Panel Data Analysis II

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- In the previous class, we discussed the uniqueness of panel data and how it allows us to relax the identification assumption.
- Two common assumptions: sequential ignorability and strict exogeneity.
- They are based on different ideal experiments.
- Under strict exogeneity, we must impose structural restrictions on the DGP.
- Then, we can rely on the TWFE model and the within estimator to estimate the causal effect.

From TWFE to DID

- Suppose there are only two periods, 1 and 2.
- There are N_1 units in the treatment group ($i \in \mathcal{T}$) and N_0 in the control group ($i \in \mathcal{C}$).
- $D_{it} = 0$ in period 1 for any i and $D_{it} = 1$ in period 2 only for units in the treatment group:

$$Y_{it} = \begin{cases} Y_{it}(1), \text{ if } i \in \mathcal{T} \text{ and } t = 2 \\ Y_{it}(0), \text{ otherwise} \end{cases}$$

- We maintain the assumptions for the TWFE model:

$$Y_{it} = \mu + \tau D_{it} + \alpha_i + \xi_t + \varepsilon_{it},$$
$$E[\varepsilon_{is} | D_{it}, \alpha_i, \xi_t] = 0 \text{ for any } s.$$

- There exists a simpler estimator for τ in this case

From TWFE to DID

- Note that strict exogeneity implies the following:

$$
\begin{aligned}
& E[Y_{i2}(0) - Y_{i1}(0) | i \in \mathcal{T}] \\
=& E[\mu + \alpha_i + \xi_2 + \varepsilon_{i2} - (\mu + \alpha_i + \xi_1 + \varepsilon_{i1}) | i \in \mathcal{T}] \\
=& \xi_2 - \xi_1 \\
=& E[Y_{i2}(0) - Y_{i1}(0) | i \in \mathcal{C}].
\end{aligned}
$$

- This assumption is known as "parallel trends."
- It means that without the treatment, the increase (trend) in the outcome would be the same across the two groups.
- Strict exogeneity is a sufficient condition for parallel trends to hold.
- This assumption implies that

$$
\begin{aligned}
& E[Y_{i2}(0) | i \in \mathcal{T}] \\
=& E[Y_{i1}(0) | i \in \mathcal{T}] + E[Y_{i2}(0) | i \in \mathcal{C}] - E[Y_{i1}(0) | i \in \mathcal{C}].
\end{aligned}
$$

From TWFE to DID

- Therefore,

$$\begin{aligned}\tau =& E[Y_{i2}(1) - Y_{i2}(0) | i \in \mathcal{T}] \\ =& E[Y_{i2}(1) | i \in \mathcal{T}] - E[Y_{i2}(0) | i \in \mathcal{T}] \\ =& E[Y_{i2}(1) | i \in \mathcal{T}] - E[Y_{i1}(0) | i \in \mathcal{T}] \\ & - (E[Y_{i2}(0) | i \in \mathcal{C}] - E[Y_{i1}(0) | i \in \mathcal{C}]) \\ =& E[Y_{i2} | i \in \mathcal{T}] - E[Y_{i1} | i \in \mathcal{T}] \\ & - (E[Y_{i2} | i \in \mathcal{C}] - E[Y_{i1} | i \in \mathcal{C}])\end{aligned}$$

- In practice, we estimate τ by

$$\hat{\tau}_{DID} = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} Y_{i2} - \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} Y_{i1} - \left(\frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} Y_{i2} - \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} Y_{i1} \right).$$

From TWFE to DID

- This is known as the difference-in-differences (DID) estimator.
- We first take the within-unit difference for each i, and then take another difference between the two average differences.
- The estimator is motivated by the TWFE model with homogeneous treatment effects.
- But it is robust to heterogeneous treatment effects.
- Let's assume that parallel trends holds and denote

$$\tau_{it} = Y_{it}(1) - Y_{it}(0)$$
$$\tau_{ATT,t} = E[\tau_{it} | i \in \mathcal{T}].$$

- We can show that

$$E[\hat{\tau}_{DID}] = E[Y_{i2}(1) - Y_{i2}(0) | i \in \mathcal{T}] = \tau_{ATT,2}.$$

- Moreover, $\hat{\tau}_{DID} = \hat{\tau}_{TWFE}$.

Validate parallel trends

- As an estimator, DID requires only parallel trends rather than strict exogeneity.
- The assumption allows us to impute the conterfactual for the treated observations.
- It is not directly testable since the definition involves $E[Y_{i2}(0)|i \in \mathcal{T}]$, which is not observable.
- But we can test its validity indirectly.
- Suppose we have another pre-treatment period, period 0.
- If the trends are parallel between periods 1 and 2, it is reasonable to expect them to be parallel between 0 and 1:

$$
\begin{aligned}
& E[Y_{i1}(0)|i \in \mathcal{T}] - E[Y_{i0}(0)|i \in \mathcal{T}] \\
=& E[Y_{i1}(0)|i \in \mathcal{C}] - E[Y_{i0}(0)|i \in \mathcal{C}].
\end{aligned}
$$

- In finite sample, it implies that

$$
\frac{1}{|\mathcal{T}|}\sum_{i \in \mathcal{T}} Y_{i1} - \frac{1}{|\mathcal{T}|}\sum_{i \in \mathcal{T}} Y_{i0} - \left(\frac{1}{|\mathcal{C}|}\sum_{i \in \mathcal{C}} Y_{i1} - \frac{1}{|\mathcal{C}|}\sum_{i \in \mathcal{C}} Y_{i0}\right) \approx 0.
$$

Conditional parallel trends

- Parallel trends may be more plausible once we focus on a smaller group:

$$E[Y_{i2}(0) - Y_{i1}(0) | i \in \mathcal{T}, \mathbf{X}_i = \mathbf{x}]$$
$$= E[Y_{i2}(0) - Y_{i1}(0) | i \in \mathcal{C}, \mathbf{X}_i = \mathbf{x}].$$

- This is known as "conditional parallel trends."
- The assumption is compatible with the following outcome model:

$$Y_{it}(0) = \mu + \delta_t \mathbf{X}_i + \alpha_i + \xi_t + \varepsilon_{it},$$
$$E[\varepsilon_{is} | D_{it}, \mathbf{X}_i, \alpha_i, \xi_t] = 0 \text{ for any } s.$$

- Can we condition on time-varying confounders, or even past outcomes?

Conditional parallel trends

- Let's define $D_i = \mathbf{1}\{i \in \mathcal{T}\}$ and $\Delta Y_i(D_i) = Y_{i2}(D_i) - Y_{i1}(D_i)$.
- The condition that $E[\Delta Y_i(0)|D_i = 1, \mathbf{X}_i = \mathbf{x}] = E[\Delta Y_i(0)|D_i = 0, \mathbf{X}_i = \mathbf{x}]$ is similar to unconfoundedness.
- It is sufficient for identifying the ATT via the IPW estimators:

$$\hat{\tau}_{SDID} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\Delta Y_i - \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}}\frac{\hat{g}(\mathbf{X}_i)\Delta Y_i}{1-\hat{g}(\mathbf{X}_i)}.$$

- This is the semiparametric DID estimator in Abadie (2005).
- We can further construct "doubly robust estimators" with the differenced outcome (Callaway and Sant'Anna 2021).

Multi-period DID

- We can extend the analysis to datasets with multiple periods.
- Suppose there are T_0 pre-treatment periods and T_1 post-treatment periods.
- For any $t \geq T_0 + 1$,

$$
\begin{aligned}
\widehat{\tau}_{DID,t} = & \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} Y_{it} - \frac{1}{|\mathcal{T}| T_0} \sum_{i \in \mathcal{T}} \sum_{s=1}^{T_0} Y_{is} \\
& - \left(\frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} Y_{it} - \frac{1}{|\mathcal{C}| T_0} \sum_{i \in \mathcal{C}} \sum_{s=1}^{T_0} Y_{is} \right)
\end{aligned}
$$

- We can then average over all the periods under treatment:

$$
\widehat{\tau}_{DID} = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \widehat{\tau}_{DID,t}.
$$

- We can similarly show that $E[\widehat{\tau}_{DID,t}] = \tau_t$ and $E[\widehat{\tau}_{DID}] = \tau$.

Multi-period DID

- The previous estimator is equivalent to the following regression model:

$$Y_{it} = \mu + \sum_{s=1}^{T} \tau_s \mathbf{1}\{t = s\}\mathbf{1}\{i \in \mathcal{T}\} + \alpha_i + \xi_t + \varepsilon_{it}.$$

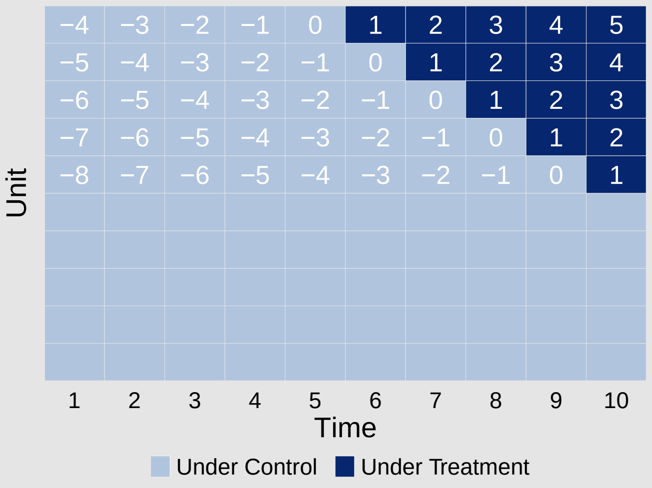
- For each treated unit, we control for the "leads and lags" of the treatment indicator on the right hand side.
- This is known as an "event study" model in the literature
- We can show that $\hat{\tau}_t = \widehat{\tau}_{DID,t}$ for $t \geq T_0 + 1$ and $E[\hat{\tau}_t] = 0$ for $t \leq T_0$.
- It generalizes our test for parallel trends in the two-period case.

Summary

- We say the data have a DID structure when all the treated units become under treatment from the same period.
- We can use the TWFE model to estimate the ATT, or use the event study method to estimate the ATT in any post-treatment period.
- Their results are identical to those from the DID estimator and robust to the heterogeneity in treatment effects.
- Both the TWFE and the event study models are justified by strict exogeneity, while the DID estimator only requires (conditional) parallel trends.
- Such an equivalence will break down when the data have a more complex structure.

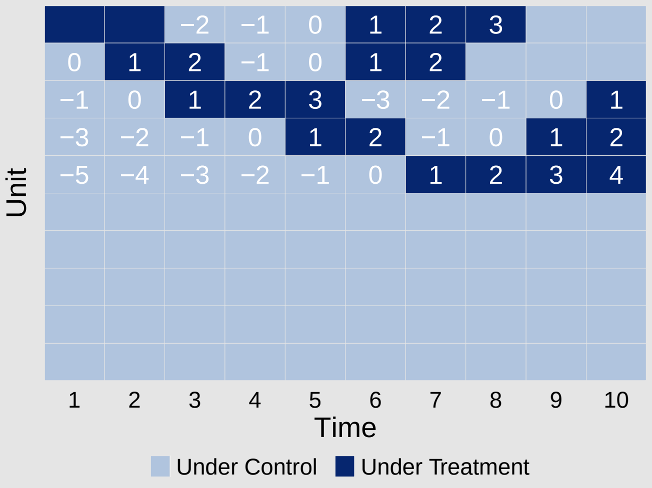
Complex data structure

- Possibility I: once treated, always treated (staggered adoption/generalized DID).



Complex data structure

- Possibility II: treatment switches on and off over periods.



Caveats of the TWFE model

- In either case, the within estimator from the TWFE model is inconsistent for the ATT.
- This problem was identified by a series of papers at the same time (Goodman-Bacon 2018; Chaisemartin and D'Haultfœuille 2020; Strezhnev 2017).
- The TWFE estimate equals a weighted average of individualistic treatment effects across the treated observations.
- The idea is similar to that in Aronow and Samii (2016), but the consequence is more severe.
- Let's denote the collection of treated observations as $\mathcal{M}$ and untreated ones as $\mathcal{O}$.
- Then,

$$\hat{\tau}_{TWFE} \rightarrow \sum_{it:(i,t)\in\mathcal{M}} w_{it}\tau_{it},$$

where each $w_{it} = \frac{\tilde{D}_{it}}{\sum_{it:(i,t)\in\mathcal{M}} \tilde{D}_{it}}$ and $\tilde{D}_{it} = D_{it} - \bar{D}_{i.} - \bar{D}_{.t} + \bar{D}$.

Caveats of the TWFE model

- Consider the following example:

		Periods 1	2	3	4	$\bar{D}_{i.}$
Units	1	0	0	0	0	0
	2	0	0	0	1	1/4
	3	0	1	1	1	3/4
	$\bar{D}_{.t}$	0	1/3	1/3	2/3	1/3

- We can show that the within estimator, $\widehat{\tau}$, converges to

$$\frac{10}{12}\left(\frac{5}{12}\tau_{24} + \frac{1}{4}\tau_{32} + \frac{1}{4}\tau_{33} - \frac{1}{12}\tau_{34}\right),$$

- Some weights can even be negative in practice, making it difficult to interpret the estimate in a causal way.
- The event study model has the same problem (Sun and Abraham 2021).
- Chiu et al. (2023) show that the problem may not be very severe.

Solutions under staggered adoption

- Note that the problem does not exist for DID. (Why?)
- Define cohort t as units whose treatment start from period $t+1$.
- We can estimate the ATT for each cohort as in the multi-period DID.
- We combine units that are treated only from period t and units that have not been treated in period t and obtain a dataset with the DID structure.
- Finally, we average over cohorts for a consistent estimate of the ATT (Goodman-Bacon 2018; Strezhnev 2017).

Solutions under staggered adoption

- In the event study model, Sun and Abraham (2021) propose a similar modification.
- Instead of just "leads and lags," we should also control for the interaction between them and the cohort indicators.
- In other words, we should estimate the effects of "leads and lags" within each cohort and then aggregate across cohorts.
- These solutions do not work when the treatment switches on and off as we no longer have cohorts.
- But the key idea still applies: do not use treated observations to estimate any parameter other than τ.

Counterfactual estimation

- Liu, Wang, and Xu (2024) extend the idea to data with treatment reversal.
- Remember how the DID estimator works:

$$
\begin{aligned}
\widehat{\tau}_{DID,t} =& \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{it} - \frac{1}{|\mathcal{T}|\, T_0}\sum_{i\in\mathcal{T}}\sum_{s=1}^{T_0} Y_{is} \\
& - \left(\frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} Y_{it} - \frac{1}{|\mathcal{C}|\, T_0}\sum_{i\in\mathcal{C}}\sum_{s=1}^{T_0} Y_{is}\right) \\
=& \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}(Y_{it} - \hat{Y}_{it}(0)).
\end{aligned}
$$

- We impute the counterfactual for treated observations $((i,t) \in \mathcal{M})$ using a transformation of the untreated observations $((i,t) \in \mathcal{M})$.

Counterfactual estimation

- Liu, Wang, and Xu (2024) combine the two-way fixed effects model with the Neyman-Rubin framework and assume that:

$$Y_{it}(0) = \mathbf{X}_{it}\beta + \alpha_i + \xi_t + e_{it},$$
$$Y_{it}(1) = Y_{it}(0) + \tau_{it}.$$

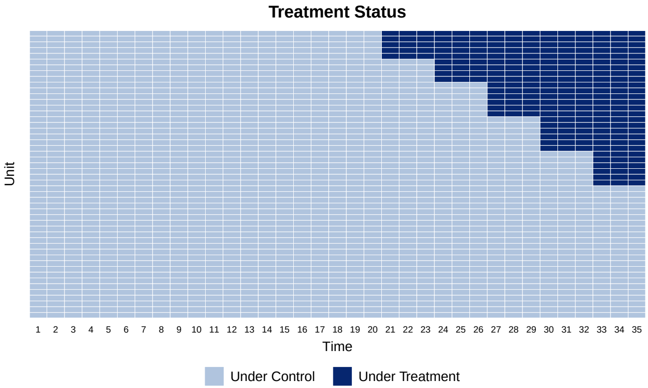
- We use untreated observations to fit a two-way fixed effects model and employ the model to predict $Y_{it}(0)$ for each treated observation.
- Clearly, $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ and

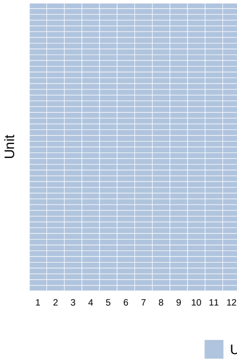
$$\widehat{\tau}_{ATT} = \frac{1}{|\mathcal{M}|} \sum_{(i,t)\in\mathcal{M}} \hat{\tau}_{it}.$$

- It is equivalent to other alternatives under staggered adoption.

Counterfactual estimation

- In a panel setting, treat $Y(1)$ as missing data
- Predict $Y(0)$ based on an outcome model
- (Use pre-treatment data for model selection)
- Estimate ATT by averaging differences between $Y(1)$ and $\hat{Y}(0)$



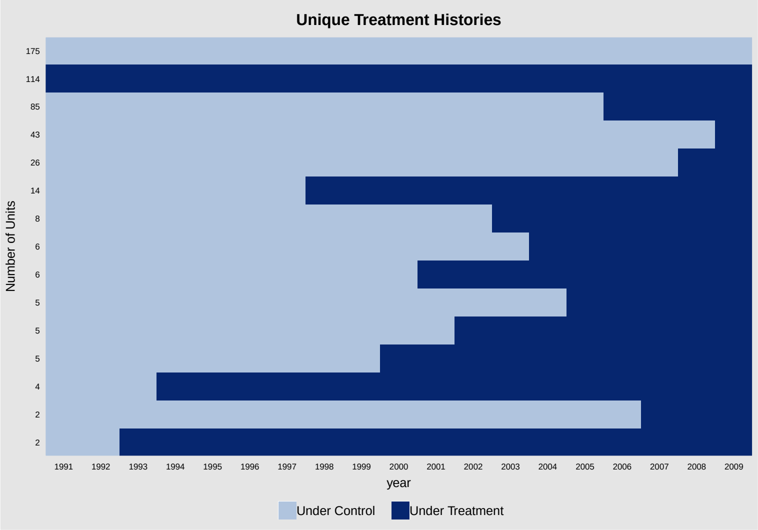


Counterfactual estimation

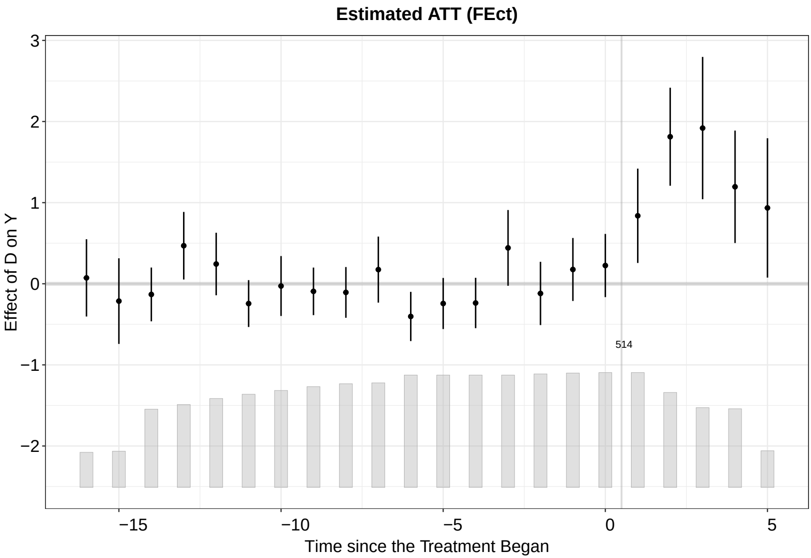
- Liu, Wang, and Xu (2024) show that the estimator is unbiased and consistent for the ATT in each period.
- The periods are now redefined relative to when the treatment kicks off.
- It thus avoids the problem of negative weights.
- It is more straightforward to conduct event study using this method.
- ATT estimates in the pre-treatment periods provide us a way to examine the assumptions.
- They rely on block bootstrap to estimate the standard errors and the confidence interval.
- The framework can incorporate more complicated models.
- It can be implemented in R with the package *fect*.

Counterfactual estimation: application

- A key step is to examine the data structure.



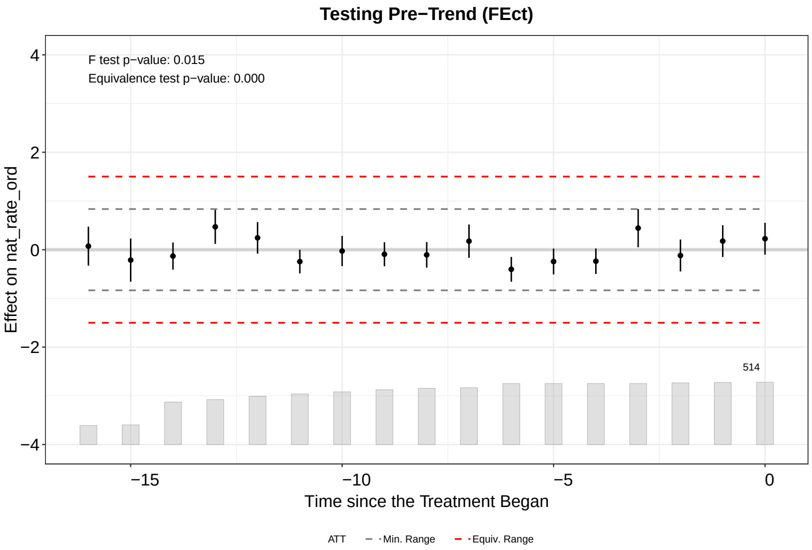
Counterfactual estimation: application



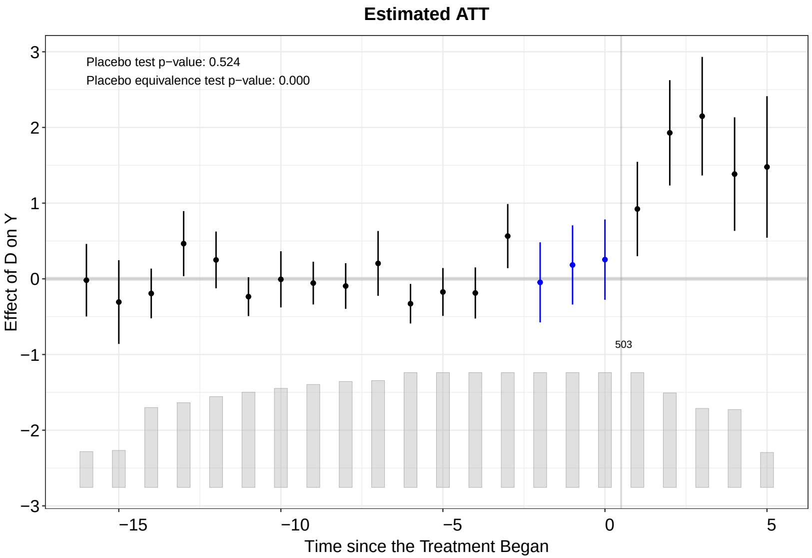
Test in counterfactual estimation

- There are tools for practitioners to evaluate the identification assumption rigorously.
- A placebo test: estimate treatment effects before the treatment's onset and test their significance.
- Idea: if we apply the estimator to period $-s$, then the result should be indistinguishable from zero.
- An equivalence test: test whether all the pre-treatment ATTs are equal to zero.
- A test on the violation of SUTVA.

Test in counterfactual estimation: application



Test in counterfactual estimation: application



Counterfactual estimation: caveats

- We should keep in mind that the validity of this approach relies on a series of assumptions.
- The model specification has to be correct:
  - Observable and unobservable confounders are separable.
  - Observable confounders affect the outcome in a linear and homogeneous manner.
  - Unobservable confounders have a low-dimensional decomposition.
- It also requires strict exogeneity and the absence of interference.

References I

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* 72 (1): 1–19.

Aronow, Peter M, and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*.

Chiu, Albert, Xingchen Lan, Ziyi Liu, and Yiqing Xu. 2023. "What to Do (and Not to Do) with Causal Panel Analysis Under Parallel Trends: Lessons from a Large Reanalysis Study." *arXiv Preprint arXiv:2309.15983*.

References II

Goodman-Bacon, Andrew. 2018. "Difference-in-Differences with Variation in Treatment Timing." National Bureau of Economic Research.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68 (1): 160–76.

Strezhnev, Anton. 2017. "Generalized Difference-in-Differences Estimands and Synthetic Controls." *Unpublished Manuscript*.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99.