Panel Data Analysis IV

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- In the previous class, we discussed methods based on factor models.

Review

- In the previous class, we discussed methods based on factor models.
- They allow for the interaction between unit and period fixed effects hence are more flexible.

Review

- In the previous class, we discussed methods based on factor models.
- They allow for the interaction between unit and period fixed effects hence are more flexible.
- They can be estimated by eigenvalue decomposition.

Review

- In the previous class, we discussed methods based on factor models.
- They allow for the interaction between unit and period fixed effects hence are more flexible.
- They can be estimated by eigenvalue decomposition.
- If there is only one treated unit, we can direct weight untreated units to eliminate the influence of factors.

Review

- In the previous class, we discussed methods based on factor models.
- They allow for the interaction between unit and period fixed effects hence are more flexible.
- They can be estimated by eigenvalue decomposition.
- If there is only one treated unit, we can direct weight untreated units to eliminate the influence of factors.
- This synthetic control method is designed for case study and does not have real confidence intervals.

Review

- In the previous class, we discussed methods based on factor models.
- They allow for the interaction between unit and period fixed effects hence are more flexible.
- They can be estimated by eigenvalue decomposition.
- If there is only one treated unit, we can direct weight untreated units to eliminate the influence of factors.
- This synthetic control method is designed for case study and does not have real confidence intervals.
- Otherwise, it is better to estimate the factors and factor loadings first, and use them to predict the counterfactual outcomes.

Carryover

- Everything becomes complicated if SUTVA is violated.

Carryover

- Everything becomes complicated if SUTVA is violated.
- Assume that the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}\left(D_{it}, \mathbf{D}_i^{1:(t-1)}\right)$.

Carryover

- Everything becomes complicated if SUTVA is violated.
- Assume that the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}\left(D_{it}, \mathbf{D}_i^{1:(t-1)}\right)$.
- This is essentially interference in the time dimension and known as carryover in the literature.

Carryover

- Everything becomes complicated if SUTVA is violated.
- Assume that the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}\left(D_{it}, \mathbf{D}_i^{1:(t-1)}\right)$.
- This is essentially interference in the time dimension and known as carryover in the literature.
- We still assume no anticipation or cross-unit interference.

Carryover

- Everything becomes complicated if SUTVA is violated.
- Assume that the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}\left(D_{it}, \mathbf{D}_i^{1:(t-1)}\right)$.
- This is essentially interference in the time dimension and known as carryover in the literature.
- We still assume no anticipation or cross-unit interference.
- If treatment assignments over periods are dependent, then past treatment status become confounders.

Carryover

- Everything becomes complicated if SUTVA is violated.
- Assume that the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}\left(D_{it}, \mathbf{D}_i^{1:(t-1)}\right)$.
- This is essentially interference in the time dimension and known as carryover in the literature.
- We still assume no anticipation or cross-unit interference.
- If treatment assignments over periods are dependent, then past treatment status become confounders.
- We must control their influence under either strict exogeneity or sequential ignorability.

Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.

Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.
- Consider the classic DID case.

Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.
- Consider the classic DID case.
- $P(D_{it} = 1) = 0$ when $t \leq T_0$ and $P(D_{it} = 1) = 1$ when $t > T_0$.

Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.
- Consider the classic DID case.
- $P(D_{it} = 1) = 0$ when $t \leq T_0$ and $P(D_{it} = 1) = 1$ when $t > T_0$.
- For any $t \leq T_0$, $\mathbf{D}_i^{1:(t-1)}$ is an all-zero vector; for any $t > T_0$, $\mathbf{D}_i^{1:(t-1)}$ do not affect the probability of being treated.

Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.
- Consider the classic DID case.
- $P(D_{it} = 1) = 0$ when $t \leq T_0$ and $P(D_{it} = 1) = 1$ when $t > T_0$.
- For any $t \leq T_0$, $\mathbf{D}_i^{1:(t-1)}$ is an all-zero vector; for any $t > T_0$, $\mathbf{D}_i^{1:(t-1)}$ do not affect the probability of being treated.
- $\mathbf{D}_i^{1:(t-1)}$ are not confounders.

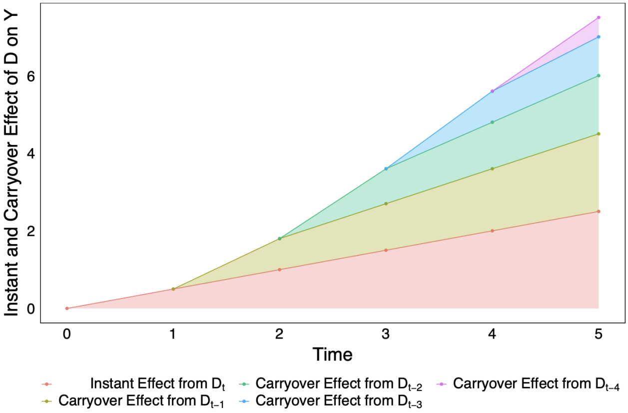
Carryover under staggered adoption

- Under staggered adoption, we should not be too concerned about carryover.
- Consider the classic DID case.
- $P(D_{it} = 1) = 0$ when $t \leq T_0$ and $P(D_{it} = 1) = 1$ when $t > T_0$.
- For any $t \leq T_0$, $\mathbf{D}_i^{1:(t-1)}$ is an all-zero vector; for any $t > T_0$, $\mathbf{D}_i^{1:(t-1)}$ do not affect the probability of being treated.
- $\mathbf{D}_i^{1:(t-1)}$ are not confounders.
- We can still estimate $\tau_{t,ATT}$ with

$$\begin{aligned}\hat{\tau}_t =& \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(Y_{it} - \hat{Y}_{it}\left(\mathbf{0}^{1:t}\right)\right).\\ =& \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(Y_{it}\left(\mathbf{0}^{1:T_0}, \mathbf{1}^{(T_0+1):t}\right) - \hat{Y}_{it}\left(\mathbf{0}^{1:t}\right)\right).\end{aligned}$$

Carryover under staggered adoption

- We can show that the estimate converges to the cumulative effect of the treatment assignment history: $\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\sum_{s=T_0+1}^{t}\tau_{is}$.



Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.
- Note that only $\mathbf{Y}_i^{1:T_0}$ can be confounders in DID.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.
- Note that only $\mathbf{Y}_i^{1:T_0}$ can be confounders in DID.
- Any Y_{it} for $t > T_0$ is a mediator rather a confounder.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.
- Note that only $\mathbf{Y}_i^{1:T_0}$ can be confounders in DID.
- Any Y_{it} for $t > T_0$ is a mediator rather a confounder.
- All we need to do is to balance $\mathbf{Y}_i^{1:T_0}$ across the two groups.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.
- Note that only $\mathbf{Y}_i^{1:T_0}$ can be confounders in DID.
- Any Y_{it} for $t > T_0$ is a mediator rather a confounder.
- All we need to do is to balance $\mathbf{Y}_i^{1:T_0}$ across the two groups.
- As suggested by the method of synthetic control, it accounts for the influence of both the outcome history and unobservable confounders.

Feedback under staggered adoption

- We have feedback if past outcomes affect treatment assignment.
- This is the case under sequential ignorability.
- A more complex scenario is the combination of strict exogeneity and sequential ignorability.
- Under staggered adoption, feedback may not be a serious problem.
- Note that only $\mathbf{Y}_i^{1:T_0}$ can be confounders in DID.
- Any Y_{it} for $t > T_0$ is a mediator rather a confounder.
- All we need to do is to balance $\mathbf{Y}_i^{1:T_0}$ across the two groups.
- As suggested by the method of synthetic control, it accounts for the influence of both the outcome history and unobservable confounders.
- α_i can be learned from $\mathbf{Y}_i^{1:T_0}$, hence controlling for $\mathbf{Y}_i^{1:T_0}$ controls for α_i.

Balance the outcome history

- Hazlett and Xu (2018) generalize the idea of synthetic control to data with a structure of staggered adoption.

Balance the outcome history

- Hazlett and Xu (2018) generalize the idea of synthetic control to data with a structure of staggered adoption.
- We search for a group of weights $\{w_i\}$ such that:

$$\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} f\left(\mathbf{Y}_i^{1:T_0}\right) = \sum_{i\in\mathcal{C}} w_i f\left(\mathbf{Y}_i^{1:T_0}\right),$$

$$\sum_{i\in\mathcal{C}} w_i = 1, w_i \geq 0.$$

Balance the outcome history

- Hazlett and Xu (2018) generalize the idea of synthetic control to data with a structure of staggered adoption.
- We search for a group of weights $\{w_i\}$ such that:

$$\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} f\left(\mathbf{Y}_i^{1:T_0}\right) = \sum_{i\in\mathcal{C}} w_i f\left(\mathbf{Y}_i^{1:T_0}\right),$$
$$\sum_{i\in\mathcal{C}} w_i = 1, w_i \geq 0.$$

- Next, we estimate the treatment effect on the treated in each period

$$\hat{\tau}_t = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{it} - \sum_{i\in\mathcal{C}} w_i Y_{it}$$

Balance the outcome history

- Hazlett and Xu (2018) generalize the idea of synthetic control to data with a structure of staggered adoption.
- We search for a group of weights $\{w_i\}$ such that:

$$\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} f\left(\mathbf{Y}_i^{1:T_0}\right) = \sum_{i\in\mathcal{C}} w_i f\left(\mathbf{Y}_i^{1:T_0}\right),$$
$$\sum_{i\in\mathcal{C}} w_i = 1, w_i \geq 0.$$

- Next, we estimate the treatment effect on the treated in each period

$$\hat{\tau}_t = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{it} - \sum_{i\in\mathcal{C}} w_i Y_{it}$$

- They show the consistency of this estimator when the potential outcomes are linear in $f\left(\mathbf{Y}_i^{1:T_0}\right)$.

Balance the outcome history

- Hazlett and Xu (2018) generalize the idea of synthetic control to data with a structure of staggered adoption.
- We search for a group of weights $\{w_i\}$ such that:

$$\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} f\left(\mathbf{Y}_i^{1:T_0}\right) = \sum_{i\in\mathcal{C}} w_i f\left(\mathbf{Y}_i^{1:T_0}\right),$$
$$\sum_{i\in\mathcal{C}} w_i = 1, w_i \geq 0.$$

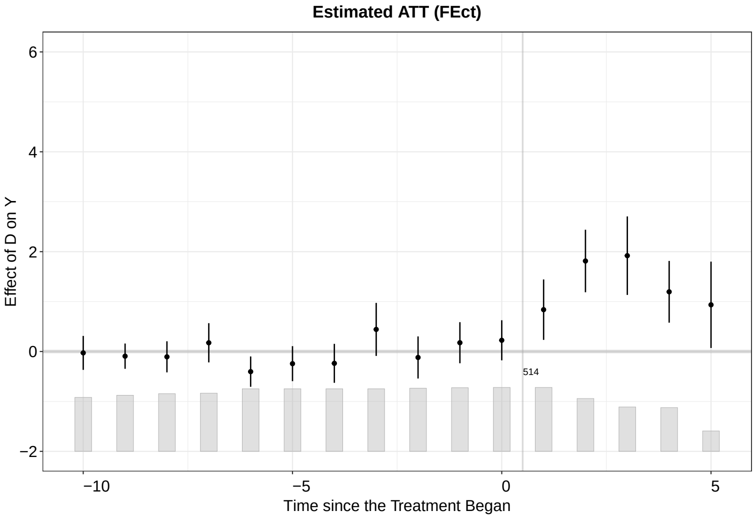
- Next, we estimate the treatment effect on the treated in each period

$$\hat{\tau}_t = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} Y_{it} - \sum_{i\in\mathcal{C}} w_i Y_{it}$$

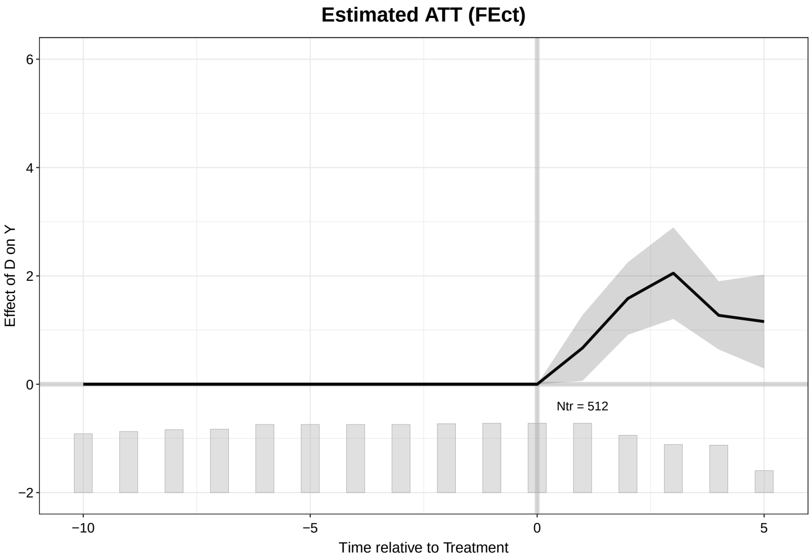
- They show the consistency of this estimator when the potential outcomes are linear in $f\left(\mathbf{Y}_i^{1:T_0}\right)$.
- Kim, Imai, and Wang (2019) argue that we can do the same via matching.

Application

- Let's revisit Hainmueller and Hangartner (2019).



Application



Carryover under treatment reversal

- In this case, our outcome model becomes

$$
\begin{aligned}
Y_{it} =& \tau_{it} D_{it} + \sum_{s=1}^{t-1} \tau_{is} D_{is} + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it} \\
=& \tau_{it} D_{it} + v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it},
\end{aligned}
$$

where $v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) = \sum_{s=1}^{t-1} \tau_{is} D_{is}$ may not have a low-dimensional decomposition.

Carryover under treatment reversal

- In this case, our outcome model becomes

$$\begin{aligned} Y_{it} =& \tau_{it} D_{it} + \sum_{s=1}^{t-1} \tau_{is} D_{is} + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it} \\ =& \tau_{it} D_{it} + v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it}, \end{aligned}$$

  where $v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) = \sum_{s=1}^{t-1} \tau_{is} D_{is}$ may not have a low-dimensional decomposition.
- Under carryover, each $Y_{it}(0)$ could be affected by both $v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right)$ and $h_t\left(\mathbf{U}_i\right)$.

Carryover under treatment reversal

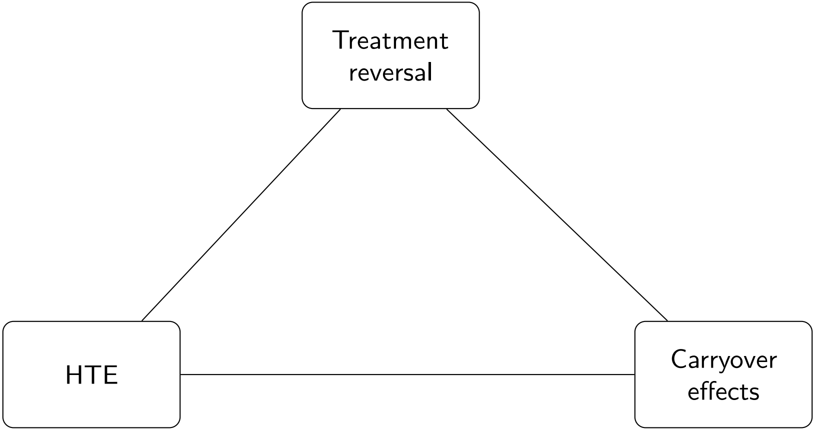
- In this case, our outcome model becomes

$$
\begin{aligned}
Y_{it} =& \tau_{it} D_{it} + \sum_{s=1}^{t-1} \tau_{is} D_{is} + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it} \\
=& \tau_{it} D_{it} + v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) + f_t\left(\mathbf{X}_{it}\right) + h_t\left(\mathbf{U}_i\right) + \varepsilon_{it},
\end{aligned}
$$

  where $v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right) = \sum_{s=1}^{t-1} \tau_{is} D_{is}$ may not have a low-dimensional decomposition.
- Under carryover, each $Y_{it}(0)$ could be affected by both $v_t\left(\mathbf{D}_i^{1:(t-1)}, \mathbf{X}_{it}, \mathbf{U}_i\right)$ and $h_t\left(\mathbf{U}_i\right)$.
- We cannot learn the unobservable confounders from untreated observations.

Carryover under treatment reversal

- Wang and Xu (2024) argue that there exists a trilemma for methods under strict exogeneity:



Carryover under treatment reversal

- Everything works well under staggered adoption.

Carryover under treatment reversal

- Everything works well under staggered adoption.
- If there is no carryover, we can still apply *fect* and other methods.

Carryover under treatment reversal

- Everything works well under staggered adoption.
- If there is no carryover, we can still apply *fect* and other methods.
- If treatment effects only vary over $\mathbf{X}_{it}$, we can model them by including interaction terms.

Carryover under treatment reversal

- Everything works well under staggered adoption.
- If there is no carryover, we can still apply *fect* and other methods.
- If treatment effects only vary over $\mathbf{X}_{it}$, we can model them by including interaction terms.
- It is necessary to test for the existence of carryover before using fixed effects models.

Carryover under treatment reversal

- Everything works well under staggered adoption.
- If there is no carryover, we can still apply *fect* and other methods.
- If treatment effects only vary over $\mathbf{X}_{it}$, we can model them by including interaction terms.
- It is necessary to test for the existence of carryover before using fixed effects models.
- Another solution is drop periods after treatment reversal.

Carryover under treatment reversal

- Everything works well under staggered adoption.
- If there is no carryover, we can still apply *fect* and other methods.
- If treatment effects only vary over $\mathbf{X}_{it}$, we can model them by including interaction terms.
- It is necessary to test for the existence of carryover before using fixed effects models.
- Another solution is drop periods after treatment reversal.
- But it changes the estimand and leads to efficiency loss.

Summary

- Always check your data structure first!

Summary

- Always check your data structure first!
- Visualization is helpful.

Summary

- Always check your data structure first!
- Visualization is helpful.
- Think about the "target trial" or "ideal experiment" before starting your analysis.

Summary

- Always check your data structure first!
- Visualization is helpful.
- Think about the "target trial" or "ideal experiment" before starting your analysis.
- Justify the structural restrictions with various methods or models.

Summary

- Always check your data structure first!
- Visualization is helpful.
- Think about the “target trial” or “ideal experiment” before starting your analysis.
- Justify the structural restrictions with various methods or models.
- Validate the identification assumptions via placebo tests or sensitivity analysis.

Summary

- Always check your data structure first!
- Visualization is helpful.
- Think about the "target trial" or "ideal experiment" before starting your analysis.
- Justify the structural restrictions with various methods or models.
- Validate the identification assumptions via placebo tests or sensitivity analysis.
- Do not seek for causality all the time.

Summary

- Always check your data structure first!
- Visualization is helpful.
- Think about the "target trial" or "ideal experiment" before starting your analysis.
- Justify the structural restrictions with various methods or models.
- Validate the identification assumptions via placebo tests or sensitivity analysis.
- Do not seek for causality all the time.
- Be honest and transparent to yourself and your readers.

References I

Hainmueller, Jens, and Dominik Hangartner. 2019. "Does Direct Democracy Hurt Immigrant Minorities? Evidence from Naturalization Decisions in Switzerland." *American Journal of Political Science* 63 (3): 530–47.

Hazlett, Chad, and Yiqing Xu. 2018. "Trajectory Balancing: A General Reweighting Approach to Causal Inference with Time-Series Cross-Sectional Data." *Available at SSRN 3214231*.

Kim, In Song, Kosuke Imai, and Erik Wang. 2019. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data."