双重差分法

汇报人：蒋涵琦

2020/12/3

双重差分法

- 第一节：单重差分法
- 第二节：双重差分法的直观理解
- 第三节：双重差分法的回归模型实例

第一节 单重差分法

- 企业1和2所在省份A受到税法改革的影响，为处置组；企业3和4所在省份B未实施税法改革，为控制组

Id	year	q	tax
1	2010	7.00	0
1	2011	7.02	0
1	2012	7.04	0
1	2013	7.06	0
1	2014	7.60	1
1	2015	7.50	1
1	2016	7.60	1
1	2017	7.70	1
2	2010	6.50	0
2	2011	6.52	0
2	2012	6.54	0
2	2013	6.56	0
2	2014	7.20	1
2	2015	7.10	1
2	2016	7.10	1
2	2017	7.00	1

Id	year	q	tax
3	2010	6.00	0
3	2011	6.02	0
3	2012	6.04	0
3	2013	6.06	0
3	2014	6.20	1
3	2015	6.20	1
3	2016	6.20	1
3	2017	6.20	1
4	2010	5.50	0
4	2011	5.52	0
4	2012	5.54	0
4	2013	5.56	0
4	2014	5.70	1
4	2015	5.70	1
4	2016	5.70	1
4	2017	5.70	1

第一节 单重差分法

- $Treat_i$表示分组虚拟变量，$After_t$表示时期虚拟变量
- $T_{after} = E(Y_{it} \mid Treat_i = 1, After_t = 1) = 7.35$
- $T_{before} = E(Y_{it} \mid Treat_i = 1, After_t = 0) = 6.78$
- $C_{after} = E(Y_{it} \mid Treat_i = 0, After_t = 1) = 5.95$
- $C_{before} = E(Y_{it} \mid Treat_i = 0, After_t = 0) = 5.78$

第一节 单重差分法

	平均业绩		
	2010~2013年	2014~2017年	横向差异
处置组	$T_{\text{before}} = 6.78$	$T_{after} = 7.35$	$T_{after} - T_{before} = 0.57$
控制组	$C_{before} = 5.78$	$C_{after} = 5.95$	$C_{after} - C_{before} = 0.17$
纵向差异	$T_{before} - C_{before} = 1$	$T_{after} - C_{after} = 1.40$	

- $ATT = T_{after} - T'_{\text{after}}$
- 估计反事实结果T'_{after}的两种方法
  - 横截面单重差分

    $T'_{after} = C_{after}$
  - 时间序列单重差分

    $T'_{after} = T_{before}$

第一节 横截面单重差分

- $Y_{it} = \beta_0 + \beta_1 Treat_i + e_{it}$，如果$After_t = 1$
- 数据的回归结果如下

q	$Coef.$	$Std.Err.$	t	$P > \|t\|$	$[95\%\ Conf.Interval]$	
Treat	1.40	0.136277	10.27	0.000	1.107715	1.692285
_cons	5.95	0.0963624	61.75	0.000	5.743323	6.156677

- $\hat{\beta}_1 = 1.40$，与$\widehat{ATT} = T_{\text{after}} - C_{after} = 1.40$ 结果一致

第一节 横截面单重差分

- $E(Y_{it} \mid Treat_i = 1) = \beta_0 + \beta_1 + E(e_{it} \mid Treat_i = 1)$
- $E(Y_{it} \mid Treat_i = 0) = \beta_0 + E(e_{it} \mid Treat_i = 0)$
- $E(Y_{it} \mid Treat_i = 1) - E(Y_{it} \mid Treat_i = 0) = \beta_1 + \underbrace{E(e_{it} \mid Treat_i = 1) - E(e_{it} \mid Treat_i = 0)}_{\text{横截面单重差分估计偏差}}$
- 添加控制变量
  - $Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Z_i + e_{it}$
  - 存在完全共线性和变量无法观测的问题

第一节 时间序列单重差分

- $Y_{it} = \beta_0 + \beta_1 After_t + e_{it}$，如果$Treat_t = 1$
- 数据的回归结果如下

q	$Coef.$	$Std.Err.$	t	$P > \|t\|$	$[95\%\ Conf.Interval]$	
after	0.57	0.1365388	4.17	0.001	0.2771533	0.8628467
_cons	6.78	0.0965475	70.22	0.000	6.572926	6.987074

- $\hat{\beta}_1 = 0.57$，与$\widehat{ATT} = T_{\text{after}} - T_{before} = 0.57$ 结果一致

第一节 时间序列单重差分

- $E(Y_{it} \mid After_t = 1) = \beta_0 + \beta_1 + E(e_{it} \mid After_t = 1)$
- $E(Y_{it} \mid After_t = 0) = \beta_0 + E(e_{it} \mid After_t = 0)$
- $E(Y_{it} \mid After_t = 1) - E(Y_{it} \mid After_t = 0) = \beta_1 + \underbrace{E(e_{it} \mid After_t = 1) - E(e_{it} \mid After_t = 0)}_{\text{时间序列单重差分估计偏差}}$
- 添加控制变量
  - $Y_{it} = \beta_0 + \beta_1 \text{After}_t + \beta_2 Z_i + e_{it}$
  - 存在完全共线性和变量无法观测的问题

第二节 双重差分法的直观理解

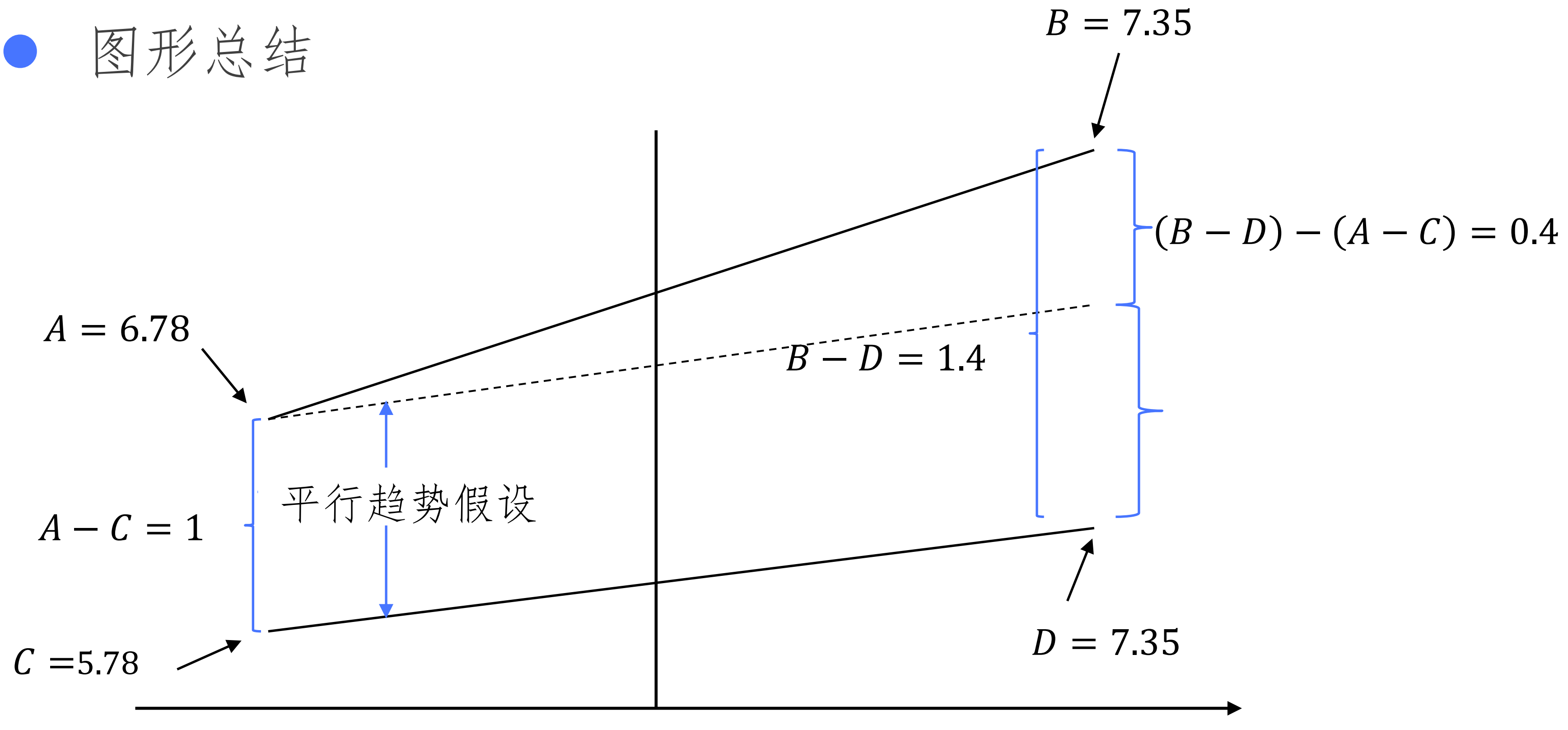
- 方法1：从横向差异理解
- 第一重差分
  - $T_{\text{after}} - T_{before}$ = 处置效应+其他因素造成的处置组在2014年前后的差异
  - $C_{\text{after}} - C_{before}$ = 其他因素造成的控制组在2014年前后的差异
- 第二重差分
  - $\left[T_{after} - T_{before}\right] - \left[C_{after} - C_{before}\right]$ = 处置效应
- 平行趋势假设
  - 其他因素造成的处置组在2014年前后的差异=其他因素造成的控制组在2014年前后的差异

第二节 双重差分法的直观理解

- 方法2：从纵向差异理解
- 第一重差分
  - $T_{\text{after}} - C_{\text{after}} =$ 处置效应+其他因素造成的处置组和控制组在2014年后的差异
  - $T_{\text{before}} - C_{before} =$ 其他因素造成的处置组和控制组在2014年前的差异
- 第二重差分
  - $\left[T_{after} - C_{after}\right] - \left[T_{before} - C_{before}\right] =$ 处置效应
- 差异不变假设
  - 其他因素造成的处置组和控制组在2014年后的差异=其他因素造成的处置组和控制组在2014年前的差异

第二节 双重差分法的直观理解

- 平行趋势假设和差异不变假设的一致性
- 图形总结



第三节 双重差分法回归模型实例

- 基本双重差分法回归模型

$$Y_{it}=\beta_0+\beta_1 Treat_i+\beta_2 After_t+\beta_3 Treat_i\times After_t+e_{it}$$

$$E(e_{it}\mid Treat_i, After_t)=0$$

- 控制组在处置事件发生前Y_{it}的均值

$$E(Y_{it}\mid Treat_i=0, After_t=0)=\beta_0$$

- 控制组在处置事件发生后Y_{it}的均值

$$E(Y_{it}\mid Treat_i=0, After_t=1)=\beta_0+\beta_2$$

第三节 双重差分法回归模型实例

- 处置组在处置事件发生前Y_{it}的均值

$$E(Y_{it} \mid Treat_i = 1, After_t = 0) = \beta_0 + \beta_1$$

- 处置组在处置事件发生后Y_{it}的均值

$$E(Y_{it} \mid Treat_i = 1, After_t = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

- 处置组和控制组在处置事件发生前Y_{it}的均值差异

$$E(Y_{it} \mid Treat_i = 1, After_t = 0) - E(Y_{it} \mid Treat_i = 0, After_t = 0) = \beta_1$$

- 控制组在处置事件发生前后Y_{it}的均值变化

$$E(Y_{it} \mid Treat_i = 0, After_t = 1) - E(Y_{it} \mid Treat_i = 0, After_t = 0) = \beta_2$$

第三节 双重差分法回归模型实例

- 交乘项$Treat_i \times After_t$的估计系数β_3的经济含义
  - 方法1：从横向差异理解

    处置组在处置前后Y_{it}的均值差异−控制组在处置前后Y_{it}的均值差异

    $= E(Y_{it} \mid Treat_i = 1, After_t = 1) - E(Y_{it} \mid Treat_i = 1, After_t = 0)$

    $-E(Y_{it} \mid Treat_i = 0, After_t = 1) - E(Y_{it} \mid Treat_i = 0, After_t = 0)$

    $= [(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1)] - [(\beta_0 + \beta_2) - \beta_0]$

    $=\beta_3$
  - 方法2：从纵向差异理解
  - β_3为双重差分估计量

第三节 双重差分回归模型实例

- 回归结果如下

q	Coef.	Std. Err.	t	p>|t|	[95% Conf. Interval]	
After	0.17	0.1352247	1.26	0.219	-0.1069952	0.4469952
Treat	1	0.1352247	7.40	0.000	0.7230048	1.276995
Treat after	0.4	0.1912366	2.09	0.046	0.0082696	0.7917304
_cons	5.78	0.0956183	60.45	0.000	5.584135	5.975865

- $\hat{\beta}_3 = 0.4 \quad SE\left(\hat{\beta}_3\right) = 0.191$

第三节 双重差分法回归模型实例

- 使用个体和时间固定效应细化模型，提高模型精度，降低估计系数方差

$$Y_{it} = \beta_3 Treat_i \times After_t + \alpha_i + Year_t + e_{it}$$

q	Coef.	Std. Err.	t	p>|t|	[95% Conf. Interval]	
Treat after	0.4	0.0295804	13.52	0.000	0.3382964	0.4617036
id1	7	0.0256174	273.25	0.000	6.946563	7.053437

- 交叉项$Treat_i \times After_t$的系数为0.4，与简单固定效应模型结果一致，但标准误降低为0.029

第三节 双重差分法回归模型实例

- 一般只需得到处置组的平均处置效应，不再将$\text{Treat}_i \times After_t$细化为$\alpha_i \times \text{Year}_t$
- 要研究事件对处置组在不同时间的影响，将交叉项$\text{Treat}_i \times After_t$中的$After_t$细化,例如

$$Y_{it} = \beta_3^1 Treat_i \times After_1 + \beta_3^2 Treat_i \times After_2 + \beta_3^3 Treat_i \times After_{3+4} + \alpha_i + Year_t + e_{it}$$

第三节 双重差分法回归模型实例

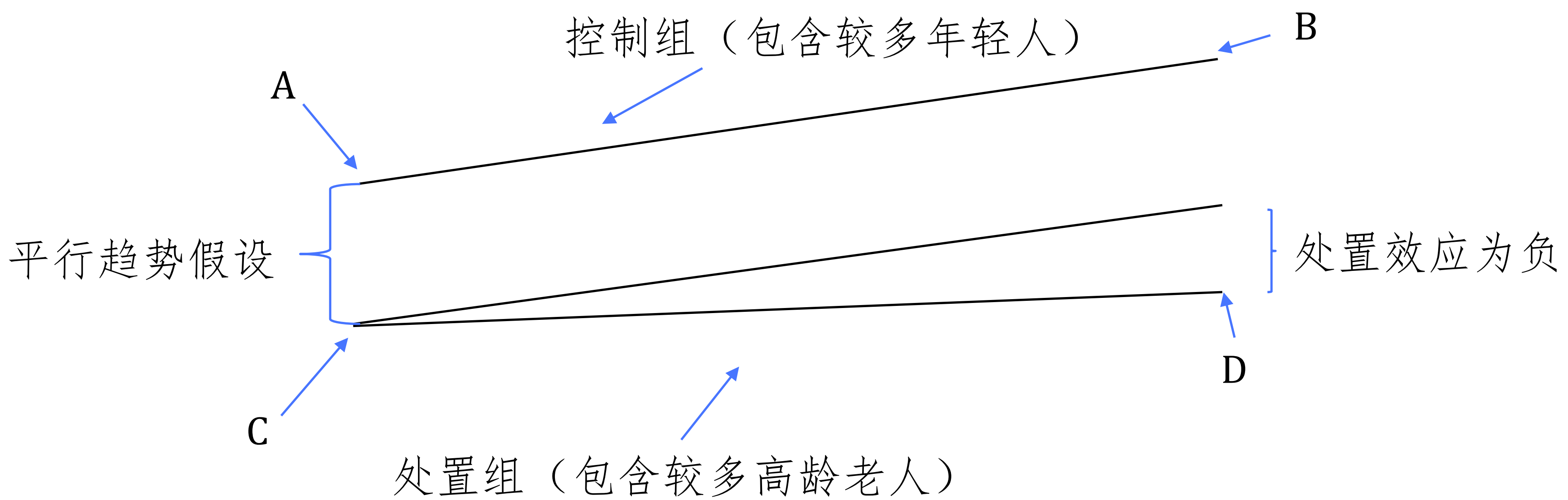
- 加入其他可观测的随时间变化的变量

$$Y_{it} = \beta_3 Treat_i \times After_t + \gamma X_{it} + \alpha_i + Year_t + e_{it}$$

- 加入控制变量X_{it}的意义及注意事项
  - 模型的隐含假设是，处置组和控制组随时间变化的特征在同一时间上的变化是相同的，故加入新的控制变量X_{it}并不会改变估计值，只是分离出e_{it}的一部分变化，降低估计值的方差
  - 基本的平行趋势假设不成立时，加入X_{it}可以降低估计误差
  - 为避免过度控制误差，加入模型的其他控制变量应该是不受事件影响的变量

第三节 双重差分法回归模型实例

- 一般而言，双重差分需要加入控制变量，保证条件独立性假设成立，从而控制选择偏误



例子：服药对身体健康的影响