GSERM - Ljubljana 2024
Analyzing Panel Data

January 18, 2024

Causation

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*

Notation

- N observations indexed by i, $i \in \{1, 2, ...N\}$
- 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:

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

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:

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

W have some missing values,
Z are "complete"

Consider a matrix **R** with:

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

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

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.

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:

$$X \longrightarrow Y$$

$$R_{(Y)}$$

Missing Completely At Random (MCAR)

Rubin Remixed (continued)

$$\begin{tikzcd} X \arrow[r] \arrow[d] & Y \\ R_{(Y)} & \end{tikzcd}$$

Missing At Random (MAR)

Rubin Remixed (continued)

$$\begin{tikzcd} X \arrow[rr] & & Y \arrow[dll] \\ R_{(Y)} & & \end{tikzcd}$$

Not Missing At Random (NMAR)

Missingness Due To Confounding

$$\begin{tikzcd}
X \arrow[r] & Y \\
R_{(Y)} & (U) \arrow[l] \arrow[u] \arrow[ul, dashed]
\end{tikzcd}$$

(Also) Not Missing At Random (NMAR)

[press "play"]

Estimating Treatment Effects

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

Neyman/Rubin/Holland: Treat inability to observed 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 that "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

SUTVA

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

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

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 $\mathbf{X}$ (continued)

Estimands:

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

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

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

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

Corresponding estimates:

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

and

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

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

Confounding

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).

Digression: DAGs

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

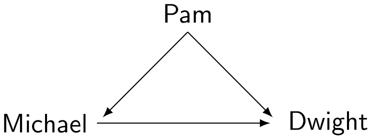


Figure: A DAG

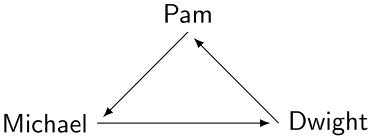


Figure: Not a DAG

DAGs and Confounding

No Confounding

$$\begin{tikzcd} W \arrow[r] & Y & X \arrow[l] \end{tikzcd}$$

A "Collider"

$$\begin{tikzcd} & X & \\ W \arrow[ur] \arrow[rr, dashed] & & Y \arrow[ul] \end{tikzcd}$$

Confounding

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ W \arrow[rr, dashed] & & Y \end{tikzcd}$$

What We're On About

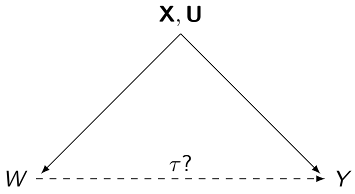


Figure: Potential Confounding

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 τ, the "treatment effect" of W on Y,
- $\mathbf{X}$ are observed confounders,
- $\mathbf{U}$ are unobserved confounders.

Things We Can Do

- **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

$$\begin{tikzcd} & \mathbf{U} \arrow[d] & \\ W \arrow[r] & Y & \mathbf{X} \arrow[l] \end{tikzcd}$$

Figure: = no confounding!

Note:

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

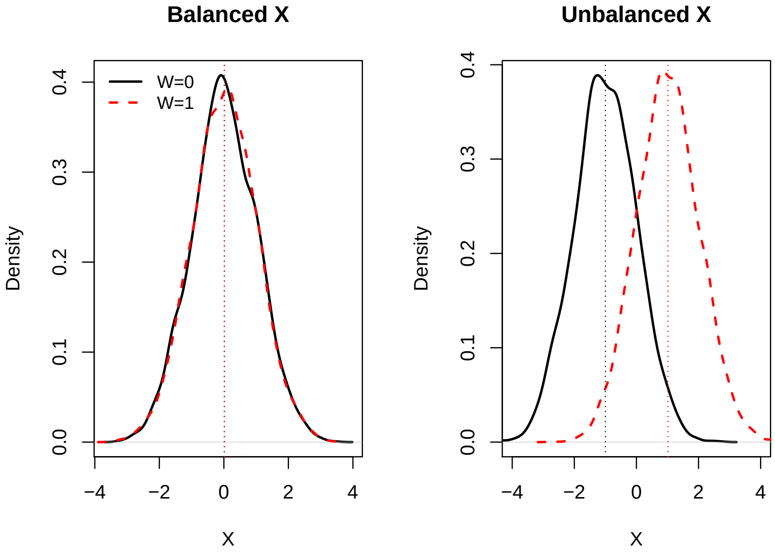
$$\mathrm{E}(\mathbf{X}_i, \mathbf{U}_i \mid W_i = 0) = \mathrm{E}(\mathbf{X}_i, \mathbf{U}_i \mid W_i = 1)$$

  or, more demandingly,

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

- <u>Can</u> yield imbalance by random chance...

Covariate Balance / Imbalance



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 τ
  - 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 **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)

A DAG:

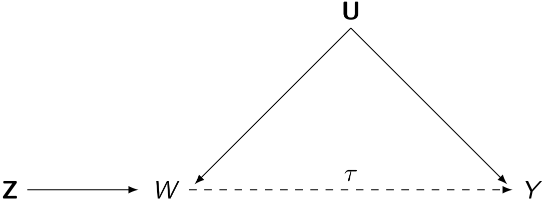


Figure: Instrumental Variables

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:

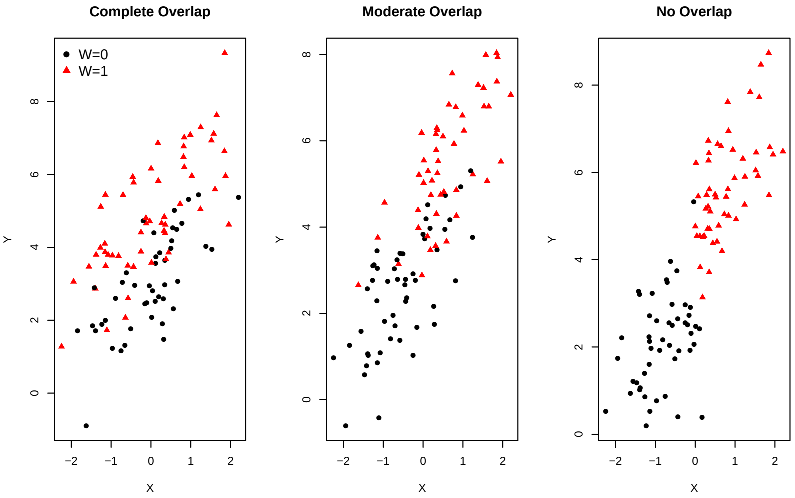
- Requires:
    1. $\text{Cov}(\mathbf{Z}, W) \neq 0)$
    2. $\mathbf{Z}$ has no independent effect on Y, except through W
    3. $\mathbf{Z}$ is exogenous [i.e., $\text{Cov}(\mathbf{Z}, \mathbf{U}) = 0$]
- Arguably most useful when treatment compliance is uncertain / driven by unmeasured factors ("intent 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 $\mathbf{U} = \varnothing$...
- 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
W=0
W=1
X
Y
−2
−1
0
1
2
0
2
4
6
8

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

Matching is a way of dealing with one of 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...

Flavors of Matching

- 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.[1]

[1] 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:

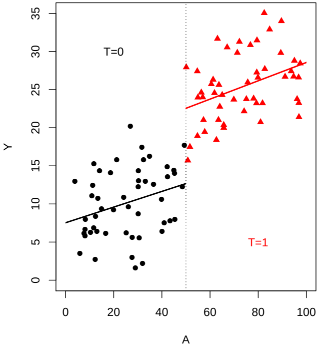
$$W_i = \begin{cases} 0 \text{ if } A_i \leq c \\ 1 \text{ if } A_i > c \end{cases}$$

- Intuition: Observations near but on either side of the threshold(s) are highly comparable, and can be used to (locally) identify τ
- 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)

- 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 ($T = 0$)	Post-Treatment ($T = 1$)
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 equal 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
- 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)
  - Also: compare the pre-treatment trend in the synthetic control to that in the treated case
  - 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

Software Matters

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 *Econometrics* and *SocialSciences* CRAN Task Views
- 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:

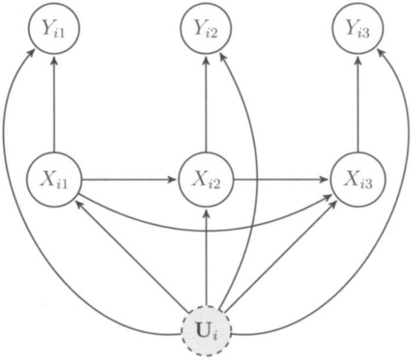
$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$

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, $\boldsymbol{\beta} = \tau$, 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).

[2] 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:

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

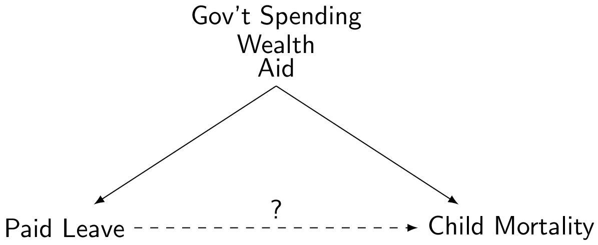
- I & K: The same is not true for $T > 2$...
- More important: two-way FEs' ability to control for unmeasured confounders depends on the (linearity of the) functional form...
- Upshot: two-way FEs aren't a (nonparametric) cure-all...
- Related: When we control for both α_i and η_t, what – exactly – is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

Back To The WDI

```
> describe(WDI,fast=TRUE,ranges=FALSE,check=TRUE)
                         vars     n            mean                sd              se
ISO3                        1 13545             NaN                NA              NA
Year                        2 13545             NaN                NA              NA
Region                      3 13545             NaN                NA              NA
country                     4 13545             NaN                NA              NA
iso3c                       5 13545             NaN                NA              NA
RuralPopulation             6 13268           48.45             25.74            0.22
UrbanPopulation             7 13268           51.55             25.74            0.22
BirthRatePer1K              8 12937           28.02             13.08            0.12
FertilityRate               9 12779            3.91              2.00            0.02
PrimarySchoolAge           10 10699            6.14              0.62            0.01
LifeExpectancy             11 12766           64.63             11.29            0.10
AgeDepRatioOld             12 13300           10.62              6.93            0.06
ChildMortality             13 11092           74.32             77.17            0.73
GDP                        14  9843 245055369928.40 1121079127717.87 11299845990.78
GDPPerCapita               15  9843        11874.12          18895.82          190.46
GDPPerCapGrowth            16  9818            1.93              6.17            0.06
TotalTrade                 17  8548           78.32             54.23            0.59
FDIIn                      18  8406            5.50             45.06            0.49
NetAidReceived             19  8907    473766874.40      900415366.59      9540632.60
MobileCellSubscriptions    20 10057           35.06             51.01            0.51
NaturalResourceRents       21  9211            6.84             11.06            0.12
GovtExpenditures           22  8197           16.26              8.17            0.09
WomenInLegislature         23  4706           17.76             11.73            0.17
PaidParentalLeave          24  9964            0.11              0.31            0.00
ColdWar                    25 13545            0.48              0.50            0.00
YearNumeric                26 13545         1991.00             18.18            0.16
```

Do paid parental leave policies decrease child mortality?

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



Preliminary Regressions

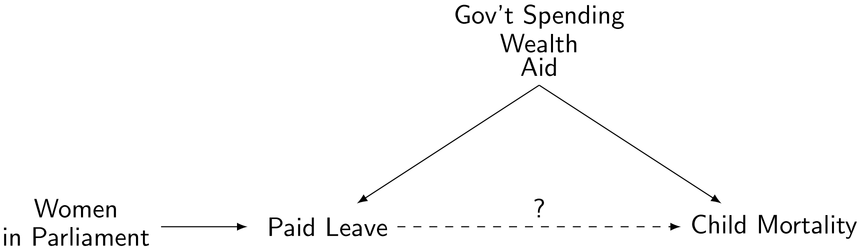
Table: Models of log(Child Mortality)

	BIV	OLS	FE.1way	FE.2way	FE.LDV
Paid Parental Leave	−1.820	−0.891***	−0.085**	−0.139***	−0.220***
	(0.035)	(0.038)	(0.042)	(0.024)	(0.026)
ln(GDP Per Capita)		−0.682***	−1.110***	−0.288***	−0.558***
		(0.009)	(0.017)	(0.013)	(0.012)
ln(Net Aid Received)		−0.079***	−0.096***	0.008**	−0.002
		(0.007)	(0.006)	(0.004)	(0.004)
Government Expenditures		−0.003***	0.001	0.002***	0.0001
		(0.001)	(0.001)	(0.001)	(0.001)
Lagged Child Mortality					0.009***
					(0.0001)
Constant	3.780*	10.800***			
	(0.011)	(0.177)			
Observations	9,357	5,184	5,184	5,184	5,180
R^2	0.222	0.588	0.488	0.117	0.805
Adjusted R^2	0.222	0.588	0.473	0.081	0.799

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

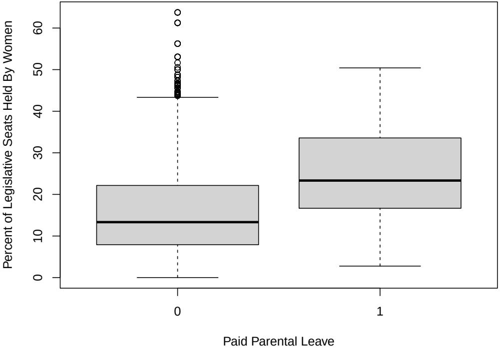
Instrumental Variables

Conceptually:



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

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

Welch Two Sample t-test

data:  WomenInLegislature by PaidParentalLeave
t = -22, df = 1326, 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:
 -9.65 -8.04
sample estimates:
mean in group 0 mean in group 1
          16.0            24.8
```



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

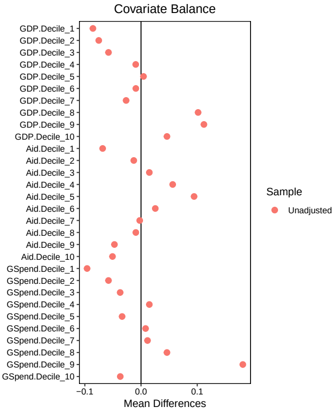
Table: IV Models of log(Child Mortality)

	OLS	FE.1way	FE.IV	RE.IV
Paid Parental Leave	−0.891	−0.085**	111.000	−5.920**
	(0.038)	(0.042)	(334.000)	(2.600)
ln(GDP Per Capita)	−0.682	−1.110***	−18.800	−0.473***
	(0.009)	(0.017)	(53.400)	(0.114)
ln(Net Aid Received)	−0.079	−0.096***	1.900	−0.021
	(0.007)	(0.006)	(5.920)	(0.033)
Government Expenditures	−0.003	0.001	−0.071	−0.002
	(0.001)	(0.001)	(0.213)	(0.003)
Constant	10.800*			8.250***
	(0.177)			(1.240)
Observations	5,184	5,184	2,680	2,680
R^2	0.588	0.488	0.00000	0.231
Adjusted R^2	0.588	0.473	-0.058	0.230

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

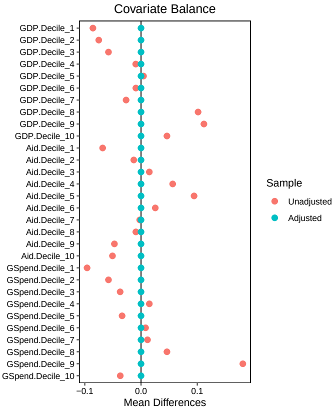
Matching: Checking Covariate Balance

```
> # Subset data a little bit:

> vars<-c("ISO3","Year","Region","country","UrbanPopulation",
          "FertilityRate","PrimarySchoolAge","ChildMortality",
          "GDPPerCapita","NetAidReceived","NaturalResourceRents",
          "GovtExpenditures","PaidParentalLeave","ColdWar",
          "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$GSpend.Decile<-as.factor(ntile(wdi$GovtExpenditures,10))


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



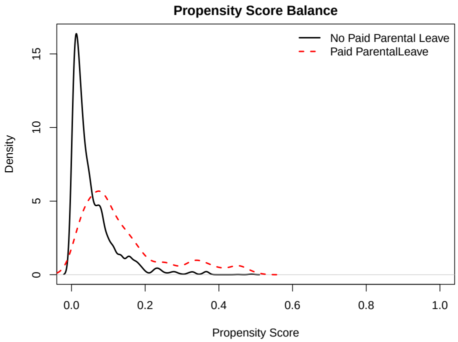
Exact Matching

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

Call:
matchit(formula = PaidParentalLeave ~ GDP.Decile + Aid.Decile +
    GSpend.Decile, data = wdi, method = "exact")

Summary of Balance for All Data:
.
.
.
Sample Sizes:
              Control Treated
All            4622.      282
Matched (ESS)   322.2     268
Matched         831.      268
Unmatched      3791.       14
Discarded         0.        0


> # Create matched data:
>
> wdi.exact <- match.data(M.exact,group="all")
> dim(wdi.exact)
[1] 1099   20
```



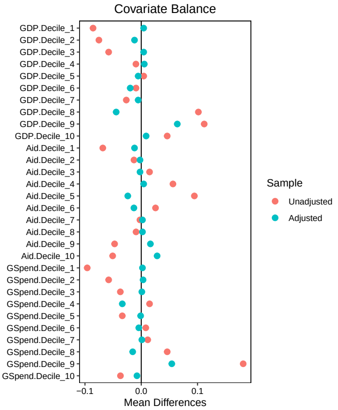
Propensity Scores

```
> # Model for propensity scores:
>
> PS.fit<-glm(PaidParentalLeave~GDP.Decile+Aid.Decile+
              GSpend.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)
```



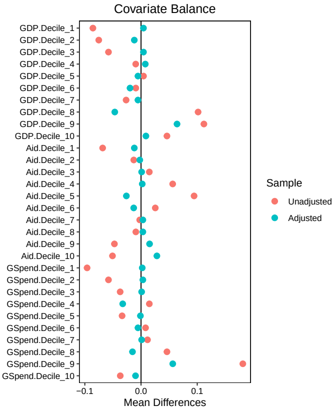
Propensity Score Matching

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

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

Summary of Balance for All Data:
.
.
.
Sample Sizes:
          Control Treated
All          4898     307
Matched       921     307
Unmatched    3977       0
Discarded       0       0


> # Matched data:
>
> wdi.ps <- match.data(M.prop,group="all")
> dim(wdi.ps)
[1] 1228   21
```



“Optimal” Matching

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

Call:
matchit(formula = PaidParentalLeave ~ GDP.Decile + Aid.Decile +
    GSpend.Decile, data = wdi, method = "optimal", ratio = 3)
.
.
.
Sample Sizes:
          Control Treated
All          4898     307
Matched       921     307
Unmatched    3977       0
Discarded       0       0

> # Matched data:
>
> wdi.opt <- match.data(M.opt,group="all")
> dim(wdi.opt)
[1] 1228   21
```



Post-Matching Regressions

Table: Models of log(Child Mortality)

	PreMatch.FE	Exact.FE	PS.FE	Optimal.FE
Paid Parental Leave	−0.084**	−0.189***	−0.220***	−0.205***
	(0.043)	(0.057)	(0.052)	(0.056)
ln(GDP Per Capita)	−1.120***	−1.040***	−1.090***	−1.110***
	(0.017)	(0.037)	(0.035)	(0.035)
ln(Net Aid Received)	−0.096***	−0.045***	−0.052***	−0.053***
	(0.007)	(0.015)	(0.015)	(0.015)
Government Expenditures	0.001	−0.001	0.005*	0.001
	(0.001)	(0.002)	(0.003)	(0.003)
Observations	5,114	1,236	1,208	1,208
R^2	0.486	0.457	0.516	0.504
Adjusted R^2	0.470	0.394	0.458	0.445

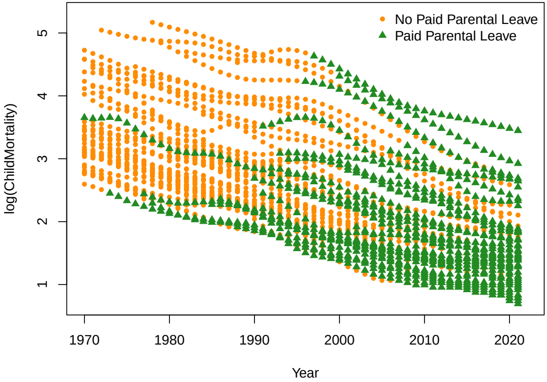
*p<0.1; **p<0.05; ***p<0.01

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	OLS	1-Way FE	1-Way FE	2-Way FE	2-Way FE
(Intercept)	53.71^{***}	312.46^{***}				
	(1.71)	(11.48)				
Paid Parental Leave	-26.48^{***}	-18.38^{*}	-15.94^{***}	-18.06^{**}	-11.87^{***}	-14.59
	(3.92)	(7.84)	(2.30)	(6.16)	(2.81)	(8.44)
Time	-0.73^{***}	-0.68^{***}	-0.83^{***}	-1.41^{***}		
	(0.04)	(0.07)	(0.02)	(0.08)		
Paid Parental Leave $\times$ Time	0.43^{***}	0.27	0.31^{***}	0.34^{**}	0.23^{***}	0.26
	(0.07)	(0.14)	(0.04)	(0.11)	(0.05)	(0.15)
log(GDPPerCapita)		-20.21^{***}		-4.05^{*}		-1.47
		(0.70)		(1.83)		(2.01)
log(NetAidReceived)		-4.17^{***}		-1.62^{***}		-2.31^{***}
		(0.44)		(0.40)		(0.46)
Govt. Expenditures		-0.15		0.93^{***}		0.86^{***}
		(0.13)		(0.13)		(0.14)
R^2	0.19	0.62	0.50	0.70	0.01	0.12
Adj. R^2	0.19	0.62	0.49	0.69	-0.03	0.00
Num. obs.	2610	726	2610	726	2610	726

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Resources

- 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 Causal Inference Book
  - Some useful notes

* I really like this one.