- Difference-in-differences models

Policy analysis with pooled cross sections

- Pooled cross sections are two or more independently sampled cross sections in different time periods. They are not necessarily the same units between the periods.

- Pooled cross sections can be used to evaluate the impact of a treatment, event, program, or policy change.

- The difference-in-differences model involves before and after comparisons in natural experiments to determine the effect of a treatment.

The natural experiments offer an alternative to controlled (laboratory or field) experiments by benefiting from historical events that create "quasi random" variations in the treatment for the analysis.

Even if these events are not intentionally designed by the researchers but still provide the opportunity to get meaningful insights.

These events include for example man-made events for example, change in policy, wars, recessions, historical episodes such as fall of communism, division of countries like British India into India and Pakistan, or natural events like floods, earthquake, drought etc.

The difference-in-differences model

- The difference-in-differences model (DID model) shows the effect of a treatment in the after period (DID effect).

- A treatment is implemented for treated units ($treated = 1$). For the control units, there is no treatment ($treated = 0$).

- Data are collected in the period after the treatment ($after = 1$) and the period before the treatment ($after = 0$).

- The interaction term is $treated * after$, which is equal to 1 for treated units in the after period.

- Difference-in-differences model:
$$y = \beta_0 + \delta_0 after + \beta_1 treated + \delta_1 after * treated + u$$

- The difference-in-differences effect δ_1 is the effect of the treatment in the after period on the outcome y.

DID effect using two models

- Difference-in-differences model:
$$y = \beta_0 + \delta_0 after + \beta_1 treated + \delta_1 after * treated + u$$
- Estimate the regression in the <u>before period</u> $(after = 0)$
$$y = \beta_0 + \beta_1 treated + u$$
- The coefficient β_1 shows the differences in outcomes between treated and control units in the before period.
- Estimate the regression in the <u>after period</u> $(after = 1)$
$$y = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)\ treated + u$$
- The coefficient $(\beta_1 + \delta_1)$ shows the differences in outcomes between the treated and control units in the after period.
- The DID effect is the difference between the two coefficients $\delta_1 = (\beta_1 + \delta_1) - \beta_1$
- The DID effect is the difference in outcomes between treated and control units in the after period and the treated and control units in the before period.

DID effect using two models

- Difference-in-differences model:
$$y = \beta_0 + \delta_0 after + \beta_1 treated + \delta_1 after * treated + u$$

- Estimate the regression <u>for the control units</u> $(treated = 0)$
$$y = \beta_0 + \delta_0 after + u$$

- The coefficient δ_0 shows the difference in outcome between after and before for the control units.

- Estimate the regression <u>for the treated units</u> $(treated = 1)$
$$y = (\beta_0 + \beta_1) + (\delta_0 + \delta_1) after + u$$

- The coefficient $(\delta_0 + \delta_1)$ shows the differences in outcomes between after and before for the treated units.

- The DID effect is the difference between the two coefficients $\delta_1 = (\delta_0 + \delta_1) - \delta_0$

- The DID effect is difference in outcomes between after and before for the treated units and after and before for the control units.

DID model

- Difference-in-differences model:
$$y = \beta_0 + \delta_0 after + \beta_1 treated + \delta_1 after * treated + u$$

- Estimate this regression model. The DID effect δ_1 is the effect of the treatment in the after period.

- Outcome for control units ($treated = 0$) in before period ($after = 0$) is $\bar{y}_{0c} = \beta_0$

- Outcome for treated units ($treated = 1$) in before period ($after = 0$) is $\bar{y}_{0t} = \beta_0 + \beta_1$

- Outcome for control units ($treated = 0$) in after period ($after = 1$) is $\bar{y}_{1c} = \beta_0 + \delta_0$

- Outcome for treated units ($treated = 1$) in after period ($after = 1$) is $\bar{y}_{1t} = \beta_0 + \delta_0 + \beta_1 + \delta_1$

- The DID effect is:

- $\delta_1 = (\bar{y}_{1t} - \bar{y}_{0t}) - (\bar{y}_{1c} - \bar{y}_{0c}) = (\bar{y}_{1t} - \bar{y}_{1c}) - (\bar{y}_{0t} - \bar{y}_{0c})$

DID effect

- Difference-in-differences model:
$$y = \beta_0 + \delta_0 after + \beta_1 treated + \delta_1 after * treated + u$$

- The DID effect δ_1 is differences in outcomes between after and before for the treated units and after and before for the control units.

- The DID effect δ_1 is also the differences in outcomes between treated and control units in the after period and the treated and control units in the before period.

Outcome		Before	After	After - Before
		after = 0	after = 1	
Control (c)	treated = 0	β_0	$\beta_0 + \delta_0$	δ_0
Treated (t)	treated = 1	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treated - Control		β_1	$\beta_1 + \delta_1$	δ_1

DID model example

- The treatment is building of a garbage incinerator (waste treatment facility) which can potentially lower the prices of houses that are near the incinerator.

- Two periods: $y81=0$ in the before period of 1978 and $y81=1$ in the after period of 1981, after the incinerator was built.

- Treated units are houses near the incinerator ($nearinc=1$) and control units are houses far from the incinerator ($nearinc=0$).

rprice		Before	After
		yr81=0	yr81=1
Control (c)	nearinc = 0	$82,517	$101,308
Treated (t)	nearinc = 1	$63,693	$70,619

- Prices for houses near the incinerator and far from the incinerator increased in the after period.

- Prices for houses near the incinerator were lower in the before period.

- Did building of the incinerator lower house prices (rprice)?

DID effect using two models

- Difference-in-differences model:
$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + u$$
- Estimate the regression in the before period ($y81 = 0$)
$$rprice = \beta_0 + \beta_1 nearinc + u$$
- The coefficient β_1 shows the differences in house prices for houses near and far from the incinerator in the before period.
- Estimate the regression in the after period ($y81 = 1$)
$$rprice = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)\, nearinc + u$$
- The coefficient $(\beta_1 + \delta_1)$ shows the differences in house prices for houses near and far from the incinerator in the after period.
- The DID effect is the difference between the two coefficients $\delta_1 = (\beta_1 + \delta_1) - \beta_1$

DID effect using two models

VARIABLES	After period rprice	Before period rprice
nearinc	-30,688***	-18,824***
	(5,828)	(4,745)
y81		
y81*nearinc		
Constant	101,308***	82,517***
	(3,093)	(2,654)

In the after period, prices for houses near the incinerator were $30,688 lower compared to prices for houses that are far from the incinerator.

In the before period, prices for houses near the incinerator were $18,824 lower compared to prices for houses that are far from the incinerator.

The DID effect is -30,688-(-18,824)= -11,864

Prices for houses near the incinerator were $11,864 lower than prices for houses far from the incinerator, after the incinerator was built.

DID effect using two models

- Difference-in-differences model:
$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + u$$
- Estimate the regression for houses far from the incinerator ($nearinc = 0$)
$$rprice = \beta_0 + \delta_0 y81 + u$$
- The coefficient δ_0 shows the difference in house prices between after and before for houses far from the incinerator.
- Estimate the regression in houses near the incinerator ($nearinc = 1$)
$$rprice = (\beta_0 + \beta_1) + (\delta_0 + \delta_1)y81 + u$$
- The coefficient $(\delta_0 + \delta_1)$ shows the difference in house prices between after and before for houses near the incinerator.
- The DID effect is the difference between the two coefficients $\delta_1 = (\delta_0 + \delta_1) - \delta_0$

DID effect using two models

VARIABLES	For houses near incinerator rprice	For houses far from incinerator rprice
nearinc		
y81	6,926	18,790***
	(8,205)	(3,383)
y81*nearinc		
Constant	63,693***	82,517***
	(5,296)	(2,278)

For houses near the incinerator, house prices were $6,926 higher in the after period compared to the before period.

For houses far from the incinerator, house prices were $18,790 higher in the after period compared to the before period.

The DID effect is 6,926 – 18,780 = -11,864 (same effect)

Prices for houses near the incinerator were $11,864 lower than prices for houses far from the incinerator, after the incinerator was built.

DID model example

- Difference-in-differences model: $rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 * nearinc + u$

- The numbers is the table show the average house prices for treated units (houses near the incinerator) and control units (houses far from the incinerator) in the before (1978) and after (1981) period.

- The DID effect δ_1 shows that houses near the incinerator had $11,864 lower prices than houses far from the incinerator, after the incinerator was built (in the after period).

		Before	After	After - Before
		after=0	after =1	
Control (c)	treated = 0	β_0=82,517	$\beta_0 + \delta_0$ = 101,308	δ_0= 18,790
Treated (t)	treated = 1	$\beta_0 + \beta_1$=63,693	$\beta_0 + \delta_0 + \beta_1 + \delta_1$=70,619	$\delta_0 + \delta_1$= 6,926
Treated - Control		β_1= -18,824	$\beta_1 + \delta_1$= -30,688	δ_1= -11,864

DID model example

VARIABLES	After period rprice	Before period rprice	For houses near incinerator rprice	For houses far from incinerator rprice	DID regression rprice
nearinc	-30,688***	-18,824***			-18,824***
	(5,828)	(4,745)			(4,875)
y81			6,926	18,790***	18,790***
			(8,205)	(3,383)	(4,050)
y81*nearinc					-11,864
					(7,457)
Constant	101,308***	82,517***	63,693***	82,517***	82,517***
	(3,093)	(2,654)	(5,296)	(2,278)	(2,727)

The coefficients in regression in the before period and for houses far from incinerator are same as in DID model.
The DID effect is -30,688 - (-18,824) = 6,926 – 18,780 = -11,864
House prices near the incinerator were $11,864 lower than prices for houses far from the incinerator, after the incinerator was built, but the effect is not significant.