TSCS Data Analysis

Ye Wang

Outline

- Today we start from comparing two basic assumptions in TSCS data analysis: strict exogeneity and sequential ignorability.

Outline

- Today we start from comparing two basic assumptions in TSCS data analysis: strict exogeneity and sequential ignorability.
- We show that classic methods under strict exogeneity have various limitations.

Outline

- Today we start from comparing two basic assumptions in TSCS data analysis: strict exogeneity and sequential ignorability.
- We show that classic methods under strict exogeneity have various limitations.
- We then introduce several approaches to fix the problems, such as counterfactual estimation.

Outline

- Today we start from comparing two basic assumptions in TSCS data analysis: strict exogeneity and sequential ignorability.
- We show that classic methods under strict exogeneity have various limitations.
- We then introduce several approaches to fix the problems, such as counterfactual estimation.
- We next review available methods under sequential ignorability.

Outline

- Today we start from comparing two basic assumptions in TSCS data analysis: strict exogeneity and sequential ignorability.
- We show that classic methods under strict exogeneity have various limitations.
- We then introduce several approaches to fix the problems, such as counterfactual estimation.
- We next review available methods under sequential ignorability.
- We conclude by discussing the idea of manifold learning and how to deal with interference in TSCS data (hopefully!).

What is unique about TSCS data?

- Why not just pooled OLS?

What is unique about TSCS data?

- Why not just pooled OLS?
  - Observations are dependent,
  - The panel structure allows us to relax the identification assumption.

What is unique about TSCS data?

- Why not just pooled OLS?
  - Observations are dependent,
  - The panel structure allows us to relax the identification assumption.
- In the cross-sectional setting, we need ignorability: $D_i \perp Y_i(D_i) | \mathbf{X}_i$, $0 < P(D_i = 1 | \mathbf{X}_i) < 1$.

What is unique about TSCS data?

- Why not just pooled OLS?
  - Observations are dependent,
  - The panel structure allows us to relax the identification assumption.
- In the cross-sectional setting, we need ignorability: $D_i \perp Y_i(D_i)|\mathbf{X}_i$, $0 < P(D_i = 1|\mathbf{X}_i) < 1$.
- Now we can relax the assumption along two possible directions (based on your belief).

What is unique about TSCS data?

- Strict exogeneity: $D_{it} \perp Y_{is}(D_{is}) | \mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}$, where $\mathbf{X}_i^{1:t}$ the history of observable confounders and $\mathbf{U}_i^{1:t}$ the history of unobservable confounders.

What is unique about TSCS data?

- Strict exogeneity: $D_{it} \perp Y_{is}(D_{is})|\mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}$, where $\mathbf{X}_i^{1:t}$ the history of observable confounders and $\mathbf{U}_i^{1:t}$ the history of unobservable confounders.
- Sequential ignorability: $D_{it} \perp Y_{it}(D_{it})|\mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$.

What is unique about TSCS data?

- Strict exogeneity: $D_{it} \perp Y_{is}(D_{is}) | \mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}$, where $\mathbf{X}_i^{1:t}$ the history of observable confounders and $\mathbf{U}_i^{1:t}$ the history of unobservable confounders.
- Sequential ignorability: $D_{it} \perp Y_{it}(D_{it}) | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$.
- Notice that we implicitly require both $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$ to be exogenous to Y_{it}.

What is unique about TSCS data?

- Strict exogeneity: $D_{it} \perp Y_{is}(D_{is}) | \mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}$, where $\mathbf{X}_i^{1:t}$ the history of observable confounders and $\mathbf{U}_i^{1:t}$ the history of unobservable confounders.
- Sequential ignorability: $D_{it} \perp Y_{it}(D_{it}) | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$.
- Notice that we implicitly require both $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$ to be exogenous to Y_{it}.
- Intuitively, their values do not hinge on the value of the outcome in history.

What is unique about TSCS data?

- Strict exogeneity: $D_{it} \perp Y_{is}(D_{is}) | \mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}$, where $\mathbf{X}_i^{1:t}$ the history of observable confounders and $\mathbf{U}_i^{1:t}$ the history of unobservable confounders.
- Sequential ignorability: $D_{it} \perp Y_{it}(D_{it}) | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$.
- Notice that we implicitly require both $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$ to be exogenous to Y_{it}.
- Intuitively, their values do not hinge on the value of the outcome in history.
- We always assume that SUTVA holds and there is no reversal causality (future treatments won't affect past outcomes).

Ideal experiments behind the two assumptions

- The two assumptions are based upon two different ideal experiments (data generating processes).

Ideal experiments behind the two assumptions

- The two assumptions are based upon two different ideal experiments (data generating processes).
- Under strict exogeneity, the experimenter observes all the unobservable attributes and pre-specifies the treatment assignment in each period for each unit before seeing the outcome.

Ideal experiments behind the two assumptions

- The two assumptions are based upon two different ideal experiments (data generating processes).
- Under strict exogeneity, the experimenter observes all the unobservable attributes and pre-specifies the treatment assignment in each period for each unit before seeing the outcome.
- In other words, the plan of assigning the treatment has been fixed before the start of the experiment.

Ideal experiments behind the two assumptions

- The two assumptions are based upon two different ideal experiments (data generating processes).
- Under strict exogeneity, the experimenter observes all the unobservable attributes and pre-specifies the treatment assignment in each period for each unit before seeing the outcome.
- In other words, the plan of assigning the treatment has been fixed before the start of the experiment.
- You will be vaccinated on Feb. 15 with probability 0.72 if you are an old Asian male who loves tequila and that day's temperature is above zero.

Ideal experiments behind the two assumptions

- The two assumptions are based upon two different ideal experiments (data generating processes).
- Under strict exogeneity, the experimenter observes all the unobservable attributes and pre-specifies the treatment assignment in each period for each unit before seeing the outcome.
- In other words, the plan of assigning the treatment has been fixed before the start of the experiment.
- You will be vaccinated on Feb. 15 with probability 0.72 if you are an old Asian male who loves tequila and that day's temperature is above zero.
- The dataset available to the analyst only includes each unit's observable attributes, health status, treatment status, and the vaccination date.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the plan of assigning the treatment will be adjusted according to the outcomes observed in the past.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the plan of assigning the treatment will be adjusted according to the outcomes observed in the past.
- Also, the experimenter does not possess the information on the unobservable attributes of a subject.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the plan of assigning the treatment will be adjusted according to the outcomes observed in the past.
- Also, the experimenter does not possess the information on the unobservable attributes of a subject.
- You will be vaccinated on Feb. 15 with probability 0.72 if you are an old Asian male who were not infected by Covid last week.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the plan of assigning the treatment will be adjusted according to the outcomes observed in the past.
- Also, the experimenter does not possess the information on the unobservable attributes of a subject.
- You will be vaccinated on Feb. 15 with probability 0.72 if you are an old Asian male who were not infected by Covid last week.
- The dataset includes the observable attributes, the history of outcomes and treatment assignments of each unit, and the vaccination date.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the analyst (us) does observe all the confounders and the only unknown part is the propensity score.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the analyst (us) does observe all the confounders and the only unknown part is the propensity score.
- All the classic methods (regression, weighting, matching, etc.) still apply with some modifications.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the analyst (us) does observe all the confounders and the only unknown part is the propensity score.
- All the classic methods (regression, weighting, matching, etc.) still apply with some modifications.
- Under strict exogeneity, we have the problem of omitted variables as part of the confounders ($\mathbf{U}_i^{1:T}$) is also unknown.

Ideal experiments behind the two assumptions

- Under sequential ignorability, the analyst (us) does observe all the confounders and the only unknown part is the propensity score.
- All the classic methods (regression, weighting, matching, etc.) still apply with some modifications.
- Under strict exogeneity, we have the problem of omitted variables as part of the confounders ($\mathbf{U}_i^{1:T}$) is also unknown.
- To eliminate the influence of the unobservables, we need stronger assumptions on the DGP, such that the values of $\mathbf{U}_i^{1:T}$ can be learned from other variables.

Estimation under strict exogeneity

- Note that the strict exogeneity assumption is satisfied by the following outcome model:

$$Y_{it} = g_{it}(D_{it}, \mathbf{X}_i^{1:t}, \mathbf{U}_i^{1:t}) + \varepsilon_{it}$$
$$\mathbb{E}[\varepsilon_{is} | D_{it}, \mathbf{X}_{it}, \mathbf{U}_{it}] = 0,$$

which is too general for identification.

Estimation under strict exogeneity

- In practice, we impose structural assumptions to simplify the model.

Estimation under strict exogeneity

- In practice, we impose structural assumptions to simplify the model.
- Only contemporary values of $\mathbf{X}$ and $\mathbf{U}$ are confounders:

$$Y_{it} = g_{it}(D_{it}, \mathbf{X}_{it}, \mathbf{U}_{it}) + \varepsilon_{it}.$$

Estimation under strict exogeneity

- In practice, we impose structural assumptions to simplify the model.
- Only contemporary values of **X** and **U** are confounders:

$$Y_{it} = g_{it}(D_{it}, \mathbf{X}_{it}, \mathbf{U}_{it}) + \varepsilon_{it}.$$

- The effects of D, **X** and **U** are additive:

$$Y_{it} = \delta_{it} D_{it} + f_{it}(\mathbf{X}_{it}) + h_{it}(\mathbf{U}_{it}) + \varepsilon_{it}.$$

Estimation under strict exogeneity

- In practice, we impose structural assumptions to simplify the model.
- Only contemporary values of **X** and **U** are confounders:

$$Y_{it} = g_{it}(D_{it}, \mathbf{X}_{it}, \mathbf{U}_{it}) + \varepsilon_{it}.$$

- The effects of D, **X** and **U** are additive:

$$Y_{it} = \delta_{it} D_{it} + f_{it}(\mathbf{X}_{it}) + h_{it}(\mathbf{U}_{it}) + \varepsilon_{it}.$$

- **X** affect Y in a linear manner and $h_{it}(\mathbf{U}_{it})$ has a low-dimensional decomposition.

Estimation under strict exogeneity

- For example, we can assume that $h_{it}(\mathbf{U}_{it}) = \mu + \alpha_i + \xi_t$, then
$$Y_{it} = \mu + \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$
- We get the two-way fixed effects model with heterogeneous treatment effects.

Estimation under strict exogeneity

- For example, we can assume that $h_{it}(\mathbf{U}_{it}) = \mu + \alpha_i + \xi_t$, then

$$Y_{it} = \mu + \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

- We get the two-way fixed effects model with heterogeneous treatment effects.
- The classic approach further assumes that the treatment effect is homogeneous:

$$Y_{it} = \mu + \delta D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

Estimation under strict exogeneity

- For example, we can assume that $h_{it}(\mathbf{U}_{it}) = \mu + \alpha_i + \xi_t$, then

$$Y_{it} = \mu + \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

- We get the two-way fixed effects model with heterogeneous treatment effects.
- The classic approach further assumes that the treatment effect is homogeneous:

$$Y_{it} = \mu + \delta D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \varepsilon_{it}.$$

- Now, the assumption of strict exogeneity becomes: $\mathbb{E}[\varepsilon_{is} | D_{it}, \mathbf{X}_{it}, \alpha_i, \xi_t] = 0$ for any s.

Estimation under strict exogeneity

- The classic two-way fixed effects model can be estimated via the within-estimator.

Estimation under strict exogeneity

- The classic two-way fixed effects model can be estimated via the within-estimator.
- We no longer have consistent estimates if treatment effects are heterogeneous.

Estimation under strict exogeneity

- The classic two-way fixed effects model can be estimated via the within-estimator.
- We no longer have consistent estimates if treatment effects are heterogeneous.
- There are several solutions to the problem.

Estimation under strict exogeneity

- The classic two-way fixed effects model can be estimated via the within-estimator.
- We no longer have consistent estimates if treatment effects are heterogeneous.
- There are several solutions to the problem.
- It is easier to fix the problem under the DID setting (Strezhnev, 2017).

Estimation under strict exogeneity

- The classic two-way fixed effects model can be estimated via the within-estimator.
- We no longer have consistent estimates if treatment effects are heterogeneous.
- There are several solutions to the problem.
- It is easier to fix the problem under the DID setting (Strezhnev, 2017).
- Otherwise, we need to modify the estimand or impose stronger assumptions.

DIDM

- Chaisemartin and D'Haultfœuille (2020) advocate that we should focus only on the instant effects generated by the treatment and estimate

$$\frac{1}{N^*} \sum_{t \geq 2, D_{it} \neq D_{i,t-1}} \delta_{it}$$

where $N^* = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\{t \geq 2, D_{it} \neq D_{i,t-1}\}$.

DIDM

- Chaisemartin and D'Haultfœuille (2020) advocate that we should focus only on the instant effects generated by the treatment and estimate

$$\frac{1}{N^*} \sum_{t \geq 2, D_{it} \neq D_{i,t-1}} \delta_{it}$$

  where $N^* = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\{t \geq 2, D_{it} \neq D_{i,t-1}\}$.
- This estimand is well-defined and can be consistently estimated.

DIDM

- Chaisemartin and D'Haultfœuille (2020) advocate that we should focus only on the instant effects generated by the treatment and estimate

$$\frac{1}{N^*} \sum_{t \geq 2, D_{it} \neq D_{i,t-1}} \delta_{it}$$

  where $N^* = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\{t \geq 2, D_{it} \neq D_{i,t-1}\}$.
- This estimand is well-defined and can be consistently estimated.
- Yet it throws away a ton of information as well thus is not very efficient.

DIDM

- Chaisemartin and D'Haultfœuille (2020) advocate that we should focus only on the instant effects generated by the treatment and estimate

$$\frac{1}{N^*} \sum_{t \geq 2, D_{it} \neq D_{i,t-1}} \delta_{it}$$

where $N^* = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\{t \geq 2, D_{it} \neq D_{i,t-1}\}$.

- This estimand is well-defined and can be consistently estimated.
- Yet it throws away a ton of information as well thus is not very efficient.
- There are more efficient approaches which require stronger assumptions on model specification.

Counterfactual estimation

- Xu (2017) and Liu, Wang, and Xu (2020) combine factor models with the Neyman-Rubin framework

$$Y_{it} = \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \mathbf{f}_t \lambda_i + e_{it},$$

Counterfactual estimation

- Xu (2017) and Liu, Wang, and Xu (2020) combine factor models with the Neyman-Rubin framework

$$Y_{it} = \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \mathbf{f}_t \lambda_i + e_{it},$$

- The basic idea is to use untreated observations ($(i,t) \in \mathcal{O}$) to fit a factor model and employ the model to predict $Y_{it}(0)$ for each treated observation ($(i,t) \in \mathcal{M}$).

Counterfactual estimation

- Xu (2017) and Liu, Wang, and Xu (2020) combine factor models with the Neyman-Rubin framework

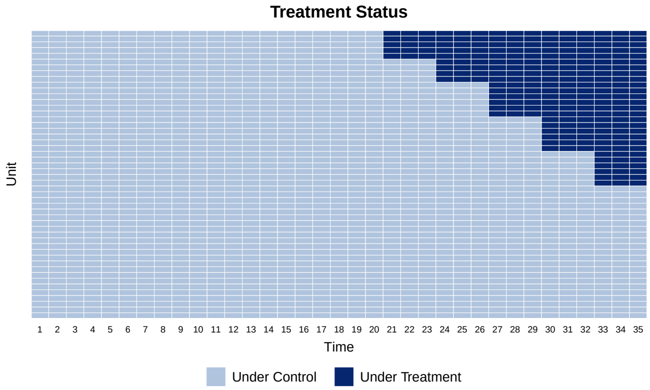
$$Y_{it} = \delta_{it} D_{it} + \mathbf{X}_{it}\beta + \alpha_i + \xi_t + \mathbf{f}_t \lambda_i + e_{it},$$

- The basic idea is to use untreated observations ($(i, t) \in \mathcal{O}$) to fit a factor model and employ the model to predict $Y_{it}(0)$ for each treated observation ($(i, t) \in \mathcal{M}$).
- Clearly, $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$ and

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

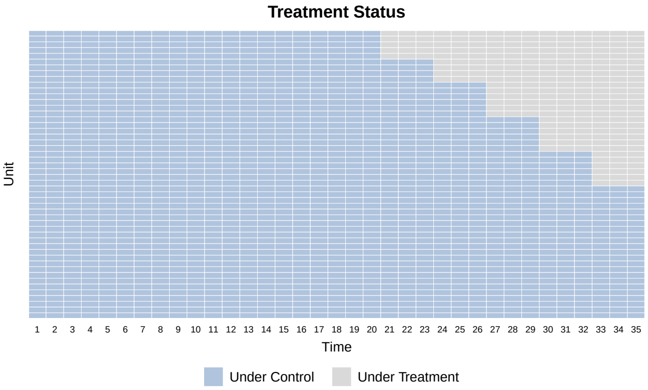
Counterfactual estimation

- In a TSCS setting, treat $Y(1)$ as missing data



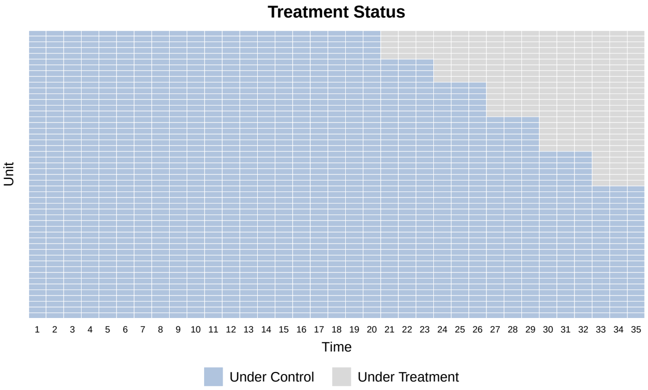
Counterfactual estimation

- In a TSCS setting, treat $Y(1)$ as missing data



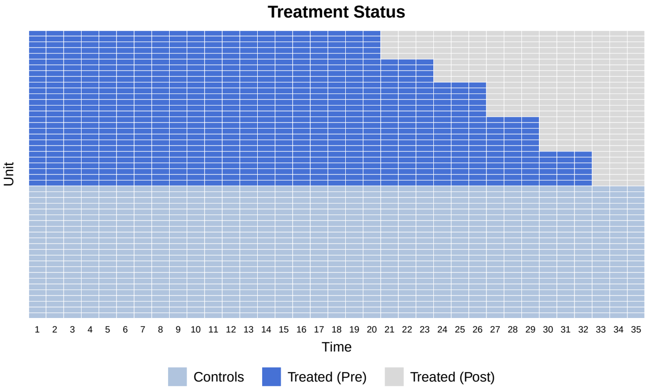
Counterfactual estimation

- In a TSCS setting, treat $Y(1)$ as missing data
- Predict $Y(0)$ based on an outcome model (FE or IFE)



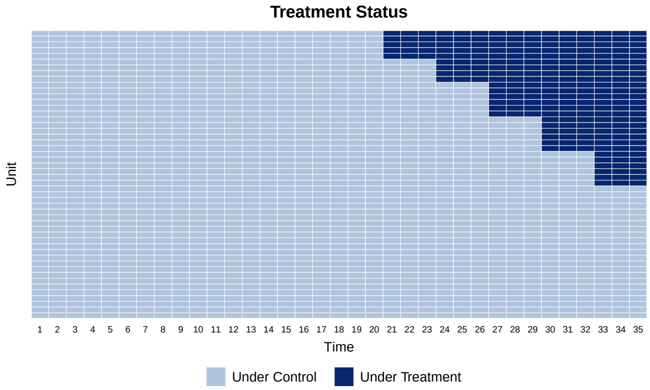
Counterfactual estimation

- In a TSCS setting, treat $Y(1)$ as missing data
- Predict $Y(0)$ based on an outcome model (FE or IFE)
- (Use pre-treatment data for model selection)



Counterfactual estimation

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



Counterfactual estimation

- We should keep in mind that the validity of this approach relies on a series of assumptions.

Counterfactual estimation

- 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 manneer.
  - Unobservable confounders have a low-dimensional decomposition.

Counterfactual estimation

- 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 manneer.
  - Unobservable confounders have a low-dimensional decomposition.
- It also requires strict exogeneity, thus excludes temporal interference.

Counterfactual estimation

- One advantage of the framework is that testing the identification assumption becomes straightforward.

Counterfactual estimation

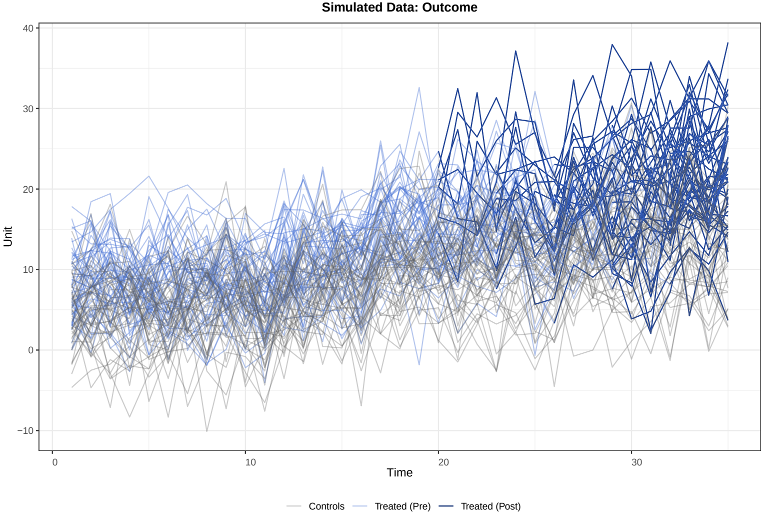
- One advantage of the framework is that testing the identification assumption becomes straightforward.
- Liu, Wang, and Xu (2020) provides a set of tools for practitioners to evaluate the identification assumption.

Counterfactual estimation

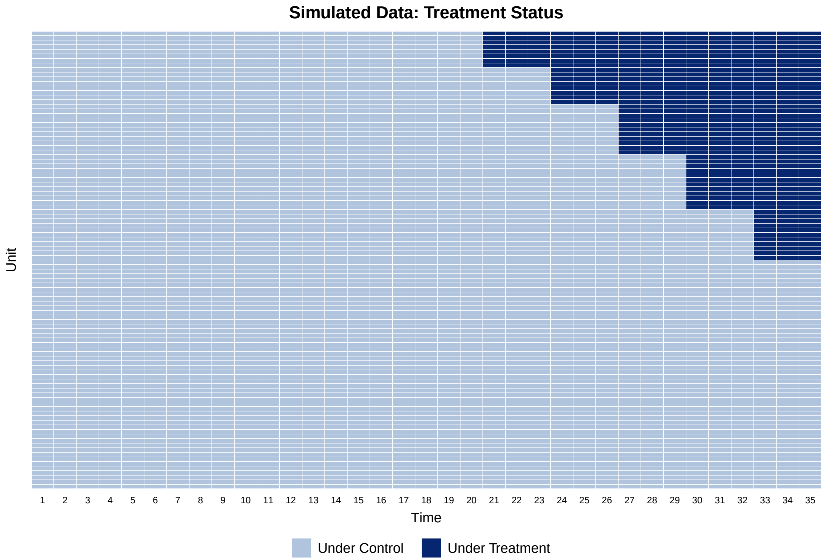
- One advantage of the framework is that testing the identification assumption becomes straightforward.
- Liu, Wang, and Xu (2020) provides a set of tools for practitioners to evaluate the identification assumption.
- A plot of dynamic effects.
- A placebo test: estimate treatment effects before the treatment's onset and test their significance.
- An equivalence test: test whether all the pre-treatment ATTs are equal to zero using TOST.

Example

- Let's consider a DGP with two factors (hence FEct will be biased).

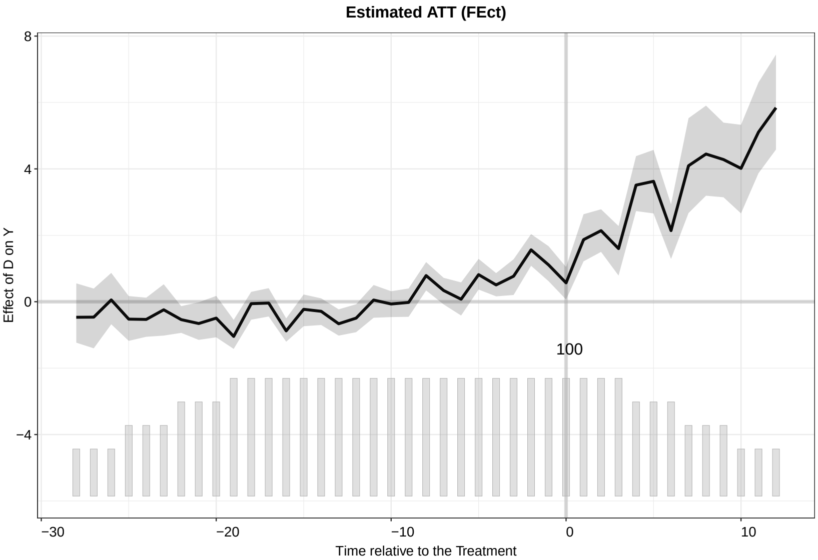


Example

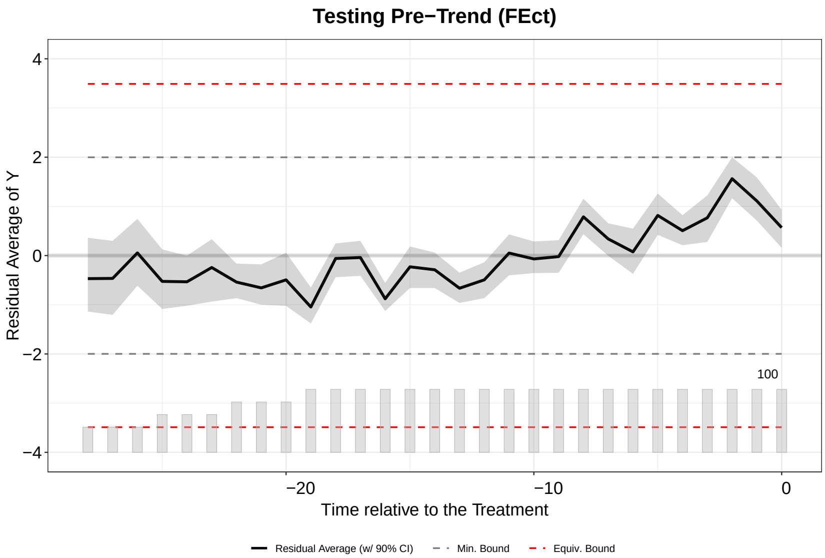

Simulated Data: Treatment Status
Unit
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Time
Under Control
Under Treatment

Example

```
## Bootstrapping for uncertainties ... .100 runs
```



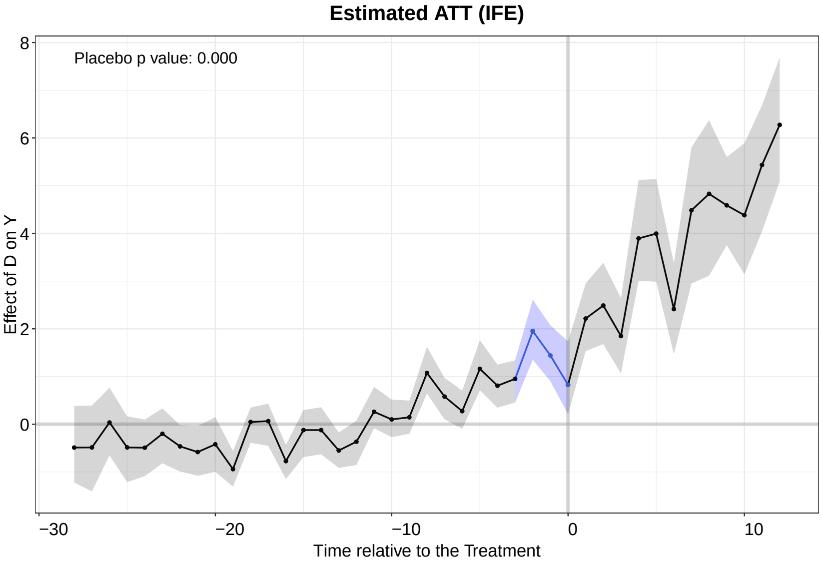
Example



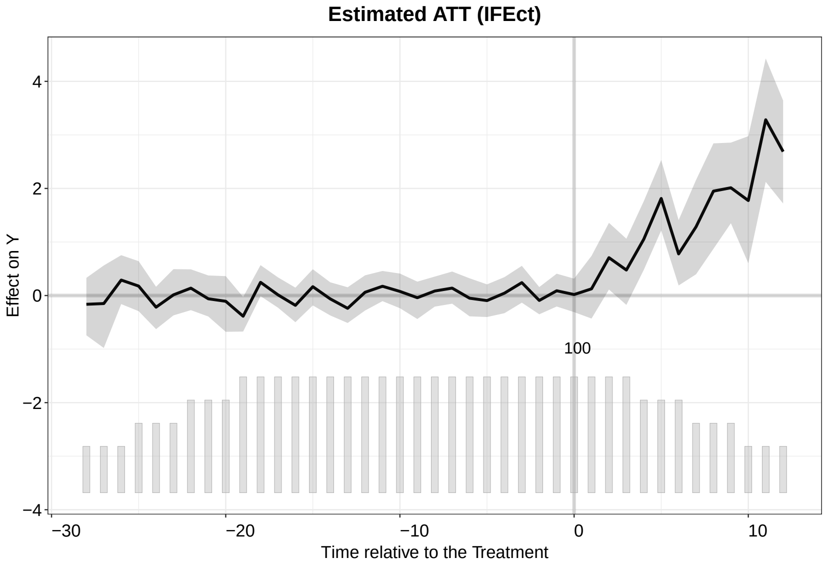
Testing Pre−Trend (FEct)
4
2
0
−2
−4
Residual Average of Y
100
−20
−10
0
Time relative to the Treatment
Residual Average (w/ 90% CI)
Min. Bound
Equiv. Bound

Example

```
## Bootstrapping for uncertainties ... .100 runs
```

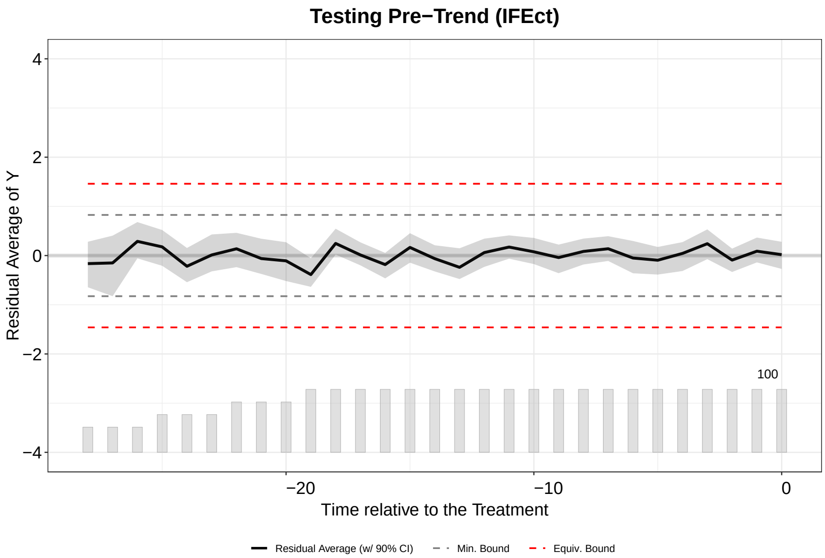


Example



Estimated ATT (IFEct)
Effect on Y
4
2
0
−2
−4
100
−30
−20
−10
0
10
Time relative to the Treatment

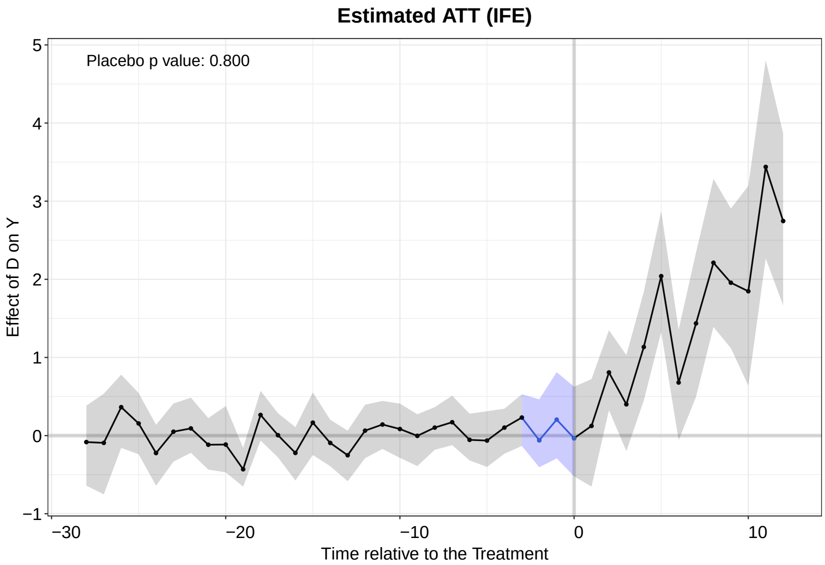
Example



Testing Pre−Trend (IFEct)
Residual Average of Y
4
2
0
−2
−4
100
−20
−10
0
Time relative to the Treatment
Residual Average (w/ 90% CI)
Min. Bound
Equiv. Bound

Example

```
## Bootstrapping for uncertainties ... .100 runs
```



Estimation under sequential ignorability

- The assumption is a weaker version of "select on observables."

Estimation under sequential ignorability

- The assumption is a weaker version of “select on observables.”
- All the techniques we have learned are applicable (regression, weighting, matching).

Estimation under sequential ignorability

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.

Estimation under sequential ignorability

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.
- A giant in this field — Neal!

Estimation under sequential ignorability

- The assumption is a weaker version of “select on observables.”
- All the techniques we have learned are applicable (regression, weighting, matching).
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.
- A giant in this field — Neal!
- The lagged dependent variable does solve a lot of problems, as the time trend is usually the most important confounder.

Estimation under sequential ignorability

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.
- A giant in this field — Neal!
- The lagged dependent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: IPTW estimators (MSMs) and doubly robust estimators.

Estimation under sequential ignorability

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.
- A giant in this field — Neal!
- The lagged dependent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: IPTW estimators (MSMs) and doubly robust estimators.
- Notice that the propensity score for unit i in period t has the form of $P(D_{it}|\mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t})$.

Weighting under sequential ignorability

- Suppose there is no interference and we are interested in the contemporary effect generated by the treatment:

$$\tau_t = \frac{1}{N}\sum_{i=1}^{N}\tau_{it}, \text{ where } \tau_{it} = Y_{it}(1) - Y_{it}(0)$$

Weighting under sequential ignorability

- Suppose there is no interference and we are interested in the contemporary effect generated by the treatment:

$$\tau_t = \frac{1}{N}\sum_{i=1}^{N}\tau_{it}, \text{ where } \tau_{it} = Y_{it}(1) - Y_{it}(0)$$

- Remember that sequential ignorability means $D_{it} \perp Y_{it}(D_{it}) | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$ for any t.

Weighting under sequential ignorability

- Suppose there is no interference and we are interested in the contemporary effect generated by the treatment:

$$\tau_t = \frac{1}{N}\sum_{i=1}^{N}\tau_{it}, \text{ where } \tau_{it} = Y_{it}(1) - Y_{it}(0)$$

- Remember that sequential ignorability means $D_{it} \perp Y_{it}(D_{it}) | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$ for any t.
- We can select a model to estimate the propensity score $e_{it} = P(D_{it} = 1 | \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t})$ and apply the IPTW estimator:

$$\hat{\tau}_t = \frac{1}{N}\sum_{i=1}^{N}\frac{D_{it}Y_{it}}{e_{it}} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_{it})Y_{it}}{1-e_{it}}.$$

Weighting under sequential ignorability

- Suppose there is no interference and we are interested in the contemporary effect generated by the treatment:

$$\tau_t = \frac{1}{N}\sum_{i=1}^{N}\tau_{it}, \text{ where } \tau_{it} = Y_{it}(1) - Y_{it}(0)$$

- Remember that sequential ignorability means $D_{it} \perp Y_{it}(D_{it})|\mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t}$ for any t.
- We can select a model to estimate the propensity score $e_{it} = P(D_{it} = 1|\mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t})$ and apply the IPTW estimator:

$$\hat{\tau}_t = \frac{1}{N}\sum_{i=1}^{N}\frac{D_{it}Y_{it}}{e_{it}} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_{it})Y_{it}}{1-e_{it}}.$$

- It is just another Horvitz-Thompson estimator.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.
- The only confounder is the history of the outcome, $\mathbf{Y}_i^{1:(t-1)}$.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.
- The only confounder is the history of the outcome, $\mathbf{Y}_i^{1:(t-1)}$.
- A natural idea is to balance the distribution of the confounder across the treatment group $\mathcal{T}$ and the control group $\mathcal{C}$, like in CBPS and entropy balancing.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.
- The only confounder is the history of the outcome, $\mathbf{Y}_i^{1:(t-1)}$.
- A natural idea is to balance the distribution of the confounder across the treatment group $\mathcal{T}$ and the control group $\mathcal{C}$, like in CBPS and entropy balancing.
- But when T is large, balancing the entire history can be challenging in practice.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.
- The only confounder is the history of the outcome, $\mathbf{Y}_i^{1:(t-1)}$.
- A natural idea is to balance the distribution of the confounder across the treatment group $\mathcal{T}$ and the control group $\mathcal{C}$, like in CBPS and entropy balancing.
- But when T is large, balancing the entire history can be challenging in practice.
- Hazlett and Xu (2018) suggest that we may instead balance kernels of the history.

Balancing under sequential ignorability

- Let's ignore the covariates from now on.
- The only confounder is the history of the outcome, $\mathbf{Y}_i^{1:(t-1)}$.
- A natural idea is to balance the distribution of the confounder across the treatment group $\mathcal{T}$ and the control group $\mathcal{C}$, like in CBPS and entropy balancing.
- But when T is large, balancing the entire history can be challenging in practice.
- Hazlett and Xu (2018) suggest that we may instead balance kernels of the history.
- The implicit assumption is that the distribution of the outcome history can be well approximated by its kernels.

Balancing under sequential ignorability

- We search for a group of positive weights $\{w_i\}$ which sum to one such that
$$\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\mathbf{K}_i = \sum_{i\in\mathcal{C}} w_i\mathbf{K}_i$$
where $\mathbf{K}_i = (k(\mathbf{Y}_i, \mathbf{Y}_1), k(\mathbf{Y}_i, \mathbf{Y}_2), \ldots, k(\mathbf{Y}_i, \mathbf{Y}_N))$, $\mathbf{Y}_i$ is the pre-treatment outcome history of unit i, and
$$k(\mathbf{Y}_i, \mathbf{Y}_j) = e^{-||\mathbf{Y}_i - \mathbf{Y}_j||^2/h}.$$

Balancing under sequential ignorability

- 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 the ATT estimator and are still working on its asymptotic normality.

Balancing under sequential ignorability

- 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 the ATT estimator and are still working on its asymptotic normality.
- Kim, Imai, and Wang (2019) show that we can do the same via matching.

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.
- Arkhangelsky and Imbens (2019) show that we can get around this problem via weighting under some circumstances.

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.
- Arkhangelsky and Imbens (2019) show that we can get around this problem via weighting under some circumstances.
- We need stronger assumptions on the assignment mechanism.

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.
- Arkhangelsky and Imbens (2019) show that we can get around this problem via weighting under some circumstances.
- We need stronger assumptions on the assignment mechanism.
- Suppose the true propensity score has the following form:

$$E\left[D_{it}|\alpha_i\right] = \frac{exp(\alpha_i + \xi_t)}{1 + exp(\alpha_i + \xi_t)}$$

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.
- Arkhangelsky and Imbens (2019) show that we can get around this problem via weighting under some circumstances.
- We need stronger assumptions on the assignment mechanism.
- Suppose the true propensity score has the following form:

$$E\left[D_{it}|\alpha_i\right] = \frac{exp(\alpha_i + \xi_t)}{1 + exp(\alpha_i + \xi_t)}$$

- They show that now $\mathbf{D}_i^T \perp \mathbf{Y}_{jt}(\mathbf{D}^T)|\frac{1}{T}\sum_{t=1}^{T} D_{it}$.

Doubly robust estimation in TSCS data

- Why can't we directly estimate propensity scores under strict exogeneity?
- We have to assume that there is a correct model with unobservables for the treatment as well.
- Arkhangelsky and Imbens (2019) show that we can get around this problem via weighting under some circumstances.
- We need stronger assumptions on the assignment mechanism.
- Suppose the true propensity score has the following form:

$$E\left[D_{it}|\alpha_i\right] = \frac{exp(\alpha_i + \xi_t)}{1 + exp(\alpha_i + \xi_t)}$$

- They show that now $\mathbf{D}_i^T \perp \mathbf{Y}_{jt}(\mathbf{D}^T) | \frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- $S_i = \frac{1}{T}\sum_{t=1}^{T} D_{it}$ is a sufficient statistic for α_i.

Manifold learning

- Arkhangelsky and Imbens (2019) further illustrate that under this assumption, all we need to do is to find a group of weights (again!) such that:

$$\{\hat{w}_{it}\} = \arg\min \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_{it}^2,$$

$$\text{s.t. } \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_{it} D_{it} = 1, \frac{1}{N} \sum_{i=1}^{N} \hat{w}_{it} = 0, \frac{1}{T} \sum_{t=1}^{T} \hat{w}_{it} = 0,$$

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_{it} \psi_{it}(S_i) = 0, \hat{w}_{it} D_{it} \geq 0.$$

- Then, the estimate of the ATE,

$$\hat{\tau} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_{it} Y_{it}$$

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.
- Feng (2020) formally develops this idea for TSCS data analysis.

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.
- Feng (2020) formally develops this idea for TSCS data analysis.
- He assumes that the only unobservable confounder is the factor loading λ_i and $Y_{it} = \zeta_t(\lambda_i) + \nu_{it}$.

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.
- Feng (2020) formally develops this idea for TSCS data analysis.
- He assumes that the only unobservable confounder is the factor loading λ_i and $Y_{it} = \zeta_t(\lambda_i) + \nu_{it}$.
- As long as we can learn the value of λ_i from Y_{it}, all the classic techniques can be applied.

Manifold learning

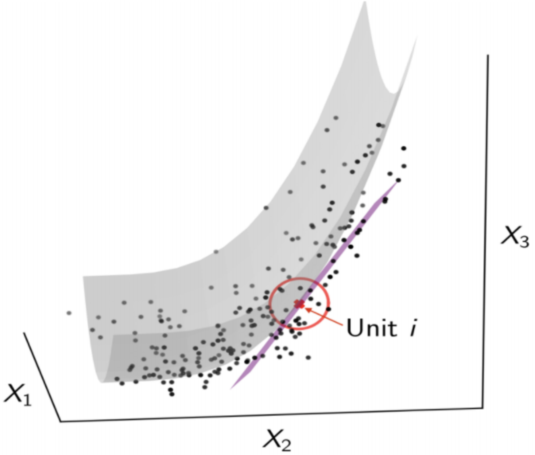
- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.
- Feng (2020) formally develops this idea for TSCS data analysis.
- He assumes that the only unobservable confounder is the factor loading λ_i and $Y_{it} = \zeta_t(\lambda_i) + \nu_{it}$.
- As long as we can learn the value of λ_i from Y_{it}, all the classic techniques can be applied.
- To estimate λ_i, we first match each unit i to K nearest neighbors.

Manifold learning

- Note that we are learning information of the unobservable α_i from the observable variable $\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- This is the idea of manifold learning: we assume that the unobservable variable is a low-dimensional manifold embeded in a high-dimensional space.
- Feng (2020) formally develops this idea for TSCS data analysis.
- He assumes that the only unobservable confounder is the factor loading λ_i and $Y_{it} = \zeta_t(\lambda_i) + \nu_{it}$.
- As long as we can learn the value of λ_i from Y_{it}, all the classic techniques can be applied.
- To estimate λ_i, we first match each unit i to K nearest neighbors.
- Next, we estimate λ_i via PCA on the $K+1$ outcome histories.

Manifold learning

Figure 2: Local Tangent Space Approximation



Temporal interference

- Everything becomes complicated if we allow for temporal interference or “carryover” of the treatment effects.

Temporal interference

- Everything becomes complicated if we allow for temporal interference or "carryover" of the treatment effects.
- Now the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}(\mathbf{D}_i^{1:t})$.

Temporal interference

- Everything becomes complicated if we allow for temporal interference or "carryover" of the treatment effects.
- Now the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}(\mathbf{D}_i^{1:t})$.
- Treatment assignments are usually temporally dependent in panel data.

Temporal interference

- Everything becomes complicated if we allow for temporal interference or "carryover" of the treatment effects.
- Now the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}(\mathbf{D}_i^{1:t})$.
- Treatment assignments are usually temporally dependent in panel data.
- Now, past treatments become confounders.

Temporal interference

- Everything becomes complicated if we allow for temporal interference or “carryover” of the treatment effects.
- Now the outcome is affected by the history of treatment assignment: $Y_{it} = Y_{it}(\mathbf{D}_i^{1:t})$.
- Treatment assignments are usually temporally dependent in panel data.
- Now, past treatments become confounders.
- We have to control their influence under either assumption.

Temporal interference under strict exogeneity

- Under strict exogeneity, temporal interference is not a big concern if we also have staggered adoption.

Temporal interference under strict exogeneity

- Under strict exogeneity, temporal interference is not a big concern if we also have staggered adoption.
- The only difference is that the estimated ATT in each period is the cumulative effect of the treatment assignment history.

Temporal interference under strict exogeneity

- Under strict exogeneity, temporal interference is not a big concern if we also have staggered adoption.
- The only difference is that the estimated ATT in each period is the cumulative effect of the treatment assignment history.
- Otherwise, there is one more reason for the invalid second difference to occur.

Temporal interference under strict exogeneity

- Under strict exogeneity, temporal interference is not a big concern if we also have staggered adoption.
- The only difference is that the estimated ATT in each period is the cumulative effect of the treatment assignment history.
- Otherwise, there is one more reason for the invalid second difference to occur.
- We can still estimate the quantity defined by Chaisemartin and D'Haultfœuille (2020), but nothing else.

Temporal interference under strict exogeneity

- Under strict exogeneity, temporal interference is not a big concern if we also have staggered adoption.
- The only difference is that the estimated ATT in each period is the cumulative effect of the treatment assignment history.
- Otherwise, there is one more reason for the invalid second difference to occur.
- We can still estimate the quantity defined by Chaisemartin and D'Haultfœuille (2020), but nothing else.
- Sequential ignorability is the more suitable assumption if we have temporal interference

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!
- We may similarly construct marginalized estimands.

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!
- We may similarly construct marginalized estimands.
- We can further construct a function to aggregate the history's effect: $\tau_{it} = g(\mathbf{D}_i^{s:t}) - g(\mathbf{D}_i^{s:t})$.

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!
- We may similarly construct marginalized estimands.
- We can further construct a function to aggregate the history's effect: $\tau_{it} = g(\mathbf{D}_i^{s:t}) - g(\mathbf{D}_i^{s:t})$.
- For example, the number of periods under the treatment: $g(\mathbf{D}_i^{s:t}) = \sum_{t'=s}^{t} \mathbf{D}_{it'}$.

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!
- We may similarly construct marginalized estimands.
- We can further construct a function to aggregate the history's effect: $\tau_{it} = g(\mathbf{D}_i^{s:t}) - g(\mathbf{D}_i^{s:t})$.
- For example, the number of periods under the treatment: $g(\mathbf{D}_i^{s:t}) = \sum_{t'=s}^{t} \mathbf{D}_{it'}$.
- We can obtain valid estimate for the history's aggregated effect after weighting observations properly.

Temporal interference under sequential ignorability

- Suppose there are T weeks, then there are 2^T possible "potential histories."
- Too many!
- We may similarly construct marginalized estimands.
- We can further construct a function to aggregate the history's effect: $\tau_{it} = g(\mathbf{D}_i^{s:t}) - g(\mathbf{D}_i^{s:t})$.
- For example, the number of periods under the treatment: $g(\mathbf{D}_i^{s:t}) = \sum_{t'=s}^{t} \mathbf{D}_{it'}$.
- We can obtain valid estimate for the history's aggregated effect after weighting observations properly.
- Now,

$$P(\mathbf{D}_i^{s:t} = \mathbf{d}^{s:t}) = \prod_{s'=s}^{t} P(D_{is'} | \mathbf{D}_i^{1:(s'-1)}, \mathbf{Y}_i^{1:(s'-1)}, \mathbf{X}_i^{1:s'}).$$

Spatial interference in TSCS data

- If we also have spatial/between-unit interference, then even the classic DID estimator will be biased.

Spatial interference in TSCS data

- If we also have spatial/between-unit interference, then even the classic DID estimator will be biased.
- Wang (2020) shows that the limit of the DID estimator has no substantive interpretation under spatial interference.

Spatial interference in TSCS data

- If we also have spatial/between-unit interference, then even the classic DID estimator will be biased.
- Wang (2020) shows that the limit of the DID estimator has no substantive interpretation under spatial interference.
- Intuitively, the control group is contaminated and we cannot eliminate the bias via outcome adjustment.

Spatial interference in TSCS data

- If we also have spatial/between-unit interference, then even the classic DID estimator will be biased.
- Wang (2020) shows that the limit of the DID estimator has no substantive interpretation under spatial interference.
- Intuitively, the control group is contaminated and we cannot eliminate the bias via outcome adjustment.
- Yet under sequential ignorability, we can combine the idea in Aronow, Samii, and Wang (2020) and MSMs to generate consistent estimates for both the direct and the indirect effects.

References I

Arkhangelsky, Dmitry, and Guido W Imbens. 2019. "Double-Robust Identification for Causal Panel Data Models." *arXiv Preprint arXiv:1909.09412*.

Aronow, Peter M, Cyrus Samii, and Ye Wang. 2020. "Design-Based Inference for Spatial Experiments with Interference." *arXiv Preprint arXiv:2010.13599*.

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

Feng, Yingjie. 2020. "Causal Inference in Possibly Nonlinear Factor Models." *arXiv Preprint arXiv:2008.13651*.

Hazlett, Chad, and Yiqing Xu. 2018. "Trajectory Balancing: A General Reweighting Approach to Causal Inference with Time-Series

References II

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."

Liu, Licheng, Ye Wang, and Yiqing Xu. 2020. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *Available at SSRN 3555463*.

Wang, Ye. 2020. "Causal Inference Under Temporal and Spatial Interference."

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1). Cambridge University Press: 57–76.