Quant II

TSCS Data Estimation

Ye Wang

3/28/2018

Outline

- Roadmap for TSCS data estimation
- Sequential experiments
- Semi-parametric models
- Trajectory balancing

Roadmap for TSCS data estimation

- What is the difference between TSCS and panel data?

Roadmap for TSCS data estimation

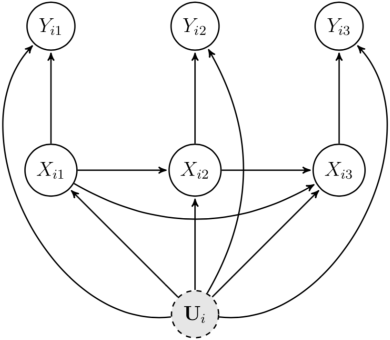
- What is the difference between TSCS and panel data?
- What is unique for TSCS data? Why not just pooled OLS?

Roadmap for TSCS data estimation

- What is the difference between TSCS and panel data?
- What is unique for TSCS data? Why not just pooled OLS? Interference in time is all that matters.

Roadmap for TSCS data estimation

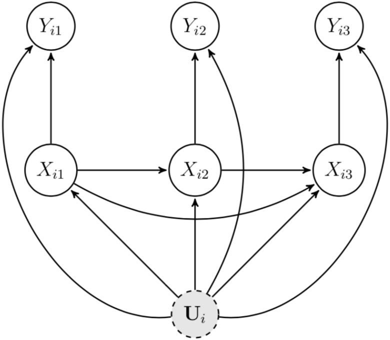
- What is the difference between TSCS and panel data?
- What is unique for TSCS data? Why not just pooled OLS? Interference in time is all that matters.



- What is the non-parametric estimator in TSCS data? How to use matching to predict counterfactual for Y_{it}?

Roadmap for TSCS data estimation

- What is the difference between TSCS and panel data?
- What is unique for TSCS data? Why not just pooled OLS?
  Interference in time is all that matters.



- What is the non-parametric estimator in TSCS data? How to use matching to predict counterfactual for Y_{it}?
  Doesn't exist. Only one observation exists for unit i, period t (Imai and Kim, 2019).

Roadmap for TSCS data estimation

- Choices must be made

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history. Sequential experiment

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history. Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Semi-parametric models (FEct, gsynth, mc, etc.)

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Semi-parametric models (FEct, gsynth, mc, etc.)
- Assumption III: The trajectories of the treatment group's and the control group's past outcome are parallel.

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Semi-parametric models (FEct, gsynth, mc, etc.)
- Assumption III: The trajectories of the treatment group's and the control group's past outcome are parallel.
  Synthetic control and weighting

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Semi-parametric models (FEct, gsynth, mc, etc.)
- Assumption III: The trajectories of the treatment group's and the control group's past outcome are parallel.
  Synthetic control and weighting
- What if I believe in none?

Roadmap for TSCS data estimation

- Choices must be made
- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Semi-parametric models (FEct, gsynth, mc, etc.)
- Assumption III: The trajectories of the treatment group's and the control group's past outcome are parallel.
  Synthetic control and weighting
- What if I believe in none?
  Make parametric assumptions

Sequential experiments: then and now

- A big name we cannot ignore in this literature

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation
  Works well in epidemiology

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation
  Works well in epidemiology
- Suppose you participate in a trial to compare two treatments

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation
  Works well in epidemiology
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation
  Works well in epidemiology
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much
  Three weeks later, you start to get treatment II

Sequential experiments: then and now

- A big name we cannot ignore in this literature
  Neal!
- Basic idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.).
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.
- Modern approach: Jim Robins's g-estimation
  Works well in epidemiology
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much
  Three weeks later, you start to get treatment II
  How to estimate the effect of both treatments?

Sequential experiments: estimation

- History decides your final health status

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.
- We can still use the conventional approach under the selection-on-observables assumption, but it will be biased.

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.
- We can still use the conventional approach under the selection-on-observables assumption, but it will be biased.
- If we control both $Y_{i,t-1}$, what will happen?

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.
- We can still use the conventional approach under the selection-on-observables assumption, but it will be biased.
- If we control both $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $X_{i,t-1}$ (but not for X_{it}).

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.
- We can still use the conventional approach under the selection-on-observables assumption, but it will be biased.
- If we control both $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $X_{i,t-1}$ (but not for X_{it}).
- What if we don't?

Sequential experiments: estimation

- History decides your final health status
- Suppose there are T weeks, then there are 2^T possible "dynamic treatment regimes."Too many!
- Usually we focus on the causal effect of switching the treatment status in the last a few periods.
- We need sequential ignorability: potential outcomes in the future are independent to the treatment conditional on history.
- We can still use the conventional approach under the selection-on-observables assumption, but it will be biased.
- If we control both $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $X_{i,t-1}$ (but not for X_{it}).
- What if we don't?
  Omitted variable bias for $X_{i,t-1}$.

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- The former models Y (response) and the latter models D (treatment)

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- The former models Y (response) and the latter models D (treatment)
- Let's assume linearity and one lag is necessary for sequential ignorability

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- The former models Y (response) and the latter models D (treatment)
- Let's assume linearity and one lag is necessary for sequential ignorability
- SNMMs:
  - Regress Y_{it} on $(Y_{i,t-1}, X_{it}, X_{i,t-1})$ and get the coefficient of X_{it}, $\hat{\gamma}_0$.
  - Transform Y_{it} into $\tilde{Y}_{it} = Y_{it} - \hat{\gamma}_0 X_{it}$
  - Regress $\tilde{Y}_{it}$ on $(Y_{i,t-2}, X_{i,t-1}, X_{i,t-2})$ and get the coefficient of $X_{i,t-1}$, $\hat{\gamma}_1$.

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- The former models Y (response) and the latter models D (treatment)
- Let's assume linearity and one lag is necessary for sequential ignorability
- SNMMs:
  - Regress Y_{it} on $(Y_{i,t-1}, X_{it}, X_{i,t-1})$ and get the coefficient of X_{it}, $\hat{\gamma}_0$.
  - Transform Y_{it} into $\tilde{Y}_{it} = Y_{it} - \hat{\gamma}_0 X_{it}$
  - Regress $\tilde{Y}_{it}$ on $(Y_{i,t-2}, X_{i,t-1}, X_{i,t-2})$ and get the coefficient of $X_{i,t-1}$, $\hat{\gamma}_1$.
- MSMs with IPTWs: estimate the propensity score for each X_{it}, and use IPTW for ATT.

$$\widehat{SW}_{it} = \prod_{t=1}^{t} \frac{\widehat{\Pr}[X_{it} \mid X_{i,t-1}; \widehat{\gamma}]}{\widehat{\Pr}[X_{it} \mid Z_{it}, Y_{i,t-1}, X_{i,t-1}; \widehat{\alpha}]}.$$

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

The only problem is that we cannot apply the within-estimator here (unbalanced panel).

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?
- From a different perspective, we are approximating the matrix $\mathbf{Y} = \{Y_{it}\}$ with the product of two low-dimension matrices: $\mathbf{Y} = \mathbf{L} + \varepsilon$, where $\mathbf{L} = \Lambda\mathbf{F}$ (matrix completion or MC).

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}
- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?
- From a different perspective, we are approximating the matrix $\mathbf{Y} = \{Y_{it}\}$ with the product of two low-dimension matrices: $\mathbf{Y} = \mathbf{L} + \varepsilon$, where $\mathbf{L} = \Lambda\mathbf{F}$ (matrix completion or MC).
  IFE and MC are equivalent when MC directly penalizes the matrix dimension r (hard-impute) rather than the magnitude of eigen values (soft-impute).

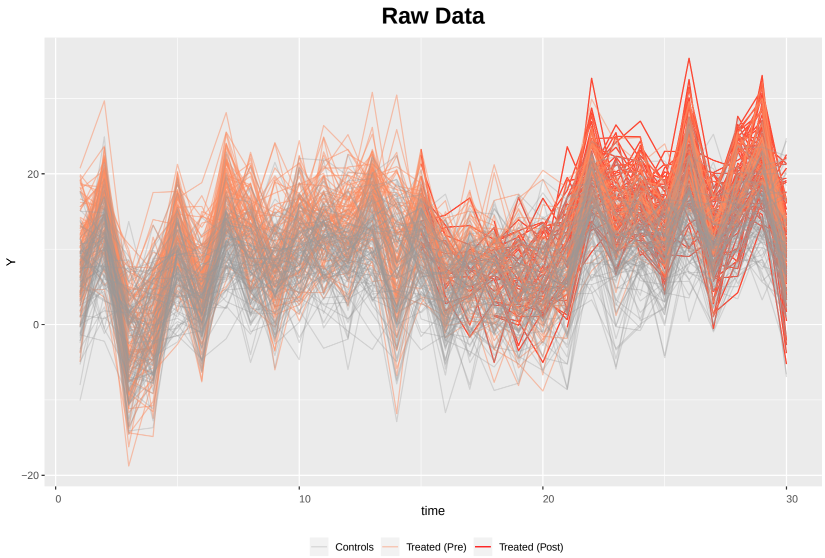
Semi-parametric models: estimation

- IFE relies on the singular value decomposition (SVD).
- We start from initial parameters, estimate β via OLS, and obtain λ_i and $\mathbf{f}_t$ via SVD of the residuals.
- r is selected by cross-validation.
- MC (soft-impute) uses the direct sum decomposition: the residual at each round can be decomposed into two orthogonal parts.
- MC (soft-impute) is biased in finite samples.
- Both work well in large samples (relative performance depends on the strength of factors).
- We can add other parts to the model such as the lagged dependent variable or ensemble methods (Athey et al., 2019)

Semi-parametric models: tools

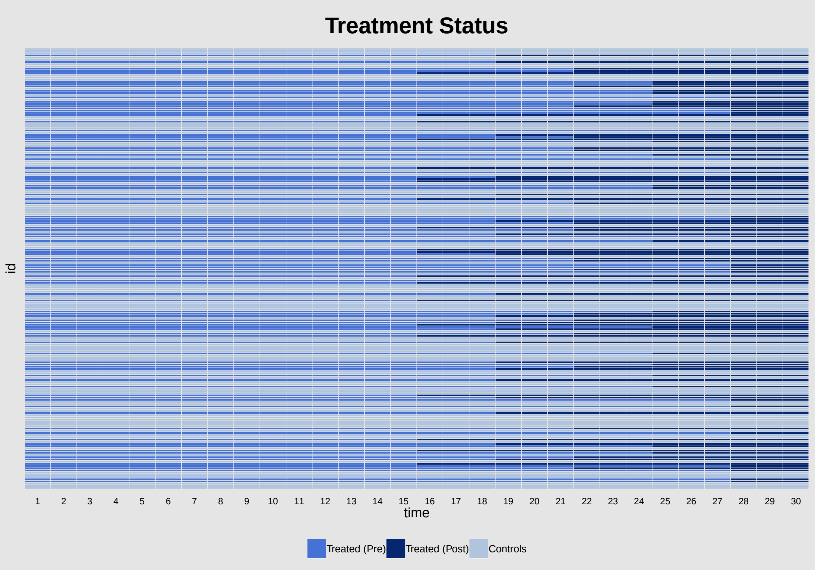
- A series of packages provided by Yiqing Xu at UCSD and his collaborators
- panelView for displaying the basic patterns
- fastplm for estimating the two-way FE models fast
- gsynth for estimating the interactive FE models
- fect to rule them all

Example



Raw Data
20
0
−20
Y
0
10
20
30
time
Controls
Treated (Pre)
Treated (Post)

Example

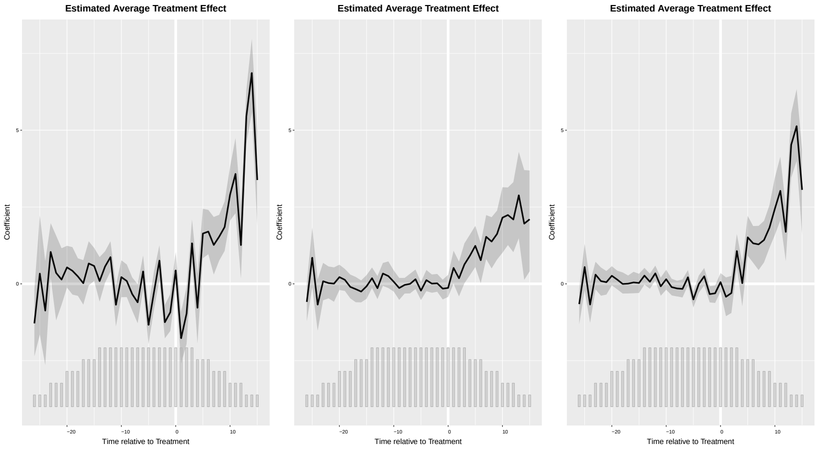


Treatment Status
id
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
time
Treated (Pre)
Treated (Post)
Controls

Example

```
##
##      Coef Std. Error t value Pr(>|t|) CI_lower CI_upper
## D  0.617      0.166   3.727        0    0.293    0.942
## X1 0.987      0.042  23.569        0    0.905    1.069
## X2 3.008      0.042  72.460        0    2.926    3.089
##
## ---
##
## Residual standard error: 3.205 on 5768 degrees of freedo
## Multiple R-squared(full model): 0.817    Adjusted R-squa
## Multiple R-squared(proj model): 0.502    Adjusted R-squa
## F-statistic(full model): 111.640 on 231 and 5768 DF, p-v
## F-statistic(proj model): 1936.160 on 3 and 5768 DF, p-va
```

Example



Estimated Average Treatment Effect
Coefficient
Time relative to Treatment
5
0
−20
−10
0
10
Estimated Average Treatment Effect
Coefficient
Time relative to Treatment
Estimated Average Treatment Effect
Coefficient
Time relative to Treatment

Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.

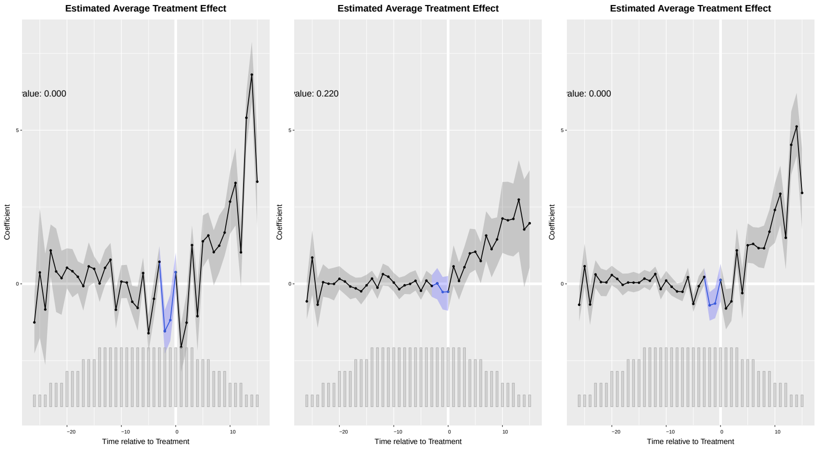
Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.
- How to test the assumption in this more general case?

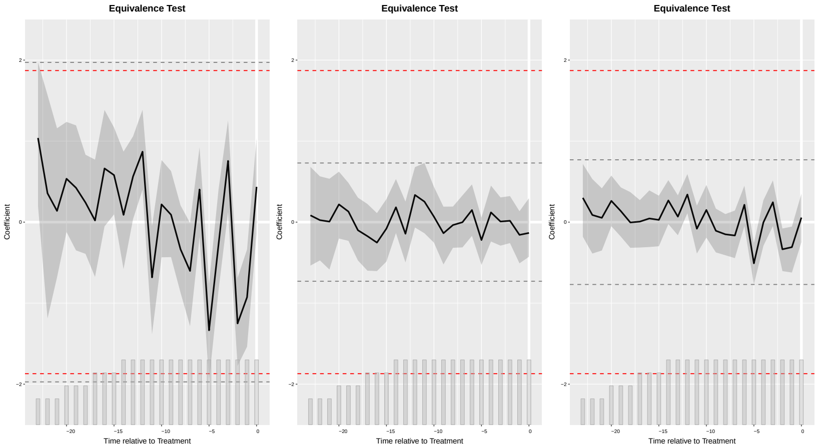
Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.
- How to test the assumption in this more general case?
- Liu, Wang, and Xu (2019): dynamic treatment effects, equivalence test and placebo test.

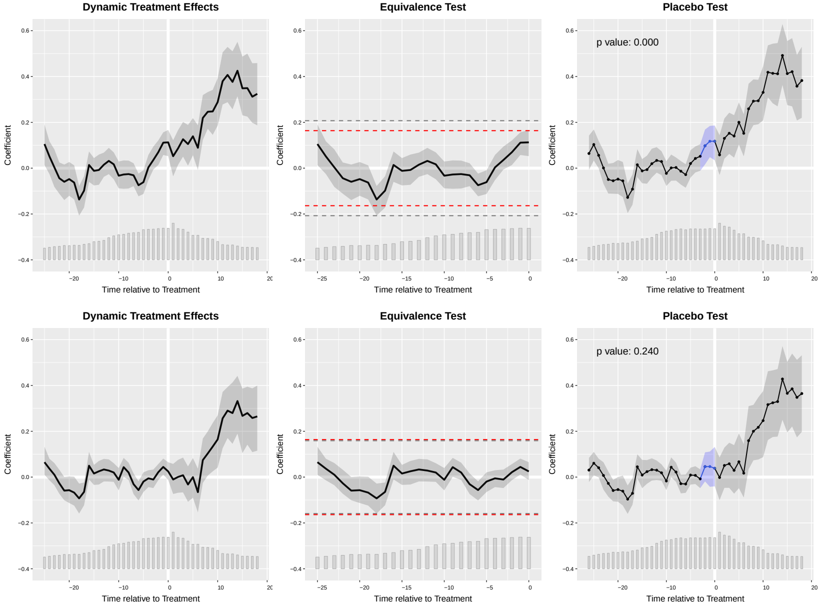
Example


Estimated Average Treatment Effect
value: 0.000
Coefficient
Time relative to Treatment
Estimated Average Treatment Effect
value: 0.220
Coefficient
Time relative to Treatment
Estimated Average Treatment Effect
value: 0.000
Coefficient
Time relative to Treatment

Example


Equivalence Test
Equivalence Test
Equivalence Test
Coefficient
Time relative to Treatment
2
0
−2
−20
−15
−10
−5
0

Example: Tomz et al. (2007)



Trajectory balancing

- Outcomes in a TSCS dataset can be divided into four parts:

$$\boldsymbol{Y}^{\text{obs}} = \begin{pmatrix} \boldsymbol{Y}^{\text{obs}}_{\text{t,post}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,post}} \\ \boldsymbol{Y}^{\text{obs}}_{\text{t,pre}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,pre}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{\text{t,post}}(1) & \boldsymbol{Y}_{\text{c,post}}(0) \\ \boldsymbol{Y}_{\text{t,pre}}(0) & \boldsymbol{Y}_{\text{c,pre}}(0) \end{pmatrix}$$

Trajectory balancing

- Outcomes in a TSCS dataset can be divided into four parts:

$$\boldsymbol{Y}^{\text{obs}} = \begin{pmatrix} \boldsymbol{Y}^{\text{obs}}_{\text{t,post}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,post}} \\ \boldsymbol{Y}^{\text{obs}}_{\text{t,pre}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,pre}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{\text{t,post}}(1) & \boldsymbol{Y}_{\text{c,post}}(0) \\ \boldsymbol{Y}_{\text{t,pre}}(0) & \boldsymbol{Y}_{\text{c,pre}}(0) \end{pmatrix}$$

- All we need to do is to predict $\mathbf{Y}_{t,post}(0)$ using the weighted sum of the other three parts.

Trajectory balancing

- Outcomes in a TSCS dataset can be divided into four parts:

$$\boldsymbol{Y}^{\text{obs}} = \begin{pmatrix} \boldsymbol{Y}^{\text{obs}}_{\text{t,post}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,post}} \\ \boldsymbol{Y}^{\text{obs}}_{\text{t,pre}} & \boldsymbol{Y}^{\text{obs}}_{\text{c,pre}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{\text{t,post}}(1) & \boldsymbol{Y}_{\text{c,post}}(0) \\ \boldsymbol{Y}_{\text{t,pre}}(0) & \boldsymbol{Y}_{\text{c,pre}}(0) \end{pmatrix}$$

- All we need to do is to predict $\mathbf{Y}_{t,post}(0)$ using the weighted sum of the other three parts.
- The weights should minimize the difference between $\mathbf{Y}_{t,pre}(0)$ and $\mathbf{Y}_{c,pre}(0)$.

Several approaches

- Classic one: Synthetic control.

Several approaches

- Classic one: Synthetic control.
- Problems: hard to get balance with many units and periods.
- Doudchenko and Imbens (2016): Directly minimize the difference with a penalty function.
- Hazlett and Xu (2017): Balance moments of the kernelized outcomes.
- Imai, Kim, and Wang (2018): Matching on the pre-treatment outcomes.

Future

- What the TA is working on...
- Clearly there are some gaps among different branches of the literature.
- How to unify sequential experiments with semi-parametric models?
- How to think about TSCS models from the design-based perspective? (Athey and Imbens, 2018)
- What is the common ground between spatial interference and temporal interference?
- For example, what can we do to estimate the dynamic effects in a network? (Egami, 2019)