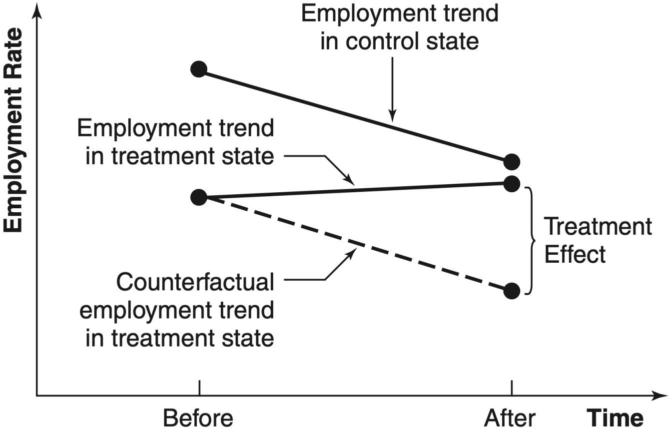
Applied Microeconometrics II , Lecture 3

Ciprian Domnisoru

Aalto University

- Difference in difference methods: extensions of the 2 X 2 model.
- DD as a form of fixed effects regression
- Panel data (multiple years and (staggered) treatments), (treatment/control unit) time trends, clustering, under treatment homogeneity and heterogeneity
- DiD functional form assumptions
- Triple differences (DDD)
- Examples: DD with spatial variation as control
- Event studies
- Synthetic control methods

Difference in differences 2X2 design



Difference in differences : multiple periods and treatments

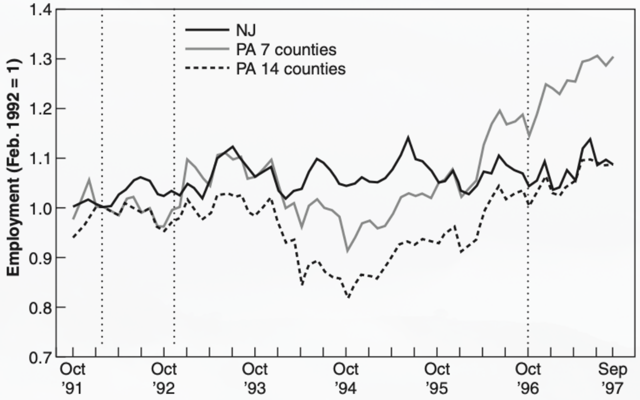


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

Differences in differences: rationale

- In a quasi-experimental setting, confounding factors are likely to bias our estimates. Think of internal validity concerns expressed last time: history (events occurring simultaneously), maturation.
- In the difference in differences strategy, the quasi-control group should ideally be affected by the same confounding effects. Ex: recessions affecting both control and treatment groups.
- The evolution of the control group serves as a counterfactual.
- We are left with many other concerns: e.g. selection into treatment.
- To some extent, these can be addressed by checking the paralell trends asumption ; assessing whether adoption was quasi-random; checking other policies were not adopted at the same time; no spillovers between treatment and control.
- We are left with concerns about time-varying unobservables. How are these concerns addressed in randomized experiments?

Differences in differences: regression model

- Regression model for entity i and time t:

$$Y_{it} = \alpha + \beta TREAT_i + \gamma POST_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

- Y_{it}= number of full-time equivalent employees working in establishment i, in state s (PA or NJ), in period t (Feb 1992, Nov 1992).
- $TREAT_i$= 0/1 variable equal to 1 for observations in NJ.
- $POST_t$=0/1 variable equal to 1 for observations in November 1992.
- $TREAT_s * POST_t$=interaction term equal to 1 in NJ in November 1992.

Differences in differences: regression model

- Regression model for entity i and time t:

$$Y_{it} = \alpha + \beta TREAT_i + \gamma POST_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

- A more flexible specification uses a dummy variable for each entity i and for each time period t

$$Y_{it} = \alpha + \sum_{j=1}^{N-1} \beta_j DEntity_{ji} + \sum_{p=1}^{T-1} \gamma_p DTime_{pt} \\ + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

- $DEntity_{ji}$ is a dummy variable equals to 1 when $i = j$ and zero otherwise (N is the number of entities – e.g. individuals, firms, counties, states, countries, etc.)
- $DTime_{pt}$ is a dummy variable equals to 1 when $t = p$ and zero otherwise (T is the number of time periods in the sample)

Regression DD with Entity and Time Fixed Effects (FE)

- These two models are identical when we have only two entities and two time periods
  - Example: minimum wage case with 2-by-2 DD table
    - Entities: NJ and PA
    - Time periods: February and November, 1992

$$Y_{it} = \alpha + \beta TREAT_i + \gamma POST_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

$$Y_{it} = \alpha + \beta_{NJ} D\_NJ_i + \gamma_{Nov} D\_Nov_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

  - β_{NJ} is what we call *state (entity) fixed effect* ($\beta_{NJ} = \beta$)
  - γ_{Nov} is what we call *month (time) fixed effect* ($\gamma_{Nov} = \gamma$)

Minimum Wage Case with 2-by-2 DD Table

First regression DD model

Entity	*Time*	Y_{it}	*Constant*	$Treat_i$	$Post_t$	$Treat_i \times Post_t$
NJ	Feb	20	1	1	0	0
NJ	Nov	21	1	1	1	1
PA	Feb	23	1	0	0	0
PA	Nov	21	1	0	1	0

Second regression DD model

Entity	*Time*	Y_{it}	*Constant*	D_NJ_i	D_Nov_t	$Treat_i \times Post_t$
NJ	Feb	20	1	1	0	0
NJ	Nov	21	1	1	1	1
PA	Feb	23	1	0	0	0
PA	Nov	21	1	0	1	0

Regression DD with Entity and Time Fixed Effects (FE)

- It's conventional to express the regression DD model with dummy variables with entity-specific coefficients and time- specific coefficients only, that is, we wouldn't write

$$Y_{it} = \alpha + \sum_{j=1}^{N-1} \beta_j DEntity_{ji} + \sum_{p=1}^{T-1} \gamma_p DTime_{pt} \\ + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

- Instead, we would express the regression DD model as

$$Y_{it} = \alpha + \beta_i + \gamma_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

  - β_i is a set of entity (e.g. state) fixed effects
  - γ_t is a set of time (e.g. month) fixed effects

Regression DD-FE

- Alright, let us express the regression DD model as

$$Y_{it} = \alpha + \beta_i + \gamma_t + \delta_{DD}(TREAT_i \times POST_t) + e_{it}$$

- A more flexible model would replace the interaction term $TREAT_i \times POST_t$ to allow
  1. the binary treatment to vary over time, or
  2. the treatment to be categorical (e.g. high, medium, low) or continuous – instead of binary (treatment vs. control)
  3. the treatment intensity – categorical or continuous – to vary over time

- To take all these cases into account, we express the regression DD model with fixed effects – regression DD-FE model – as

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

  - it is also conventional to write the FE's next to the error term

Minimum Wage Case - *Panel Data*

To run the regression DD model in Stata, we need a panel dataset

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

$$or \qquad Y_{it} = \alpha + \delta_{DD} Wage_{it} + \beta_i + \gamma_t + e_{it}$$

State	*Year*	Y_{it}	*DNJ*	*DNY*	*D93*	*D94*	$Treat_{it}$	$Wage_{it}$
NJ	1992	20	1	0	0	0	0	4.25
NJ	1993	21	1	0	1	0	1	5.05
NJ	1994	22	1	0	0	1	1	5.75
NY	1992	22	0	1	0	0	0	4.25
NY	1993	23	0	1	1	0	1	4.75
NY	1994	22	0	1	0	1	0	4.25
PA	1992	19	0	0	0	0	0	4.25
PA	1993	19	0	0	1	0	0	4.25
PA	1994	19	0	0	0	1	0	4.25

Regression DD-FE: Treatment at More Aggregated Level

- In the general regression DD model, the treatment happens at the entity i level

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

- Sometimes, however, the treatment happens at a higher entity level s, but we observe information at a more disaggregated level i. In this case, we can write the model as

$$Y_{ist} = \alpha + \delta_{DD} TREAT_{st} + \beta_i + \gamma_t + e_{ist}$$

  - for example, the minimum wage changes at the state level, but we observe information at the individual level

Panel (or FE) Regression

- It turns out that the regression DD-FE model solves the omitted variable bias problem for a number of *unobserved* explanatory variables

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

  - the set of fixed effects β_i controls for any variables associated with entity i that we do not observe, and do not vary over time, such as innate ability for individual i
  - the set of fixed effects γ_t controls for any unobserved variables that affect simultaneously all entities at each time t, such as macroeconomic conditions
  - recall that these FE's are dummy variables that incorporate any information – *observed* or *unobserved* – specific to entity i or time t

Panel (or FE) Regression

- The model below is more general than the regression DD-FE

$$Y_{it} = \alpha + \delta Z_{it} + \theta X_{it} + \beta_i + \gamma_t + e_{it}$$

  - Z_{it} is the explanatory variable of interest
  - X_{it} is a vector of other explanatory variables (*control* variables).

- Nevertheless, without a *natural* experiment (as in the DD setting – e.g. policy change), the evidence of this regression model may be less convincing than regression DD-FE model
  - Z_{it} may still be correlated to time-varying unobserved variables

- Anyways, the cost of this *panel-regression (or FE) model* is that we cannot estimate the coefficients of variables that are constant over time for entity i, such as gender or race, or vary over time equally for all entities, such as the Central Bank interest rate

Panel (or FE) Regression: FE Estimation

- To see how the panel regression model solves OVB for time-constant unobserved explanatory variables, let us discuss an alternative way to estimate that model
- To simplify the discussion, let us assume that we have a panel regression model with entity FE's only (no time FE's)

$$Y_{it} = \alpha + \delta Z_{it} + \theta X_{it} + \beta_i + e_{it}$$

- First, take the average over time for each entity (or within each group defined by the entity), for each variable in that equation, that is,

$$\bar{Y}_i = \alpha + \delta \bar{Z}_i + \theta \bar{X}_i + \beta_i + \bar{e}_i$$

- Now, subtract the second equation from the first equation

$$(Y_{it} - \bar{Y}_i) = \delta(Z_{it} - \bar{Z}_i) + \theta(X_{it} - \bar{X}_i) + (e_{it} - \bar{e}_i)$$

Panel (or FE) Regression: FE Estimation

- The resulting FE model uses only demeaned variables

$$(Y_{it} - \bar{Y}_i) = \delta(Z_{it} - \bar{Z}_i) + \theta(X_{it} - \bar{X}_i) + (e_{it} - \bar{e}_i)$$

  - it eliminates the intercept and entity FE's (and any other time-constant explanatory variables such as gender and race)
  - FE estimation: run an OLS regression on this demeaned model instead of the original model
  - FE estimation exploits only the variation in the demeaned variables – *within-group* variation
    - the R^2 of this demeaned model will reflect only the predictive power of the demeaned variables, not the variation explained by the original variables
    - that is, do not draw comparisons between the R^2 from OLS on the original model and this one – the *within-group* R^2

Estimation in Stata

- Panel-regression model (or FE model)

$$Y_{it} = \alpha + \delta Z_{it} + \theta X_{it} + \beta_i + e_{it}$$

  - first option: *reg Y Z X i.EntityVar*
  - second option: *areg Y Z X, absorb(EntityVar)*

- The FE model with demeaned variables

$$(Y_{it} - \bar{Y}_i) = \delta(Z_{it} - \bar{Z}_i) + \theta(X_{it} - \bar{X}_i) + (e_{it} - \bar{e}_i)$$

  - first option: *reg* $(Y - \bar{Y})$ $(Z - \bar{Z})$ $(X - \bar{X})$
  - second option:
    - *xtset EntityVar*
    - *xtreg Y Z X, fe*

Probing DD Assumptions: Impact of MLDA on Mortality

- The multistate regression DD model to estimate the impact of MLDA on death rates can be expressed as

$$Y_{st} = \alpha + \delta_{DD} Legal_{st} + \beta_s + \gamma_t + e_{st}$$

  - Y_{st} represents death rate in state s and year t
  - $Legal_{st}$ measures the proportion of 18-20-years-olds allowed to drink in state s and year t
  - the set of fixed effects β_s controls for any state-specific variables – *observed* and *unobserved* – that do not vary over time, such as topography, climate, and average road quality
  - the set of fixed effects γ_t controls for any *observed* and *unobserved* variables that affect simultaneously all states at each time t, such as interest rates, consumer expectations, and other macroeconomic conditions

Probing DD Assumptions: Impact of MLDA on Mortality

Table 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	.59	.47	1.49	1.26
	(.59)	(.79)	(.88)	(.89)
All internal causes	1.33	.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds. The table shows coefficients on the proportion of legal drinkers by state and year from models controlling for state and year effects. The models used to construct the estimates in columns (2) and (4) include state-specific linear time trends. Columns (3) and (4) show weighted least squares estimates, weighting by state population. The sample size is 714. Standard errors are reported in parentheses.

Probing DD Assumptions: Impact of MLDA on Mortality

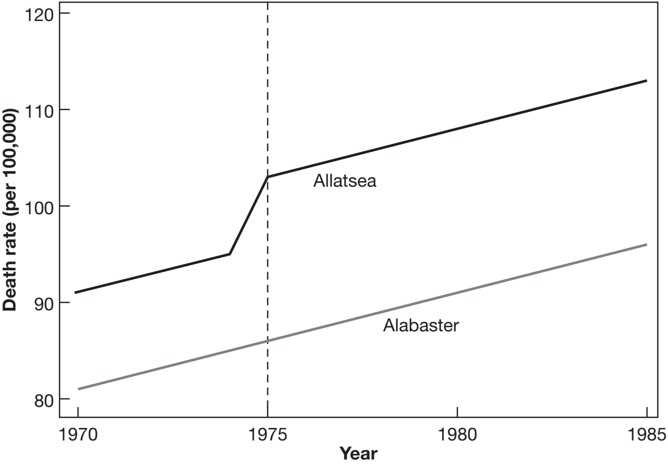
- Causal interpretation relies on *common trends* assumption
- Samples that include many states and years allow us to relax that assumption
  - i.e., allow us to introduce a degree of nonparallel evolution in outcomes between states in the absence of a treatment effect
- A regression DD model with controls for state-specific trends looks like

$$Y_{st} = \alpha + \delta_{DD} Legal_{st} + \beta_s + (\theta_s \times t) + \gamma_t + e_{st}$$

  - this model presumes that in the absence of a treatment effect, death rates in state s deviate from common year effects by following the linear trend captured by the coefficient θ_s

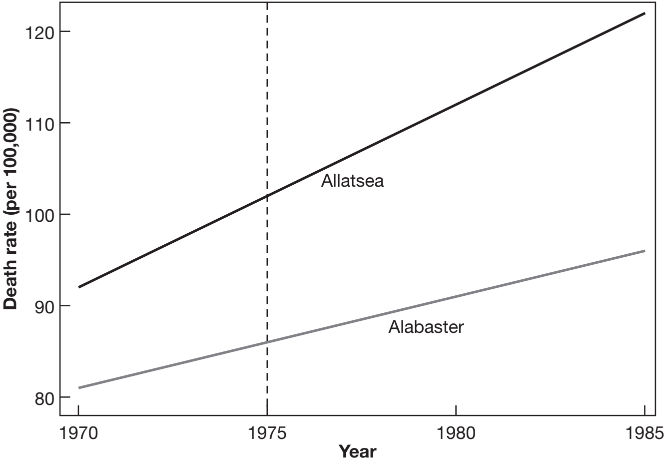
DD Estimate with Parallel Trends

Figure 5.4
An MLDA effect in states with parallel trends



Spurious DD Estimate: Trends Are *Not* Parallel

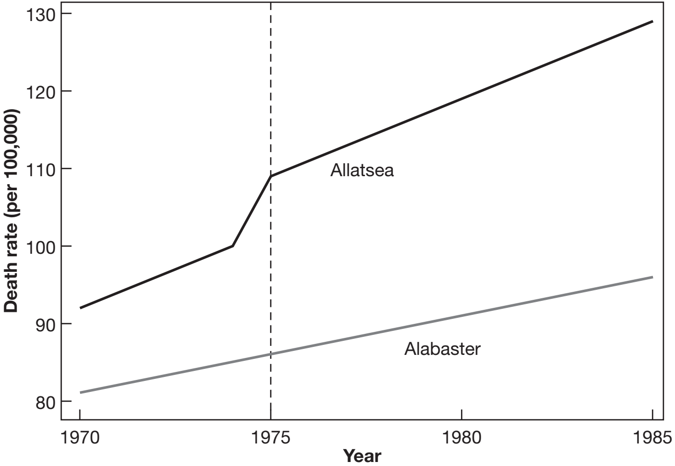
Figure 5.5
A spurious MLDA effect in states where trends are not parallel



DD Estimate with State-Specific Linear Trends

FIGURE 5.6

A real MLDA effect, visible even though trends are not parallel



DD Estimate with State-Specific Linear Trends

- In models that control for state-specific linear trends, evidence for MLDA effects comes from sharp deviations from otherwise smooth trends, even where the trends are not common
  - the coefficient on $Legal_{st}$ picks this up, while the model allows for the fact that death rates were on different trajectories from the get-go
- Models with state-specific linear trends provide an important check on the causal interpretation of any set of regression DD estimates using multiperiod data
  - for a coherent causal DD analysis of MLDA effects, the introduction of state-specific trends should have little effect on the regression DD estimates
  - notwithstanding, the addition of trends may increase standard errors – the loss of precision is due to the fact that treatment effects may emerge only gradually

DD Estimates with Competing Policy Changes

- State policymaking is a messy business, with frequent changes on many fronts
- DD estimates of MLDA effects, with or without state-specific trends, may be biased by contemporaneous policy changes in other areas
  - an important consideration in research on alcohol, for example, is the price of a drink
  - taxes are the most powerful tool the government uses to affect the price of beverages
  - many states levy a heavy tax on beer, measured in dollars per gallon of alcohol content
- Ideally, regression DD models that include controls for state beer taxes would generate MLDA estimates similar to those without such controls
  - if that's the case, the MLDA estimates are *robust* to the inclusion of a beer tax control

DD Estimates with Competing Policy Changes

TABLE 5.3
Regression DD estimates of MLDA effects controlling for beer taxes

	Without trends		With trends	
	Fraction legal	Beer tax	Fraction legal	Beer tax
Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.98	1.51	10.03	−5.52
	(4.69)	(9.07)	(4.92)	(32.24)
Motor vehicle accidents	7.59	3.82	6.89	26.88
	(2.56)	(5.40)	(2.66)	(20.12)
Suicide	.45	−3.05	.38	−12.13
	(.60)	(1.63)	(.77)	(8.82)
Internal causes	1.46	−1.36	.88	−10.31
	(1.61)	(3.07)	(1.81)	(11.64)

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds, controlling for state beer taxes. The table shows coefficients on the proportion of legal drinkers by state and year and the beer tax by state and year, from models controlling for state and year effects. The fraction legal and beer tax variables are included in a single regression model, estimated without trends to produce the estimates in columns (1) and (2) and estimated with state-specific linear trends to produce the estimates in columns (3) and (4). The sample size is 700. Standard errors are reported in parentheses.

Weights in difference-in-differences designs

- The DD estimates of MLDA effects were found giving all observations equal weight, as if data from each state were equally valuable
- States are not created equal, however, in at least one important respect – population
  - some states, like Texas and California, are bigger than most countries, while others, like Vermont and Wyoming, have populations smaller than those of many American cities
  - we may prefer estimates that reflect this fact by giving more populous states more weight
- We should then use the regression procedure called *weighted least squares* (WLS)
  - the standard OLS estimator fits a line by minimizing the sample average of squared residuals, with each squared residual getting equal weight in the sum
  - as the name suggests, WLS weights each term in the residual sum of squares by population size or some other researcher- chosen weight

Population Weighting Has Two Consequences

- *First*, population weighting generates a people-weighted average
  - regression models of treatment effects capture a weighted average of effects for the groups or cells represented in the dataset
  - in a state-year panel, these groups are states, so OLS produces estimates of average causal effects that ignore population size
    - resulting estimates are averages over states, not over people
  - with population weighting, causal effects for states like Texas get more weight than those for states like Vermont
- Population weighting may sound appealing, but it need not be
  - the typical citizen is more likely to live in Texas than Vermont, but changes in the Vermont MLDA provide variation that may be just as useful as changes in Texas
  - you should hope, therefore, that regression estimates from your state-year are not highly sensitive to weighting

Population Weighting Has Two Consequences

- *Second*, population weighting may also increase the precision of regression estimates
  - with far fewer drivers in Vermont than in Texas, MVA death rates in Vermont are likely to be more variable from year to year than those in Texas
  - in a statistical sense, the data from Texas are more reliable and therefore, perhaps, worthy of higher weight
- The best scenario is a set of findings (estimates and standard errors) that are reasonably insensitive to weighting
  - in Table 5.2, weighting by state population aged 18-20 does not matter much for the estimates of MLDA effects

DiD functional form assumptions

Despite its apparent closeness to the gold standard of randomized experiments, many inherent assumptions :

- Is the outcome variable in logs or levels? Paralell trends unlikely to hold in both.
- Treatment effect homogeneity over time? Issue in staggered designs.
- Does turning off a treatment have the same magnitude effect as turning on a treatment?
- With a continous treatment, is the effect of the treatment on the outcome linear ? Could test non-linearity by splitting the continous measure into quartiles and estimating the effect for the four different quartiles.

Treatment effect heterogeneity

Difference in differences : multiple periods and treatments

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

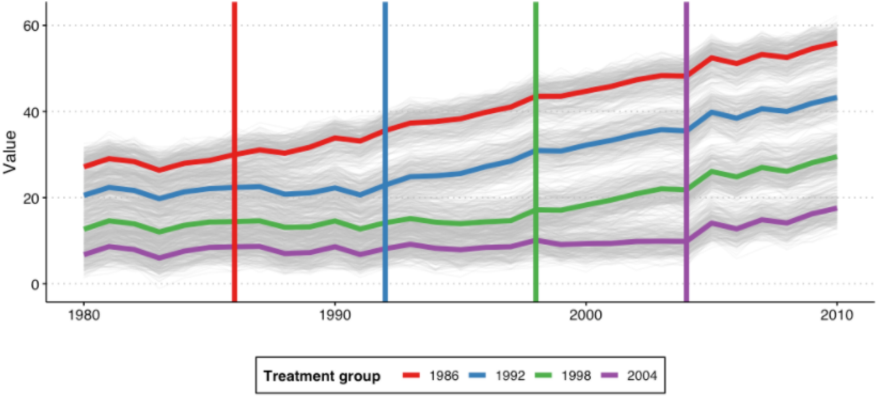


Figure source: Callaway and Sant'Anna (2020)

What do difference in difference estimates with multiple treatments measure?

$$Y_{it} = \alpha + \delta_{DD} TREAT_{it} + \beta_i + \gamma_t + e_{it}$$

- With constant treatment effects and paralell trends holding, we'd like the DD coefficient to measure the Average Treatment on the Treated:

$$\delta_{DD} = ATT = E(Y_{i,1}(1) - Y_{i,1}(0)|Treat = 1) =$$

E($Y_{i,1}(1) \mid Treat = 1) - E(Y_{i,1}(0) \mid Treat = 1)$

- Paralell trends assumption:

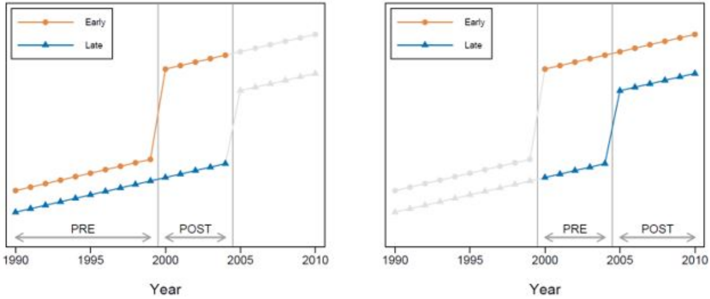
$$E[Y_{i,1}(0) - Y_{i,0}(0) \mid Treat = 1] = E[Y_{i,1}(0) - Y_{i,0}(0) \mid Treat = 0]$$

- $\widehat{ATT} = E[Y_1 - Y_0 \mid Treat = 1] - E[Y_1 - Y_0 \mid Treat = 0]$
- Constant TE assumption particularly problematic when treatments are canceled/reversed. But also if we have treatment heterogeneity within unit over time.

What do difference in difference estimates with multiple treatments measure?

If treatment effects are heterogeneous within unit, you won't be estimating exactly the ATT, several problems arise:

- If treatment effects vary over time within unit, this introduces biases in your overall DD estimate, because the already treated units serve as controls for units treated later.



- Andrew Goodman Bacon. Difference-in-Differences with variation in treatment timing, *Journal of Econometrics*, 2021.

Treatment effect heterogeneity issues and solutions

- With staggered treatments, your DD estimator is a weighted average of 2X2 comparisons. Weights are positive if effects are time-invariant, but can become negative under time-varying treatments.

$$DID = \Delta Y_{New\_treat} - \Delta Y_{Already\_treat} =$$

$$\mathrm{TE}_{NEW,t} - TE_{Already,t} + TE_{Already,t-1}$$

- OLS overweights entities with more variance in treatment status to achieve a more precise estimate of the treatment effect. (Units in the middle receive a higher weight- think of the binomial distribution variance: is it larger for a 50-50 control/treatment split or a 20-80 control/treatment split?). Implication: even if treatments are homogeneous, you will still not recover an average ATT, but a weighted average skewed towards the ATTs of the treatments in the middle.

Corrections: ongoing research

- stack DD estimates as in Cengiz et al. (2019): ensure no previously treated units enter as controls, for each treatment cohort create a dataset with k periods before and after treatment. Run TWFE regression with interactions for cohort-specific dataset with all FE and controls. Stata: **stackedev**
- Break down into 2X2 problems, address the fact that early treated units get overweighted, address multiple hypothesis testing problems
- **Callaway and Sant'Anna (2020)**: did in R, csdid in Stata. Subset the data into many 2x2. Staggered treatments turn on and need to stay on (in their implementation).
- Roth, Sant'Anna, Bilinkski and Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature"