A practical guide to the current era of difference in differences designs

Zachary Porreca

CLEAN Unit
Bocconi University

Guest Lecture, University of Sassari, March 2024

Bocconi Università Commerciale Luigi Bocconi

What is DiD?

- What is Difference-in-Differences?
- A simple motivating logic: the “difference” in the “differences” in pre/post outcomes between treatment/control units
- Basic setting: $(\bar{y}_{post;treated} - \bar{y}_{pre;treated}) - (\bar{y}_{post;control} - \bar{y}_{pre;control})$

Main elements of a DiD setup

- Units
  - "Control" and "Treatment" groups
  - States, countries, provinces, individuals, cities......whatever you desire to be the unit of analysis
- Outcomes
  - What are we interested in?
  - Your "y" variable
- Treatment
  - A policy, a certain type of event, a change of some sort
  - This needs to be something that is time varying.
  - It does not *always* effect all units

The most famous example

- Card and Kreuger's 1994 AER
    - David Card received the 2021 Nobel Prize, in part, for introducing economics to DiD in this paper
- A US state, New Jersey, increases the minimum wage. What happens to employment?
- States are the units, New Jersey=treated and Pennsylvania=control
- Employment (various measurements) in the fast food industry are is the outcome of interest
- Treatment is an increase in minimum wage in New Jersey, while it stays the same in Pennsylvania
- $\beta_{DiD} = \Delta(y_{NJ}) - \Delta(y_{PA})$

The Canonical Simultaneous Adoption Format

$$
\begin{aligned}
y_{it} = \beta_0 + \beta_1(\mathbb{I}(i \in treated)) + \beta_2(\mathbb{I}(t \in post)) + \\
\beta_3(\mathbb{I}(i \in treated \ \& \ t \in post)) + \epsilon_{it}
\end{aligned} \tag{1}
$$

β_3 is the ATE, the "impact" of the treatment

Assumptions

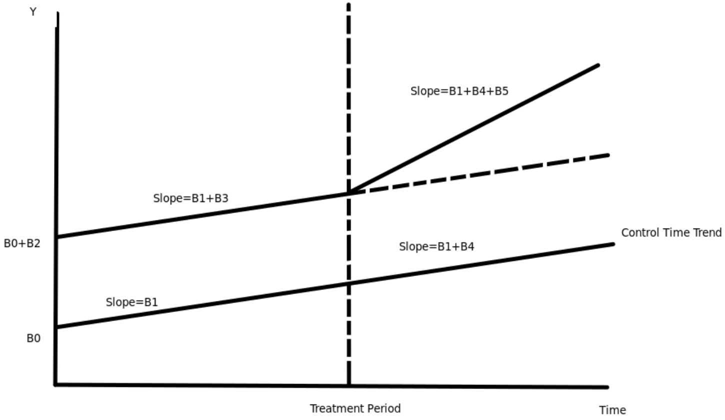
- We're going to go more in depth on these later
- Main questions to ask yourself, necessary to have DiD identify the causal effect you are hoping for
- Parallel trends
  - Would our treated units have evolved similarly to our control units in the absence of treatment?
  - Does our control represent a valid "counterfactual?"
- SUTVA
  - "Stable unit treatment value assumption"
  - Does the treatment impact units that shouldn't be considered "treated"?
  - This stresses the importance of considering spillovers and biases....

Parallel trends and an easy way to think about DiD

$$
\begin{aligned}
y_{it} = \beta_0 + \beta_1(\textit{time trend}) + \beta_2(\textit{treated}) + \beta_3(\textit{time trend by treated}) \\
+\beta_4(\textit{time trend by post}) + \beta_5(\textit{time trend by post by treated})
\end{aligned} \tag{2}
$$

Insignificance of β_3 provides evidence of parallel trends assumption holding

Parallel trends



Example (Old Bailey)

McCannon and Porreca (2023)

- 1800s London, the right to representation for felony defendants introduced.
- Individuals accused of *felonies* are treatment group
- Individuals accused of *misdemeanors* are control group
- Law passing is treatment
- Outcome is binary indicator for conviction

Example (Old Bailey)

	Baseline	Alternative Time Windows		
coverage:	±40 years	±30 years	±50 years	BH window
years:	[1796-1876]	[1806-1866]	[1786-1886]	[1803-1871]
	[1]	[2]	[3]	[4]
Post x Treated	0.0237 ***	0.0227 ***	0.0179 **	0.0249 ***
	(0.0063)	(0.0063)	(0.0079)	(0.0062)
Crime Fixed Effects?	Yes	Yes	Yes	Yes
Year Fixed Effects?	Yes	Yes	Yes	Yes
Judge Fixed Effects?	Yes	Yes	Yes	Yes
Controls?	Yes	Yes	Yes	Yes
R^2	0.530	0.570	0.476	0.545
AIC	54,538	33,823	81,777	44,464
N	135,363	117,229	153,046	125,315
# clusters	81	61	98	69
DV μ	0.7523	0.7617	0.7440	0.7583

Motivation

- The same (or similar) policies can be adopted in different locales (or among different populations) at different times
- If multiple states/countries/individuals adopt the same policy at the same time, we can use our standard approach....but how can we handle the situation when there is not a clear "post" period?

An example

- Kong and Qin (2021)- “China’s Anticorruption Campaign and Entrepreneurship”
- Does corruption hinder entrepreneurship?
- Authors want to exploit a series of anticorruption investigations to see if these probes have any effect on new business formation
    - These investigations are all from the same government initiative but occur in multiple years in different states
- The “post” period differs for different treated units

From Simultaneous to Staggered Adoption

For simultaneous adoption:
$y = \alpha + \beta_1(post) + \beta_2(treatment\ \ group) + \beta_3(post * treatment\ \ group) + \mu$

For staggered adoption:
$y_{it} = \alpha_i + \lambda_t + \tau(treated_{it}) + \mu_{it}$

Relating the two estimators:
$\alpha_i \approx \beta_2(treatment\ \ group)$
$\lambda_t \approx \beta_1(post)$
$(treated_{it}) \approx (post * treatment\ \ group)$

Two way fixed effects functional form

- Vector of unit fixed effects
  - Dummy variables for each unit (state/country/individual/etc)
  - These control for time-invariant unit-specific characteristics
- Vector of time fixed effects
  - Dummy variables for each time period
  - These control for unit-invariant time period-specific characteristics
  - Control for "global" shocks
- A treatment variable that varies within unit and across time
  - "Turns on" when a particular unit receives its initial treatment
  - Coefficient this estimates with be the "average treatment effect on the treated" (ATT)- or the estimated impact of a policy or intervention
- Time-varying covariates can also be included

Back to our Chinese Anticorruption example

- Provinces are our units
- Observations are yearly, with 2012 through 2016 covered
- Treatment indicator is equal to one once a corruption probe has been instigated for a particular province
- Outcome variable is the log of 1 plus the number of new enterprises per 10,000 people

$$Entrepreneurship_{it} = \alpha + \beta(Investigation)_{jt} + \gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad (3)$$

Back to our Chinese Anticorruption example

Anticorruption Campaign and Entrepreneurship

	Without Controls	With Controls
InvestigationAft	.084**	.092**
	(7.733)	(8.291)
LnGDP		.183**
		(3.565)
GDP2%		.018**
		(4.735)
GDP3%		.015**
		(3.581)
CPI		.033**
		(4.608)
LnPopulation		−.938**
		(−13.184)
Adjusted R^2	.955	.957

Note. The dependent variable is Entrepreneurship. All regressions include year and county fixed effects and control for local economic level and other factors. The t-statistics reported in parentheses are based on standard errors clustered at the county level. $N = 18{,}721$.

** $p < .01$.

Assumptions

- Exogeneity of treatment
- Stable unit treatment value assumption (SUTVA)
- Parallel pre-treatment trends (and no treatment anticipation)

Treatment Exogeneity

- Does some missing variable determine treatment status?
- Is treatment status correlated with the error term?
- Is treatment status effectively as good as random?
- Is the eventual treatment status correlated with the outcome variable in the pre-treatment periods?
- These are BIG QUESTIONS!
  - Validity of any quasi-experimental design in our causal inference world depends on the validity of this assumption

Testing for treatment Endogeneity

- Our Chinese Corruption paper does not address this issue
- Potential tests
  - Point biserial correlation test between error term and treatment variable
  - Correlation between outcome variable in pre-treatment periods and an indicator for treatment selection
    - Looking for selection into treatment
    - Can also be done with an event study (discussed later)
  - Demonstrate balance between control and treated observations in pre-treatment periods
  - Analytical arguments

If your treatment is endogenous.....

- Craft an argument for the direction of bias this introduces?
- Instrument for treatment status?
- Models built on latent-factor/ interactive fixed effects models (like Synthetic Difference-in-Differences or Brown and Butts (2023)) that allow for identification in the presence of unobserved time-variant global shocks or time-invariant unobserved unit characteristics (see Bai 2009, Arkhangelsky et al. 2021, Porreca 2022, Brown and Butts (2023))

SUTVA

- "Stable Unit Treatment Value Assumption" (Rubin 1980)
- The treatment status of a particular unit is not correlated with that of other units
- The outcome for one unit only depends on it's own treatment status- not that of other
  - Treatment does not spill over to other units
- Typically we are left to analytical arguments here

Addressing SUTVA

- If displacement or spillovers exist, redefine the treatment to capture these displacements...
- Example
  - Porreca (2023) explicitly examines the spillover effects of the redevelopment of neighborhoods on violence in surrounding neighborhoods
  - Treatment and the units of analysis are redefined (into a network in this case) so that units are treated *if they're neighbors are treated*
- Corrections like this are simple, but do take some thought.
- How can we redesign or data and our treatment to capture these spillovers?

Parallel Pre-Treatment Trends

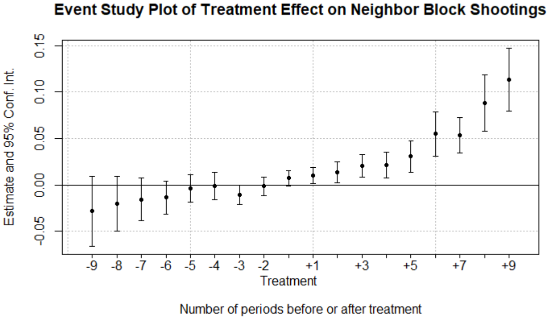
- Perhaps most important of assumptions
- In the absence of treatment, the evolution of both control and treatment group outcomes would be identical
- Harder to visualize in staggered setting
- $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across different g (From de Chaistemartin and Haultfoeuille (2022))

Event Studies and Testing for Violations

- Decompose treatment indicator into a series of treatment leads and lags
- Formally:
  - $$Y_{it} = \sum_{k=-K, k\neq -1}^{k=K} D_{it}^{k} \cdot \delta_k + \gamma_t + \psi_i + \epsilon_{it} \tag{4}$$
  - The coefficients of interest are δ_k
  - D_{it}^{k} represents a vector of dummy variables equal to one, if unit i in period t is k periods away from initial treatment
  - $k = 0$ in the initial treatment period
  - As in He and Wang (2017), $k = -1$ is omitted so that post-treatment event study estimators are relative to the period immediately before treatment.

Event Study Graph Example

Our example paper does not include a graph, so here is an example from Porreca (2023)



Possibilities for correction?

- New methods relying on latent factor models/interactive fixed effects like Bai 2009, Arkhangelsky et al. 2021, and Butts and Brown (2022), Porreca 2022 allow for identification with this assumption violated
- Standard OLS based DiD methods will fail to identify ATT with this assumption violated, however

Treatment Effect Heterogeneity

- Does the effect vary between units?
- Does the effect vary over time?
- Does the effect vary among treatment cohorts?

Basic logic of effect variation between units

- Not all units are the same, does the impact of the intervention change with that variation?
- Perhaps these differential effects are the parameter of interest?
- Porreca (2023) looks at effect of urban redevelopment on gun violence- of interest is how does that effect vary between high drug crime blocks and low drug crime blocks
- Decompose treatment effect between various types of units

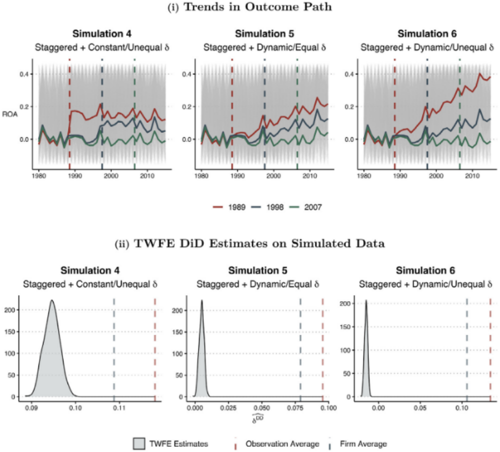
Strategy

- For example: two types of treated units
- $$y_{it} = \alpha_i + \lambda_t + \tau_1(\textit{type 1 treated}_{it}) + \tau_2(\textit{type 2 treated}_{it}) + \mu_{it}$$
- New treatment variables are interactions between treatment status and an indicator for which group of units the observation falls into
- The equality of the τ_i coefficients can be compared with a Wald Chi Square test
- Differences in effect size magnitude, significance, and sign between unit types can provide valuable information
- This same logic can easily be extended to more than two types

Basic Logic of Variation with Time or Cohort

- Great summary in Goodman-Bacon (2019): "So You've Been Told to Do My Difference-in-Differences Thing: A Guide"
- Staggered DiD estimator is a *weighted composite* of various 2x2 DiD estimators (two units, two time periods)
- Those weights come from size of the subgroups and effect size variance
- Treatment effects put units on different trends- This can introduce biases into those 2x2 estimates
- Staggered DiD is a "variance weighted average treatment effect" which is not necessarily the same as the average treatment effect on the treated
- "The dynamics of their treatment can curdle the milk and so we avoid it at all cost."

Illustration of Bias



Bacon Decomposition

- Tool to diagnose *which 2x2 estimates* matter the most in your DiD estimate
- Can show if the bulk of your estimate is being derived from *untreated versus treated units*, or if it is being derived from comparisons between units treated at different time periods
- Also, can provide weights that can be useful in de-biasing the TWFE-DiD estimate with removal of treatment timing comparisons
- Easy implementation in R or Stata

Callaway and Sant'Anna Estimator

- Estimate individual ATT for each treatment cohort- called aggregated ATT
  - ATT(g, t)
- Each cohort's effect is estimated against groups who are never treated and/or groups who are *not yet treated*
- Cohort ATT's can be averaged together to provide a single estimate of the ATT (not the VWATT of TWFE)
- Also able to provide estimates of treatment effect variation with length of exposure
- Easy Stata or R implementation

Some Other Issues/Extensions

- Multiple Treatments
- Continuous Treatments
- Lack of valid control group
  - We will discuss more later in our "synthetics" section
- Spatial Spillovers
  - Butts (2023), non-parametric estimation treatment effects minus biases from neighbor units being "treated" by spillovers and semi-parametric estimation of actual spillover effects

Continuous Treatment

- Callaway et al. (2024)
  - TWFE will fail to provide usefully interperetable estimates
  - Suggest a non-parametric estimator for interperetable results
  - Introduce "level treatment effect" (dose *d* compared to untreated counterfactual) and "causal response" (marginal impact of change in dose against counterfactual)
- Logic is MUCH less intuitive here
  - Concept of dosage of treatment is crucial here
- Scott Cunningham blog link
- Related is de Chaisemartin and D'Haultfeuille (2018): *Fuzzy* DiD
  - identification from changes in dosage when all units are partially treated and *treatment* group sees changes in dosage
  - It is a binary treatment impacting an entire population at different rates

Multiple Treatments

- What happens when there are multiple *different* treatments?
- Naive estimates suffer from "contamination" bias- other treatments' effects impact estimates of other treatments' effects
- de Chaisemartin and D'Haultfeuille (2022)
- Suggested solution is to estimate treatments *separately* with subsamples of the data
  - Simple example: two treatments, multiple groups, staggered adoption, treatment one always precedes treatment two
  - Estimate treatment one's effect on the sub-sample for which treatment two is equal to zero
  - Estimate treatment two's effect on the sub-sample for which treatment one is equal to one

Checklist

From Roth et al. (2023)

A checklist for DiD practitioners.

- Is everyone treated at the same time?

If yes, and panel is balanced, estimation with TWFE specifications such as (5) or (7) yield easily interpretable estimates.

If no, consider using a "heterogeneity-robust" estimator for staggered treatment timing as described in Section 3. The appropriate estimator will depend on whether treatment turns on/off and which parallel trends assumption you're willing to impose. Use TWFE only if you're willing to restrict treatment effect heterogeneity.

- Are you sure about the validity of the parallel trends assumption?

If yes, explain why, including a justification for your choice of functional form. If the justification is (quasi-)random treatment timing, consider using a more efficient estimator as discussed in Section 6.

If no, consider the following steps:

1. If parallel trends would be more plausible conditional on covariates, consider a method that conditions on covariates, as described in Section 4.2.
2. Assess the plausibility of the parallel trends assumption by constructing an event-study plot. If there is a common treatment date and you're using an unconditional parallel trends assumption, plot the coefficients from a specification like (16). If not, then see Section 4.3 for recommendations on event-plot construction.
3. Accompany the event-study plot with diagnostics of the power of the pre-test against relevant alternatives and/or non-inferiority tests, as described in Section 4.4.1.
4. Report formal sensitivity analyses that describe the robustness of the conclusions to potential violations of parallel trends, as described in Section 4.5.

- Do you have a large number of treated and untreated clusters sampled from a super-population?

If yes, then use cluster-robust methods at the cluster level. A good rule of thumb is to cluster at the level at which treatment is independently assigned (e.g. at the state level when policy is determined at the state level); see Section 5.2.

If you have a small number of treated clusters, consider using one of the alternative inference methods described in Section 5.1.

If you can't imagine the super-population, consider a design-based justification for inference instead, as discussed in Section 5.2.

Is everyone treated at the same time?

If yes, the OLS specifications from earlier are applicable......can even decompose into a dynamic event study type estimator
If no......

Heterogeneity Robust Estimators

- Callaway and Sant'Anna Estimator
  - As discussed previously
  - "Group time average treatment effects"....ATT(g,t)
  - Can use never-treated or not-yet-treated units as comparions
- Imputation Estimators (Boryusak et al. (2021))
  - Multi-step procedure
  - More on next slide
- Others:
  - de Chaisemartin and D'Haultfoeuille (2020)
    - Similar approach but allows "switchers"- basically a different weighting scheme but similar to CS
  - Sun and Abraham (2020)
    - Similar approach but uses never-treated or "last to be treated" units as comparison
  - Gardner (2021)
    - Two-step procedure
    - First: $y_{gpit} = \lambda_g + \gamma_p + \epsilon_{gpit}$
    - Second: Regress $y_{gpit} - \hat{\lambda}_g - \hat{\gamma}_p = D_{gp}$

Imputation Estimators

- Step 1: TWFE regression on not-yet-treated sample
  - $y_{it} = \alpha_i + \lambda_t + \epsilon_{it}$
- Step 2: Impute counterfactuals for treated units
  - $\hat{y}_{it}(D = 0)$
- Step 3: Compute individual treatment effect estimates
  - $y_{it}(D = 1) - \hat{y}_{it}(D = 0)$
- Step 4: Aggregate estimates as in Callaway Sant'Anna

Does Parallel Trends Assumption Hold?

If not....

- Condition on covariates?
  - pre-treatment vector of covariates
- Get creative and demonstrate sensitivity of results to potential violations
- Interactive Fixed Effects Models
  - Brown and Butts (2023)
  - Similar approach to synthetic DiD
  - Latent factor/ interactive fixed effects model allows for unobserved global time-period specific shocks that can vary in intensity by unit- effectively allowing unobserved unit specific shocks
  - Remember in TWFE, unit fixed effects capture time invariant unit specific unobserved effects while time fixed effects capture unobserved time period specific global shocks
- Synthetic estimators
  - Next slide

Synthetic Estimators

- Useful for both simultaneous and staggered adoption
- Logic: create a weighted average of control units to force a counterfactual that follows the pre-treatment trajectory of treatment group

$$\left(\hat{\tau}^{\text{did}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta,\mu,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \right\}$$

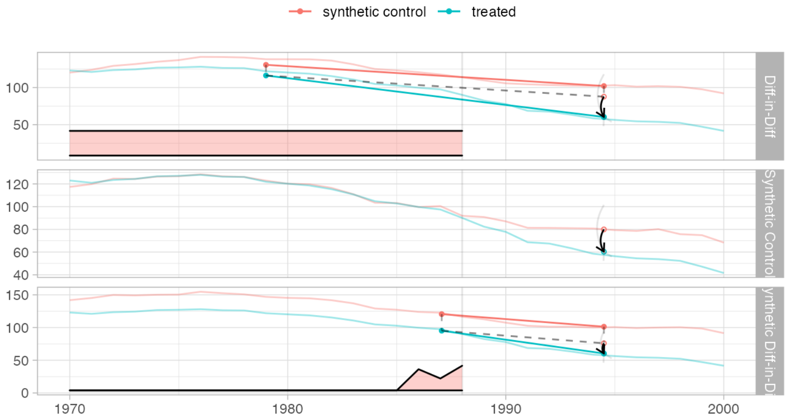
$$\left(\hat{\tau}^{\text{sc}}, \hat{\mu}, \hat{\beta}\right) = \underset{\mu,\beta,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \beta_t - W_{it}\tau \right)^2 \hat{\omega}_i^{\text{sc}} \right\}$$

$$\left(\hat{\tau}^{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_t^{\text{sdid}} \right\}$$

Synthetic Estimators

- SC creates a synthetic that rests *completely* on top of the pre-treatment trend
  - SC is typically used for a single treatment unit. Ben-Michael et al. (2021) extend this to multiple units and staggered timing by pooling units
- SDiD forces *parallel* trends, by allowing an intercept in the weights- simply forcing the control and treatment units to evolve similarly
- Porreca (2022) formalizes the extension of SDiD to staggered settings with a logic similar to the Callaway Sant'Anna estimator

Synthetic Estimators



Is there a large number of treated and untreated units? Thoughts about "super populations"?

- Yes? Cluster standard errors at the level of treatment assignment
- Few treated clusters? This is tricky, there's a lot of different strategies here. See Roth et al. (2023) for a discussion of solutions (wild bootstrap, permutation approaches, different assumptions)
- Is there no "super population" your units are drawn from? Envision the treatment as random (not the units in your sample) and cluster at unit of treatment level.

Package List

From Roth et al. (2023)

Statistical packages for recent DiD methods.

Heterogeneity Robust Estimators for Staggered Treatment Timing		
Package	Software	Description
did, csdid	R, Stata	Implements Callaway and Sant'Anna (2021)
did2s	R, Stata	Implements Gardner (2021), Borusyak et al. (2021), Sun and Abraham (2021), Callaway and Sant'Anna (2021), Roth and Sant'Anna (2021)
didimputation, did_imputation	R, Stata	Implements Borusyak et al. (2021)
DIDmultiplegt, did_multiplegt	R, Stata	Implements de Chaisemartin and D'Haultfoeuille (2020)
eventstudyinteract	Stata	Implements Sun and Abraham (2021)
flexpaneldid	Stata	Implements Dettmann (2020), based on Heckman et al. (1998)
fixest	R	Implements Sun and Abraham (2021)
stackedev	Stata	Implements stacking approach in Cengiz et al. (2019)
staggered	R	Implements Roth and Sant'Anna (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021)
xtevent	Stata	Implements Freyaldenhoven et al. (2019)
DiD with Covariates		
Package	Software	Description
DRDID, drdid	R, Stata	Implements Sant'Anna and Zhao (2020)
Diagnostics for TWFE with Staggered Timing		
Package	Software	Description
bacondecomp, ddtiming	R, Stata	Diagnostics from Goodman-Bacon (2021)
TwoWayFEWeights	R, Stata	Diagnostics from de Chaisemartin and D'Haultfoeuille (2020)
Diagnostic/ Sensitivity for Violations of Parallel Trends		
Package	Software	Description
honestDiD	R, Stata	Implements Rambachan and Roth (2022b)
pretrends	R	Diagnostics from Roth (2022)

Note: This table lists R and Stata packages for recent DiD methods, and is based on Asjad Naqvi's repository at https://asjadnaqvi.github.io/DiD/. Several of the packages listed under "Heterogeneity Robust Estimators" also accommodate covariates.

Summary

- Overview of the connection between standard DiD and the TWFE DiD estimator for staggered adoption
- Overview of the assumptions needed for this estimator to identify ATT (VWATT)
- Discussed basic issues with effect heterogeneity
- Outlines several extensions to the basic framework

Articles Referenced- Links

- de Chaisemartin and D'Haulfeuille (2018)- Fuzzy Did
- de Chaisemartin and D'Haulfeuille (2022)- TWFE Mutiple Treatments
- Callaway and Sant'Anna (2021)- Their estimator
- Goodman-Bacon (2021)- Bacon Decomposition
- Goodman-Bacon (2019)- Decomposition explanation
- Bai (2009)- Interactive Fixed Effects
- Arkhangelsky et al. (2021)-Synthetic Difference-in-Differences
- Porreca (2022)- Staggered SynthDiD
- Kong and Qin (2021)- Chinese anticorruption example
- Porreca (2023)- Staggered paper example

More Articles References

- Callaway et al. (2024)- Coninuous DiD
- Sun and Abraham (2020)
- de Chaisemartin and D'Haulfeuille (2020)
- Gardern (2021)- 2 stage DiD
- Boryusak et al. (2021)- DiD Imputation Estimator
- Brown and Butts (2023)
- Butts (2023)- Spatial Spillovers
- Minton and Mulligan (2024)- DiD Semi-structural
- Ben-Michael et al. (2021)- Staggered Synthetic Control
- McCannon and Porreca (2023)- Old Bailey Courts Paper

Additional Resources

- Roth et al. (2023)- Review of Recent DiD Literature
- Andrew Baker Youtube Video on DiD Issues and Solutions
- Baker et al. (2022)- Demonstrations of Bias in TWFE DiD

Questions/ Contact Info

Thank you! Please reach out to me via email at zachary.porreca@unibocconi.it or at @zachporreca on Twitter