

双重差分模型

李德山

¹ 西南科技大学 经济管理学院

²School of Economics and Management
Southwest University of Science and Technology

2021 年 3 月 19 日

① 双重差分模型原理

主要内容

- ① 双重差分模型原理
- ② 双重差分模型实例

主要内容

- ① 双重差分模型原理
- ② 双重差分模型实例
- ③ 双重差分法假设条件检验
- ④ 三重差分模型原理

双重差分模型原理

- 在自然实验中，个体接受处置与否是随机分配的，但是我们并不能控制这种随机性。
- 利用双重差分（difference in difference）来研究政策问题：使用群体层面的固定效应来解决地域或年份层面上出现的遗漏变量偏误。
- 双重差分的思想最早是由物理学家 John Snow(1855) 提出的。他研究了 19 世纪中期伦敦市的霍乱传染问题（是由受污染的水传染的，而不是“糟糕的空气”）。
- 相比由 A 水厂供水的地区，迁往上游的 B 水厂供水地区的霍乱死亡率急剧下降。

双重差分模型原理

- 政策评估常用的估计方法有：**双重差分法**、倾向匹配得分法、断点回归法、合成控制法和 Hsiao 面板数据评估法，等。
- 这些方法经常被称为“准自然实验”。它与随机控制实验的区别是前者的干预行为不是随机的，因此接受干预的处置组和控制组的特征分布并不相同。
- 准自然实验的干预行为是外生的，它不受个体特征的影响，因此它不存在自选择问题。但是在社会科学领域，干预行为的外生性通常不是那么“干净”。

双重差分模型原理

- 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。新税法于 2014 年开始实行。

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

双重差分模型原理

- $Treat_i$ 表示分组虚拟变量, $After_t$ 表示时期虚拟变量。

$$T_{after} = E(Y_{it}|Treat_i = 1, After_t = 1) = 7.35$$

$$T_{before} = E(Y_{it}|Treat_i = 1, After_t = 0) = 6.78$$

$$C_{after} = E(Y_{it}|Treat_i = 0, After_t = 1) = 5.95$$

$$C_{before} = E(Y_{it}|Treat_i = 0, After_t = 0) = 5.78$$

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

双重差分模型原理

- 我们无法观测到处置组在 2014 年后税改未实施情况下的平均业绩 T'_{after} , 也就无法直接计算税改对处置组的处置效应:

$$ATT = T_{after} - T'_{after}$$

- 估计反事实 T'_{after} 有两种方法: 横截面单重差分; 时间序列单重差分。
- 横截面单重差分: $T'_{after} = C_{after}$, 即通过比较处置组和控制组在事件发生后的平均结果差异估计处置效应。用 2014 年后的数据进行回归:

$$Y_{it} = \beta_0 + \beta_1 Treat_i + e_{it}, After_t = 1$$

- 回归结果得到的 $\hat{\beta}_1$ 与前面的表格结果一致, 都是 1.4。

q	<i>Coef.</i>	<i>Std. Err.</i>	t	$P > t $	[95% <i>Conf. Interval</i>]	
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

双重差分模型原理

- 我们通过条件期望值来看横截面差分模型的系数 β_1 是否反映了事件的处置效应：

$$E(Y_{it}|Treat_i = 1) = \beta_0 + \beta_1 + E(e_{it}|Treat_i = 1)$$

$$E(Y_{it}|Treat_i = 0) = \beta_0 + E(e_{it}|Treat_i = 0)$$

$$E(Y_{it}|Treat_i = 1) - E(Y_{it}|Treat_i = 0)$$

$$= \beta_1 + \underbrace{E(e_{it}|Treat_i = 1) - E(e_{it}|Treat_i = 0)}$$

横截面单重差分估计偏差

- 如果偏差 = 0, 那么 $\hat{\beta}_1$ 是 β_1 的无偏估计量。但是, 这个条件很难成立。

双重差分模型原理

- 上述回归模型存在遗漏变量偏误问题。模型中应该加上变量 Z_i 来控制处置组和控制组存在的其他差异（如企业所在省份的经济发展水平）：

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Z_i + e_{it}$$

- 但是依然存在加入的控制变量可能和分组变量共线性问题，还有就是控制组 and 处置组之间的某些差异也可能是无法观测，因此无法通过横截面数据“控制”掉。

双重差分模型原理

- 时间序列单重差分: $T'_{after} = T_{before}$, 即考虑处置组在事件前后的差异, 这样可以避免处置组和控制组差异所带来的问题。

$$Y_{it} = \beta_0 + \beta_1 After_t + e_{it}, Treat_i = 1$$

- 回归结果得到的 $\hat{\beta}_1$ 与前面的表格结果一致, 都是 0.57。

<i>q</i>	<i>Coef.</i>	<i>Std.Err.</i>	<i>t</i>	<i>P > t </i>	<i>[95% Conf. Interval]</i>	
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

双重差分模型原理

- 我们通过条件期望值来看时间序列差分模型的系数 β_1 是否反映了事件的处置效应：

$$E(Y_{it}|After_t = 1) = \beta_0 + \beta_1 + E(e_{it}|After_t = 1)$$

$$E(Y_{it}|After_t = 0) = \beta_0 + E(e_{it}|After_t = 0)$$

$$\begin{aligned} & E(Y_{it}|After_t = 1) - E(Y_{it}|After_t = 0) \\ &= \beta_1 + \underbrace{E(e_{it}|After_t = 1) - E(e_{it}|After_t = 0)}_{\text{时间序列单重差分估计偏差}} \end{aligned}$$

- 如果偏差 = 0，那么 $\hat{\beta}_1$ 是 β_1 的无偏估计量。但是，这个条件很难成立。
- 因为某些宏观因素可能和税改在同一时点发生。例如，在 2014 年后国家放松信贷政策。本质上，上述模型也存在遗漏变量问题。加入控制变量 Z_t 来控制其他导致处置组业绩在 2014 年前后发生变化的因素：

$$Y_{it} = \beta_0 + \beta_1 After_t + \beta_2 Z_t + e_{it}$$

- 同样存在加入控制变量可能和时期变量共线性问题以及无法观测的问题。

双重差分模型原理

- 方法 1：从横向差异来直观理解
- 第一重差分
 - 处置组在事件发生的 2014 年前后的差异 $= T_{after} - T_{before}$ = 实施新税法造成的差异（处置效应）+ 其他因素造成的处置组在 2014 年前后的差异。
 - 控制组在事件发生的 2014 年前后的差异 $= C_{after} - C_{before}$ = 其他因素造成的控制组在 2014 年前后的差异。
- 第二重差分
 - 处置组在事件发生的 2014 年前后的差异—控制组在事件发生的 2014 年前后的差异 $= [T_{after} - T_{before}] - [C_{after} - C_{before}]$ = 处置效应 + “0”
- 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。即在不存在税改的情况下，控制组和处置组的平均业绩随时间变化的趋势是平行的。

双重差分模型原理

- 方法 2：从纵向差异来直观理解
- 第一重差分
 - 处置组和控制组在事件发生的 2014 年后的差异 $= T_{after} - C_{after}$ = 实施新税法造成的差异（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。
 - 处置组和控制组在事件发生的 2014 年前的差异 $= T_{before} - C_{before}$ = 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
 - 处置组和控制组在事件发生的 2014 年后的差异—处置组和控制组在事件发生的 2014 年前的差异 $= [T_{after} - C_{after}] - [T_{before} - C_{before}]$ = 处置效应 + “0”
- 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。即在不存在税改的情况下，控制组和处置组在 2014 年前后的差异是相同的。

双重差分模型原理

- 平行趋势假设和差异不变假设是一致的：“平行”也就意味着两组的平均业绩差异保持不变。

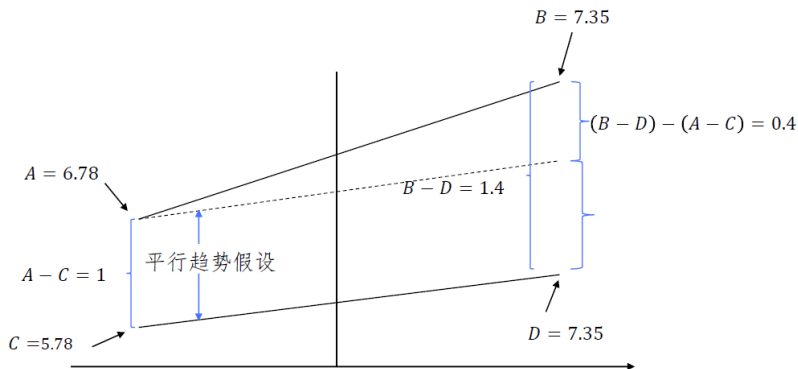


Figure: 双重差分模型的图形解释

双重差分模型实例

- 基本双重差分回归模型

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 After_t + \beta_3 Treat_i \times After_t + e_{it} \quad (1)$$

- 处置事件是外生的，意味着模型满足 $E(e_{it}|Treat_i, After_t) = 0$
 - 控制组在处置事件发生前 Y_{it} 的均值（C 点）：

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

- 控制组在处置事件发生后 Y_{it} 的均值（D 点）：

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

双重差分模型实例

- 处置组在处置事件发生前 Y_{it} 的均值 (A 点):

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

- 处置组在处置事件发生后 Y_{it} 的均值 (B 点):

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

- 处置组和控制组在处置事件发生前 Y_{it} 的均值差异 (A-C):

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

- 控制组在处置事件发生前后 Y_{it} 的均值差异 (D-C):

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

双重差分模型实例

- 交互项的估计系数 β_3 的经济含义
 - 方法 1: 从横向差异理解
 - 处置组在处置事件发生前后 Y_{it} 的均值差异—控制组在处置事件发生前后 Y_{it} 的均值差异

$$\begin{aligned} &= (B - A) - (D - C) \\ &= E(Y_{it}|Treat_i = 1, After_t = 1) - E(Y_{it}|Treat_i = 1, After_t = 0) \\ &\quad - E(Y_{it}|Treat_i = 0, After_t = 1) - E(Y_{it}|Treat_i = 0, After_t = 0) \\ &= [(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1)] - [(\beta_0 + \beta_2) - \beta_0] \\ &= (\beta_2 + \beta_3) - \beta_2 \\ &= \beta_3 \end{aligned}$$

双重差分模型实例

- 方法 2: 从纵向差异理解
- 处置组和控制组在处置事件发生后 Y_{it} 的均值差异—处置组和控制组在处置事件发生前 Y_{it} 的均值差异

$$\begin{aligned} &= (B - D) - (A - C) \\ &= E(Y_{it}|Treat_i = 1, After_t = 1) - E(Y_{it}|Treat_i = 0, After_t = 1) \\ &\quad - E(Y_{it}|Treat_i = 1, After_t = 0) - E(Y_{it}|Treat_i = 0, After_t = 0) \\ &= [(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2)] - [(\beta_0 + \beta_1) - \beta_0] \\ &= (\beta_1 + \beta_3) - \beta_1 \\ &= \beta_3 \end{aligned}$$

双重差分模型实例

- 回归结果如下:

- $\hat{\beta}_3 = 0.4, SE(\hat{\beta}_3) = 0.191$

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

Figure: 双重差分法的回归结果

双重差分模型实例

- 使用个体和时间固定效应

- 基本的双重差分法也是一个种固定效应模型。如果有同一个个体在不同时间信息的面板数据，就可以使用个体和时间固定效应细化模型，提高模型精度，降低估计系数的方差。

$$Y_{it} = \beta_3 Treat_i \times After_t + \alpha_i + Year_t + e_{it} \quad (2)$$

- 这个模型相对于基本模型的优点是： α_i 和 $Year_t$ 比 β_0 、 β_1 和 β_2 更精细的固定效应。
- α_i 是个体的固定效应，而基本模型里 β_0 、 β_1 是组固定效应。组固定效应是个体固定效应的平均值。同理， $Year_t$ 是每年的固定效应， β_2 是时期固定效应。时期固定效应是时期内每年固定效应的平均值。

双重差分模型实例

- 使用个体和时间固定效应的 DID 回归结果：
 - 交互项的回归系数为 0.4 与基本模型结果一致，但是标准误为 0.029，比基本模型的标准误要小了很多。

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

Figure: 使用个体和时间固定效应的 DID 回归结果

双重差分模型实例

- 当然，我们也可以将 $Treat_i \times After_t$ 细化为 $\alpha_i \times After_t$ ，这样可以得到事件对每个处置个体在事件后每一年的处置效应。但是如果个体数量太多，那么交互项估计系数就很多。所以，一般不再细化。
- 但是，如果要研究事件对处置组在不同时间的影响，可以将交互项中的 $After_t$ 细化。例如：

$$Y_{it} = \beta_3^1 Treat_i \times After_1 + \beta_3^2 Treat_i \times After_2 + \beta_3^3 Treat_i \times After_3 + \alpha_i + Year_t + e_{it}$$

双重差分模型实例

- 加入控制变量

- 加入其它可观测的随时间变化的变量，模型扩展为：

$$Y_{it} = \beta_3 Treat_i \times After_t + \gamma X_{it} + \alpha_i + Year_t + e_{it} \quad (3)$$

- 模型隐含的假设是，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，因此能被同一个时间固定效应 $Year_t$ 所控制，从而使得新加入的控制变量 X 并不会改变估计值，而只是分离出干扰项的一部分变化，从而降低了干扰项的“噪声”。
- 如果加入控制变量 X 后，估计值改变了，可能是基本模型存在遗漏变量问题。即存在不可观测的随时间变化的变量与处置变量和结果变量之间存在相关性。
- 加入模型的控制变量应该是不受事件影响的变量，否则会造成过度控制误差。

双重差分模型假设条件检验

- 平行趋势假设是使用双重差分估计处置效应的关键假设
 - 平行趋势假设：在没有处置事件的情况下，控制组和处置组的被解释变量的均值差异在不同时间内保持一致。
 - 除处置事件外的其他因素对处置组和控制组的影响是相同。

双重差分模型假设条件检验

- 比较处置组和控制组在事件发生前的趋势差异

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

- β_1 是处置组和控制组在事件发生前的平均差异。我们可以进一步分析处置组和控制组在事件发生前每年的差异，如果平行趋势成立，那么两组之间每年的差异应该没有显著区别：

$$\begin{aligned} Y_{it} = & \beta_0 + \beta_1^{2010} Treat_i + \beta_1^{2011} Treat_i Year_{2011} \\ & + \beta_1^{2012} Treat_i Year_{2012} + \beta_1^{2013} Treat_i Year_{2013} \\ & + \beta_2 After_t + \beta_3 Treat_i After_t + e_{it} \end{aligned}$$

双重差分模型假设条件检验

- 如果时间 t 是税改前的 n 年 ($n = 2011, 2012, 2013$), $Year_t = 1$; 否则, $Year_t = 0$ 。
- 如果个体 i 属于处置组, 并且观测值在税改发生前的 n 年, 那么 $Treat_i Year_n = 1$; 否则, $Treat_i Year_n = 0$ 。
- β_1^{2010} 是处置组和控制组在税改发生前的 2010 年的差异 (基准差异)。
- β_1^{2011} 、 β_1^{2012} 、 β_1^{2013} 反映了税改发生前的 2011、2012、2013 年处置组和控制组的差异相较于它们在 2010 年差异的变化。
- 如果平行趋势假设成立 (差异不变), 那么:

$$\beta_1^{2011} = \beta_1^{2012} = \beta_1^{2013} = 0$$

- β_3 是税改发生后处置组和控制组的差异相较于在 2010 年差异的变化。

双重差分模型假设条件检验

- Stata 命令

```
gen treat=id<=2 产生组虚拟变量
gen after=year>=2014 产生时期虚拟变量
gen treatafter=after*treat 产生组和时期交叉变量
tab year, gen(year) 年虚拟变量
gen treatyear2=treat*year2
gen treatyear3=treat*year3
gen treatyear4=treat*year4
reg q treat treatyear2-treatyear4 after treatafter
```

- 回归结果：

q	Coef.	Std.Err
treatyear2	.02	.2859371
treatyear3	.04	.2859371
treatyear4	.06	.2859371

双重差分模型假设条件检验

- 各系数并不显著，说明在税改发生前的三年，两组的差异相较于它们在 2010 年差异的变化基本为 0，满足平行趋势假设。
- 平行趋势假设检验只能检验事件发生前两组的趋势是否一致。事件发生后，我们无法获知，如果事件不发生，两者的趋势是否仍然一致。因此，平行趋势假设本质上是无法检验的。我们只能希望如果事件发生时间点前的趋势是一致的，那么它们在事件发生的时间点后仍然是一致的。

双重差分模型假设条件检验

- 检验不影响变量或组的结果
 - 安慰剂检验（Pacebo Test）：如果发现事件对不该受影响的变量或组有影响，说明事件包含了其他作用。
 - 例如，在税法改革的 A 省，受影响的只是省属企业，对央属企业并没有影响。将样本分为省属企业和央属企业，并对每个子样本适用双重差分法估计事件的影响。对于央属企业的双重差分得到的处置效应应为 0。如果不为 0，则说明 A 省在 2014 年前后发生了与税改无关的事件并由此影响了企业业绩。这意味着前面发现的税改对业绩的影响结果受到了其他因素的干扰，并无法证明税改对企业业绩有因果关系。
 - 另一个检验方法是用事件发生前年份的数据做双重差分分析。例如，可以随意选择事件发生前的一个年份作为“假设事件”发生年。如果发现“假设事件”有显著作用，那么就说明即使在未实施新税法的情况下，我们也会发现处置组和控制组有差异。

三重差分模型

- 双重差分法最有可能面临的问题是，存在其他混淆事件，并只发生在处置组，由此导致平行趋势假设不成立。在这种情况下，即使处置事件未发生，其他混淆事件也会导致处置组和控制组有差异。
- 例如，如果 2014 年后，A 省不仅通过了新税法，还有其他利好因素发生。即使没有新税法，A 省企业业绩在 2014 年后也会比 B 省企业业绩增长得快：平行趋势假设不成立，双重差分法无法识别出新税法对业绩的作用。
- 如果新税法并不影响原本就有免税优惠的企业，可以将样本分为普通企业和免税企业，并对每个子样本使用双重差分法。
- 对普通企业样本得到的双重差分结果反映了新税法和其他利好因素的作用；通过免税企业样本得到的双重差分结果反映了其他利好因素的作用。将前一个结果减去后一个结果，即得到了**三重差分**，反映了新税法对普通企业业绩的影响。

三重差分模型

- 基本三重差分回归模型:

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 After_t + \beta_3 High_i + \beta_4 Treat_i \times After_t \\ + \beta_5 Treat_i \times High_i + \beta_6 High_i \times After_t + \beta_7 Treat_i \times High_i \times After_t + e_{it}$$

如果个体受事件影响, 则 $High_i = 1$; 否则 $High_i = 0$ 。

- β_4 代表其他因素对处置组的影响; β_7 代表事件对处置组的处置效应。
- 三重差分法包括了 4 个单变量系数 (包含常数项)、3 个两变量交互项系数、1 个三变量交互项系数。
- 第三重差分变量 $High_i$ 必须是外生的。
- 该模型同样可以进行扩展, 比如加上时间个体固定效应、控制变量和时间趋势项。
- 三重差分还可以估计事件对不同个体处置效应的差异性。对第三重差分变量重新定义。例如, 如果是省属企业 $High_i = 1$; 否则, $High_i = 0$ 。这时候的 β_7 代表事件 (如税法) 对省属和央企业绩影响的差异。

三重差分模型

免税企业双重差分 ($High_i=0$)	Before	After	横向差异
处置组 (T)	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_4$	$\beta_2 + \beta_4$
控制组 (C)	β_0	$\beta_0 + \beta_2$	β_2

普通企业双重差分 ($High_i=1$)	Before	After	横向差异
处置组 (T)	$\beta_0 + \beta_1 + \beta_3 + \beta_5$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7$	$\beta_2 + \beta_4 + \beta_6 + \beta_7$
控制组 (C)	$\beta_0 + \beta_3$	$\beta_0 + \beta_2 + \beta_3 + \beta_6$	$\beta_2 + \beta_6$

参考文献:

Angrist and Pischke(2008). Mostly harmless econometrics: An empiricist' s companion. Princeton university press.

陈强. 高级计量经济学及 Statat 应用 [M]. 高等教育出版社, 2014.

<https://www.zhihu.com/question/24322044>