Frontier Topics in Empirical Economics: Week 9
Causal Inference with Panel Data II

Zibin Huang [1]

[1]College of Business, Shanghai University of Finance and Economics

November 18, 2024

Staggered DID: Settings

- Up until now, we consider a plain vanilla DID setting
- Some policy is implemented at time t_0 in one set of units (treated group), but not the other set of units (untreated group)
- Let's go to more general case of TWFE $\Longrightarrow$ Staggered DID
- Policy can be rolled out in different places at different time

Staggered DID: Settings

- Up until now, we consider a plain vanilla DID setting
- Some policy is implemented at time t_0 in one set of units (treated group), but not the other set of units (untreated group)
- Let's go to more general case of TWFE $\Longrightarrow$ Staggered DID
- Policy can be rolled out in different places at different time

Staggered DID: Settings

- Up until now, we consider a plain vanilla DID setting
- Some policy is implemented at time t_0 in one set of units (treated group), but not the other set of units (untreated group)
- Let's go to more general case of TWFE $\Rightarrow$ Staggered DID
- Policy can be rolled out in different places at different time

Staggered DID: Settings

- Up until now, we consider a plain vanilla DID setting
- Some policy is implemented at time t_0 in one set of units (treated group), but not the other set of units (untreated group)
- Let's go to more general case of TWFE $\Rightarrow$ Staggered DID
- Policy can be rolled out in different places at different time

Staggered DID: Settings

- Up until now, we consider a plain vanilla DID setting
- Some policy is implemented at time t_0 in one set of units (treated group), but not the other set of units (untreated group)
- Let's go to more general case of TWFE $\Rightarrow$ Staggered DID
- Policy can be rolled out in different places at different time

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \quad (1)$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

Staggered DID: Settings

- Assume we have policy implemented at group g level, rolling out in different periods
- Let's run the same TWFE regression for individual i in group g at time t:

$$Y_{igt} = \gamma_g + \lambda_t + \delta D_{gt} + \epsilon_{igt} \tag{1}$$

- $D_{gt} = 1$ if group g is treated at time t
- In homogeneous treatment effect case: δ is TE
- In heterogeneous treatment effect case: δ is a weighted sum of ATE in each group and period
- But, is this in general a good estimator? NO!

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

TWFE: Issues

De Chaisemartin and d'Haultfoeuille (2020) Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

- CD(2020) proposes a main issue of the TWFE when treatment effect is heterogeneous
- TWFE identifies a weighted average TE
- But some of the weights can be negative
- Thus, the weighted average may be negative even if signs of all cell TEs are positive
- Let's see why (Please read CD(2020), this is important!)

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

Setup in CD(2020)

- Three-level of data: individual i, group g, period t
- $N_{g,t}$ is the number of observations in cell (g,t), N is the total number of samples
- Assume $D_{i,g,t}$ is a binary treatment, $Y_{i,g,t}(0)$ and $Y_{i,g,t}(1)$ are potential outcomes
- We have (g,t) cell-level average variables as:

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(0)$$

$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} Y_{i,g,t}$$

CD(2020): Settings

- Assumption 1: For all (g,t), $N_{g,t} > 0$ (Full Support)
- Assumption 2: For all (g,t), $D_{i,g,t} = D_{g,t}$ (Sharp Design)
  Treatment is identical for everyone in the same group-time cell
- Assumption 3: Vectors $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{1 \le t \le T}$ are mutually independent
  No correlation across groups (correlations within group across time is allowed)
  Weaker version of iid

CD(2020): Settings

- Assumption 1: For all (g, t), $N_{g,t} > 0$ (Full Support)
- Assumption 2: For all (g, t), $D_{i,g,t} = D_{g,t}$ (Sharp Design)
  Treatment is identical for everyone in the same group-time cell
- Assumption 3: Vectors $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{1 \leq t \leq T}$ are mutually independent
  No correlation across groups (correlations within group across time is allowed)
  Weaker version of iid

CD(2020): Settings

- Assumption 1: For all (g, t), $N_{g,t} > 0$ (Full Support)
- Assumption 2: For all (g, t), $D_{i,g,t} = D_{g,t}$ (Sharp Design)
  Treatment is identical for everyone in the same group-time cell
- Assumption 3: Vectors $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{1 \le t \le T}$ are mutually independent
  No correlation across groups (correlations within group across time is allowed)
  Weaker version of iid

CD(2020): Settings

- Assumption 1: For all (g, t), $N_{g,t} > 0$ (Full Support)
- Assumption 2: For all (g, t), $D_{i,g,t} = D_{g,t}$ (Sharp Design)
  Treatment is identical for everyone in the same group-time cell
- Assumption 3: Vectors $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{1 \leq t \leq T}$ are mutually independent
  No correlation across groups (correlations within group across time is allowed)
  Weaker version of iid

CD(2020): Settings

- Assumption 4:
  For all (g,t), $E(Y_{g,t}(0) - Y_{g,t-1}(0) | D_{g,1}, \ldots, D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - a. This is called "Strong exogeneity"
  - a. Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - a. No Ashenfelter dip
  - a. Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- Assumption 4:
  For all (g, t), $E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - This is called "Strong exogeneity".
  - Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - No Ashenfelter dip
  - Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- Assumption 4:
  For all (g, t), $E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - This is called "Strong exogeneity".
  - Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - No Ashenfelter dip
  - Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- Assumption 4:
  For all (g, t), $E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - This is called "Strong exogeneity".
  - Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - No Ashenfelter dip
  - Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- Assumption 4:
  For all (g, t), $E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - This is called "Strong exogeneity".
  - Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - No Ashenfelter dip
  - Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- Assumption 4:
  For all (g, t), $E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1}, ..., D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0))$
  - This is called "Strong exogeneity".
  - Treatment cannot be assigned to some group because they experienced a negative/positive shock
  - No Ashenfelter dip
  - Exogenous treatment group assignment and treatment timing

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + c_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g
  Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + \epsilon_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + \epsilon_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g
  Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + \epsilon_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g
  Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + \epsilon_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): Settings

- In a more general case than staggered DID, exit is also allowed
- Some groups can cancel the policy after some periods, then we have $D_{g,t} = 1, D_{g,t+1} = 0$
- Let's run the TWFE regression:

$$Y_{igt} = \gamma_g + \lambda_t + \beta^{fe} D_{gt} + \epsilon_{igt}$$

- Assumption 5: For $t \geq 2$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g Common trends
- Potential outcomes without treatment evolve identically across groups.

CD(2020): ATT and TWFE Estimator

- ATT: $\delta^{TR} = E[Y_{i,g,t}(1) - Y_{i,g,t}(0)|D_{g,t} = 1]$
- Cell average TE: $\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$
- Thus, we have ATT to be an expected weighted average of cell averages:

$$\delta^{TR} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t}] \quad (2)$$

- Let $\hat{\beta}^{fe}$ be the TWFE estimator and $\beta^{fe} = E(\hat{\beta}^{fe})$

CD(2020): ATT and TWFE Estimator

- ATT: $\delta^{TR} = E[Y_{i,g,t}(1) - Y_{i,g,t}(0)|D_{g,t} = 1]$
- Cell average TE: $\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$
- Thus, we have ATT to be an expected weighted average of cell averages:

$$\delta^{TR} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t}] \tag{2}$$

- Let $\hat{\beta}^{fe}$ be the TWFE estimator and $\beta^{fe} = E(\hat{\beta}^{fe})$

CD(2020): ATT and TWFE Estimator

- ATT: $\delta^{TR} = E[Y_{i,g,t}(1) - Y_{i,g,t}(0)|D_{g,t} = 1]$
- Cell average TE: $\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$
- Thus, we have ATT to be an expected weighted average of cell averages:

$$\delta^{TR} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t}] \quad (2)$$

- Let $\hat{\beta}^{fe}$ be the TWFE estimator and $\beta^{fe} = E(\hat{\beta}^{fe})$

CD(2020): ATT and TWFE Estimator

- ATT: $\delta^{TR} = E[Y_{i,g,t}(1) - Y_{i,g,t}(0)|D_{g,t} = 1]$
- Cell average TE: $\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$
- Thus, we have ATT to be an expected weighted average of cell averages:

$$\delta^{TR} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t}] \tag{2}$$

- Let $\hat{\beta}^{fe}$ be the TWFE estimator and $\beta^{fe} = E(\hat{\beta}^{fe})$

CD(2020): ATT and TWFE Estimator

- ATT: $\delta^{TR} = E[Y_{i,g,t}(1) - Y_{i,g,t}(0)|D_{g,t} = 1]$
- Cell average TE: $\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i \in (g,t)} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$
- Thus, we have ATT to be an expected weighted average of cell averages:

$$\delta^{TR} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t}] \tag{2}$$

- Let $\hat{\beta}^{fe}$ be the TWFE estimator and $\beta^{fe} = E(\hat{\beta}^{fe})$

CD(2020): ATT and TWFE Estimator

- Let $\epsilon_{g,t}$ denote the residual of the regression of $D_{g,t}$ on group and period FE

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{3}$$

Theorem 1 in CD(2020)

If we have Assumption 1-5, then

$$\beta^{fe} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} w_{g,t} \Delta_{g,t}]$$

$$w_{g,t} = \frac{\epsilon_{g,t}}{\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \epsilon_{g,t}}$$

The TWFE estimator is a weighted average of cell-level ATE, with $w_{g,t}$ as weights.

CD(2020): ATT and TWFE Estimator

- Let $\epsilon_{g,t}$ denote the residual of the regression of $D_{g,t}$ on group and period FE

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{3}$$

Theorem 1 in CD(2020)

If we have Assumption 1-5, then

$$\beta^{fe} = E[\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} w_{g,t} \Delta_{g,t}]$$

$$w_{g,t} = \frac{\epsilon_{g,t}}{\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \epsilon_{g,t}}$$

The TWFE estimator is a weighted average of cell-level ATE, with $w_{g,t}$ as weights.

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): ATT and TWFE Estimator

- How to interpret this weight $w_{g,t}$?
- You assign more weights to cells (g, t) deviating from the average treatment level
  - of all cells in group g
  - of all cells at time t
- If everyone in this group, or every one in this year is not treated, but you are treated, then you will be assigned a large weight
- Seems OK to you? A big issue is negative weight

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,\cdot} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{\cdot,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{\cdot,\cdot} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$

$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$

$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,\cdot} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{\cdot,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{\cdot,\cdot} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,\cdot} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{\cdot,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{\cdot,\cdot} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,.} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{.,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{.,.} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,.} - D_{.,t} + D_{.,.} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,\cdot} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{\cdot,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{\cdot,\cdot} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- Consider a simple case with 2 groups and 3 periods with equal group size
- Group 1 gets treated at periods 3; group 2 gets treated at periods 2 and 3
- Let $D_{g,\cdot} = \sum_t \frac{N_{g,t}}{N_g} D_{g,t}$, $D_{\cdot,t} = \sum_g \frac{N_{g,t}}{N_t} D_{g,t}$, $D_{\cdot,\cdot} = \sum_{(g,t)} \frac{N_{g,t}}{N} D_{g,t}$:

$$\epsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot} \tag{4}$$

- Thus, we have:

$$\epsilon_{1,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$
$$\epsilon_{2,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3}$$
$$\epsilon_{2,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6}$$

- $\epsilon_{2,3} < 0$!!!

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1$, $E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1, E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1, E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1, E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1, E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): An Example

- In this special case, we have:

$$\beta^{fe} = \frac{1}{2}E[\Delta_{1,3}] + E[\Delta_{2,2}] - \frac{1}{2}E[\Delta_{2,3}]$$

- We assign negative weight to $\Delta_{2,3}$
- Negative weight can make results weird
- If $E[\Delta_{1,3}] = E[\Delta_{2,2}] = 1, E[\Delta_{2,3}] = 4$, we have:

$$\beta^{fe} = \frac{1}{2} \times 1 + 1 - \frac{1}{2} \times 4 = -\frac{1}{2}$$

- We have a negative TWFE estimation, when all treatment effects are positive

CD(2020): Negative Weights

- Let's see in more details why there is negative weight
- In this case, we have two switches: Group 1 at period 3, and group 2 at period 2
- Thus, we have two DID comparisons
- It can be proved that β^{fe} is the average of these two:

$$DID_1 = E(Y_{2,2}) - E(Y_{2,1}) - [E(Y_{1,2}) - E(Y_{1,1})]$$
$$DID_2 = E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})]$$
$$\beta^{fe} = \frac{1}{2}(DID_1 + DID_2)$$

CD(2020): Negative Weights

- Let's see in more details why there is negative weight
- In this case, we have two switches: Group 1 at period 3, and group 2 at period 2
- Thus, we have two DID comparisons
- It can be proved that β^{fe} is the average of these two:

$$DID_1 = E(Y_{2,2}) - E(Y_{2,1}) - [E(Y_{1,2}) - E(Y_{1,1})]$$
$$DID_2 = E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})]$$
$$\beta^{fe} = \frac{1}{2}(DID_1 + DID_2)$$

CD(2020): Negative Weights

- Let's see in more details why there is negative weight
- In this case, we have two switches: Group 1 at period 3, and group 2 at period 2
- Thus, we have two DID comparisons
- It can be proved that β^{fe} is the average of these two:

$$DID_1 = E(Y_{2,2}) - E(Y_{2,1}) - [E(Y_{1,2}) - E(Y_{1,1})]$$
$$DID_2 = E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})]$$
$$\beta^{fe} = \frac{1}{2}(DID_1 + DID_2)$$

CD(2020): Negative Weights

- Let's see in more details why there is negative weight
- In this case, we have two switches: Group 1 at period 3, and group 2 at period 2
- Thus, we have two DID comparisons
- It can be proved that β^{fe} is the average of these two:

$$DID_1 = E(Y_{2,2}) - E(Y_{2,1}) - [E(Y_{1,2}) - E(Y_{1,1})]$$
$$DID_2 = E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})]$$
$$\beta^{fe} = \frac{1}{2}(DID_1 + DID_2)$$

CD(2020): Negative Weights

- Let's see in more details why there is negative weight
- In this case, we have two switches: Group 1 at period 3, and group 2 at period 2
- Thus, we have two DID comparisons
- It can be proved that β^{fe} is the average of these two:

$$DID_1 = E(Y_{2,2}) - E(Y_{2,1}) - [E(Y_{1,2}) - E(Y_{1,1})]$$
$$DID_2 = E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})]$$
$$\beta^{fe} = \frac{1}{2}(DID_1 + DID_2)$$

CD(2020): Negative Weights

- We can show that $DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$, NOT $DID_2 = E[\Delta_{1,3}]$

Proof:

$$
\begin{aligned}
DID_2 &= E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})] \\
&= E(Y_{1,3}(1)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))] \\
&= E(Y_{1,3}(1)) - E(Y_{1,3}(0)) + \underbrace{E(Y_{1,3}(0)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))]}_{\text{You cannot cancel this in a forbidden comparison!}} \\
&= E[\Delta_{1,3}] + E(Y_{1,3}(0)) - E(Y_{1,2}(0)) \\
&\quad - [E(Y_{2,3}(1)) - E(Y_{2,3}(0)) + E(Y_{2,3}(0)) - E(Y_{2,2}(1))] \\
&= E[\Delta_{1,3}] - E[\Delta_{2,3}] \\
&\quad + [E(Y_{2,2}(1)) - E(Y_{2,2}(0)) + \underbrace{E(Y_{2,2}(0)) - E(Y_{2,3}(0)) - [E(Y_{1,2}(0)) - E(Y_{1,3}(0))]}_{=0}] \\
&= E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}]
\end{aligned}
$$

CD(2020): Negative Weights

- We can show that $DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$, NOT $DID_2 = E[\Delta_{1,3}]$

Proof:

$$
\begin{aligned}
DID_2 &= E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})] \\
&= E(Y_{1,3}(1)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))] \\
&= E(Y_{1,3}(1)) - E(Y_{1,3}(0)) + \underbrace{E(Y_{1,3}(0)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))]}_{\text{You cannot cancel this in a forbidden comparison!}} \\
&= E[\Delta_{1,3}] + E(Y_{1,3}(0)) - E(Y_{1,2}(0)) \\
&\quad - [E(Y_{2,3}(1)) - E(Y_{2,3}(0)) + E(Y_{2,3}(0)) - E(Y_{2,2}(1))] \\
&= E[\Delta_{1,3}] - E[\Delta_{2,3}] \\
&\quad + [E(Y_{2,2}(1)) - E(Y_{2,2}(0)) + \underbrace{E(Y_{2,2}(0)) - E(Y_{2,3}(0)) - [E(Y_{1,2}(0)) - E(Y_{1,3}(0))]}_{=0}] \\
&= E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}]
\end{aligned}
$$

CD(2020): Negative Weights

- We can show that $DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$, NOT $DID_2 = E[\Delta_{1,3}]$

Proof:

$$
\begin{aligned}
DID_2 &= E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})] \\
&= E(Y_{1,3}(1)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))] \\
&= E(Y_{1,3}(1)) - E(Y_{1,3}(0)) + \underbrace{E(Y_{1,3}(0)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))]}_{\text{You cannot cancel this in a forbidden comparison!}} \\
&= E[\Delta_{1,3}] + E(Y_{1,3}(0)) - E(Y_{1,2}(0)) \\
&\quad - [E(Y_{2,3}(1)) - E(Y_{2,3}(0)) + E(Y_{2,3}(0)) - E(Y_{2,2}(1))] \\
&= E[\Delta_{1,3}] - E[\Delta_{2,3}] \\
&\quad + [E(Y_{2,2}(1)) - E(Y_{2,2}(0)) + \underbrace{E(Y_{2,2}(0)) - E(Y_{2,3}(0)) - [E(Y_{1,2}(0)) - E(Y_{1,3}(0))]}_{=0}] \\
&= E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}]
\end{aligned}
$$

CD(2020): Negative Weights

- We can show that $DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$, NOT $DID_2 = E[\Delta_{1,3}]$

Proof:

$$
\begin{aligned}
DID_2 &= E(Y_{1,3}) - E(Y_{1,2}) - [E(Y_{2,3}) - E(Y_{2,2})] \\
&= E(Y_{1,3}(1)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))] \\
&= E(Y_{1,3}(1)) - E(Y_{1,3}(0)) + \underbrace{E(Y_{1,3}(0)) - E(Y_{1,2}(0)) - [E(Y_{2,3}(1)) - E(Y_{2,2}(1))]}_{\text{You cannot cancel this in a forbidden comparison!}} \\
&= E[\Delta_{1,3}] + E(Y_{1,3}(0)) - E(Y_{1,2}(0)) \\
&\quad - [E(Y_{2,3}(1)) - E(Y_{2,3}(0)) + E(Y_{2,3}(0)) - E(Y_{2,2}(1))] \\
&= E[\Delta_{1,3}] - E[\Delta_{2,3}] \\
&\quad + [E(Y_{2,2}(1)) - E(Y_{2,2}(0)) + \underbrace{E(Y_{2,2}(0)) - E(Y_{2,3}(0)) - [E(Y_{1,2}(0)) - E(Y_{1,3}(0))]}_{=0}] \\
&= E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}]
\end{aligned}
$$

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

$DID_2 = E[\Delta_{1,3}] - (E[\Delta_{2,3}] - E[\Delta_{2,2}])$

- Then, $E[\Delta_{2,3}]$ enters into β^{fe} with a negative weight
- What does this equation mean?
- It means that this DID comparison, is ATE in group 1 period 3, minus changes in group 2's ATE between period 2 and 3
- You are using treated cells as the "control" group!! $\Leftarrow [E(Y_{2,3}) - E(Y_{2,2})]$
- For this already treated group, although there is no treatment status change from period 2 to 3, the outcome change of $Y(1)$ is not comparable to that of $Y(0)$!
- That's why you may assign negative weights to some cell ATEs

CD(2020): Negative Weights

- Do not use "continuously treated groups" as "control groups"
- This is called forbidden comparison

CD(2020): Negative Weights

- Do not use "continuously treated groups" as "control groups"
- This is called forbidden comparison

CD(2020): Negative Weights

- Do not use "continuously treated groups" as "control groups"
- This is called forbidden comparison

CD(2020): Negative Weights

- Homework : Show what is DID_1, express it in terms of $E[\Delta_{g,t}]$, tell me what is the difference between the derivation of DID_1 and DID_2. Why some terms can be canceled out in the derivation of DID_1, but not DID_2
- This homework can make you have a deeper understanding of why using treated cells as "controls" can be dangerous!

CD(2020): Negative Weights

- Homework : Show what is DID_1, express it in terms of $E[\Delta_{g,t}]$, tell me what is the difference between the derivation of DID_1 and DID_2. Why some terms can be canceled out in the derivation of DID_1, but not DID_2
- This homework can make you have a deeper understanding of why using treated cells as "controls" can be dangerous!

CD(2020): Negative Weights

- Homework : Show what is DID_1, express it in terms of $E[\Delta_{g,t}]$, tell me what is the difference between the derivation of DID_1 and DID_2. Why some terms can be canceled out in the derivation of DID_1, but not DID_2
- This homework can make you have a deeper understanding of why using treated cells as "controls" can be dangerous!

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods.

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- In general, which kind of cells are more likely to have negative weights?
- Let's go back to the function of weight

$$D_{g,t} = \alpha + \gamma_g + \lambda_t + \epsilon_{g,t} \tag{5}$$

- When will $\epsilon_{g,t}$ become negative?
- A cell is more likely to have negative weight if
  - (1) At a period when many groups are treated;
  - (2) In a group where it is treated for many periods;

CD(2020): Negative Weights

- This is the opposite of who are assigned larger weight
- If everyone in this group, or every one in this year are treated, then you are assigned negative weight
- Just like group 2 in period 3 in this example

CD(2020): Negative Weights

- This is the opposite of who are assigned larger weight
- If everyone in this group, or every one in this year are treated, then you are assigned negative weight
- Just like group 2 in period 3 in this example

CD(2020): Negative Weights

- This is the opposite of who are assigned larger weight
- If everyone in this group, or every one in this year are treated, then you are assigned negative weight
- Just like group 2 in period 3 in this example

CD(2020): Negative Weights

- This is the opposite of who are assigned larger weight
- If everyone in this group, or every one in this year are treated, then you are assigned negative weight
- Just like group 2 in period 3 in this example

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Negative Weights

- The intuition is that they are more likely to take treated cells as "control"
- Many chances for you to do forbidden comparisons in these cells
- Specifically, in staggered adoption case, these are very dangerous
  - Groups adopting treatment/policy earlier
  - Periods that are very late

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by standard of the weights, you can derive the maximal value of the standard dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}^{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - e.g. (1) Compute the weights, see whether some of them are negative
  - e.g. (2) Dividing $|\hat{\beta}^{fe}|$ by standard deviation of the weights, you can derive the maximal value of the standard deviation of ATE across (g, t) cells under which ATE may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): Sensitivity Check

So, what should we do?

- Step 1: Check how sensitive your result is to treatment effect heterogeneity
  - (1) Compute the weights, see whether some of them are negative
  - (2) Dividing $|\hat{\beta}^{fe}|$ by std dev of the weights, you can derive the minimal value of the std dev of ATE across (g, t) cells under which ATT may have the opposite sign
- If there are many negative weights, or $\frac{|\hat{\beta}_{fe}|}{sd(w)}$ is small, do not use TWFE
- Since TWFE estimator is vulnerable to treatment effect heterogeneity in this case
- In practice, you can use *twowayfeweights* Stata package

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

- Step 2: If you have many negative weights, or the threshold value of the flipped sign is small, using a new estimator
- CD(2020) constructs a new estimator for TWFE regression, called DID_M
- We define a new average treatment effect:

$$\delta^s = E[\frac{1}{N_s} \sum_{(i,g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]]$$

- $N_s = \sum_{(g,t):t\geq 2, D_{g,t}\neq D_{g,t-1}} N_{g,t}$, number of obs changing their treatment status from $t-1$ to t
- δ^s is the ATE of all switching cells
- Switching cells include joiners and leavers

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1) | D_{g,1}, ... D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 9: Strong Exogeneity for $Y(1)$ (corresponding to A4)
  For all (g,t), $E(Y_{g,t}(1) - Y_{g,t-1}(1)|D_{g,1},...D_{g,T}) = E(Y_{g,t}(1) - Y_{g,t-1}(1))$
- Assumption 10: Common Trends for $Y(1)$ (corresponding to A5)
  For $t \geq 2$, $E(Y_{g,t}(1) - Y_{g,t-1}(1))$ does not vary across g
- A10 sets some homogeneity for the treatment effect
- These two assumptions ensure one to reconstruct the potential outcomes of leavers
- They are needed only when you have exits
- They are not necessary if we have a staggered adoption (no leavers)

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  (a) If there is g such that $D_{g,t-1}=0, D_{g,t}=1$, there exists g^* such that $D_{g^*,t-1}=D_{g^*,t}=0$.
  (b) If there is g such that $D_{g,t-1}=1, D_{g,t}=0$, there exists g^* such that $D_{g^*,t-1}=D_{g^*,t}=1$.
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g)$, $E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  (a) If there is g such that $D_{g,t-1}=0, D_{g,t}=1$, there exists g^* such that $D_{g^*,t-1}=D_{g^*,t}=0$.
  (b) If there is g such that $D_{g,t-1}=1, D_{g,t}=0$, there exists g^* such that $D_{g^*,t-1}=D_{g^*,t}=1$.
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g)$, $E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  - (i) If there is g such that $D_{g,t-1} = 0, D_{g,t} = 1$, there exists g' such that $D_{g',t-1} = D_{g',t} = 0$.
  - (ii) If there is g such that $D_{g,t-1} = 1, D_{g,t} = 0$, there exists g' such that $D_{g',t-1} = D_{g',t} = 1$
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g), E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  - (i) If there is g such that $D_{g,t-1} = 0, D_{g,t} = 1$, there exists g' such that $D_{g',t-1} = D_{g',t} = 0$.
  - (ii) If there is g such that $D_{g,t-1} = 1, D_{g,t} = 0$, there exists g' such that $D_{g',t-1} = D_{g',t} = 1$
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g), E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  - (i) If there is g such that $D_{g,t-1} = 0, D_{g,t} = 1$, there exists g' such that $D_{g',t-1} = D_{g',t} = 0$.
  - (ii) If there is g such that $D_{g,t-1} = 1, D_{g,t} = 0$, there exists g' such that $D_{g',t-1} = D_{g',t} = 1$
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g), E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

Additional assumptions

- Assumption 11: Existence of "Stable" Groups (existence of control groups)
  - (i) If there is g such that $D_{g,t-1} = 0, D_{g,t} = 1$, there exists g' such that $D_{g',t-1} = D_{g',t} = 0$.
  - (ii) If there is g such that $D_{g,t-1} = 1, D_{g,t} = 0$, there exists g' such that $D_{g',t-1} = D_{g',t} = 1$
- Assumption 12: Mean Independence between a group's outcome and other groups' treatment (No spillover)
  For all g, t, $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g), E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$

CD(2020): New Estimator

- Let's define the DID_M estimator
- Let $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$, that is, number of obs with treatment d at t and d' at $t-1$
- Let's define two parts of DID comparisons:

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

- DID_+ is DID for joiners vs untreated, DID_- is DID for leavers vs treated

CD(2020): New Estimator

- Let's define the DID_M estimator
- Let $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$, that is, number of obs with treatment d at t and d' at $t-1$
- Let's define two parts of DID comparisons:

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

- DID_+ is DID for joiners vs untreated, DID_- is DID for leavers vs treated

CD(2020): New Estimator

- Let's define the DID_M estimator
- Let $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$, that is, number of obs with treatment d at t and d' at $t-1$
- Let's define two parts of DID comparisons:

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

- DID_+ is DID for joiners vs untreated, DID_- is DID for leavers vs treated

CD(2020): New Estimator

- Let's define the DID_M estimator
- Let $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$, that is, number of obs with treatment d at t and d' at $t-1$
- Let's define two parts of DID comparisons:

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

- DID_+ is DID for joiners vs untreated, DID_- is DID for leavers vs treated

CD(2020): New Estimator

- Let's define the DID_M estimator
- Let $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$, that is, number of obs with treatment d at t and d' at $t-1$
- Let's define two parts of DID comparisons:

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

- DID_+ is DID for joiners vs untreated, DID_- is DID for leavers vs treated

CD(2020): New Estimator

- DID_M estimator is defined as

$$DID_M = \sum_{t=2}^{T} \left(\frac{N_{1,0,t}}{N_s} DID_{+,t} + \frac{N_{0,1,t}}{N_s} DID_{-,t} \right) \quad (6)$$

Theorem 3 in CD(2020)

If we have Assumption 1,2,4,5, and 9-12 then

$$E[DID_M] = \delta^s$$

The DID_M estimator is a weighted average of joiners' and leavers' treatment effect. It is an unbiased estimator of δ^s, that is, the ATE of all switching cells.

CD(2020): New Estimator

- DID_M estimator is defined as

$$DID_M = \sum_{t=2}^{T} \left(\frac{N_{1,0,t}}{N_s} DID_{+,t} + \frac{N_{0,1,t}}{N_s} DID_{-,t} \right) \tag{6}$$

Theorem 3 in CD(2020)

If we have Assumption 1,2,4,5, and 9-12 then

$$E[DID_M] = \delta^s$$

The DID_M estimator is a weighted average of joiners' and leavers' treatment effect. It is an unbiased estimator of δ^s, that is, the ATE of all switching cells.

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): New Estimator

- DID_M is also consistent and asymptotically normal
- DID_M is nonparametric, thus, less efficient than TWFE (bias-variance tradeoff)
- A placebo test can be constructed, to check the parallel trend assumption
- Basic idea: Compare outcome's evolution from $t-2$ to $t-1$ for groups which change their treatments from $t-1$ to t (pre-trend test)
- Calculate $DID_{+,t}$ and $DID_{-,t}$ by changing $Y_{g,t}$ to be $Y_{g,t-1}$ and $Y_{g,t-1}$ to be $Y_{g,t-2}$
- You should expect zeros for $DID_{+,t}$ and $DID_{-,t}$ in this placebo
- You can use Stata packages *fuzzydid* or *did_multiplegt*

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
    - If there are negative weights, or the threshold is small, use DID_g rather than TWFE
    - If use DID_g, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

Main takeaways of De Chaisemartin and d'Haultfoeuille (2020)

- In general, TWFE is not a good estimator in settings with heterogeneous treatment effect
- It may assign negative weights to some group-period ATEs
- If you have periods when many groups are treated, or groups treated for many periods, be careful!
- In practice, here are the things you can do
  - Calculate weights and the threshold value of flipping the sign
  - If there are negative weights, or the threshold is small, use DID_M rather than TWFE
  - If use DID_M, implement placebo test to validate the parallel trend assumption
- In general, the argument can be applied to any ordered treatment D

CD(2020): Conclusion

- Homework 2: Will plain vanilla DID (like Card and Krueger (1994)) suffer from the same issue when we use TWFE estimator? Explain your answer in words. (Do not use math)

CD(2020): Conclusion

- Homework 2: Will plain vanilla DID (like Card and Krueger (1994)) suffer from the same issue when we use TWFE estimator? Explain your answer in words. (Do not use math)

Continuous DID

- We only considered discrete treatment variables in previous lectures
- What about continuous cases? They are also very common
- For example, the effect of US-China trade war tariff on China's employment

Continuous DID

- We only considered discrete treatment variables in previous lectures
- What about continuous cases? They are also very common
- For example, the effect of US-China trade war tariff on China's employment

Continuous DID

- We only considered discrete treatment variables in previous lectures
- What about continuous cases? They are also very common
- For example, the effect of US-China trade war tariff on China's employment

Continuous DID

- We only considered discrete treatment variables in previous lectures
- What about continuous cases? They are also very common
- For example, the effect of US-China trade war tariff on China's employment

Continuous DID

- In general, we will see that TWFE with continuous treatment is much more complicated
- We need additional assumptions to identify some meaningful causal effects
- It also suffers from forbidden comparison issue in staggered DID case
- More generally, we will see how to interpret continuous treatment in the potential outcome framework (usually complicated)

Continuous DID

- In general, we will see that TWFE with continuous treatment is much more complicated
- We need additional assumptions to identify some meaningful causal effects
- It also suffers from forbidden comparison issue in staggered DID case
- More generally, we will see how to interpret continuous treatment in the potential outcome framework (usually complicated)

Continuous DID

- In general, we will see that TWFE with continuous treatment is much more complicated
- We need additional assumptions to identify some meaningful causal effects
- It also suffers from forbidden comparison issue in staggered DID case
- More generally, we will see how to interpret continuous treatment in the potential outcome framework (usually complicated)

Continuous DID

- In general, we will see that TWFE with continuous treatment is much more complicated
- We need additional assumptions to identify some meaningful causal effects
- It also suffers from forbidden comparison issue in staggered DID case
- More generally, we will see how to interpret continuous treatment in the potential outcome framework (usually complicated)

Continuous DID

- In general, we will see that TWFE with continuous treatment is much more complicated
- We need additional assumptions to identify some meaningful causal effects
- It also suffers from forbidden comparison issue in staggered DID case
- More generally, we will see how to interpret continuous treatment in the potential outcome framework (usually complicated)

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - a. Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - b. Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - a. Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - b. Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Introduction

Callaway, Goodman-Bacon, and Sant'Anna (2021) Difference-in-Differences with a Continuous Treatment

- Define two types of causal effects: level effect (d vs 0) and slope effect (d vs d')
- Consider a vanilla two-period DID case
  - Discuss assumptions that are needed for the identification of these causal effects (non-parametrically)
  - Analyze TWFE estimator: relation to causal effects
- Extend results to more than two periods and staggered DID

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1}=0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D=\{0\}\cup D_+, P(D=0)>0, dF_D(d)>0, \forall\, d\in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1}=Y_{it-1}(0), Y_{it}=Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+=[d_L,d_U], 0<d_L<d_U<\infty, P(D=0)>0, f_D(d)>0 \forall\, d\in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1}=0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D.
  $\mathcal{D} = \{0\} \cup \mathcal{D}_+, P(D=0) > 0, dF_D(d) > 0, \forall d \in \mathcal{D}_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $\mathcal{D}_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D=0) > 0, f_D(d) > 0 \forall d \in \mathcal{D}_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1}=0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D=\{0\}\cup D_+.P(D=0)>0, dF_D(d)>0, \forall d\in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1}=Y_{it-1}(0), Y_{it}=Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+=[d_L,d_U], 0<d_L<d_U<\infty, P(D=0)>0, f_D(d)>0 \forall d\in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

Two-period, one-time policy treatment

- We have two periods, t and $t-1$
- Units receive a treatment dose D_i in t, not $t-1$ ($D_{it-1} = 0$)
- Potential outcome of individual i at time s receiving d is $Y_{is}(d)$
- Assumption 1: We have i.i.d. samples.
- Assumption 2: Full support of D,
  $D = \{0\} \cup D_+ . P(D = 0) > 0, dF_D(d) > 0, \forall d \in D_+$. No units are treated in $t-1$.
- Assumption 3: No anticipation effect. $Y_{it-1} = Y_{it-1}(0), Y_{it} = Y_{it}(D_i)$
- Assumption 4: Continuous treatment.
  $D_+ = [d_L, d_U], 0 < d_L < d_U < \infty, P(D = 0) > 0, f_D(d) > 0 \forall d \in D_+$

Callaway et al(2021): Settings

- The definition of causal effect can be much more complicated in continuous treatment case
- Since you are not only comparing d and 0, but also d and d'
- 1. Level effect: $Y_t(d) - Y_t(0)$
  Difference between effect of some dose level d and no treatment
- 2. Slope effect: $Y_t'(d)$
  The derivative of the potential outcome function. The marginal increase in the effect when dose is increased.

Callaway et al(2021): Settings

- The definition of causal effect can be much more complicated in continuous treatment case
- Since you are not only comparing d and 0, but also d and d'
- 1. Level effect: $Y_t(d) - Y_t(0)$
  Difference between effect of some dose level d and no treatment
- 2. Slope effect: $Y_t^l(d)$
  The derivative of the potential outcome function. The marginal increase in the effect when dose is increased.

Callaway et al(2021): Settings

- The definition of causal effect can be much more complicated in continuous treatment case
- Since you are not only comparing d and 0, but also d and d'
- 1. Level effect: $Y_t(d) - Y_t(0)$
  Difference between effect of some dose level d and no treatment
- 2. Slope effect: $Y_t^l(d)$
  The derivative of the potential outcome function. The marginal increase in the effect when dose is increased.

Callaway et al(2021): Settings

- The definition of causal effect can be much more complicated in continuous treatment case
- Since you are not only comparing d and 0, but also d and d'
- 1. Level effect: $Y_t(d) - Y_t(0)$
  Difference between effect of some dose level d and no treatment
- 2. Slope effect: $Y_t^l(d)$
  The derivative of the potential outcome function. The marginal increase in the effect when dose is increased.

Callaway et al(2021): Settings

- The definition of causal effect can be much more complicated in continuous treatment case
- Since you are not only comparing d and 0, but also d and d'
- 1. Level effect: $Y_t(d) - Y_t(0)$
  Difference between effect of some dose level d and no treatment
- 2. Slope effect: $Y_t^l(d)$
  The derivative of the potential outcome function. The marginal increase in the effect when dose is increased.

Callaway et al(2021): Settings

- We define average **level effects** as:

$$ATT(a|b) = E[Y_t(a) - Y_t(0)|D = b], \quad ATE(d) = E[Y_t(d) - Y_t(0)]$$

- $ATT(a|b)$: Average effect of dose a on units that who actually experience dose b
- a is potential treatment, b is real treatment
- $ATE(d)$: Average effect of dose d on all units

Callaway et al(2021): Settings

- We define average **level effects** as:

$$ATT(a|b) = E[Y_t(a) - Y_t(0)|D = b], \quad ATE(d) = E[Y_t(d) - Y_t(0)]$$

- $ATT(a|b)$: Average effect of dose a on units that who actually experience dose b
- a is potential treatment, b is real treatment
- $ATE(d)$: Average effect of dose d on all units

Callaway et al(2021): Settings

- We define average **level effects** as:

$$ATT(a|b) = E[Y_t(a) - Y_t(0)|D = b], \quad ATE(d) = E[Y_t(d) - Y_t(0)]$$

- $ATT(a|b)$: Average effect of dose a on units that who actually experience dose b
- a is potential treatment, b is real treatment
- $ATE(d)$: Average effect of dose d on all units

Callaway et al(2021): Settings

- We define average **level effects** as:

$$ATT(a|b) = E[Y_t(a) - Y_t(0)|D = b], \quad ATE(d) = E[Y_t(d) - Y_t(0)]$$

- $ATT(a|b)$: Average effect of dose a on units that who actually experience dose b
- a is potential treatment, b is real treatment
- $ATE(d)$: Average effect of dose d on all units

Callaway et al(2021): Settings

- We define average **level effects** as:

$$ATT(a|b) = E[Y_t(a) - Y_t(0)|D = b], \quad ATE(d) = E[Y_t(d) - Y_t(0)]$$

- $ATT(a|b)$: Average effect of dose a on units that who actually experience dose b
- a is potential treatment, b is real treatment
- $ATE(d)$: Average effect of dose d on all units

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Settings

- We define average **slope effects** as:

$$ACRT(d|d) = \frac{\partial E[Y_t(l)|D = d]}{\partial l}\Big|_{l=d}, \quad ACR(d) = \frac{\partial E[Y_t(d)]}{\partial d}$$

- We call them Average Causal Response Function
- $ACRT(d|d)$: Average causal response of a small change in dose d, for the group of units who actually experience dose d
- What is the impact for people who get dose d to get a little bit more dose
- $ACR(d)$: Average causal response of a small change in dose d for everyone

Callaway et al(2021): Non-parametric Identification

- Assumption 4: Parallel Trends.
  $\forall d, E[Y_t(0) - Y_{t-1}(0)|D = d] = E[Y_t(0) - Y_{t-1}(0)|D = 0]$
- It says that the path of untreated potential outcomes would have been the same for untreated group and treated group with any dose level

Under Assumptions 1 to 4, $ATT(d|d)$ is identified for all $d \in \mathcal{D}_+$

$$ATT(d|d) = E[\Delta Y_t | D = d] - E[\Delta Y_t | D = 0]$$

where $\Delta Y_t = Y_t - Y_{t-1}$

- We can non-parametrically identify ATT (level effect) under parallel trend assumption in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 4: Parallel Trends.
  $\forall d, E[Y_t(0) - Y_{t-1}(0)|D = d] = E[Y_t(0) - Y_{t-1}(0)|D = 0]$
- It says that the path of untreated potential outcomes would have been the same for untreated group and treated group with any dose level

Theorem 1 in Callaway et al(2021)

Under Assumptions 1 to 4, $ATT(d|d)$ is identified for all $d \in D$:

$$ATT(d|d) = E[\Delta Y_t|D = d] - E[\Delta Y_t|D = 0]$$

where $\Delta Y_t = Y_t - Y_{t-1}$

- We can non-parametrically identify ATT (level effect) under parallel trend assumption in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 4: Parallel Trends.
  $\forall d, E[Y_t(0) - Y_{t-1}(0)|D = d] = E[Y_t(0) - Y_{t-1}(0)|D = 0]$
- It says that the path of untreated potential outcomes would have been the same for untreated group and treated group with any dose level

Theorem 1 in Callaway et al(2021)

Under Assumptions 1 to 4, $ATT(d|d)$ is identified for all $d \in D$:

$$ATT(d|d) = E[\Delta Y_t|D = d] - E[\Delta Y_t|D = 0]$$

where $\Delta Y_t = Y_t - Y_{t-1}$

- We can non-parametrically identify ATT (level effect) under parallel trend assumption in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 4: Parallel Trends.
  $\forall d, E[Y_t(0) - Y_{t-1}(0)|D = d] = E[Y_t(0) - Y_{t-1}(0)|D = 0]$
- It says that the path of untreated potential outcomes would have been the same for untreated group and treated group with any dose level

Theorem 1 in Callaway et al(2021)

Under Assumptions 1 to 4, $ATT(d|d)$ is identified for all $d \in D$:

$$ATT(d|d) = E[\Delta Y_t|D = d] - E[\Delta Y_t|D = 0]$$

where $\Delta Y_t = Y_t - Y_{t-1}$

- We can non-parametrically identify ATT (level effect) under parallel trend assumption in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 4: Parallel Trends.
  $\forall d, E[Y_t(0) - Y_{t-1}(0)|D = d] = E[Y_t(0) - Y_{t-1}(0)|D = 0]$
- It says that the path of untreated potential outcomes would have been the same for untreated group and treated group with any dose level

Theorem 1 in Callaway et al(2021)

Under Assumptions 1 to 4, $ATT(d|d)$ is identified for all $d \in D$:

$$ATT(d|d) = E[\Delta Y_t|D = d] - E[\Delta Y_t|D = 0]$$

where $\Delta Y_t = Y_t - Y_{t-1}$

- We can non-parametrically identify ATT (level effect) under parallel trend assumption in a DID fashion.

Callaway et al(2021): Non-parametric Identification

Proposition 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4, generally, $ACRT(d|d)$ is NOT identified:

$$\frac{\partial E[\Delta Y_t | D = d]}{\partial d} = ACRT(d|d) + \underbrace{\left.\frac{\partial ATT(d|l)}{\partial l}\right|_{l=d}}_{\text{Selection bias}}$$

- Under traditional parallel trend assumption, local comparisons of paths of outcomes mix $ACRT(d|d)$ and a selection bias term
- The bias is the marginal change in ATT of group $D = l$ if they get dose d
- ACRT CANNOT be identified with traditional parallel trend assumption in a DID fashion! Why?

Callaway et al(2021): Non-parametric Identification

Proposition 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4, generally, $ACRT(d|d)$ is NOT identified:

$$\frac{\partial E[\Delta Y_t | D = d]}{\partial d} = ACRT(d|d) + \underbrace{\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d}}_{\text{Selection bias}}$$

- Under traditional parallel trend assumption, local comparisons of paths of outcomes mix $ACRT(d|d)$ and a selection bias term
- The bias is the marginal change in ATT of group $D = l$ if they get dose d
- ACRT CANNOT be identified with traditional parallel trend assumption in a DID fashion! Why?

Callaway et al(2021): Non-parametric Identification

Proposition 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4, generally, $ACRT(d|d)$ is NOT identified:

$$\frac{\partial E[\Delta Y_t|D=d]}{\partial d} = ACRT(d|d) + \underbrace{\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d}}_{\text{Selection bias}}$$

- Under traditional parallel trend assumption, local comparisons of paths of outcomes mix $ACRT(d|d)$ and a selection bias term
- The bias is the marginal change in ATT of group $D = l$ if they get dose d
- ACRT CANNOT be identified with traditional parallel trend assumption in a DID fashion! Why?

Callaway et al(2021): Non-parametric Identification

Proposition 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4, generally, $ACRT(d|d)$ is NOT identified:

$$\frac{\partial E[\Delta Y_t | D = d]}{\partial d} = ACRT(d|d) + \underbrace{\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d}}_{\text{Selection bias}}$$

- Under traditional parallel trend assumption, local comparisons of paths of outcomes mix $ACRT(d|d)$ and a selection bias term
- The bias is the marginal change in ATT of group $D = l$ if they get dose d
- ACRT CANNOT be identified with traditional parallel trend assumption in a DID fashion! Why?

Callaway et al(2021): Non-parametric Identification

Proposition 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4, generally, $ACRT(d|d)$ is NOT identified:

$$\frac{\partial E[\Delta Y_t | D = d]}{\partial d} = ACRT(d|d) + \underbrace{\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d}}_{\text{Selection bias}}$$

- Under traditional parallel trend assumption, local comparisons of paths of outcomes mix $ACRT(d|d)$ and a selection bias term
- The bias is the marginal change in ATT of group $D = l$ if they get dose d
- ACRT CANNOT be identified with traditional parallel trend assumption in a DID fashion! Why?

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- For $ACRT(d|d)$, you consider a marginal increase in d to l
- You are comparing d and l, but not d and 0!
- We assume parallel trends only for group $D = d$ and group $D = l$ if they are not treated $(Y_t(0) - Y_{t-1}(0)|D)$
- But not parallel trends for group $D = d$ and group $D = l$ if they are treated at level d $(Y_t(d) - Y_{t-1}(0)|D)$
- Parallel trends only for untreated potential outcome $(d = 0)$, not any dose level $(d = d)$
- You need some exogeneity/homogeneity about the dose assignment

Callaway et al(2021): Non-parametric Identification

- Let's write the selection bias in the form of local difference

$$\begin{aligned}\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d} &= E[\Delta Y_t(d)|D=d'] - E[\Delta Y_t(d)|D=d] \\ &= E[Y_t(d) - Y_{t-1}(0)|D=d'] - E[Y_t(d) - Y_{t-1}(0)|D=d]\end{aligned}$$

- When subtracting observed average outcome of $D = d$ from $D = d'$, we have both causal effect for group $D = d$, and differences in effects for group $D = d$ and $D = d'$
- We need an additional assumption to eliminate this
- To assume a parallel trend for group $D = d$ and group $D = d'$ if they are assigned dose d

Callaway et al(2021): Non-parametric Identification

- Let's write the selection bias in the form of local difference

$$\begin{aligned}\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d} &= E[\Delta Y_t(d)|D = d^l] - E[\Delta Y_t(d)|D = d] \\ &= E[Y_t(d) - Y_{t-1}(0)|D = d^l] - E[Y_t(d) - Y_{t-1}(0)|D = d]\end{aligned}$$

- When subtracting observed average outcome of $D = d$ from $D = d^l$, we have both causal effect for group $D = d$, and differences in effects for group $D = d$ and $D = d^l$
- We need an additional assumption to eliminate this
- To assume a parallel trend for group $D = d$ and group $D = d^l$ if they are assigned dose d

Callaway et al(2021): Non-parametric Identification

- Let's write the selection bias in the form of local difference

$$\begin{aligned}\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d} &= E[\Delta Y_t(d)|D = d^l] - E[\Delta Y_t(d)|D = d] \\ &= E[Y_t(d) - Y_{t-1}(0)|D = d^l] - E[Y_t(d) - Y_{t-1}(0)|D = d]\end{aligned}$$

- When subtracting observed average outcome of $D = d$ from $D = d^l$, we have both causal effect for group $D = d$, and differences in effects for group $D = d$ and $D = d^l$
- We need an additional assumption to eliminate this
- To assume a parallel trend for group $D = d$ and group $D = d^l$ if they are assigned dose d

Callaway et al(2021): Non-parametric Identification

- Let's write the selection bias in the form of local difference

$$\begin{aligned}\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d} &= E[\Delta Y_t(d)|D = d^l] - E[\Delta Y_t(d)|D = d] \\ &= E[Y_t(d) - Y_{t-1}(0)|D = d^l] - E[Y_t(d) - Y_{t-1}(0)|D = d]\end{aligned}$$

- When subtracting observed average outcome of $D = d$ from $D = d^l$, we have both causal effect for group $D = d$, and differences in effects for group $D = d$ and $D = d^l$
- We need an additional assumption to eliminate this
- To assume a parallel trend for group $D = d$ and group $D = d^l$ if they are assigned dose d

Callaway et al(2021): Non-parametric Identification

- Let's write the selection bias in the form of local difference

$$\begin{aligned}\frac{\partial ATT(d|l)}{\partial l}\Big|_{l=d} &= E[\Delta Y_t(d)|D = d^l] - E[\Delta Y_t(d)|D = d] \\ &= E[Y_t(d) - Y_{t-1}(0)|D = d^l] - E[Y_t(d) - Y_{t-1}(0)|D = d]\end{aligned}$$

- When subtracting observed average outcome of $D = d$ from $D = d^l$, we have both causal effect for group $D = d$, and differences in effects for group $D = d$ and $D = d^l$
- We need an additional assumption to eliminate this
- To assume a parallel trend for group $D = d$ and group $D = d^l$ if they are assigned dose d

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Under Assumptions 1 to 3 and 5, $ACR(d|d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_s|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Proposition 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 3 and 5, $ACR(d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_t|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Proposition 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 3 and 5, $ACR(d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_t|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Proposition 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 3 and 5, $ACR(d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_t|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Proposition 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 3 and 5, $ACR(d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_t|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Non-parametric Identification

- Assumption 5: Strong Parallel Trends.
  $\forall d, E[Y_t(d) - Y_{t-1}(0)] = E[Y_t(d) - Y_{t-1}(0)|D = d]$
- It says that for all doses, the average change in outcomes over time across all units if they had been assigned dose d, is the same as those actually experienced dose d.
- It imposes some homogeneity on treatment effect

Proposition 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 3 and 5, $ACR(d)$ and $ACRT(d|d)$ is identified:

$$\frac{\partial E[\Delta Y_t|D = d]}{\partial d} = ACRT(d|d) = ACR(d)$$

- We can non-parametrically identify ACRT under strong parallel trend assumption, in a DID fashion.

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Under Assumptions 1 to 4,

$$\beta^{sharp} = \int_{d_L}^{d_U} w_1(l)ACRT(l|l)dl + \frac{\partial ATT(l|l)}{\partial l}\Big|_{l=d_L} [l-d_L]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (l)w_1(l) \geq 0, w_0 > 0, \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Theorem 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (i) w_1(l) \geq 0, w_0 > 0, (ii) \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Theorem 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (i) w_1(l) \geq 0, w_0 > 0, (ii) \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Theorem 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (i) w_1(l) \geq 0, w_0 > 0, (ii) \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Theorem 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (i) w_1(l) \geq 0, w_0 > 0, (ii) \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

- Now we consider the causal interpretation of the traditional TWFE Estimator

Theorem 3 (a) in Callaway et al(2021)

Under Assumptions 1 to 4,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACRT(l|l) + \frac{\partial ATT(l|h)}{\partial h}\Big|_{h=l}]dl + w_0 \frac{ATT(d_L|d_L)}{d_L}$$

$$\text{where, } (i) w_1(l) \geq 0, w_0 > 0, (ii) \int_{d_L}^{d_U} w_1(l)dl + w_0 = 1$$

- The first term is the average causal effect of running from d_L to d_U
- The third term is the causal effect of having the lowest dose (d_L vs 0)
- The second term is the selection bias (without Assumption 5)

Callaway et al(2021): Causal Effect and TWFE Estimator

Theorem 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 5,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACR(l)]dl + w_0 \frac{ATE(d_L)}{d_L}$$

- Under strong parallel trend assumption, we eliminate the selection bias

Callaway et al(2021): Causal Effect and TWFE Estimator

Theorem 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 5,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACR(l)]dl + w_0 \frac{ATE(d_L)}{d_L}$$

- Under strong parallel trend assumption, we eliminate the selection bias

Callaway et al(2021): Causal Effect and TWFE Estimator

Theorem 3 (b) in Callaway et al(2021)

Under Assumptions 1 to 5,

$$\beta^{twfe} = \int_{d_L}^{d_U} w_1(l)[ACR(l)]dl + w_0 \frac{ATE(d_L)}{d_L}$$

- Under strong parallel trend assumption, we eliminate the selection bias

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) path of outcomes for units treated at the same time but with different doses
  - (ii) path of outcomes for early treated relative to later treated groups in periods before later is treated
  - (iii) path of outcomes between later treated and already treated groups in post-treatment periods for later treated group
  - (iv) path of outcomes between early treated and later treated group in the post-treatment periods relative to their pre-treatment periods
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Causal Effect and TWFE Estimator

- What if we extend it to multiple periods and staggered DID?
- Under strong parallel trend assumption, β^{twfe} is composed of four comparisons
  - (i) paths of outcomes for units treated at the same time but with different doses;
  - (ii) paths of outcomes in early-treated relative to later-treated groups in periods before later is treated;
  - (iii) paths of outcomes between later-treated and already treated groups in post-treatment period of the later-treated group;
  - (iv) paths of outcomes between early-treated and later-treated groups in their common post treatment periods relative to their common pre periods.
- First two are fine. (iii) and (iv) are forbidden comparison!
- The control group is not $Y_t(0) - Y_{t-1}(0)$
- Same issue as in CD(2020)

Callaway et al(2021): Conclusion

- In general, it is hard to identify meaningful causal effects using DID fashion in continuous treatment case
- Causal level effects are non-parametrically identified under common parallel trend assumption
- But causal slope effects are non-parametrically identified (in a DID fashion) only under strong parallel trend assumption, which is not testable by pre-trend
- You need not only random assignment of treatment, but also random assignment of dose amount

Callaway et al(2021): Conclusion

- In general, it is hard to identify meaningful causal effects using DID fashion in continuous treatment case
- Causal level effects are non-parametrically identified under common parallel trend assumption
- But causal slope effects are non-parametrically identified (in a DID fashion) only under strong parallel trend assumption, which is not testable by pre-trend
- You need not only random assignment of treatment, but also random assignment of dose amount

Callaway et al(2021): Conclusion

- In general, it is hard to identify meaningful causal effects using DID fashion in continuous treatment case
- Causal level effects are non-parametrically identified under common parallel trend assumption
- But causal slope effects are non-parametrically identified (in a DID fashion) only under strong parallel trend assumption, which is not testable by pre-trend
- You need not only random assignment of treatment, but also random assignment of dose amount

Callaway et al(2021): Conclusion

- In general, it is hard to identify meaningful causal effects using DID fashion in continuous treatment case
- Causal level effects are non-parametrically identified under common parallel trend assumption
- But causal slope effects are non-parametrically identified (in a DID fashion) only under strong parallel trend assumption, which is not testable by pre-trend
- You need not only random assignment of treatment, but also random assignment of dose amount

Callaway et al(2021): Conclusion

- In general, it is hard to identify meaningful causal effects using DID fashion in continuous treatment case
- Causal level effects are non-parametrically identified under common parallel trend assumption
- But causal slope effects are non-parametrically identified (in a DID fashion) only under strong parallel trend assumption, which is not testable by pre-trend
- You need not only random assignment of treatment, but also random assignment of dose amount

Callaway et al(2021): Conclusion

- In a simple two period case, TWFE estimator delivers a weighted average of causal responses only under strong parallel trend assumption
- In a staggered DID case, TWFE estimator suffers from negative weight and forbidden comparison issue even under strong parallel trend assumption

Callaway et al(2021): Conclusion

- In a simple two period case, TWFE estimator delivers a weighted average of causal responses only under strong parallel trend assumption
- In a staggered DID case, TWFE estimator suffers from negative weight and forbidden comparison issue even under strong parallel trend assumption

Callaway et al(2021): Conclusion

- In a simple two period case, TWFE estimator delivers a weighted average of causal responses only under strong parallel trend assumption
- In a staggered DID case, TWFE estimator suffers from negative weight and forbidden comparison issue even under strong parallel trend assumption

Callaway et al(2021): Conclusion

Bad News! What should we do?

- Using non-parametric method to estimate the effect (Callaway, Goodman-Bacon, and Sant'Anna (2021) does not give the Stata Package)
- Using structural method or theoretical models to help you to interpret your results
- Be careful about the strong parallel trend assumption you have to impose

Callaway et al(2021): Conclusion

Bad News! What should we do?

- Using non-parametric method to estimate the effect (Callaway, Goodman-Bacon, and Sant'Anna (2021) does not give the Stata Package)
- Using structural method or theoretical models to help you to interpret your results
- Be careful about the strong parallel trend assumption you have to impose

Callaway et al(2021): Conclusion

Bad News! What should we do?

- Using non-parametric method to estimate the effect (Callaway, Goodman-Bacon, and Sant'Anna (2021) does not give the Stata Package)
- Using structural method or theoretical models to help you to interpret your results
- Be careful about the strong parallel trend assumption you have to impose

Callaway et al(2021): Conclusion

Bad News! What should we do?

- Using non-parametric method to estimate the effect (Callaway, Goodman-Bacon, and Sant'Anna (2021) does not give the Stata Package)
- Using structural method or theoretical models to help you to interpret your results
- Be careful about the strong parallel trend assumption you have to impose

Callaway et al(2021): Conclusion

Bad News! What should we do?

- Using non-parametric method to estimate the effect (Callaway, Goodman-Bacon, and Sant'Anna (2021) does not give the Stata Package)
- Using structural method or theoretical models to help you to interpret your results
- Be careful about the strong parallel trend assumption you have to impose

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

Final Conclusion

- Linear regression is, after all, a parametric method, which imposes strong functional form assumptions
- It is a simple, elegant, and good statistical tool
- But when things become more and more complicated (heterogeneous, dynamic, continuous...), regression may not be capable to capture many data patterns and give weird results
- Fundamental solution 1: Non-parametric tools which gives you enough flexibility to capture complicated patterns
- Fundamental solution 2: Structural model which helps you to regulate and rationalize the data with economic theories

References

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with a Continuous Treatment." *arXiv preprint arXiv:2107.02637* .

Card, David and Alan B Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4):772–793.

De Chaisemartin, Clément and Xavier d'Haultfoeuille. 2020. "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–2996.