Q3
Time
Actual Path
Predictec Path

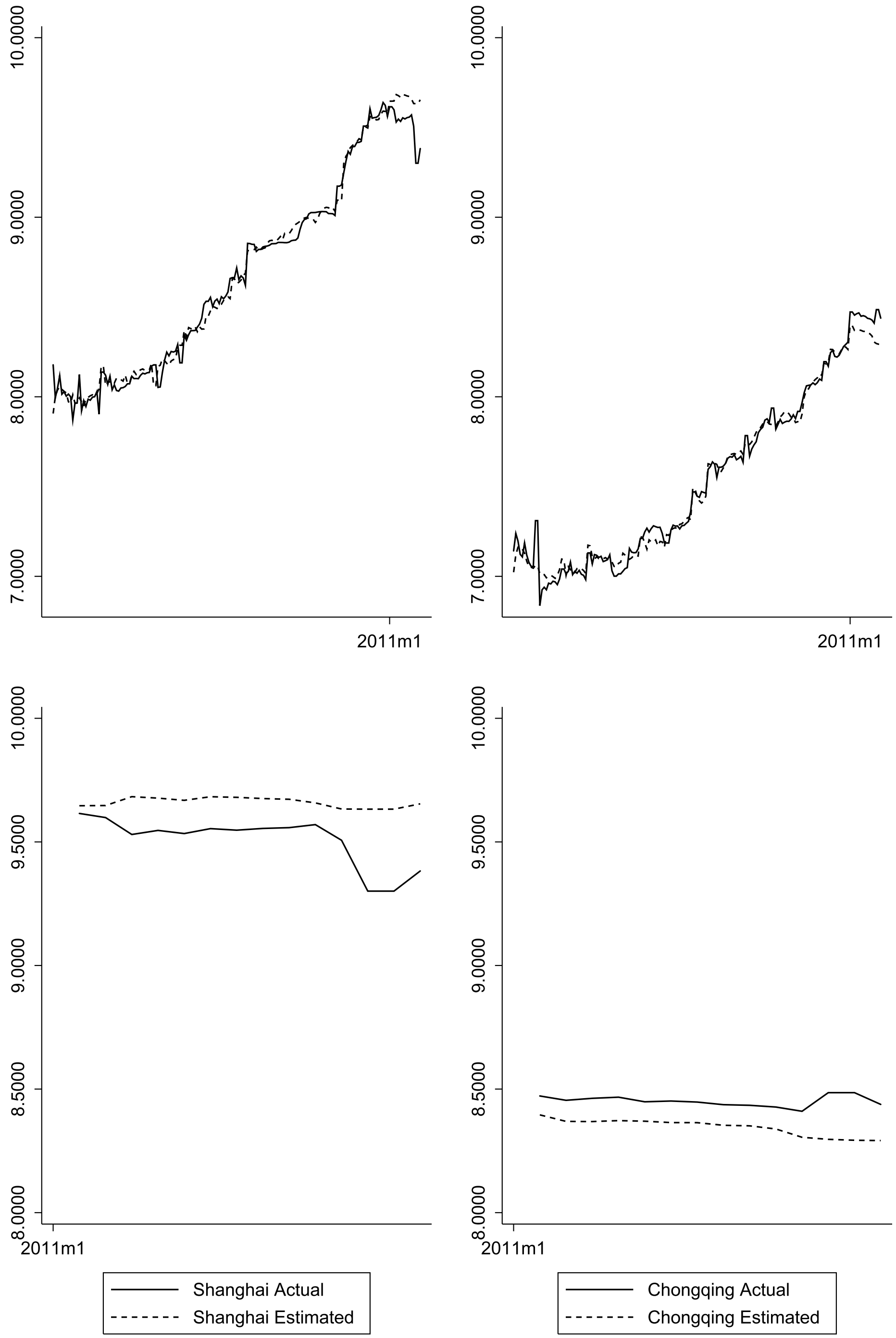


Y-o-Y Real GDP Growth Rate
0.14
0.10
0.06
0.02
-0.02
-0.06
-0.10
94Q1
95Q1
96Q1
97Q1
98Q1
99Q1
00Q1
01Q1
02Q1
03Q1
Time
Actual Path
Predicted Path


Y-o-Y Real GDP Growth Rate
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.04
04Q1
05Q1
06Q1
07Q1
08Q1
Time
Actual Path
Predicted Path

示例 30. 房产税与房价 (Bai et al., 2014, *JOE*).

Provinces/cities	Weights	SD	T	Weights	SD	T
Panel A: Shanghai as the treatment city						
	$R^2 = 0.9907$ F-stats $= 5299.69$			$R^2 = 0.9910$ F-stats $= 4145.84$		
Jiangsu	0.7503***	0.0684	10.96	0.6453***	0.0909	7.10
Zhejiang	0.2908***	0.0668	4.35	0.2944***	0.0650	4.53
Heilongjiang	0.1976***	0.0551	3.59	0.1633***	0.0565	2.89
Sichuan	−0.2350***	0.0388	−6.06	−0.2157***	0.0385	−5.61
Post-2008 dummy	0.1682***	0.0170	9.87	0.1753***	0.0175	10.00
Time trend				0.0010	0.0007	1.50
Panel B: Chongqing as the treatment city						
	$R^2 = 0.9795$ F-stats $= 4135.50$			$R^2 = 0.9837$ F-stats $= 3300.26$		
Jiangsu	0.3973***	0.0688	5.77	0.7441***	0.0834	8.92
Zhejiang	0.1806***	0.0630	2.87	0.2044***	0.0601	3.40
Beijing	0.1232*	0.0656	1.88	0.1071*	0.0629	1.70
Sichuan	0.1342*	0.0810	1.66	0.1195*	0.0706	1.69
Post-2008 dummy	0.0653***	0.0129	5.05	0.0623***	0.0132	4.73
Time trend				−0.0035***	0.0007	−4.95


10.0000
9.0000
8.0000
7.0000
2011m1
10.0000
9.0000
8.0000
7.0000
2011m1
10.0000
9.5000
9.0000
8.5000
8.0000
2011m1
10.0000
9.5000
9.0000
8.5000
8.0000
2011m1
Shanghai Actual
Shanghai Estimated
Chongqing Actual
Chongqing Estimated