# 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: July 4, 2020

# Lecture 6　面板数据的参数和非参数方法

## 6.1　面板数据固定效应模型

- 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，

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

此时真正隐含的关键假设仍然是“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)。

- 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。

本文使用省级(包括直辖市)党代会数据和地级市本级政府财政数据考察中国地方政府是否存在政治预算周期,设定了如下的估计模型:

$$g_{it} = \alpha_0 + \alpha_1 \mathrm{CCP4}_{it} + \alpha_2 \mathrm{CCP0}_{it} + \alpha_3 \mathrm{CCP1}_{it} + \alpha_4 \mathrm{CCP2}_{it} + X'\beta + \lambda_i + \mu_t + \varepsilon_{it}$$

其中,角标 $i$ 表示地级市,角标 $t$ 表示时间,被解释变量 $g_{it}$ 表示市本级政府一般预算支出的增长率,计算公式为:当年一般预算支出增长率 $=\ln$(当年一般预算支出/上年一般预算支出)。在后文中,由于研究需要,我们还会将被解释变量换为一般预算收入增长率等增长率变量,具体的被解释变量将在回归部分予以说明。$\mathrm{CCP4}_{it}$ 表示党代会召开前一年(本届党代会召开前一年即上一届党代会开完的第四年),$\mathrm{CCP0}_{it}$ 表示党代会召开当年,$\mathrm{CCP1}_{it}$ 和 $\mathrm{CCP2}_{it}$ 分别表示党代会召开后第一年和党代会召开后第二年。回归方程中没有加入的党代会召开后第三年($\mathrm{CCP3}_{it}$),以其为比较的基准。$X$ 为表示其他控制变量的矩阵,在地级

本文将采用1999年至2011年281个城市的面板数据，考察省级党代会的召开对土地出让面积的影响，基本的计量模型设定如下：

$$\ln land_{it} = \alpha_0 + \alpha_1 PPCpre2_{it} + \alpha_2 PPCpre1_{it} + \alpha_3 PPC_{it} + \alpha_4 PPCpost1_{it} + X_{it}\beta + \mu_i + \alpha_5 trend_t + \varepsilon_{it} \quad (9)$$

其中，$\ln land_{it}$为地级市 $i$ 在 $t$ 年国有土地出让总面积对数值。$PPCpre2_{it}$、$PPCpre1_{it}$、$PPC_{it}$和 $PPCpost1_{it}$是四个虚拟变量，用于刻画省党代会（Provincial Party Congress）的影响。如果省党代会在当年召开，$PPC_{it}=1$，否则取0；省党代会在未来两年召开，$PPCpre2_{it}=1$，否则取0；省党代会在未来一年召开，$PPCpre1_{it}=1$，否则取0；省党代会在去年召开，$PPCpost1_{it}=1$，否则取0。$X_{it}$是一组可能影响土地出让

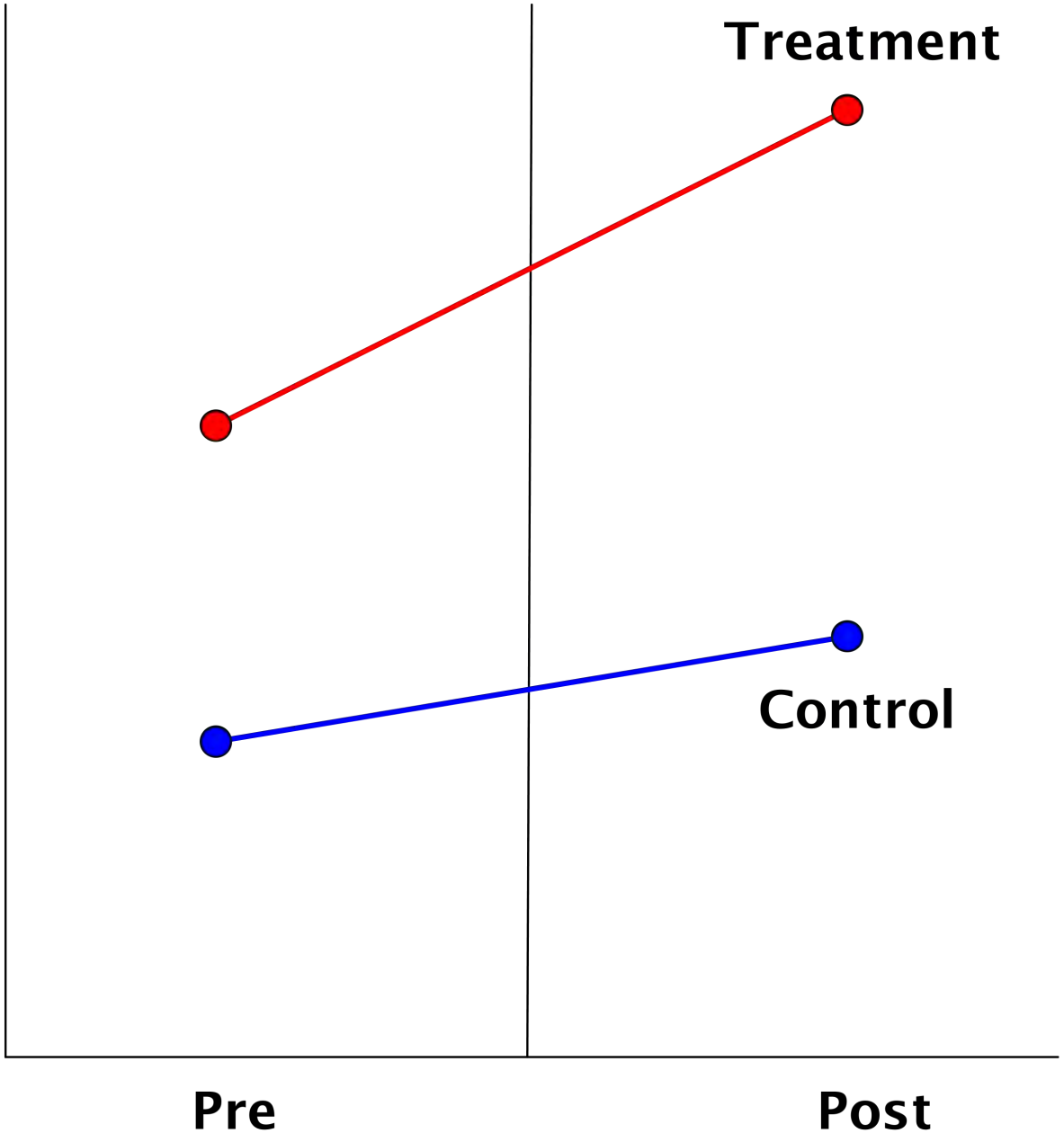
① 为了考察省党代会的影响，年份效应没有在回归模型中控制。原因在于如果在回归模型中同时放入省党代会虚拟变量和年份虚拟变量，我们无法区分省党代会的影响。因此，模型控制时间趋势，而没有像通常情况下控制年份虚拟变量。

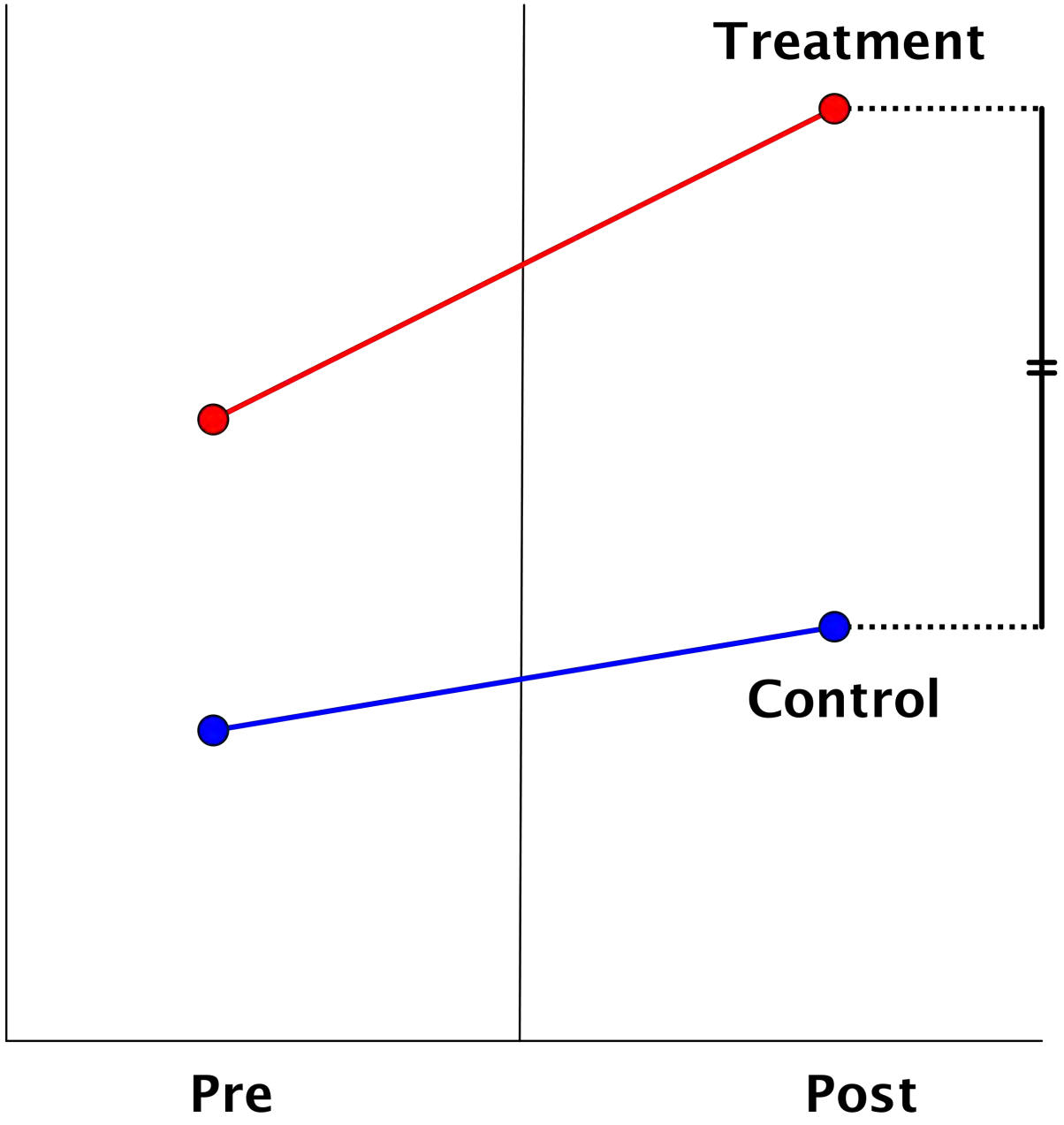
- 报告 $R^2$：可以报告 `xtreg` 的 within $R^2$, 也可以报告 `reghdfe` 的 $R^2$, 但不建议报告 `xtreg` 的 overall $R^2$.

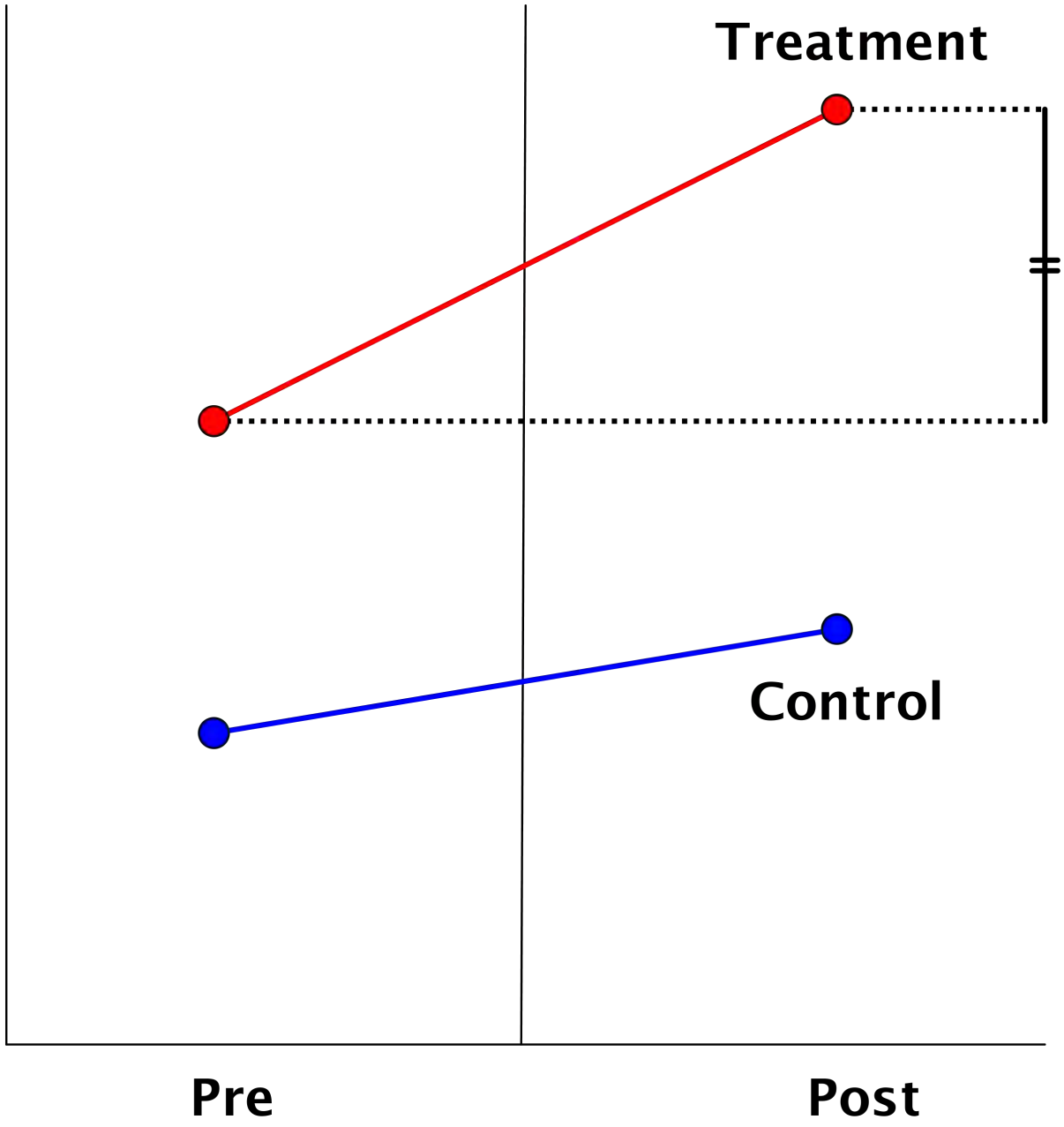
$$\begin{aligned}
\text{within } R^2 &= \rho^2(Y_{it} - \bar{Y}_i, \hat{\beta} D_{it} - \hat{\beta} \bar{D}_i) \\
\text{between } R^2 &= \rho^2(\bar{Y}_i, \hat{\beta} \bar{D}_i) \\
\text{overall } R^2 &= \rho^2(Y_{it}, \hat{\beta} D_{it})
\end{aligned}$$

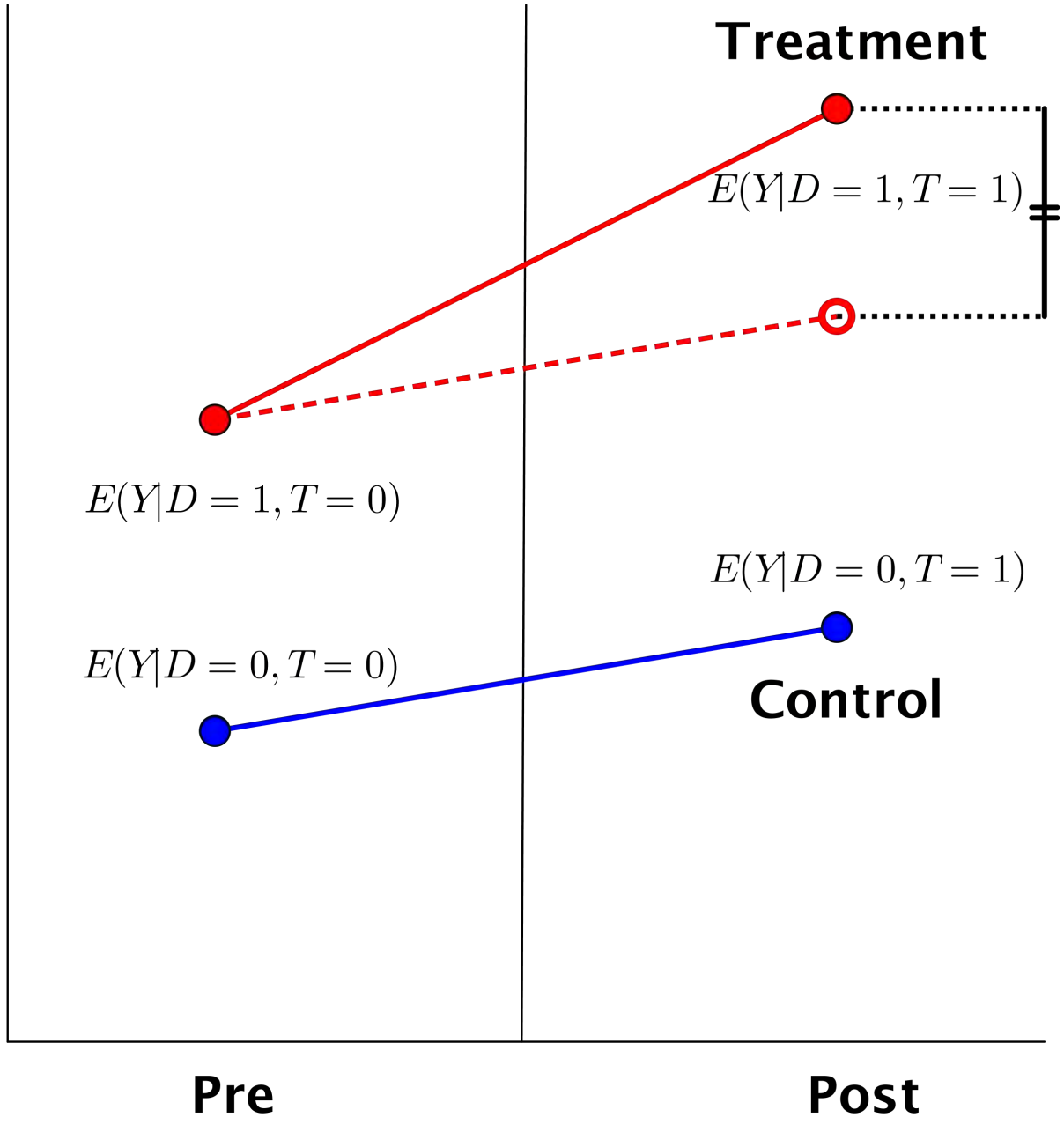
- 不能通过将解释变量替换成其滞后项来解决内生性问题，也不能使用解释变量的滞后项作为其工具变量。
- 对于非线性模型，消除固定效应的方法失效，此时要得到解释变量系数的一致估计需要限制性很强的假设，因此对于结构复杂的面板数据，研究者更多地采用线性概率模型。

# 6.2 双重差分








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?)
  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 $X$，$\Delta Y^0$ 均值独立于 $D$，即关于 $\Delta Y^0$ 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）。

# 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

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

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

- 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

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

- 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

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

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

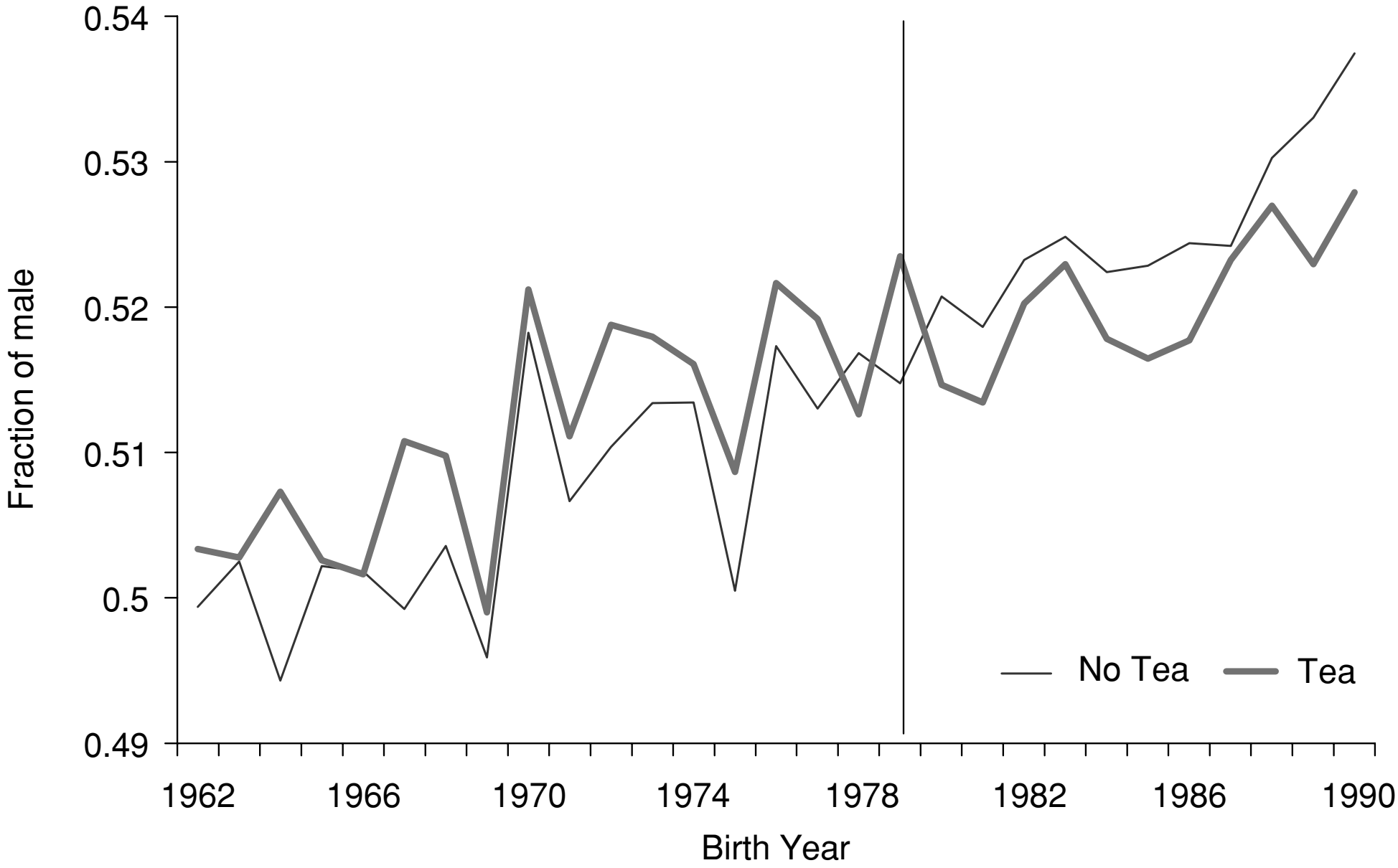
我们应该期望得到

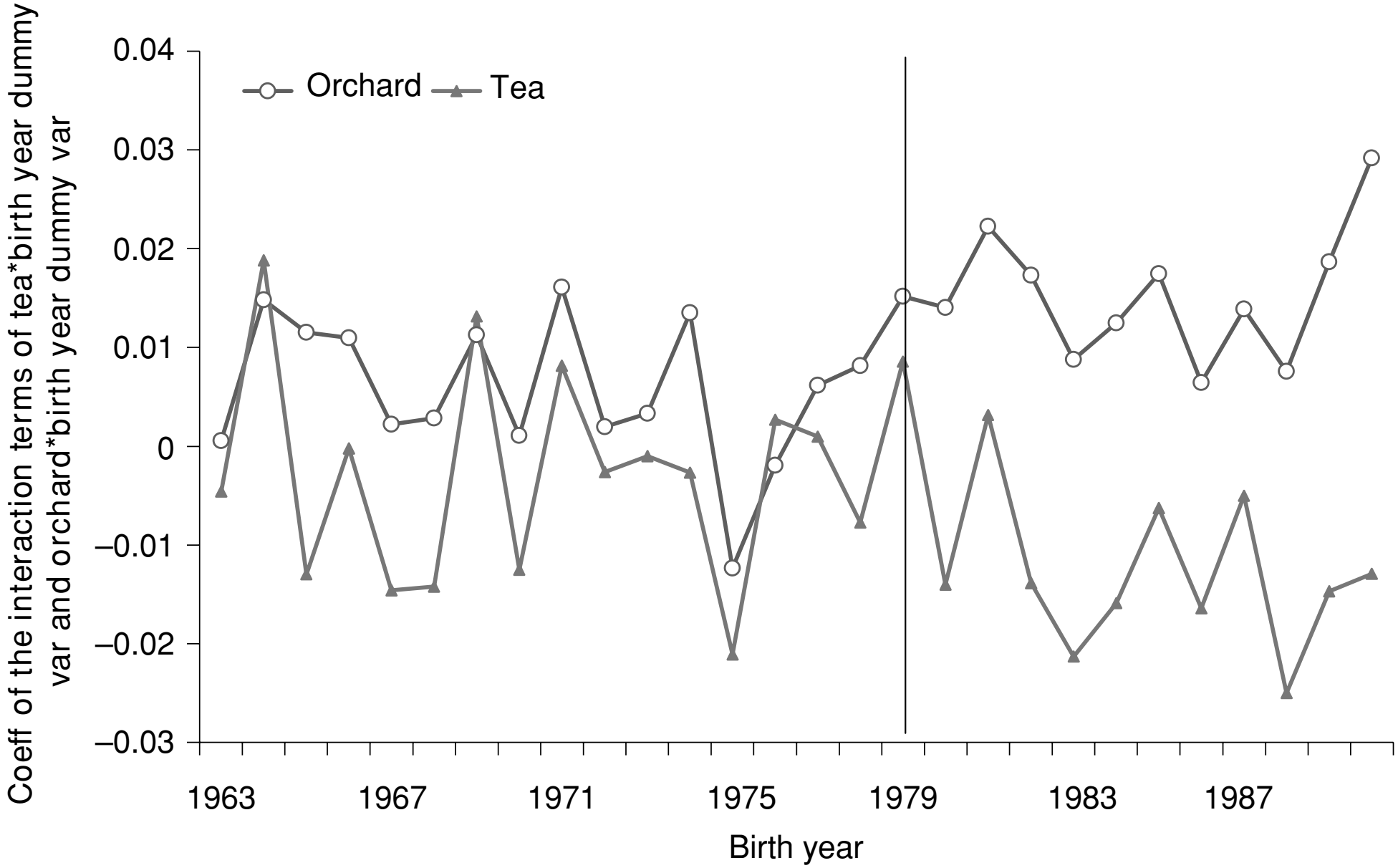
$$\beta_2 \approx \beta_3 \approx \cdots \beta_{T^*-1} \approx 0;\ \beta_{T^*},\ \beta_{T^*+1},\ \cdots,\ \beta_T > 0$$

**示例 24.** 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。

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



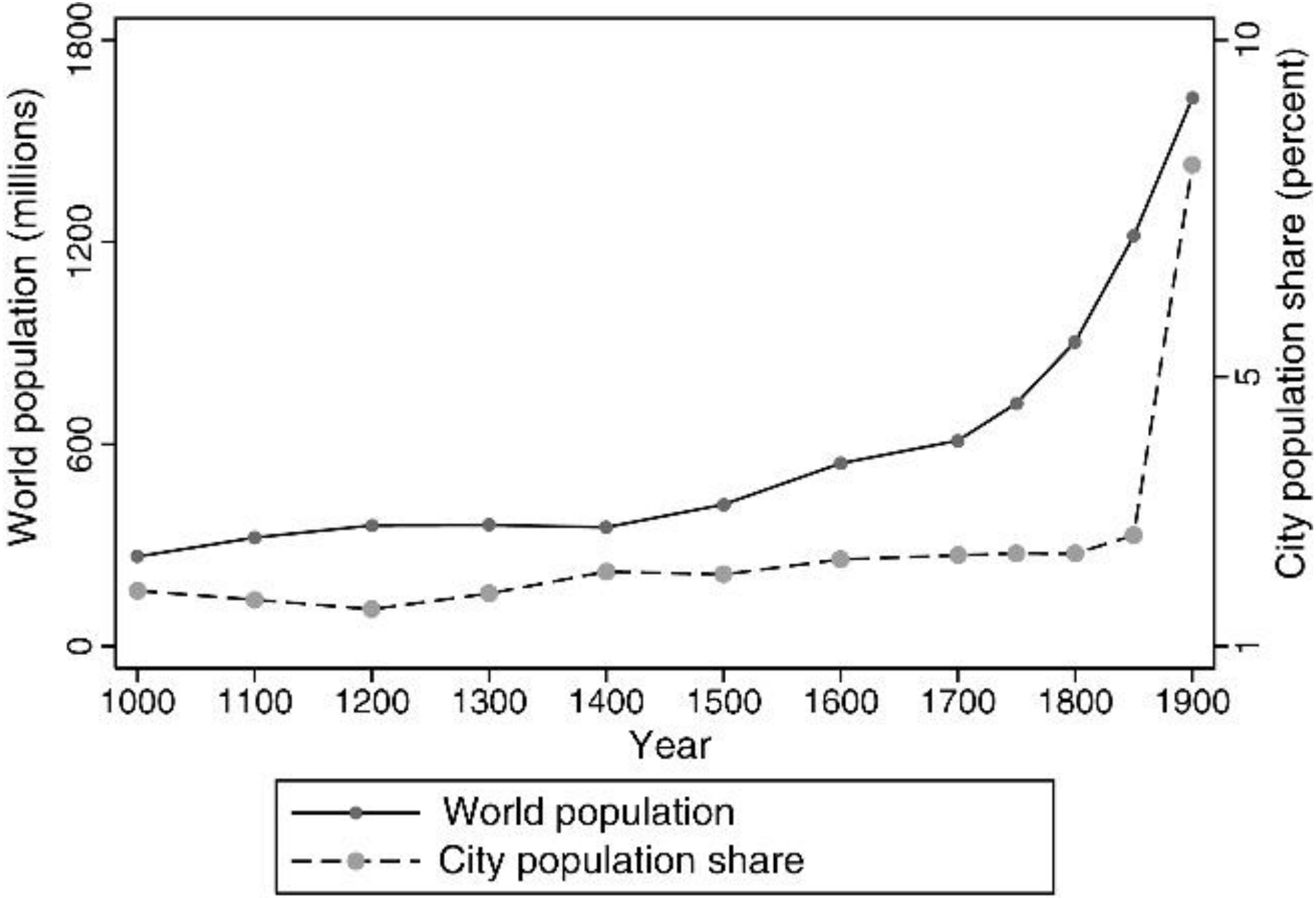


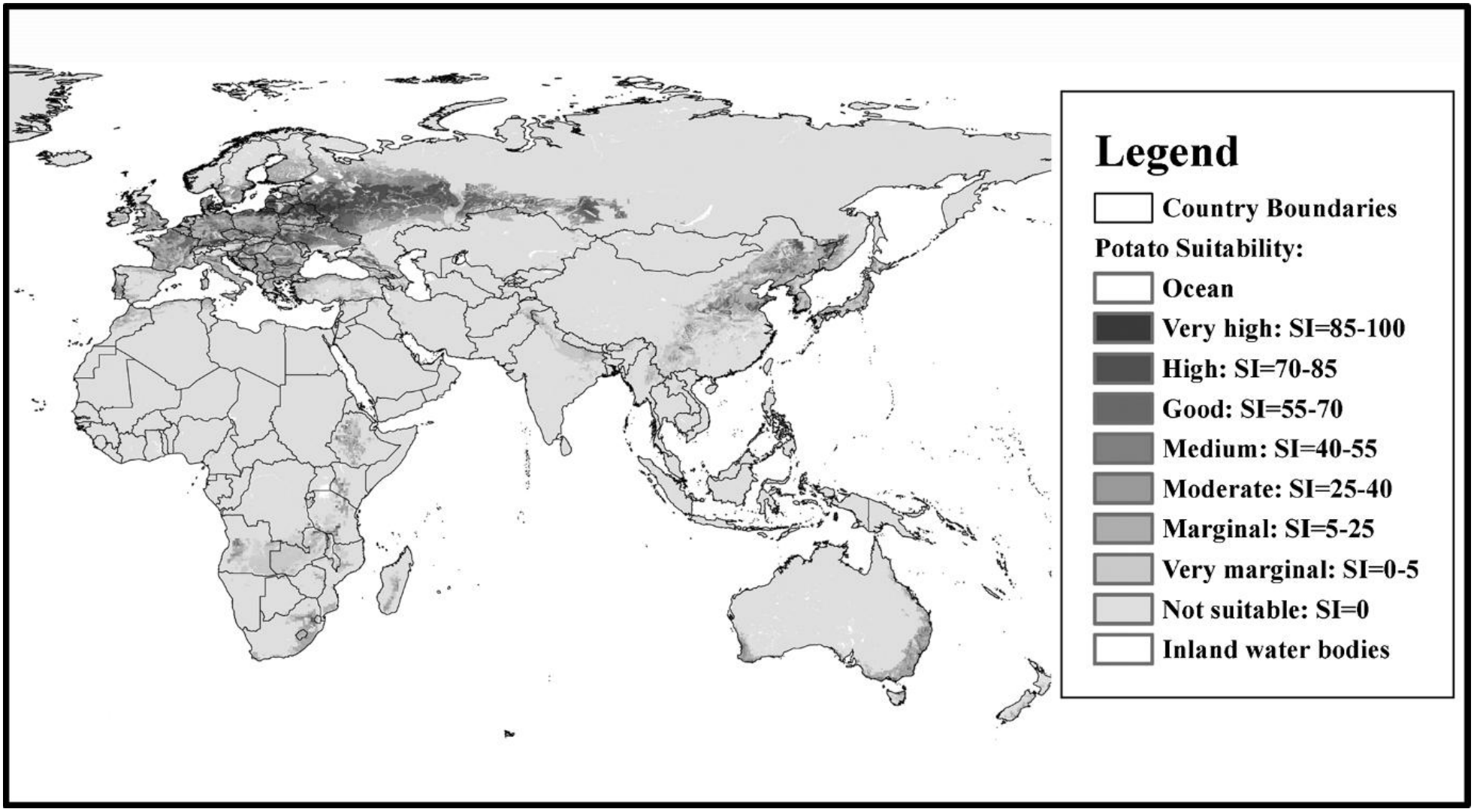
- 如果所有个体都在同一时间接受了处理，则 $D_i \times T_t = T_t$ 与时间固定效应共线性，无法识别。此时需要尝试构建处理强度指标 (treatment intensity) 指标，可以称之为连续处理的 DID 或连续型 DID。

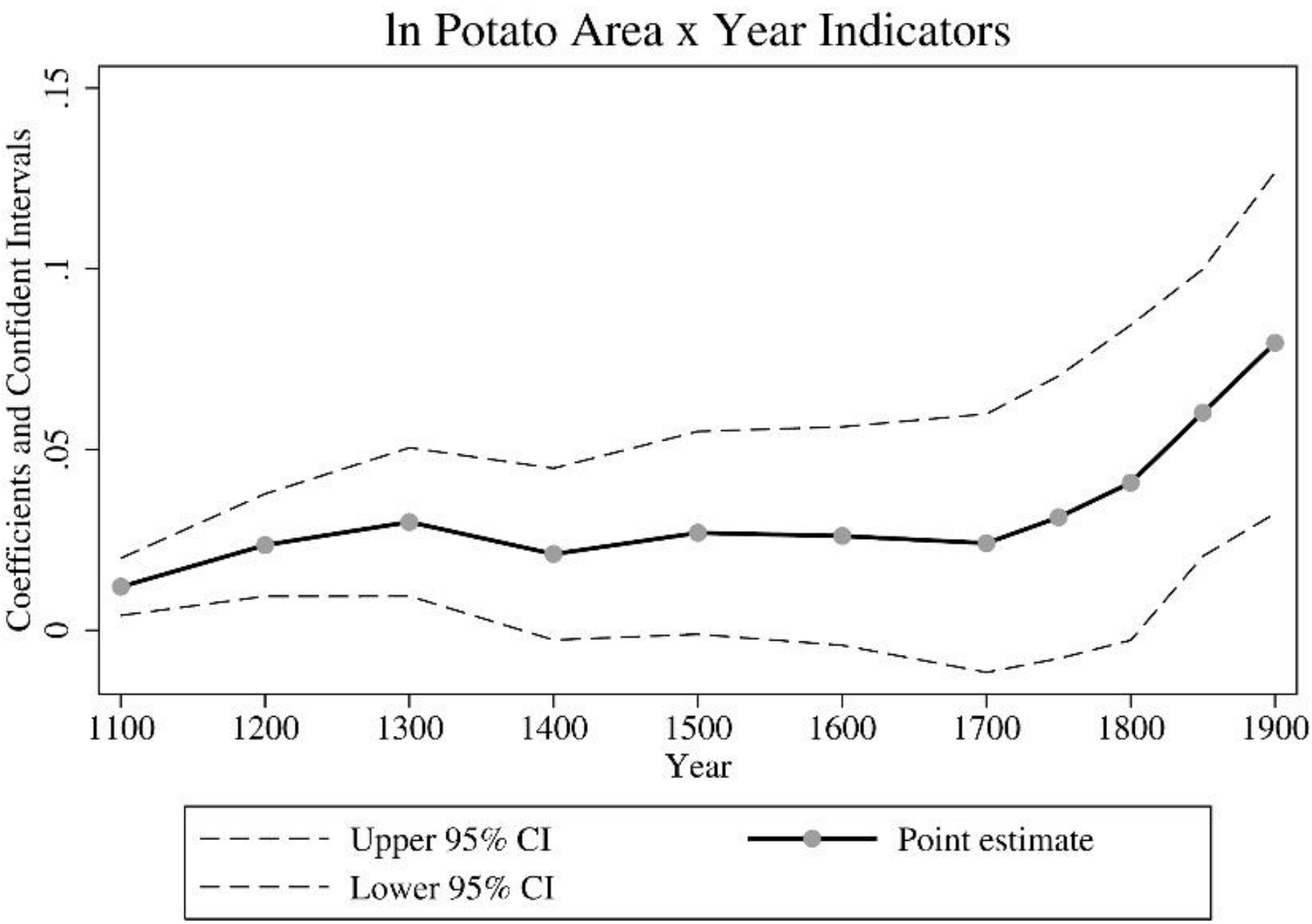
  **示例 25.** 土豆与人口增长 (Nunn and Qian, 2011, *QJE*).

$$y_{it} = \beta \ln Potato\,Area_i \cdot I_t^{Post} + \sum_{j=1100}^{1900} \mathbf{X}'_i \mathbf{I}_t^j \Phi_j + \sum_c \gamma_c I_i^c + \sum_{j=1100}^{1900} \rho_j I_t^j + \varepsilon_{it}$$

$$y_{it} = \sum_{j=1100}^{1900} \beta_j \ln Potato\ \ Area_i \cdot I_t^j + \sum_{j=1100}^{1900} \mathbf{X}'_i \mathbf{I}_t^j \Phi_j + \sum_c \gamma_c I_i^c + \sum_{j=1100}^{1900} \rho_j I_t^j + \varepsilon_{it}$$








(a) ln Total Population

- 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。

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

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

# DID 与匹配的结合

- 先通过匹配方法构造控制组，然后进行 DID 估计。
  - $1:1$ 匹配的样本，进行简单 OLS 回归。
    **示例 26.** 医院兼并对竞争的影响 (Schmitt, 2018, *AER*)
  - $1:m$ 匹配的样本，进行加权的简单 OLS 回归。
    **示例 27.** 厂网分离对竞争的影响 (Cicala, 2015, *AER*)
- 先差分，然后对差分结果进行匹配估计。
  再看**示例 7**。
- 匹配、DID、匹配 +DID 的区别与联系。

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.