

# **GSERM - St. Gallen 2024**

## Analyzing Panel Data

June 13, 2024

The goal: **Making causal inferences from observational data.**

- Establish and measure the *causal* relationship between variables in a non-experimental setting.

- The *fundamental problem of causal inference*:

*It is impossible to observe the causal effect of a treatment or a predictor on a single unit.*

- Specific challenges:
  - *Confounding*
  - *Selection bias*
  - *Heterogenous treatment effects*

# Causation and Counterfactuals

## Causal statements imply counterfactual reasoning.

- “If the cause(s) had been different, the outcome(s) would be different, too.”
- Conditioning, probabilistic and causal:

| Probabilistic conditioning                      | Causal conditioning                                      |
|---|--|
| $\Pr(Y X = x)$                                  | $\Pr[Y do(X = x)]$                                       |
| Factual   | Counterfactual   |
| Select a sub-population                         | Generate a new population                                |
| Predicts passive observation                    | Predicts active manipulation                             |
| Calculate from full DAG*                        | Calculate from surgically-altered DAG*                   |
| Always identifiable when X and Y are observable | Not always identifiable even when X and Y are observable |

\*See below. Source: Swiped from Shalizi, “Advanced Data Analysis from an Elementary Point of View”, Table 23.1.

- Causality (typically) implies / requires:
  - *Temporal ordering*
  - *Mechanism*
  - *Correlation*

# The Counterfactual Paradigm

## Notation

- $N$  observations indexed by  $i$ ,  $i \in \{1, 2, \dots, N\}$
- Outcome variable  $Y$
- Interest: the effect on  $Y$  of a treatment variable  $W$ :
  - $W_i = 1 \leftrightarrow$  observation  $i$  is “treated”
  - $W_i = 0 \leftrightarrow$  observation  $i$  is “control”

## Potential Outcomes

- $Y_{0i}$  = the value of  $Y_i$  if  $W_i = 0$
- $Y_{1i}$  = the value of  $Y_i$  if  $W_i = 1$
- $\delta_i = (Y_{1i} - Y_{0i})$  = the treatment effect of  $W$

The average treatment effect (ATE) is just:

$$\begin{aligned} \text{ATE} \equiv \bar{\delta} &= E(Y_{1i} - Y_{0i}) \\ &= \frac{1}{N} \sum_{i=1}^N Y_{1i} - Y_{0i}. \end{aligned}$$

BUT we observe only  $Y_i$ :

$$Y_i = \begin{cases} Y_{0i} & \text{if } W_i = 0, \\ Y_{1i} & \text{if } W_i = 1. \end{cases}$$

or (equivalently)

$$Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}.$$

# Estimating Treatment Effects

Key to estimating treatment effects: **Assignment mechanism for  $W$** .

Neyman/Rubin/Holland: Treat inability to observe  $Y_{0i}$  /  $Y_{1i}$  as a missing data problem.

[press “pause”]

Notation:

$$\mathbf{X}_{N \times K} \cup \{\mathbf{W}, \mathbf{Z}\}$$

**W** have some missing values,  
**Z** are “complete”

Consider a matrix **R** with:

$$R_{ik} = \begin{cases} 1 & \text{if } X_{ik} \text{ is missing,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\pi_{ik} = \Pr(R_{ik} = 1)$$

# Missing Data (continued)

## Rubin's flavors of missingness:

- Missing completely at random (“MCAR”) (= “ignorable”):

$$\mathbf{R} \perp \{\mathbf{Z}, \mathbf{W}\}$$

- Missing at random (“MAR”) (conditionally “ignorable”):

$$\mathbf{R} \perp \mathbf{W} | \mathbf{Z}$$

- Anything else is “informatively” (or “non-ignorably”) missing (“MNAR”).



# Rubin's Flavors Remix

Suppose we have two variables, an outcome  $Y$  and a covariate / predictor  $X$ . Define  $R_{(Y)}$  as the vector of missing data indicators for  $Y$  (analogously to above).

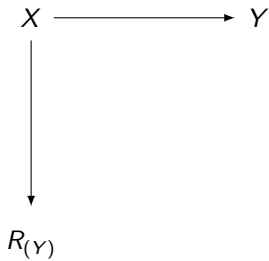
Then:



$R_{(Y)}$

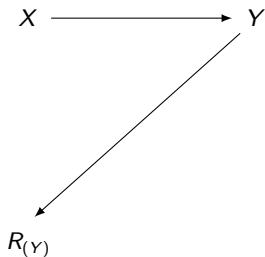
Missing Completely At Random (MCAR)

## Rubin Remixed (continued)



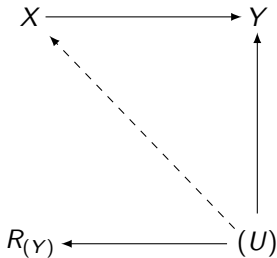
Missing At Random (MAR)

## Rubin Remixed (continued)



Missing Not At Random (MNAR)

# Missingness Due To Confounding



(Also) Missing Not At Random (MNAR)

[press “play”]

# Estimating Treatment Effects

Key to estimating treatment effects: **Assignment mechanism for  $W$** .

Neyman/Rubin/Holland: Treat inability to observe  $Y_{0i}$  /  $Y_{1i}$  as a missing data problem.

- If the “missingness” due to the value of  $W_i$  is orthogonal to the values of  $Y$ , then it is ignorable. Formally:

$$\Pr(W_i | \mathbf{X}_i, Y_{0i}, Y_{1i}) = \Pr(W_i | \mathbf{X}_i)$$

- If the “missingness” is non-orthogonal, then it is not ignorable, and can lead to bias in estimation
- Non-ignorable assignment of  $W$  requires understanding (and accounting for) the mechanism by which that assignment occurs

One more thing: the stable unit-treatment value assumption (“SUTVA”)

- Requires that there be two and only two possible values of  $Y$  for each observation  $i$ ...
- “the observation (of  $Y_i$ ) on one unit should be unaffected by the particular assignment of treatments to the other units.”
- $\equiv$  the “assumption of no interference between units,” meaning:
  - Values of  $Y$  for any two  $i, j$  ( $i \neq j$ ) observations do not depend on each other
  - Treatment effects (for observation  $i$ ) are *homogenous* within categories defined by  $W$

# Treatment Effects Under Randomization of $W$

If  $W_i$  is assigned randomly, then:

$$\Pr(W_i) \perp Y_{0i}, Y_{1i}$$

and so:

$$\Pr(W_i | Y_{0i}, Y_{1i}) = \Pr(W_i) \forall Y_{0i}, Y_{1i}.$$

This means that the “missing” data on  $Y_0/Y_1$  are ignorable (here, in the special case where the  $\mathbf{X}_i$  on which  $W_i$  depends is null). This in turn means that:

$$f(Y_{0i} | W_i = 0) = f(Y_{0i} | W_i = 1) = f(Y_i | W_i = 0) = f(Y_i | W_i = 1)$$

and

$$f(Y_{1i} | W_i = 0) = f(Y_{1i} | W_i = 1) = f(Y_i | W_i = 0) = f(Y_i | W_i = 1)$$

## Randomized $W$ (continued)

Implication:  $Y_{0i}$  and  $Y_{1i}$  are (not identical but) *exchangeable*...

This in turn means that:

$$E(Y_{0i}|W_i) = E(Y_{1i}|W_i)$$

and so

$$\begin{aligned}\widehat{ATE} &= E(Y_i|W_i = 1) - E(Y_i|W_i = 0) \\ &= \bar{Y}_{W=1} - \bar{Y}_{W=0}.\end{aligned}$$

will be an unbiased estimate of the ATE.



# Observational Data: $W$ Depends on $\mathbf{X}$

Formally,

$$Y_{0i}, Y_{1i} \perp W_i | \mathbf{X}_i.$$

Here,

- $\mathbf{X}$  are *known confounders* that (stochastically) determine the value of  $W_i$ ,
- Conditioning on  $\mathbf{X}$  is necessary to achieve exchangeability.

So long as  $W$  is entirely due to  $\mathbf{X}$ , we can condition:

$$f(Y_{1i} | \mathbf{X}_i, W_i = 1) = f(Y_{1i} | \mathbf{X}_i, W_i = 0) = f(Y_i | \mathbf{X}_i, W_i)$$

and similarly for  $Y_{0i}$ .

## W Depends on **X** (continued)

### Estimands:

- the *average treatment effect for the treated* (ATT):

$$ATT = E(Y_{1i}|W_i = 1) - E(Y_{0i}|W_i = 1).$$

- the *average treatment effect for the controls* (ATC):

$$ATC = E(Y_{1i}|W_i = 0) - E(Y_{0i}|W_i = 0).$$

### Corresponding estimates:

$$\widehat{ATT} = \mathbf{E}\{[E(Y_i|\mathbf{X}_i, W_i = 1) - E(Y_i|\mathbf{X}_i, W_i = 0)]|W_i = 1\}.$$

and

$$\widehat{ATC} = \mathbf{E}\{[E(Y_i|\mathbf{X}_i, W_i = 1) - E(Y_i|\mathbf{X}_i, W_i = 0)]|W_i = 0\}.$$

Note that in both cases **the expectation of the whole term is conditioned on  $W_i$ .**

Confounding occurs when one or more observed or unobserved factors  $\mathbf{X}$  affect the causal relationship between  $W$  and  $Y$ .

Formally, confounding requires that:

- $\text{Cov}(\mathbf{X}, W) \neq 0$  (the confounder is associated with the “treatment”)
- $\text{Cov}(\mathbf{X}, Y) \neq 0$  (the confounder is associated with the outcome)
- $\mathbf{X}$  does not “lie on the path” between  $W$  and  $Z$  (that is,  $\mathbf{X}$  is not affected by either  $W$  or  $Y$ ).

Directed acyclic graphs (DAGs) are a tool for visualizing and interpreting structural/causal phenomena.

- DAGs comprise:
  - Nodes (typically, variables / phenomena) and
  - Edges (or lines; typically, relationships/causal paths).
- Directed means each edge is *unidirectional*.
- Acyclical means exactly what it suggests: If a graph has a “feedback loop,” it is not a DAG.
- Read more at the [Wikipedia page](#), or at this useful [page](#).

# Know your DAG

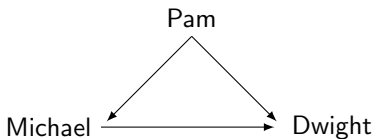


Figure: A DAG

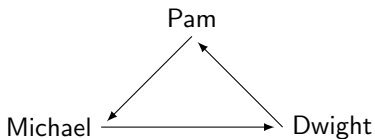
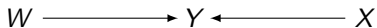


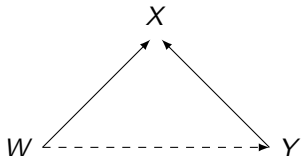
Figure: Not a DAG

# DAGs and Confounding

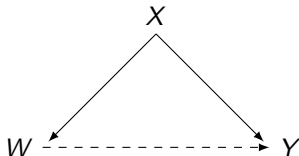
No Confounding



A "Collider"



Confounding



# What We're On About

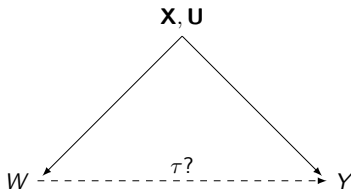


Figure: Potential Confounding

Here:

- $Y$  is the outcome of interest,
- $W$  is the primary predictor / covariate ("treatment") of interest,
- $T_i$  is the "treatment indicator" for observation  $i$ ,
- We're interested in estimating  $\tau$ , the "treatment effect" of  $W$  on  $Y$ ,
- $\mathbf{X}$  are observed confounders,
- $\mathbf{U}$  are unobserved confounders.

- **Randomize**

(or...)

- Instrumental Variables Approaches
- Selection on Observables:
  - Regression / Weighting
  - Matching (propensity scores, multivariate/minimum-distance, genetic, etc.)
- Regression Discontinuity Designs (“RDD”)
- Differences-In-Differences (“DiD”)
- Synthetic Controls
- Others...



# Under Randomization

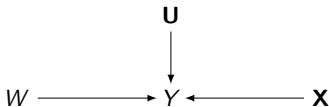


Figure: = no confounding!

## Note:

- Randomized assignment of  $W$  “balances” covariate values – both observed and unobserved – *on average*...
- That is, under randomization of  $W$ :

$$E(\mathbf{X}_i, \mathbf{U}_i \mid W_i = 0) = E(\mathbf{X}_i, \mathbf{U}_i \mid W_i = 1)$$

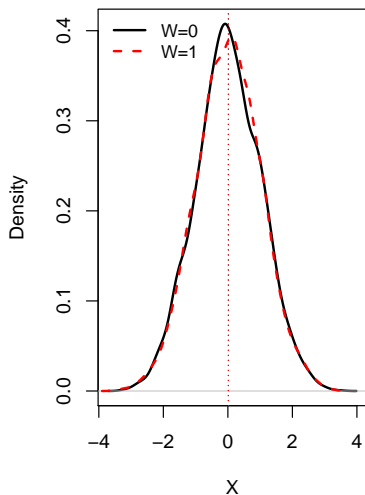
or, more demandinglly,

$$E[f(\mathbf{X}, \mathbf{U}) \mid W_i = 0] = E[f(\mathbf{X}, \mathbf{U}) \mid W_i = 1]$$

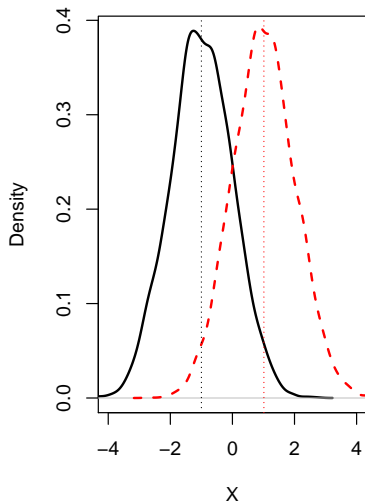
- Can yield imbalance by random chance...

# Covariate Balance / Imbalance

**Balanced X**



**Unbalanced X**



# Nonrandom Assignment of $W_i$

Valid causal inference requires  $Y_{0i}, Y_{1i} \perp W_i | \mathbf{X}_i, \mathbf{U}_i$

- That is, treatment assignment  $W_i$  is *conditionally ignorable*

## “What if I have unmeasured confounders?”

- In general, that's a bad thing.
- One approach: obtain *bounds* on possible values of  $\tau$ 
  - Assume you have one or more unmeasured confounders
  - Undertake one of the methods described below to get  $\hat{\tau}$
  - Calculate the range of values for  $\hat{\tau}$  that could occur, depending on the degree and direction of confounding bias
  - Or ask: How strong would the effect of the  $\mathbf{U}$ s have to be to make  $\hat{\tau} \rightarrow 0$ ?
- Some useful cites:
  - Rosenbaum and Rubin (1983)
  - Rosenbaum (2002)
  - DiPrete and Gangl (2004)
  - Liu et al. (2013)
  - Ding and VanderWeele (2016)

# Digression: Instrumental Variables

A DAG:

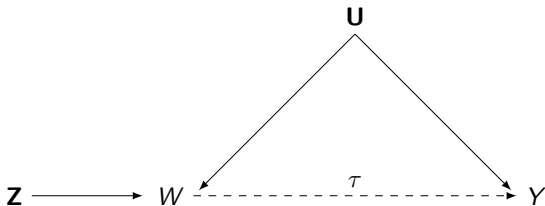


Figure: Instrumental Variables

As in the more general regression case where we have  $\text{Cov}(\mathbf{X}, \mathbf{u}) \neq 0$ , instrumental variables can be used to address confounding in causal analyses.

# Instrumental Variables (continued)

## Considerations:

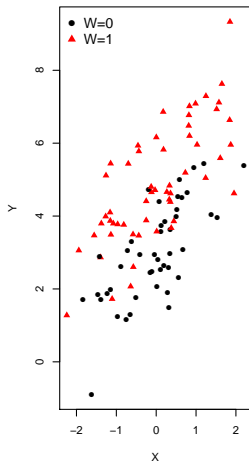
- Requires:
  1.  $\text{Cov}(\mathbf{Z}, W) \neq 0$
  2.  $\mathbf{Z}$  has no independent effect on  $Y$ , except through  $W$
  3.  $\mathbf{Z}$  is exogenous [i.e.,  $\text{Cov}(\mathbf{Z}, \mathbf{U}) = 0$ ]
- Arguably most useful when treatment compliance is uncertain / driven by unmeasured factors (“intent to treat” analyses)
- Mostly, they’re not that useful at all...
  - [Bound et al. \(1995\)](#): Weak instruments are worse than endogeneity bias
  - [Young \(2020\)](#): Inferences in published IV work (in economics) are wrong and terrible
  - [Shalizi \(2020, chapters 20-21\)](#): Gathers all the issues together, sometimes hilariously
- Other useful references:
  - [Imbens et al. \(1996\)](#) (the overly-cited one)
  - [Hernan and Robins \(2006\)](#) (making sense of things)
  - [Lousdal \(2018\)](#) (a good intuitive introduction)

# Nonrandom Assignment of $W_i$ (continued)

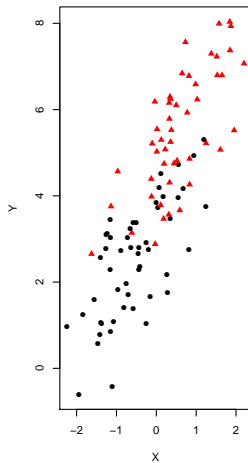
So...

- Causal inference with observational data typically requires that  $\mathbf{U} = \emptyset$ ...
- This typically requires a strong theoretical motivation in order to assume that the specification conditioning on the observed  $\mathbf{X}$  exhausts the list of possible confounders.
- **Even if** this assumption is reasonable, there are two (related) important concerns:
  - Lack of *covariate balance* (as above)
  - Lack of *overlap* among observations with  $W_i = 0$  vs.  $W_i = 1$
  - The latter is related to *positivity*, the requirement that each observation's probability of receiving (or not receiving) the treatment is greater than zero

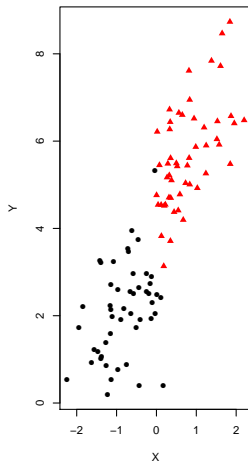
**Complete Overlap**



**Moderate Overlap**



**No Overlap**



## In general:

- Ensuring overlap allows us to make counterfactual statements from observational data
  - Requires that we have comparable  $W_i = 0$  and  $W_i = 1$  units
  - It's *necessary* – no overlap means any counterfactual statements are based on assumption
  - Think of this as an aspect of *model identification* (Crump et al. 2009)
  - Most often handled via matching
- Ensuring covariate balance corrects potential bias in  $\hat{\tau}$  due to (observed) confounding
  - This can be done a number of different ways: stratification, weighting, regression...
  - Key: Adjusting for (observable) differences across groups defined by values of  $W$
- In general, we usually address overlap first, then balance...



Matching is a way of dealing with one or both of covariate overlap and (im)balance.

The process, generally:

1. Choose the  $\mathbf{X}$  on which the observations will be matched, and the matching procedure;
2. Match the observations with  $W_i = 0$  and  $W_i = 1$ ;
3. Check for balance in  $\mathbf{X}_i$ ; and
4. Estimate  $\hat{\tau}$  using the matched pairs.

Variants / considerations:

- 1:1 vs. 1:k matching
- “Greedy” vs. “Optimal” matching (see [Gu and Rosenbaum 1993](#))
- Distances, calipers, and “common support”
- Post-matching: Balance checking...

- Simplest: Exact Matching

- For each of the  $n$  observations  $i$  with  $W = 1$ , find a corresponding observation  $j$  with  $W = 0$  that has identical values of  $\mathbf{X}$
- Calculate  $\hat{\tau} = \frac{1}{n} \sum (Y_i - Y_j)$
- Generally not practical, especially for high-dimensional  $\mathbf{X}$
- Variants: “coarsened” exact matching (e.g., [Iacus et al. 2011](#))

- Multivariate Matching

- Match each observation  $i$  which has  $W = 1$  with a corresponding observation  $j$  with  $W = 0$ , and whose values on  $\mathbf{X}_j$  are the most similar to  $\mathbf{X}_i$
- One example: Mahalanobis distance matching, based on the distance:

$$d_M(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)' \mathbf{S}^{-1} (\mathbf{X}_i - \mathbf{X}_j)}.$$

# Flavors of Matching (continued)

- Propensity Score Matching
  - Match observation  $i$  which has  $W = 1$  with observation  $j$  having  $W = 0$  based on the closeness of their *propensity score*
  - The propensity score is,  $\Pr(W_i = 1|\mathbf{X}_i)$ , typically calculated as the predicted value of  $T_i$  (the treatment indicator) from a logistic (or other) regression of  $T$  on  $\mathbf{X}$ .
  - The assumptions about matching [that  $Y$  is orthogonal to  $W|\mathbf{X}$  and that  $\Pr(W_i = 1|\mathbf{X}_i) \in (0, 1)$ ] mean that  $Y \perp W | \Pr(T|\mathbf{X})$ .
  - In practice: [read this...](#)
- Other variants: Genetic matching ([Diamond and Sekhon 2013](#)), etc.<sup>1</sup>

---

<sup>1</sup>Shalizi (2016) notes that "(A)pproximate matching is implicitly doing nonparametric regression by a nearest-neighbor method," and that "(M)aybe it is easier to get doctors and economists to swallow "matching" than "nonparametric nearest neighbor regression"; this is not much of a reason to present the subject as though nonparametric smoothing did not exist, or had nothing to teach us about causal inference."

Interestingly, quite a few of the good matching programs written for R have been written by political scientists...

- the `Match` package (does propensity score,  $M$ -distance, and genetic matching, plus balance checking and other diagnostics)
- the `MatchIt` package (for pre-analysis matching; also has nice options for checking balance)
- the `optmatch` package (suite for 1:1 and 1: $k$  matching via propensity scores,  $M$ -distance, and optimum balancing)
- `matching` (in the `arm` package)

# Regression Discontinuity Designs

“RDD”:

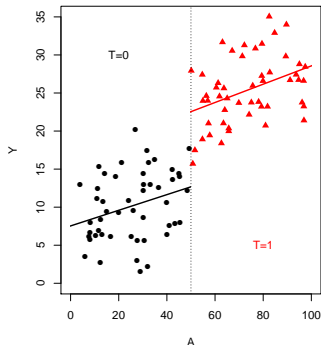
- Treatment changes abruptly [usually at some threshold(s)] according to the value(s) of some measured, continuous, pre-treatment variable(s)
  - This is known as the “assignment” or “forcing variable(s),” sometimes denoted **A**
  - Formally:

$$W_i = \begin{cases} 0 & \text{if } A_i \leq c \\ 1 & \text{if } A_i > c \end{cases}$$

- Intuition: Observations near but on either side of the threshold(s) are highly comparable, and can be used to (locally) identify  $\tau$
- This is because variation in  $W_i$  near the threshold is effectively random (a “local randomized experiment”)
- E.g. [Carpenter and Dobkin \(2011\)](#) (on the relationship between the legal drinking age and public health outcomes like accidental deaths)

# RDD (continued)

- Pluses:
  - Can be estimated straightforwardly, as:
$$Y_i = \beta_0 + \beta_1 A_i + \tau W_i + \gamma A_i W_i + \epsilon_i$$
  - Generally requires fewer assumptions than IV or DiD (and those assumptions are easier to observe and test)
- Minuses:
  - Provides only an estimate of a local treatment effect
  - Fails if (say) subjects can manipulate  $A$  in the vicinity of  $c$
- [Lee and Lemieux \(2010\)](#) is an excellent (if fanboi-ish) review
- R packages: `rddtools`, `rdd`, `rdrobust`, `rdpower`, `rdmulti`



# Panel Data Approaches: Differences-In-Differences

“DiD”:

- Leverages two-group, two-period data ( $T = 2$ ):

|                       | Pre-Treatment<br>( $T = 0$ ) | Post-Treatment<br>( $T = 1$ ) |
|-----------------------|------------------------------|-------------------------------|
| Treated ( $W = 1$ )   | A                            | B                             |
| Untreated ( $W = 0$ ) | C                            | D                             |

- Process (simple version):
  - Calculate the pre- vs. post-treatment difference for the treated group ( $B - A$ )
  - Calculate the pre- vs. post-treatment difference for the untreated group ( $D - C$ )
  - Calculate the differences between the differences [ $DiD = (B - A) - (D - C)$ ]
  - This is the same as fitting the regression:

$$Y_{it} = \beta_0 + \beta_1 W_{it} + \beta_2 T_{it} + \beta_3 W_{it} T_{it} + u_{it}$$

- Validity depends on (a) all the usual assumptions required by OLS, plus (b) the parallel trends assumption – that there are no time-varying differences between the two groups as we go from  $T = 0$  to  $T = 1$ .
- Resources:
  - Our old friend [Wikipedia](#)
  - Pischke's [slides on DiD](#)
  - R: package [did](#)
  - Stata: [ieddtab](#) in the [ietoolkit](#)

# Panel Data Approaches: Synthetic Controls

The “synthetic control method” (SCM):

- Addresses situations in which we have a single treated case (or small number of them)...
- Requires at least one (and ideally more) repeated measurements over time on the outcome of interest, and
- Also requires multiple (but not *too* many) non-treated cases
- Assumptions:
  - Possible control units are similar
  - Lack of spillover between treated and potential control units
  - Lack of exogenous shocks to potential control units



# Synthetic Controls (continued)

## SCM details:

- Intuition:
  - Create a counterfactual “control” unit that is as similar to the (pre-treatment) treated case as possible
  - Do so by weighting the observed predictors across “control” cases to minimize the difference (in a MSE sense)
  - Compare the pre-treatment trends in the synthetic control and treated cases
  - The weights are then used to create a post-treatment trend for the synthetic control
  - Inference is via placebo methods (varying the timing of the intervention)
- Advantages:
  - Works with (very) small  $N$
  - Doesn't require parallel trends (a la DiD)
  - Abadie et al. claim that SCM controls for both observed and unobserved time-varying confounders
- A few references:
  - A nontechnical [introduction](#) in the *BMJ*
  - [Method of the Month](#) Blog
  - The [Development Impact](#) blog post on SCM

In general:

- R
  - Packages for matching are listed above (Matching, MatchIt, etc.)
  - Similarly for RDD (rddtools, rdd, etc.) and DiD (did)
  - IV regression: ivreg (in AER), tsls (in sem), others
  - Synthetic controls are in Synth and MicroSynth
  - See generally the CRAN Task View on *Causal Inference*.
- Stata also has a large suite of routines for attempting causal inference with observational data...
- And there's a pretty good NumPy/SciPy-dependent package for Python, called (creatively) *CausalInference*

# Causal Inference: One-Way (FE) Models

Imai and Kim (2019):

- The punch line first: “(t)he ability of unit fixed effects regression models to adjust for unobserved time-invariant confounders comes at the expense of dynamic causal relationships between treatment and outcome variables.”
- Also dependent on functional form assumptions (specifically, linearity)

Intuition: For the model:

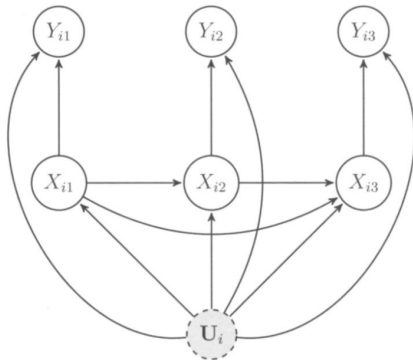
$$Y_{it} = \mathbf{X}_{it}\beta + \alpha_i + u_{it}$$

where (for simplicity)  $X$  is a binary treatment for which we want to know a causal effect on  $Y$ :

- Identification is via  $\text{Cov}[(\mathbf{X}_{it}, \alpha_i), u_{it}] = 0$
- In this framework,  $\beta = \tau$ , the typical causal estimand (that is, the expected difference between  $Y_{it}(0)$  and  $Y_{it}(1)$ )

A more flexible approach is to think of a FE model as a DAG...

# Fixed-Effects DAG



Source: Imai and Kim (2019).

Summarizing Imai and Kim (2019):

- Three key identifying assumptions for FE models:
  - No unobserved time-varying confounders
  - Past treatments / values of  $\mathbf{X}$  do not affect current values of  $Y^2$
  - Past outcomes  $Y$  do not affect current values of  $\mathbf{X}$ .
- Alternatively, one can select on observables (a la Blackwell and Glynn 2018) and model dynamics (albeit at the cost of failing to control for unobserved time-constant confounders).

*“...researchers must choose either to adjust for unobserved time-invariant confounders through unit fixed effects models or to model dynamic causal relationships between treatment and outcome under a selection-on-observables approach. No existing method can achieve both objectives without additional assumptions” (Imai and Kim 2019, 484).*

---

<sup>2</sup>Can be relaxed via IV, but that requires independence of past and present values of  $Y$ .

Imai and Kim redux (2020):

- In the simple  $T = 2$  case, DiD is equivalent to a two-way FE model:

$$Y_{it} = \mathbf{X}_{it}\beta + \alpha_i + \eta_t + u_{it}$$

- Imai & Kim: The same is not true for  $T > 2$ ...
- More important: two-way FEs' ability to control for unmeasured confounders depends on the (linearity of the) functional form...
- Upshot: two-way FEs aren't a (nonparametric) cure-all...
- Related: When we control for both  $\alpha_i$  and  $\eta_t$ , what – exactly – is the counterfactual?

# Back To The WDI

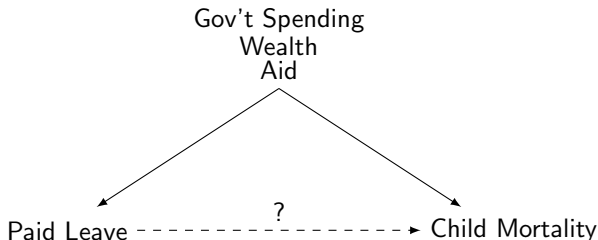
```
> describe(WDI,fast=TRUE,ranges=FALSE,check=TRUE)
```

|                         | vars | n     | mean            | sd               | skew  | kurtosis | se             |
|-------------------------|------|-------|-----------------|------------------|-------|----------|----------------|
| IS03                    | 1    | 13760 | NaN             | NA               | NA    | NA       | NA             |
| Year                    | 2    | 13760 | NaN             | NA               | NA    | NA       | NA             |
| Region                  | 3    | 13760 | NaN             | NA               | NA    | NA       | NA             |
| country                 | 4    | 13760 | NaN             | NA               | NA    | NA       | NA             |
| iso3c                   | 5    | 13760 | NaN             | NA               | NA    | NA       | NA             |
| RuralPopulation         | 6    | 13482 | 48.28           | 25.74            | -0.12 | -1.00    | 0.22           |
| UrbanPopulation         | 7    | 13482 | 51.72           | 25.74            | 0.12  | -1.00    | 0.22           |
| BirthRatePer1K          | 8    | 12937 | 28.02           | 13.08            | 0.21  | -1.25    | 0.12           |
| FertilityRate           | 9    | 12779 | 3.91            | 2.00             | 0.38  | -1.23    | 0.02           |
| PrimarySchoolAge        | 10   | 10896 | 6.14            | 0.61             | -0.04 | 0.11     | 0.01           |
| LifeExpectancy          | 11   | 12766 | 64.63           | 11.29            | -0.73 | -0.03    | 0.10           |
| AgeDepRatioOld          | 12   | 13515 | 10.70           | 7.04             | 1.74  | 4.57     | 0.06           |
| ChildMortality          | 13   | 11092 | 74.32           | 77.17            | 1.46  | 1.67     | 0.73           |
| GDP                     | 14   | 10099 | 250284546944.35 | 1140901242824.16 | 11.01 | 146.93   | 11352953710.99 |
| GDPPerCapita            | 15   | 10103 | 12112.45        | 19135.45         | 3.19  | 14.79    | 190.38         |
| GDPPerCapGrowth         | 16   | 10074 | 1.95            | 6.21             | 1.84  | 47.90    | 0.06           |
| TotalTrade              | 17   | 8622  | 78.38           | 53.99            | 2.99  | 17.71    | 0.58           |
| FDIIn                   | 18   | 8484  | 5.49            | 45.03            | 15.71 | 572.23   | 0.49           |
| NetAidReceived          | 19   | 9043  | 506951242.00    | 997064633.65     | 8.32  | 157.34   | 10484966.48    |
| MobileCellSubscriptions | 20   | 10212 | 36.32           | 51.76            | 1.29  | 1.14     | 0.51           |
| NaturalResourceRents    | 21   | 9211  | 6.85            | 11.06            | 2.60  | 8.04     | 0.12           |
| GovtExpenditures        | 22   | 8280  | 16.33           | 8.23             | 3.82  | 34.97    | 0.09           |
| WomenInLegislature      | 23   | 4706  | 17.76           | 11.73            | 0.72  | 0.12     | 0.17           |
| PaidParentalLeave       | 24   | 10152 | 0.11            | 0.31             | 2.50  | 4.27     | 0.00           |
| PostColdWar             | 25   | 13760 | 0.53            | 0.50             | -0.13 | -1.98    | 0.00           |
| lnGDPPerCap             | 26   | 10103 | 8.38            | 1.50             | 0.12  | -0.88    | 0.01           |
| lnNetAidReceived        | 27   | 8876  | 18.81           | 1.97             | -1.06 | 1.99     | 0.02           |
| YearNumeric             | 28   | 13760 | 1991.50         | 18.47            | 0.00  | -1.20    | 0.16           |

# A New Question

## Do paid parental leave policies decrease child mortality?

- $Y = \text{ChildMortality}$  ( $N$  of deaths of children under 5 per 1000 live births) (**logged**)
- $T = \text{PaidParentalLeave}$  (1 if provided, 0 if not)
- $X_s$ :
  - $\text{GDPPerCapita}$  (Wealth; in constant \$US) (logged)
  - $\text{NetAidReceived}$  (Net official development aid received; in constant \$US) (logged)
  - $\text{GovtExpenditures}$  (Government Expenditures, as a percent of GDP)





# Preliminary Regressions

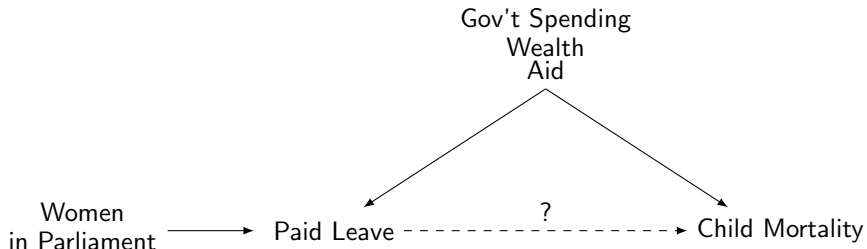
Table: Models of log(Child Mortality)

|                         | Bivariate OLS     | OLS                  | One-Way FE           | Two-Way FE           | FE w.Lagged Y        |
|-------------------------|-------------------|----------------------|----------------------|----------------------|----------------------|
| Paid Parental Leave     | -1.820<br>(0.035) | -0.900***<br>(0.037) | -0.066<br>(0.043)    | -0.139***<br>(0.024) | -0.206***<br>(0.027) |
| ln(GDP Per Capita)      |                   | -0.683***<br>(0.009) | -1.120***<br>(0.017) | -0.291***<br>(0.013) | -0.564***<br>(0.012) |
| ln(Net Aid Received)    |                   | -0.082***<br>(0.007) | -0.096***<br>(0.006) | 0.005<br>(0.004)     | -0.003<br>(0.004)    |
| Government Expenditures |                   | -0.002*<br>(0.001)   | 0.001<br>(0.001)     | 0.002***<br>(0.001)  | -0.0003<br>(0.001)   |
| Lagged Child Mortality  |                   |                      |                      |                      | 0.009***<br>(0.0001) |
| Constant                | 3.790*<br>(0.011) | 10.900***<br>(0.179) |                      |                      |                      |
| Observations            | 9,357             | 5,110                | 5,110                | 5,110                | 5,106                |
| R <sup>2</sup>          | 0.224             | 0.585                | 0.486                | 0.118                | 0.809                |
| Adjusted R <sup>2</sup> | 0.224             | 0.585                | 0.471                | 0.082                | 0.803                |

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

# Instrumental Variables

Conceptually:



# Instrumental Variables (continued)

Assessing  $\text{Cov}(W, Z)$ :

```
> with(WDI, t.test(WomenInLegislature ~ PaidParentalLeave))
```

Welch Two Sample t-test

data: WomenInLegislature by PaidParentalLeave

t = -21, df = 1330, p-value <0.0000000000000002

alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0

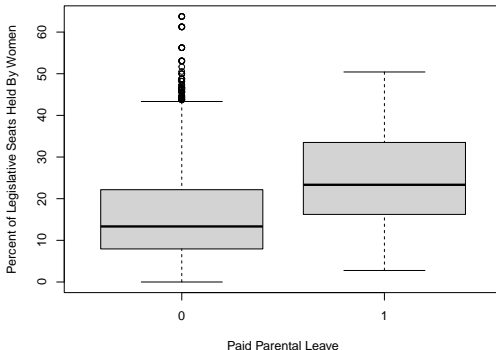
95 percent confidence interval:

-9.46 -7.84

sample estimates:

mean in group 0 mean in group 1

16.0 24.6



# Instrumental Variables: Syntax

E.g., one-way fixed effects with IV:

```
FE.IV<-plm(lnCM~PaidParentalLeave+log(GDPPerCapita)+  
            log(NetAidReceived)+GovtExpenditures |  
            . - PaidParentalLeave+WomenInLegislature,  
            data=WDI,effect="individual",model="within")
```

# Instrumental Variable Results

Table: IV Models of log(Child Mortality)

|                         | OLS                | One-Way FE           | FE w/IV              | RE w/IV              |
|-------------------------|--------------------|----------------------|----------------------|----------------------|
| Paid Parental Leave     | -0.900<br>(0.037)  | -0.066<br>(0.043)    | 131.000<br>(466.000) | -5.210**<br>(2.110)  |
| ln(GDP Per Capita)      | -0.683<br>(0.009)  | -1.120***<br>(0.017) | -21.800<br>(73.600)  | -0.510***<br>(0.093) |
| ln(Net Aid Received)    | -0.082<br>(0.007)  | -0.096***<br>(0.006) | 1.780<br>(6.600)     | -0.041<br>(0.028)    |
| Government Expenditures | -0.002<br>(0.001)  | 0.001<br>(0.001)     | -0.080<br>(0.283)    | -0.002<br>(0.003)    |
| Constant                | 10.900*<br>(0.179) |                      |                      | 8.880***<br>(1.010)  |
| Observations            | 5,110              | 5,110                | 2,630                | 2,630                |
| R <sup>2</sup>          | 0.585              | 0.486                | 0.00000              | 0.259                |
| Adjusted R <sup>2</sup> | 0.585              | 0.471                | -0.058               | 0.258                |

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

# Matching: Checking Covariate Balance

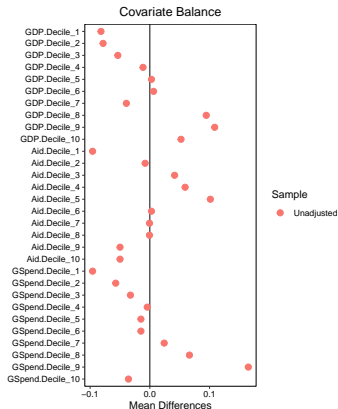
```
> # Subset data a little bit:

> vars<-c("ISO3", "Year", "Region", "country", "UrbanPopulation", "FertilityRate",
+         "PrimarySchoolAge", "ChildMortality", "lnGDPPerCap",
+         "lnNetAidReceived", "NaturalResourceRents", "GovtExpenditures",
+         "PaidParentalLeave", "PostColdWar", "lnCM")
> wdi<-WDI[vars]
> wdi<-na.omit(wdi)

> # Create discrete-valued variables (i.e., coarsen) for
> # matching on continuous predictors:

> wdi$GDP.Decile<-as.factor(ntile(wdi$GDPPerCapita,10))
> wdi$Aid.Decile<-as.factor(ntile(wdi$NetAidReceived,10))
> wdi$GSpent.Decile<-as.factor(ntile(wdi$GovtExpenditures,10))

> # Pre-match balance statistics...
>
> BeforeBal<-bal.tab(PaidParentalLeave~GDP.Decile+
+                   Aid.Decile+GSpent.Decile,data=wdi,
+                   stats=c("mean.diffs", "ks.statistics"))
```



# Exact Matching

```
> M.exact <- matchit(PaidParentalLeave~GDP.Decile+Aid.Decile+
+                   GSpending.Decile,data=wdi,method="exact")
> summary(M.exact)
```

Call:

```
matchit(formula = PaidParentalLeave ~ GDP.Decile + Aid.Decile +
        GSpending.Decile, data = wdi, method = "exact")
```

Summary of Balance for All Data:

.  
.  
.

Sample Sizes:

|               | Control | Treated |
|---------------|---------|---------|
| All           | 4734    | 302     |
| Matched (ESS) | 346     | 287     |
| Matched       | 898     | 287     |
| Unmatched     | 3836    | 15      |
| Discarded     | 0       | 0       |

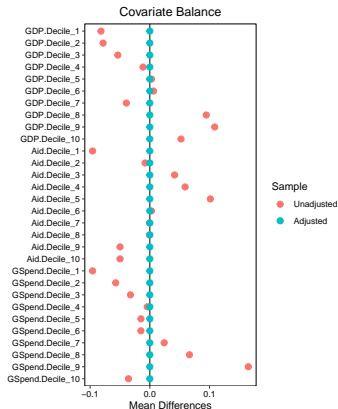
```
> # Create matched data:
```

```
>
```

```
> wdi.exact <- match.data(M.exact,group="all")
```

```
> dim(wdi.exact)
```

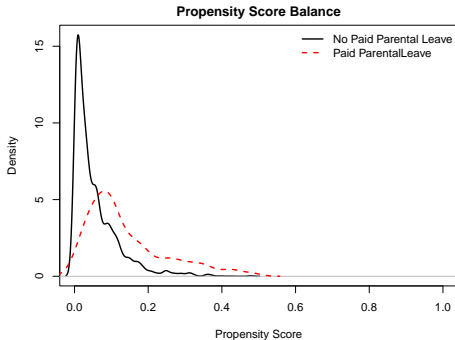
```
[1] 1185 20
```



# Propensity Scores

```
> PS.fit<-glm(PaidParentalLeave~GDP.Decile+Aid.Decile+
+             GSpending.Decile,data=wdi,
+             family=binomial(link="logit"))

> # Generate scores & check common support:
>
> PS.df<-data.frame(PS = predict(PS.fit,type="response"),
+                   PaidParentalLeave=PS.fit$model$PaidParentalLeave)
```





# Propensity Score Matching

```
> M.prop <- matchit(PaidParentalLeave~GDP.Decile+Aid.Decile+
+                   GSpent.Decile,data=wdi,method="nearest",
+                   ratio=3)
> summary(M.prop)
```

```
Call:
matchit(formula = PaidParentalLeave ~ GDP.Decile + Aid.Decile +
        GSpent.Decile, data = wdi, method = "nearest", ratio = 3)
```

Summary of Balance for All Data:

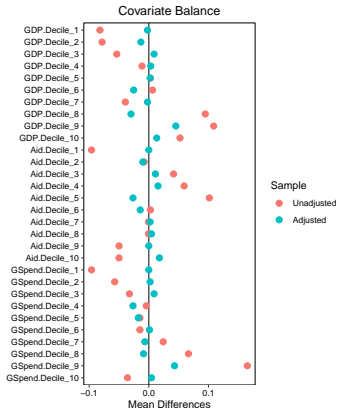
.  
.
.

Sample Sizes:

|           | Control | Treated |
|-----------|---------|---------|
| All       | 4734    | 302     |
| Matched   | 906     | 302     |
| Unmatched | 3828    | 0       |
| Discarded | 0       | 0       |

```
> # Matched data:
```

```
> wdi.ps <- match.data(M.prop,group="all")
> dim(wdi.ps)
[1] 1208  21
```



# “Optimal” Matching

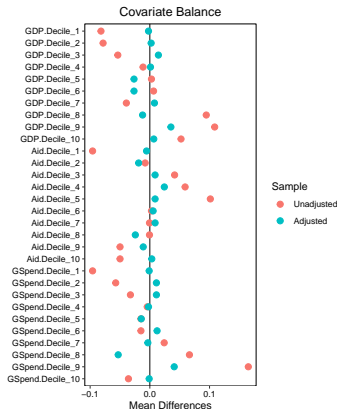
```
> M.opt <- matchit(PaidParentalLeave ~ GDP.Decile + Aid.Decile +  
+                 GSpending.Decile, data = wdi, method = "optimal",  
+                 ratio = 3)  
> summary(M.opt)
```

```
Call:  
matchit(formula = PaidParentalLeave ~ GDP.Decile + Aid.Decile +  
        GSpending.Decile, data = wdi, method = "optimal", ratio = 3)
```

Summary of Balance for All Data:

```
.  
.  
.  
Sample Sizes:  
      Control Treated  
All      4734      302  
Matched   906      302  
Unmatched 3828       0  
Discarded  0       0
```

```
> # Matched data:  
>  
> wdi.opt <- match.data(M.opt, group = "all")  
> dim(wdi.opt)  
[1] 1208  21
```



# Post-Matching Regressions

Table: FE Models of log(Child Mortality): Matched Data

|                         | Pre-Matching         | Exact                | Prop. Score          | Optimal              |
|-------------------------|----------------------|----------------------|----------------------|----------------------|
| Paid Parental Leave     | −0.065<br>(0.044)    | −0.155**<br>(0.060)  | −0.148***<br>(0.054) | −0.171***<br>(0.052) |
| ln(GDP Per Capita)      | −1.120***<br>(0.017) | −1.080***<br>(0.037) | −1.180***<br>(0.037) | −1.230***<br>(0.034) |
| ln(Net Aid Received)    | −0.096***<br>(0.007) | −0.064***<br>(0.017) | −0.076***<br>(0.017) | −0.048***<br>(0.015) |
| Government Expenditures | 0.001<br>(0.001)     | 0.004<br>(0.003)     | 0.006**<br>(0.003)   | 0.007**<br>(0.003)   |
| Observations            | 5,036                | 1,185                | 1,208                | 1,208                |
| R <sup>2</sup>          | 0.483                | 0.480                | 0.519                | 0.574                |
| Adjusted R <sup>2</sup> | 0.468                | 0.418                | 0.465                | 0.528                |

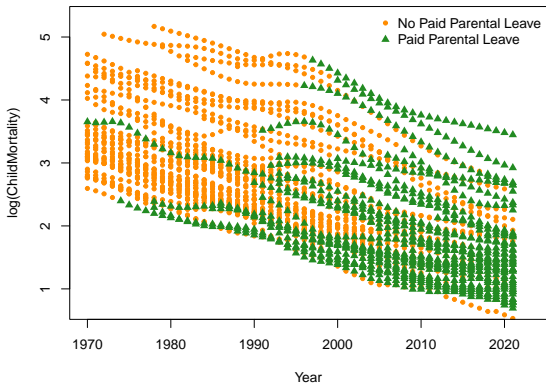
\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

# Another Approach: RDD

**Intuition:** Compare the child mortality “trajectories” of countries before and after they implement paid parental leave policies.

The model is:

$$\begin{aligned}\text{Child Mortality}_{it} &= \beta_0 + \beta_1(\text{Paid Parental Leave}_{it}) + \beta_2(\text{Time}_t) + \\ &= \beta_3(\text{Paid Parental Leave}_{it} \times \text{Time}_t) + (\text{confounders}) + u_{it}\end{aligned}$$



# RDD Regressions

RDD Models of log(Child Mortality)

|                            | OLS #1                 | OLS #2                 | One-Way FE #1          | One-Way FE #2          | Two-Way FE #1          | Two-Way FE #2          |
|----------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| (Intercept)                | 4.3543***<br>(0.0570)  | 11.6056***<br>(0.3046) |                        |                        |                        |                        |
| Paid Parental Leave        | -0.7302***<br>(0.1349) | 0.1293<br>(0.2028)     | -0.0254<br>(0.0420)    | 0.1994*<br>(0.0829)    | -9.1081***<br>(2.7357) | -19.4622*<br>(8.0378)  |
| Time (1950=0)              | -0.0394***<br>(0.0013) | -0.0224***<br>(0.0018) | -0.0423***<br>(0.0004) | -0.0427***<br>(0.0012) |                        |                        |
| Paid Parental Leave x Time | 0.0102***<br>(0.0025)  | -0.0045<br>(0.0036)    | 0.0007<br>(0.0007)     | -0.0038**<br>(0.0014)  | 0.1768***<br>(0.0501)  | 0.3325*<br>(0.1438)    |
| ln(GDP Per Capita)         |                        | -0.6974***<br>(0.0179) |                        | -0.1994***<br>(0.0252) |                        | -1.6885<br>(1.9381)    |
| ln(Net Aid Received)       |                        | -0.0633***<br>(0.0116) |                        | 0.0099+<br>(0.0056)    |                        | -2.5245***<br>(0.4528) |
| Government Expenditures    |                        | -0.0291***<br>(0.0035) |                        | 0.0100***<br>(0.0018)  |                        | 0.8411***<br>(0.1364)  |
| Num.Obs.                   | 2698                   | 759                    | 2698                   | 759                    | 2698                   | 759                    |
| R2                         | 0.405                  | 0.746                  | 0.905                  | 0.928                  | 0.005                  | 0.134                  |
| R2 Adj.                    | 0.405                  | 0.744                  | 0.903                  | 0.925                  | -0.036                 | 0.022                  |

Note: + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001.

# Differences-in-Differences

## Challenges:

- Multiple periods (years) per unit (country), both before and after “treatment”
- “Staggered” treatment timing (adoption of *Paid Parental Leave*)

## One approach:

Callaway, Brantley, and Pedro H.C. Sant’Anna. 2021. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics* 225:200-230.

## Details:

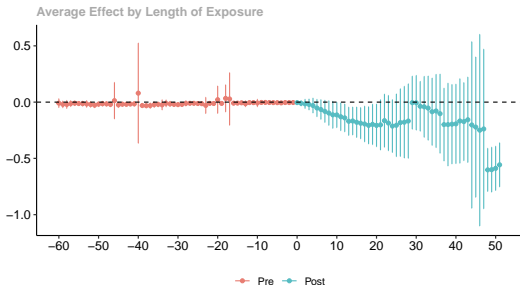
- Deals with the issues related above
- Flexibly fit / interpreted using the `did` package

# Differences-in-Differences via did

Simple bivariate model (no controls):

```
> DiD.fit1<-att_gt(yname = "lnCM",gname = "YearPPL",idname = "ID",  
+                 tname = "YearNumeric",allow_unbalanced_panel = TRUE,  
+                 xformula = ~1,data = WDI,est_method = "reg")  
  
> # Event study object:  
>  
> DiD.ev1 <- aggte(DiD.fit1,type="dynamic",na.rm=TRUE)
```

Plot the event study results:



# ATTs by "Group"

```
> DiD.grp1<-aggte(DiD.fit1,type="group",na.rm=TRUE)
> summary(DiD.grp1)
```

```
Call:
aggte(MP = DiD.fit1, type = "group", na.rm = TRUE)
```

Overall summary of ATT's based on group/cohort aggregation:

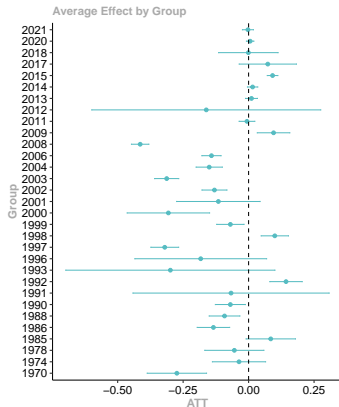
| ATT     | Std. Error | [ 95% Conf. Int.] |
|---------|------------|-------------------|
| -0.0888 | 0.0205     | -0.129 -0.0487 *  |

Group Effects:

| Group | Estimate | Std. Error | [95% Simult. Conf. Band] |
|-------|----------|------------|--------------------------|
| 1970  | -0.2739  | 0.0431     | -0.3873 -0.1605 *        |
| 1974  | -0.0366  | 0.0386     | -0.1383 0.0650           |
| 1978  | -0.0544  | 0.0430     | -0.1675 0.0587           |
| 1985  | 0.0846   | 0.0359     | -0.0098 0.1791           |
| 1986  | -0.1343  | 0.0236     | -0.1964 -0.0723 *        |
| 1988  | -0.0919  | 0.0224     | -0.1507 -0.0330 *        |
| 1990  | -0.0696  | 0.0218     | -0.1270 -0.0123 *        |
| 1991  | -0.0670  | 0.1425     | -0.4418 0.3078           |
| .     |          |            |                          |
| .     |          |            |                          |
| .     |          |            |                          |
| 2018  | -0.0009  | 0.0432     | -0.1145 0.1128           |
| 2020  | 0.0060   | 0.0054     | -0.0082 0.0203           |
| 2021  | -0.0024  | 0.0080     | -0.0235 0.0186           |

---  
Signif. codes: '\*' confidence band does not cover 0

Control Group: Never Treated, Anticipation Periods: 0  
Estimation Method: Outcome Regression



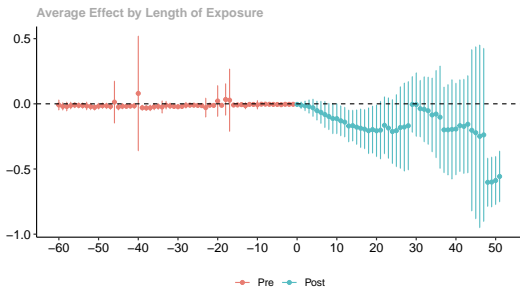


# Differences-in-Differences with Controls

Adding control variables:

```
> DiD.fit2<-att_gt(yname = "lnCM",gname = "YearPPL",idname = "ID",  
+                 tname = "YearNumeric",allow_unbalanced_panel = TRUE,  
+                 xformula = ~lnGDPPerCap+lnNetAidReceived+GovtExpenditures,  
+                 data = WDI, est_method = "reg")  
>  
> # Event study object:  
>  
> DiD.ev2 <- aggte(DiD.fit2,type="dynamic",na.rm=TRUE)
```

Plot the event study results:



# ATTs by "Group" (with controls)

```
> DiD.grp2<-aggte(DiD.fit2,type="group",na.rm=TRUE)
> summary(DiD.grp2)
```

```
Call:
aggte(MP = DiD.fit2, type = "group", na.rm = TRUE)
```

Overall summary of ATT's based on group/cohort aggregation:

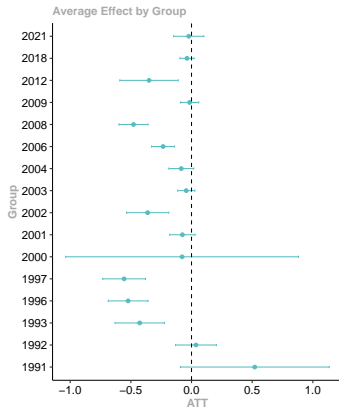
| ATT    | Std. Error | [ 95% Conf. Int.] |
|--------|------------|-------------------|
| -0.153 | 0.0442     | -0.24 -0.0664 *   |

Group Effects:

| Group | Estimate | Std. Error | [95% Simult. Conf. Band] |
|-------|----------|------------|--------------------------|
| 1991  | 0.5214   | 0.2267     | -0.0918 1.1347           |
| 1992  | 0.0368   | 0.0622     | -0.1313 0.2049           |
| 1993  | -0.4267  | 0.0753     | -0.6304 -0.2230 *        |
| 1996  | -0.5226  | 0.0605     | -0.6863 -0.3590 *        |
| 1997  | -0.5557  | 0.0651     | -0.7318 -0.3795 *        |
| 2000  | -0.0778  | 0.3545     | -1.0369 0.8812           |
| 2001  | -0.0753  | 0.0387     | -0.1801 0.0295           |
| 2002  | -0.3619  | 0.0644     | -0.5361 -0.1877 *        |
| 2003  | -0.0435  | 0.0262     | -0.1144 0.0273           |
| 2004  | -0.0845  | 0.0378     | -0.1867 0.0178           |
| 2006  | -0.2344  | 0.0350     | -0.3291 -0.1397 *        |
| 2008  | -0.4783  | 0.0444     | -0.5983 -0.3583 *        |
| 2009  | -0.0159  | 0.0279     | -0.0912 0.0595           |
| 2012  | -0.3502  | 0.0890     | -0.5910 -0.1094 *        |
| 2018  | -0.0373  | 0.0216     | -0.0959 0.0212           |
| 2021  | -0.0228  | 0.0461     | -0.1474 0.1018           |

---  
Signif. codes: '\*' confidence band does not cover 0

Control Group: Never Treated, Anticipation Periods: 0  
Estimation Method: Outcome Regression



- Good references:
  - [Freedman \(2012\)](#)\*
  - [Shalizi \(someday\)](#)\*
  - [Morgan and Winship \(2014\)](#)
  - [Pearl et al. \(2016\)](#)
  - [Peters et al. \(2017\)](#)
- Courses / syllabi (a sampling):
  - [Eggers \(2019\)](#)
  - [Frey \(2023\)](#)
  - [Imai \(2023\)](#)
  - [Munger \(2023\)](#)
  - [Xu \(2018, 2023\)](#).
  - [Yamamoto \(2022\)](#)
- Other useful things:
  - [The CRAN task view on causal inference](#)
  - [The Causal Inference Book](#)
  - [Some useful notes](#)

\* I really like this one.