

Frontier Topics in Empirical Economics: Week 8

Causal Inference with Panel Data I

Zibin Huang¹

¹College of Business, Shanghai University of Finance and Economics

November 11, 2024

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider only cross-sectional data
- What if we have one more dimension: Time?
- We call it Panel Data
- We can exploit variations across time for the same individual (unit)

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider only cross-sectional data
- What if we have one more dimension: Time?
- We call it Panel Data
- We can exploit variations across time for the same individual (unit)

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider only cross-sectional data
- What if we have one more dimension: Time?
- We call it Panel Data
- We can exploit variations across time for the same individual (unit)

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider only cross-sectional data
- What if we have one more dimension: Time?
- We call it Panel Data
- We can exploit variations across time for the same individual (unit)

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider only cross-sectional data
- What if we have one more dimension: Time?
- We call it Panel Data
- We can exploit variations across time for the same individual (unit)

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \beta D_{it} + \alpha A_i + \gamma X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}; D_{it}) = \beta D_{it} + \alpha A_i + \gamma X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \in D_{it} | A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \beta D_{it} + \alpha A_i + \gamma X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}, D_{it}) = \beta D_{it} + \alpha A_i + \gamma X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \perp D_{it} | A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}, D_{it}) = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \perp D_{it} | A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}, D_{it}) = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \in D_{it} \perp A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}, D_{it}) = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \perp D_{it} | A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Settings

- What is the impact of military service on wages?
- Person i , time t , wage Y_{it} , military service status D_{it} , ability A_i , covariates X_{it}
- Assume constant TE and linear CEF, we have:

$$Y_{it} = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} + \epsilon_{it} \quad (1)$$

$$E(Y_{it} | A_i, X_{it}, D_{it}) = \alpha + \beta D_{it} + \gamma A_i + \delta X_{it} \quad (2)$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \perp D_{it} | A_i, X_{it}$
- How to estimate β ? Three simple ways

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$Y_{it} = D_{it} A_i + X_{it} \beta + \epsilon_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_i + X_{it} - \bar{X}_{it} \beta + \epsilon_{it} - \bar{\epsilon}_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} X_{it} - \bar{X}_{it} \beta + \epsilon_{it} - \bar{\epsilon}_{it} \quad (4)$$

- Unobserved time-invariant A_i is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$Y_{it} = D_{it} A_i + X_{it} \beta + \epsilon_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_i = D_{it} - \bar{D}_i + A_i - A_i + X_{it} - \bar{X}_i + \epsilon_{it} - \bar{\epsilon}_i \quad (3)$$

$$D_{it} - \bar{D}_i + X_{it} - \bar{X}_i + \epsilon_{it} - \bar{\epsilon}_i \quad (4)$$

- Unobserved time-invariant A_i is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$\bar{Y}_{it} = \bar{D}_{it} A_j + \bar{X}_{it} \beta_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_j + X_{it} - \bar{X}_{it} \beta_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} = X_{it} - \bar{X}_{it} \beta_{it} \quad (4)$$

- Unobserved time-invariant A_j is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$\bar{Y}_{it} = \bar{D}_{it} A_j + \bar{X}_{it} \beta_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_j + X_{it} - \bar{X}_{it} \beta_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} = X_{it} - \bar{X}_{it} \beta_{it} \quad (4)$$

- Unobserved time-invariant A_j is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$\bar{Y}_{it} = \bar{D}_{it} A_j + \bar{X}_{it} \beta_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_j + X_{it} - \bar{X}_{it} \beta_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} = X_{it} - \bar{X}_{it} \beta_{it} \quad (4)$$

- Unobserved time-invariant A_j is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$\bar{Y}_{it} = \bar{D}_{it} A_i + \bar{X}_{it} \beta_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_i + X_{it} - \bar{X}_{it} \beta_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} = X_{it} - \bar{X}_{it} \beta_{it} \quad (4)$$

- Unobserved time-invariant A_i is canceled out
- Just run regression (4) and get

Fixed Effect: FE Estimator

Method 1: Fixed Effect Estimator

- FE Estimator is a deviation-from-mean estimator
- Step 1: Take individual-level means of both sides of the regression

$$\bar{Y}_{it} = \bar{D}_{it} A_i + \bar{X}_{it} \beta_{it}$$

- Step 2: Subtract the mean from the original regression

$$Y_{it} - \bar{Y}_{it} = D_{it} - \bar{D}_{it} A_i + X_{it} - \bar{X}_{it} \beta_{it} \quad (3)$$

$$D_{it} - \bar{D}_{it} = X_{it} - \bar{X}_{it} \beta_{it} \quad (4)$$

- Unobserved time-invariant A_i is canceled out
- Just run regression (4) and get

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

- We can add a set of individual dummies
- Saturate across the individual dimension

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (5)$$

- Unobserved α_i is absorbed in dummy α_i
- Just run regression (5) and get
- Dummy regression is identical to FE regression

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t - 1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t - 1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t - 1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FD Estimator

Method 3: FD Estimator

- We can run the regression using differencing (across time) variables
- Assume that $\Delta Y_{it} = Y_{it} - Y_{it-1}$ means time difference
- Subtracting regression in t by $t - 1$, we have:

$$\Delta Y_{it} = \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \quad (6)$$

- Unobserved A_i is canceled out by the differencing

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
 - Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

Fixed Effect: FE, Dummy, and FD Estimator

- All of them employ the variations across time for the same person
- FE and Dummy estimators are identical
Their point estimations, std errs, and other main statistics are the same
- So we usually call FE and Dummy estimators "FE Model"
- FE and FD are same in two-period case
- FE and FD are different when $T \neq 2$
- When ϵ_{it} are uncorrelated shocks, FE is more efficient than FD
- Since FD will create serial correlation
- But when ϵ_{it} follows random walk, FD is better since difference is now uncorrelated

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = D_{it} + X_{it} + \alpha_i + \alpha_t + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \mu_t + \nu_i + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + X_{it}'\gamma + \mu_i + \nu_t + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \mu_t + \nu_i + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \mu_i + \nu_t + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \mu_t + \nu_i + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

- For panel data, usually we can control for both individual and time FE

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \mu_t + \nu_i + \epsilon_{it} \quad (7)$$

- This is called Two-way Fixed Effect Model (TWFE)
- Difference-in-Differences (DID) is a special case of TWFE model
- In DID, usually some policy is implemented at higher level (Province, City...)
- D_{it} is binary (whether individual i at time t is treated by the policy)
- We control for Individual/Province/City level FE and time FE

DID: Settings

- Example: Card and Krueger (1994) Effects of minimum wage on employment
- On April 1, 1992, New Jersey raised the state minimum wage
- But in its neighbouring state of Pennsylvania, nothing happened
- Card and Krueger collected employment data in fast food restaurants in NJ and PA in Feb 1992 and Nov 1992

- Example: Card and Krueger (1994) Effects of minimum wage on employment
- On April 1, 1992, New Jersey raised the state minimum wage
- But in its neighbouring state of Pennsylvania, nothing happened
- Card and Krueger collected employment data in fast food restaurants in NJ and PA in Feb 1992 and Nov 1992

- Example: Card and Krueger (1994) Effects of minimum wage on employment
- On April 1, 1992, New Jersey raised the state minimum wage
- But in its neighbouring state of Pennsylvania, nothing happened
- Card and Krueger collected employment data in fast food restaurants in NJ and PA in Feb 1992 and Nov 1992

- Example: Card and Krueger (1994) Effects of minimum wage on employment
- On April 1, 1992, New Jersey raised the state minimum wage
- But in its neighbouring state of Pennsylvania, nothing happened
- Card and Krueger collected employment data in fast food restaurants in NJ and PA in Feb 1992 and Nov 1992

- Example: Card and Krueger (1994) Effects of minimum wage on employment
- On April 1, 1992, New Jersey raised the state minimum wage
- But in its neighbouring state of Pennsylvania, nothing happened
- Card and Krueger collected employment data in fast food restaurants in NJ and PA in Feb 1992 and Nov 1992

DID: Parallel Trend Assumption

- For restaurant l in state s at time t , we denote:
employment Y_{lst} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = \mathbb{1}_{NJ_s} d_t$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1lst} - Y_{0lst} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1lst} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- For restaurant i in state s at time t , we denote:
employment Y_{ist} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = NJ_s d_t$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1ist} - Y_{0ist} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1ist} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- For restaurant i in state s at time t , we denote:
employment Y_{ist} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = \mathbb{1}_{\{s = NJ, t \geq t_0\}}$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1ist} - Y_{0ist} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1ist} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- For restaurant i in state s at time t , we denote:
employment Y_{ist} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = \mathbb{1}_{\{s = NJ, t \geq t_0\}}$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1ist} - Y_{0ist} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1ist} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- For restaurant i in state s at time t , we denote:
employment Y_{ist} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = \mathbb{1}_{\{s = NJ, t \geq t_0\}}$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1ist} - Y_{0ist} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1ist} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- For restaurant i in state s at time t , we denote:
employment Y_{ist} , minimum wage policy change dummy D_{st}
- In this case, $D_{st} = \mathbb{1}_{\{s = NJ, t \geq t_0\}}$, if t is after the policy change, $d_t = 1$
- Our target: $E[Y_{1ist} - Y_{0ist} | D_{st} = 1]$ (ATT)
- Question: We only observe Y_{1ist} for restaurants in NJ (treated state) after policy
- How would the employment evolve without the policy in NJ?

DID: Parallel Trend Assumption

- Let's use restaurants in PA (untreated state) as the control group
- Parallel Trend Assumption: there is no different trends across treated/non-treated states if none of them experienced policy changes

$$E[Y_{0st} | s; t] = \alpha_s + \beta_t \quad (8)$$

- The no treatment potential outcome Y_0 does not vary across dimension $s \quad t$
- No terms like γ_{st} in $E[Y_{0st} | s; t]$

DID: Parallel Trend Assumption

- Let's use restaurants in PA (untreated state) as the control group
- Parallel Trend Assumption: there is no different trends across treated/non-treated states if none of them experienced policy changes

$$E Y_{0ist} \parallel s; t \quad s \quad t \quad (8)$$

- The no treatment potential outcome Y_0 does not vary across dimension $s \quad t$
- No terms like $_{st}$ in $E Y_{0ist} \parallel s; t$

DID: Parallel Trend Assumption

- Let's use restaurants in PA (untreated state) as the control group
- Parallel Trend Assumption: there is no different trends across treated/non-treated states if none of them experienced policy changes

$$E Y_{0ist} \parallel s; t \quad s \quad t \quad (8)$$

- The no treatment potential outcome Y_0 does not vary across dimension $s \quad t$
- No terms like $_{st}$ in $E Y_{0ist} \parallel s; t$

DID: Parallel Trend Assumption

- Let's use restaurants in PA (untreated state) as the control group
- Parallel Trend Assumption: there is no different trends across treated/non-treated states if none of them experienced policy changes

$$E Y_{0ist} \parallel s; t \quad s \quad t \quad (8)$$

- The no treatment potential outcome Y_0 does not vary across dimension $s \quad t$
- No terms like $_{st}$ in $E Y_{0ist} \parallel s; t$

DID: Parallel Trend Assumption

- Let's use restaurants in PA (untreated state) as the control group
- Parallel Trend Assumption: there is no different trends across treated/non-treated states if none of them experienced policy changes

$$E Y_{0ist} \parallel s; t \quad s \quad t \quad (8)$$

- The no treatment potential outcome Y_0 does not vary across dimension $s \quad t$
- No terms like st in $E Y_{0ist} \parallel s; t$

DID: Identification

- With the parallel trend assumption, we can identify the policy effect by running:

$$Y_{ist} = \alpha_s + \beta_t + \gamma_{st} + \delta_{ist} + \epsilon_{ist} \quad (9)$$

- First difference: For same state, dif across time

$$E[Y_{ist} | s = PA; t = Nov] - E[Y_{ist} | s = PA; t = Feb] = \alpha_s + \beta_{Nov} - \alpha_s - \beta_{Feb} = \beta_{Nov} - \beta_{Feb} \quad (10)$$

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = NJ; t = Feb] = \alpha_s + \beta_{Nov} - \alpha_s - \beta_{Feb} = \beta_{Nov} - \beta_{Feb} \quad (11)$$

- Second difference: Difference in trends across states

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = PA; t = Nov] - (E[Y_{ist} | s = NJ; t = Feb] - E[Y_{ist} | s = PA; t = Feb]) = \beta_{Nov} - \beta_{Nov} = 0 \quad (12)$$

DID: Identification

- With the parallel trend assumption, we can identify the policy effect by running:

$$Y_{ist} = \alpha_s + \beta_t + \gamma_{st} + D_{st} + \epsilon_{ist} \quad (9)$$

- First difference: For same state, dif across time

$$E[Y_{ist} | s = PA; t = Nov] - E[Y_{ist} | s = PA; t = Feb] = \alpha_s + \beta_{Nov} - \alpha_s - \beta_{Feb} = \beta_{Nov} - \beta_{Feb} \quad (10)$$

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = NJ; t = Feb] = \alpha_s + \beta_{Nov} - \alpha_s - \beta_{Feb} = \beta_{Nov} - \beta_{Feb} \quad (11)$$

- Second difference: Difference in trends across states

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = PA; t = Nov] - (E[Y_{ist} | s = NJ; t = Feb] - E[Y_{ist} | s = PA; t = Feb]) = \alpha_{NJ} + \beta_{Nov} - \alpha_{PA} - \beta_{Nov} - (\alpha_{NJ} + \beta_{Feb} - \alpha_{PA} - \beta_{Feb}) = \alpha_{NJ} - \alpha_{PA} - \alpha_{NJ} + \alpha_{PA} = 0 \quad (12)$$

DID: Identification

- With the parallel trend assumption, we can identify the policy effect by running:

$$Y_{ist} = \alpha_s + \beta_t + \gamma D_{st} + \epsilon_{ist} \quad (9)$$

- First difference: For same state, dif across time

$$E[Y_{ist} | s = PA; t = Nov] - E[Y_{ist} | s = PA; t = Feb] = \beta_{Nov} - \beta_{Feb} \quad (10)$$

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = NJ; t = Feb] = \beta_{Nov} - \beta_{Feb} \quad (11)$$

- Second difference: Difference in trends across states

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = PA; t = Nov] - (E[Y_{ist} | s = NJ; t = Feb] - E[Y_{ist} | s = PA; t = Feb]) = \beta_{Nov} - \beta_{Feb} - (\beta_{Nov} - \beta_{Feb}) = 0 \quad (12)$$

DID: Identification

- With the parallel trend assumption, we can identify the policy effect by running:

$$Y_{ist} = \alpha_s + \beta_t + \gamma_{st} + D_{st} + \epsilon_{ist} \quad (9)$$

- First difference: For same state, dif across time

$$E[Y_{ist} | s = PA; t = Nov] - E[Y_{ist} | s = PA; t = Feb] = \beta_{Nov} - \beta_{Feb} \quad (10)$$

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = NJ; t = Feb] = \beta_{Nov} - \beta_{Feb} \quad (11)$$

- Second difference: Difference in trends across states

$$E[Y_{ist} | s = NJ; t = Nov] - E[Y_{ist} | s = PA; t = Nov] - (E[Y_{ist} | s = NJ; t = Feb] - E[Y_{ist} | s = PA; t = Feb]) = \beta_{Nov} - \beta_{Feb} - (\beta_{Nov} - \beta_{Feb}) = 0 \quad (12)$$

DID: Identification

We are taking untreated group as the control!

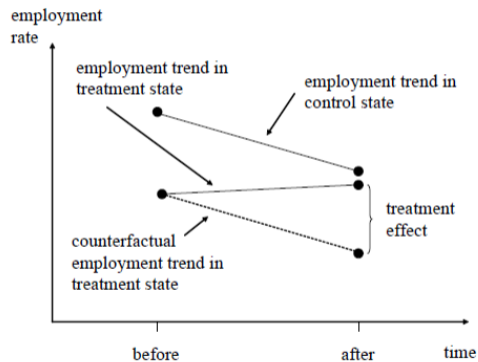


Figure 5.2.1: Causal effects in the differences-in-differences model

DID: Test of Parallel Trend

- After the implementation of the policy at t_0 , we can no longer observe Y_{0j} for the treated group
- Thus, we cannot test parallel trend after t_0
- We test parallel trend before t_0 : Pre-trend test
- There are two simple ways to do that

DID: Test of Parallel Trend

- After the implementation of the policy at t_0 , we can no longer observe Y_{0i} for the treated group
- Thus, we cannot test parallel trend after t_0
- We test parallel trend before t_0 : Pre-trend test
- There are two simple ways to do that

DID: Test of Parallel Trend

- After the implementation of the policy at t_0 , we can no longer observe Y_{0i} for the treated group
- Thus, we cannot test parallel trend after t_0
- We test parallel trend before t_0 : Pre-trend test
- There are two simple ways to do that

DID: Test of Parallel Trend

- After the implementation of the policy at t_0 , we can no longer observe Y_{0i} for the treated group
- Thus, we cannot test parallel trend after t_0
- We test parallel trend before t_0 : Pre-trend test
- There are two simple ways to do that

DID: Test of Parallel Trend

- After the implementation of the policy at t_0 , we can no longer observe Y_{0i} for the treated group
- Thus, we cannot test parallel trend after t_0
- We test parallel trend before t_0 : Pre-trend test
- There are two simple ways to do that

DID: Test of Parallel Trend

1. Draw the changes in Y across time directly
 - Is this a good pre-trend? (Before the first vertical line)

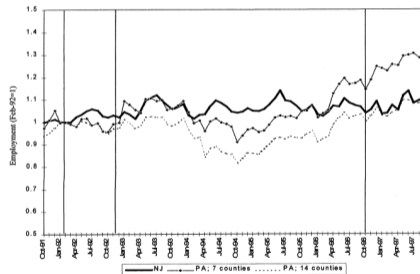


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

DID: Test of Parallel Trend

1. Draw the changes in Y across time directly

- Is this a good pre-trend? (Before the first vertical line)

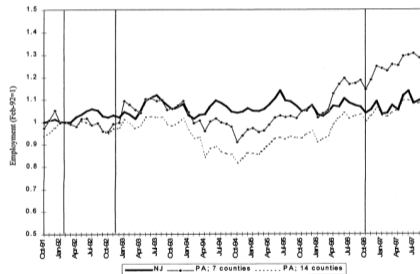


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

DID: Test of Parallel Trend

1. Draw the changes in Y across time directly
 - Is this a good pre-trend? (Before the first vertical line)

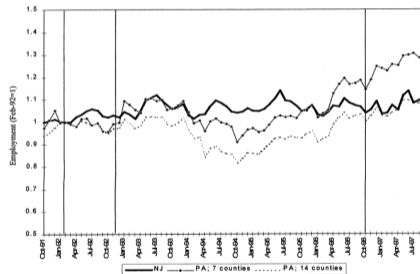


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

DID: Test of Parallel Trend

■ What about this?

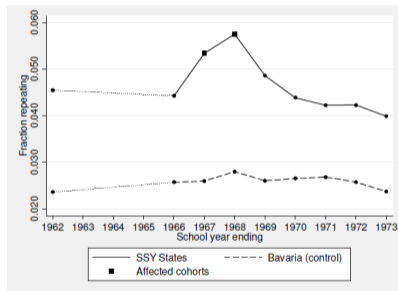


Figure 5.2.3: Average rates of grade repetition in second grade for treatment and control schools in Germany (from Pischke 2007). The data span a period before and after a change in term length for students outside of Bavaria.

DID: Test of Parallel Trend

- What about this?

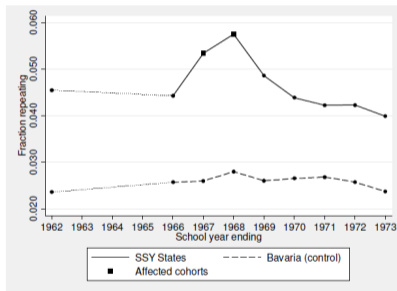


Figure 5.2.3: Average rates of grade repetition in second grade for treatment and control schools in Germany (from Pischke 2007). The data span a period before and after a change in term length for students outside of Bavaria.

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{jt} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \alpha_s + \beta_t + \sum_{j=1}^J \gamma_j D_{s,ist} \quad (13)$$

- γ_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{jt} is implemented at $t = 0$
- Let D_{st} be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \alpha_s + \beta_t + \sum_{j=1}^T \delta_j D_{st} + \epsilon_{ist} \quad (13)$$

- δ_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

2. Event Study Regression

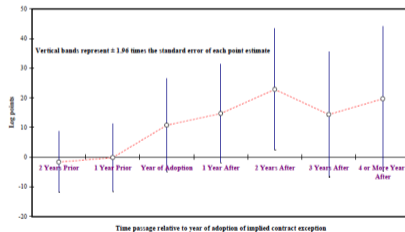
- If we have data from $-T$ to T' , and the policy D_{it} is implemented at $t = 0$
- Let D_s be the dummy of whether in the treated group
- Run the following regression

$$Y_{ist} = \sum_{j=1}^J \beta_j D_{st} + \epsilon_{ist} \quad (13)$$

- β_j shows the changes of differences in trends between treated and untreated groups
- In DID, we separate the time into two parts: Before $t = 0$ and after $t = 0$
We have $D_{st} = D_s d_t = 1$ only for treated group after policy implementation
- In event study, we give each time point (year/month) a parameter

DID: Test of Parallel Trend

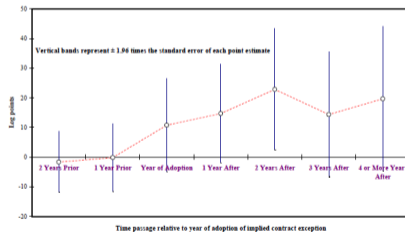
- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of $\Delta \ln y_{it}$ for each time period and have:



- Points before $t = 0$ are not significant Pre-trend is parallel
- Points after $t = 0$ shows the policy effect

DID: Test of Parallel Trend

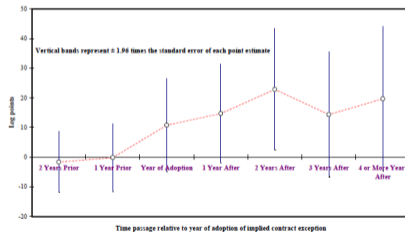
- Usually $t = 1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of $\Delta \ln y_{it}$ for each time period and have:



- Points before $t = 0$ are not significant Pre-trend is parallel
- Points after $t = 0$ shows the policy effect

DID: Test of Parallel Trend

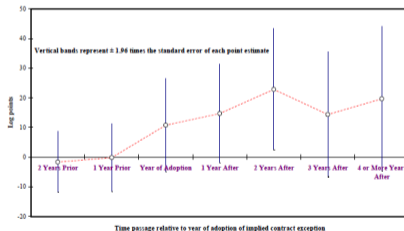
- Usually $t = 1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of $\Delta \ln y_{it}$ for each time period and have:



- Points before $t = 0$ are not significant Pre-trend is parallel
- Points after $t = 0$ shows the policy effect

DID: Test of Parallel Trend

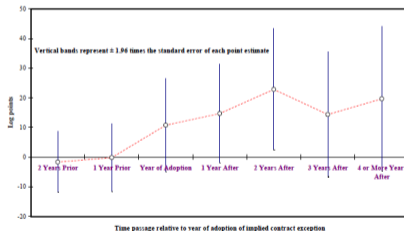
- Usually $t = 1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of $\Delta \ln y_{it}$ for each time period and have:



- Points before $t = 0$ are not significant Pre-trend is parallel
- Points after $t = 0$ shows the policy effect

DID: Test of Parallel Trend

- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of $\Delta \ln y_{it}$ for each time period and have:



- Points before $t = 0$ are not significant Pre-trend is parallel
- Points after $t = 0$ shows the policy effect

DID: A Traditional Procedure

- 1. Draw changes of Y as a descriptive evidence
- 2. Run your main DID regression
- 3. Run event study regression to check the pre-trend and the dynamic effect
- 4. Remember to cluster your standard errors (More details in the following lectures)

DID: A Traditional Procedure

- 1. Draw changes of Y as a descriptive evidence
- 2. Run your main DID regression
- 3. Run event study regression to check the pre-trend and the dynamic effect
- 4. Remember to cluster your standard errors (More details in the following lectures)

DID: A Traditional Procedure

- 1. Draw changes of Y as a descriptive evidence
- 2. Run your main DID regression
- 3. Run event study regression to check the pre-trend and the dynamic effect
- 4. Remember to cluster your standard errors (More details in the following lectures)

DID: A Traditional Procedure

- 1. Draw changes of Y as a descriptive evidence
- 2. Run your main DID regression
- 3. Run event study regression to check the pre-trend and the dynamic effect
- 4. Remember to cluster your standard errors (More details in the following lectures)

DID: A Traditional Procedure

- 1. Draw changes of Y as a descriptive evidence
- 2. Run your main DID regression
- 3. Run event study regression to check the pre-trend and the dynamic effect
- 4. Remember to cluster your standard errors (More details in the following lectures)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- When you are doing causal research, a central question is:
What kind of variations are used to identify the causal effect?
- It is very very very important!!!
- It determines how you can interpret your results
- It determines which assumption you are using
- Then also determines how you should defend your research (why your assumption is reasonable)

Identification Source

- It becomes complicated in panel data → more dimensions
- Sometimes, people control many FEs at different levels
- Some are even combined with IV, RD, or other regression structure
- Still, you should always be very clear about your identification source

Identification Source

- It becomes complicated in panel data more dimensions
- Sometimes, people control many FEs at different levels
- Some are even combined with IV, RD, or other regression structure
- Still, you should always be very clear about your identification source

Identification Source

- It becomes complicated in panel data more dimensions
- Sometimes, people control many FEs at different levels
- Some are even combined with IV, RD, or other regression structure
- Still, you should always be very clear about your identification source

Identification Source

- It becomes complicated in panel data more dimensions
- Sometimes, people control many FEs at different levels
- Some are even combined with IV, RD, or other regression structure
- Still, you should always be very clear about your identification source

Identification Source

- It becomes complicated in panel data more dimensions
- Sometimes, people control many FEs at different levels
- Some are even combined with IV, RD, or other regression structure
- Still, you should always be very clear about your identification source

Identification Source

- Let's consider a simple case: effects of working experience on wage
- For individual l from family j at time t :

$$wage_{ijt} = \alpha_0 + \alpha_1 exp_{ijt} + \epsilon_{ijt} \quad (14)$$

Identification Source

- Let's consider a simple case: effects of working experience on wage
- For individual i from family j at time t :

$$wage_{ijt} = \alpha_0 + \alpha_1 exp_{ijt} + \epsilon_{ijt} \quad (14)$$

Identification Source

- Let's consider a simple case: effects of working experience on wage
- For individual i from family j at time t :

$$wage_{ijt} = \alpha_0 + \alpha_1 exp_{ijt} + \epsilon_{ijt} \quad (14)$$

Identification Source

- When controlling for time FE, you are using variations across individuals and families ($i;j$ level) in the same year
- When controlling for individual FE, you are using variations across time (t level) for the same people
- When controlling for family FE and time FE, you are using variations across individuals within the same family ($i|j$ level)
- When controlling for individual FE and time FE, you are using variations of time trends for different people ($i \times t$ level)

Identification Source

- When controlling for time FE, you are using variations across individuals and families ($i; j$ level) in the same year
- When controlling for individual FE, you are using variations across time (t level) for the same people
- When controlling for family FE and time FE, you are using variations across individuals within the same family ($i; j$ level)
- When controlling for individual FE and time FE, you are using variations of time trends for different people ($i \quad t$ level)

Identification Source

- When controlling for time FE, you are using variations across individuals and families ($i;j$ level) in the same year
- When controlling for individual FE, you are using variations across time (t level) for the same people
- When controlling for family FE and time FE, you are using variations across individuals within the same family ($i||j$ level)
- When controlling for individual FE and time FE, you are using variations of time trends for different people ($i \quad t$ level)

Identification Source

- When controlling for time FE, you are using variations across individuals and families ($i;j$ level) in the same year
- When controlling for individual FE, you are using variations across time (t level) for the same people
- When controlling for family FE and time FE, you are using variations across individuals within the same family ($i||j$ level)
- When controlling for individual FE and time FE, you are using variations of time trends for different people ($i \quad t$ level)

Identification Source

- When controlling for time FE, you are using variations across individuals and families ($i;j$ level) in the same year
- When controlling for individual FE, you are using variations across time (t level) for the same people
- When controlling for family FE and time FE, you are using variations across individuals within the same family ($i|j$ level)
- When controlling for individual FE and time FE, you are using variations of time trends for different people ($i \quad t$ level)

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in product testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complex

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in pre-trend testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complicated

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in pre-trend testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complicated

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in pre-trend testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complicated

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in pre-trend testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complicated

Further Topics in Panel Data: Extension of DID

- We have introduced basic methods of causal inference in panel data
- Now we go to three important extensions
 - Recent development in pre-trend testing
 - Synthetic Control Method: When you do not have parallel trend
 - Staggered DID: When policy implementation scheme is complicated

Pre-trend Testing: New Development

- Is event study a perfect tool to test parallel pre-trend?
- It's good, but far from perfect
- Roth (2022) Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends
- It also has a latest published version in AER Insights

Pre-trend Testing: New Development

- Is event study a perfect tool to test parallel pre-trend?
- It's good, but far from perfect
- Roth (2022) Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends
- It also has a latest published version in AER Insights

Pre-trend Testing: New Development

- Is event study a perfect tool to test parallel pre-trend?
- It's good, but far from perfect
- Roth (2022) Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends
- It also has a latest published version in AER Insights

Pre-trend Testing: New Development

- Is event study a perfect tool to test parallel pre-trend?
- It's good, but far from perfect
- Roth (2022) Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends
- It also has a latest published version in AER Insights

Pre-trend Testing: New Development

- Is event study a perfect tool to test parallel pre-trend?
- It's good, but far from perfect
- Roth (2022) Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends
- It also has a latest published version in AER Insights

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error

- Pre-existing trends that produce meaningful bias may not be detected
- Assuming a linear violation of parallel trend: $y_t = \alpha_0 + \alpha_1 t$
- Roth implements some Monte Carlo Simulation using data from 70 papers
- He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
- The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error

- Pre-existing trends that produce meaningful bias may not be detected
- Assuming a linear violation of parallel trend: $y_t = \alpha_0 + \alpha_1 t$
- Roth implements some Monte Carlo Simulation using data from 70 papers
- He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
- The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error

- Pre-existing trends that produce meaningful bias may not be detected
- Assuming a linear violation of parallel trend: $\beta_1 t - \beta_0 t - t$
- Roth implements some Monte Carlo Simulation using data from 70 papers
- He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
- The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error
 - Pre-existing trends that produce meaningful bias may not be detected
 - Assuming a linear violation of parallel trend: $\beta_1 t - \beta_0 t - t$
 - Roth implements some Monte Carlo Simulation using data from 70 papers
 - He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
 - The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error
 - Pre-existing trends that produce meaningful bias may not be detected
 - Assuming a linear violation of parallel trend: $y_{1t} - y_{0t} = \alpha t$
 - Roth implements some Monte Carlo Simulation using data from 70 papers
 - He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
 - The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error
 - Pre-existing trends that produce meaningful bias may not be detected
 - Assuming a linear violation of parallel trend: $\beta_1 t + \beta_0 + \gamma t$
 - Roth implements some Monte Carlo Simulation using data from 70 papers
 - He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
 - The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

1. Statistical power is low: Likely to have type-II error
 - Pre-existing trends that produce meaningful bias may not be detected
 - Assuming a linear violation of parallel trend: $\beta_1 t + \beta_0 + \gamma t$
 - Roth implements some Monte Carlo Simulation using data from 70 papers
 - He finds that if you want to detect this violation for 80% of the time, the bias has to be as large as the estimated TE! (100% bias)
 - The bias has to be very large for you to detect it!

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Why? It goes back to the nature of the statistical test

- Type I error: H_0 is true but we reject it,
- Type II error: H_0 is false but we do not reject it,
- Significance level: Probability of committing Type I error,
- Power: Probability of rejecting H_0 if it is false, 1

Pre-trend Testing: New Development

Tradeo !!!

- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger ; Go right, you have larger

Pre-trend Testing: New Development

Tradeo !!!

- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger ; Go right, you have larger

Pre-trend Testing: New Development

Tradeo !!!

- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger β ; Go right, you have larger α

Pre-trend Testing: New Development

Tradeo !!!

- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger ; Go right, you have larger

Pre-trend Testing: New Development

Tradeo !!!

- You can decrease T1ER by decreasing α increasing β
- You can decrease T2ER by decreasing α increasing β
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative

Pre-trend Testing: New Development

Tradeo !!!

- You can decrease T1ER by decreasing α increasing β
- You can decrease T2ER by decreasing α increasing β
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative

Pre-trend Testing: New Development

Tradeo !!!

- You can decrease T1ER by decreasing α increasing β
- You can decrease T2ER by decreasing α increasing β
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative

Pre-trend Testing: New Development

Tradeo !!!

- You can decrease T1ER by decreasing α increasing β
- You can decrease T2ER by decreasing β increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative

Pre-trend Testing: New Development

Tradeo !!!

- You can decrease T1ER by decreasing α increasing β
- You can decrease T2ER by decreasing β increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative

Pre-trend Testing: New Development

- In traditional testing, we try to be conservative about rejecting H_0
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large β small power
- But in pre-trend testing, actually we care more about power
- We want to be more conservative about NOT rejecting H_0
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H_0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large α small power
- But in pre-trend testing, actually we care more about power
- We want to be more conservative about NOT rejecting H_0
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H_0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large α ! small power
- But in pre-trend testing, actually we care more about power
- We want to be more conservative about NOT rejecting H_0
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large β ! small power
- But in pre-trend testing, actually we care more about power
- We want to be more conservative about NOT rejecting H0
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H_0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large ! small power
- **But in pre-trend testing, actually we care more about power**
- We want to be more conservative about NOT rejecting H_0
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large ! small power
- **But in pre-trend testing, actually we care more about power**
- **We want to be more conservative about NOT rejecting H0**
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

- In traditional testing, we try to be **conservative about rejecting H_0**
- Minimize Type I error probability to be smaller than some level (10%, 5%, 1%)
- It then leads to large ! small power
- **But in pre-trend testing, actually we care more about power**
- **We want to be more conservative about NOT rejecting H_0**
- If you make rejection very hard, of course it is very likely that you have good pre-trend

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
- 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)

■ 2c. The bias is certainly exacerbated in common cases (multiple trends and non-constant, non-independent errors)

■ 2d. Thus, the use of pre-trend testing can be ambiguous. It may be useful in some cases, but it may also be harmful in others.

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
- 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)

→ The bias is certainly exacerbated in common cases (including trends and heteroskedastic errors)

→ Thus, the use of pre-trend testing can be ambiguous

→ The best solution is to use a test that does not require conditioning on passing the test

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
- 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)
 - The bias is certainly exacerbated in common cases (monotone trends and homoskedastic errors)
 - Thus, the effect of pre-trend testing can be ambiguous
Reject non-parallel cases (good) vs. Increasing bias if there is bias (bad)

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
 - 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)
 - The bias is certainly exacerbated in common cases (monotone trends and homoskedastic errors)
 - Thus, the effect of pre-trend testing can be ambiguous
- Reject non-parallel cases (good) vs. Increasing bias if there is bias (bad)

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
 - 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)
 - The bias is certainly exacerbated in common cases (monotone trends and homoskedastic errors)
 - Thus, the effect of pre-trend testing can be ambiguous
- Reject non-parallel cases (good) vs. Increasing bias if there is bias (bad)

Pre-trend Testing: New Development

By selecting samples that can pass the test

- 2a. Underestimate the variance of the estimation
- 2b. If there is bias, conditioning on passing the event study test may exacerbate it (Adding bias to point estimation)
 - The bias is certainly exacerbated in common cases (monotone trends and homoskedastic errors)
 - Thus, the effect of pre-trend testing can be ambiguous
Reject non-parallel cases (good) vs. Increasing bias if there is bias (bad)

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
 - Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do power calculations against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
 - Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do power calculations against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do **power calculations** against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- **Do not think you are safe when event study does not detect anything**
- Do **power calculations** against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do power calculations against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do power calculations against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Pre-trend Testing: New Development

Practical suggestions proposed by Roth

- Most important advice:
Always use your economic knowledge to verify the parallel trend assumption!
- Do not think you are safe when event study does not detect anything
- Do power calculations against economically relevant violations of parallel trends (R package, but not Stata...)
- If you know the functional form of the differences in trends, control it
- Sensitivity analysis using Rambachan and Roth (2023)
Calculate the bounds of your estimates if there is some violation

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

- The critical assumption for DID is parallel trend
- What if we do not have it?
- What if treated and control provinces have different trends?
- Let's create one control group! Synthetic Control
- Synthetic control is a matching method

Synthetic Control: Main Idea

All the following contexts come from Abadie, Diamond, and Hainmueller (2010, 2015); Abadie (2021)

- The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unselected units often provides a more appropriate comparison than any single unselected unit alone.

Synthetic Control: Main Idea

All the following contexts come from Abadie, Diamond, and Hainmueller (2010, 2015); Abadie (2021)

- The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unselected units often provides a more appropriate comparison than any single unselected unit alone.

Synthetic Control: Main Idea

All the following contexts come from Abadie, Diamond, and Hainmueller (2010, 2015); Abadie (2021)

- The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, **a combination of unselected units** often provides a more appropriate comparison than any single unselected unit alone.

Synthetic Control: Main Idea

- Take Abadie, Diamond, and Hainmueller (2010) as an example
- California implemented Proposition 99 in 1988
- It is a large-scale tobacco control program
 - A 25-cent per pack excise tax on the sale of tobacco cigarettes, cigars and chewing tobacco
 - A ban on cigarette vending machines in public areas accessible by juveniles

Synthetic Control: Main Idea

- But it seems that pre-trends are very different across states

Synthetic Control: Main Idea

- Even when you average over all control states, you have this

Synthetic Control: Main Idea

- Then you have to combine them to create a "synthetic" control state
- A man-made "synthetic" California

(a)

(b)

Synthetic Control: Settings

- Suppose we have $j = 1, 2, \dots, J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, \dots, J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1; 2; \dots; J - 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period $j = 1$ is the treated unit
- We call $j = 2; 3; \dots; J - 1$ as "donor pool"
- We will create the synthetic control group from units in this "donor pool"
- X_{kj} are observed characteristics, which can include pre-treatment values of
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $\tau_{jt} = Y_{jt}^I - Y_{jt}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$Y_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$\tau_{1t} = Y_{1t} - Y_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- Define potential outcome: $Y_{jt}^I; Y_{jt}^N$
- Treatment effect of interest: $Y_{1t}^I - Y_{1t}^N$ for $t \geq T_0$
- Treatment effect can vary across time
- A synthetic control is defined as a weighted average of the units in the donor pool:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J-1} w_j Y_{jt}$$

- w_j is the weight assigned to donor j
- Then we can estimate the treatment effect:

$$Y_{1t}^I - \hat{Y}_{1t}^N$$

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k (w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1})^2 \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1} \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1} \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1} \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1} \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- How to define the weights?
- We minimize the following:

$$\frac{1}{2} \|X_1 - X_0 W\|_2^2 = \sum_{h=1}^k w_h X_{h1} + w_2 X_{h2} + \dots + w_{J-1} X_{hJ-1} \quad (15)$$

- This is the weighted euclidean distance between X_1 and X_0
- We try to find a combination of donors that can mimic our treated group the best
- Watch out: the difference between weights w and weights w

Synthetic Control: Settings

- w is the weight assigned to each unit (state) when we want to create a synthetic control group (state)
- v is the weight assigned to each characteristic when we try to calculate

Synthetic Control: Settings

- w is the weight assigned to each unit (state) when we want to create a synthetic control group (state)
- v is the weight assigned to each characteristic when we try to calculate

Synthetic Control: Settings

- w is the weight assigned to each unit (state) when we want to create a synthetic control group (state)
- v is the weight assigned to each characteristic when we try to calculate w

Synthetic Control: An Example

- How to estimate the effect of the 1990 German reunification
- Treated: West Germany; Untreated: Other OECD countries

Synthetic Control: An Example

- How to estimate the effect of the 1990 German reunification
- Treated: West Germany; Untreated: Other OECD countries

Synthetic Control: An Example

- How to estimate the effect of the 1990 German reunification
- Treated: West Germany; Untreated: Other OECD countries

Synthetic Control: An Example

- v are weights for economic predictors: The importance of each predictor for the match procedure

Synthetic Control: An Example

- w are weights for compared countries: the importance of each country in forming the synthetic Germany

Synthetic Control: Procedure

How to determine \hat{v} ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1; \dots; t_0$; and Validation: $t = t_0 + 1; \dots; T_0$
- Step 2: For every possible \hat{v} , calculate \hat{w} by minimizing equation (15) using training data $\hat{w} = \hat{w}(V)$
- Step 3: Find the best \hat{v}^* that minimizes MSPE using validation data:

$$\hat{v}^* = \underset{\hat{v}}{\text{argmin}} \sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \hat{w}_2(V) Y_{2t} - \dots - \hat{w}_{J-1}(V) Y_{J-1t} \right)^2$$

- Step 4: Using \hat{v}^* and validation data to calculate \hat{v}

Synthetic Control: Procedure

How to determine \hat{v} ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1, \dots, t_0$; and Validation: $t = t_0 + 1, \dots, T_0$
- Step 2: For every possible \hat{v} , calculate \hat{w} by minimizing equation (15) using training data $\hat{w} = \hat{w}^T(\hat{v})$
- Step 3: Find the best \hat{v}^* that minimizes MSPE using validation data:

$$\hat{v}^* = \underset{\hat{v}}{\text{argmin}} \sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \hat{w}_2^T(\hat{v}) Y_{2t} - \dots - \hat{w}_{J-1}^T(\hat{v}) Y_{J-1t} \right)^2$$

- Step 4: Using \hat{v}^* and validation data to calculate \hat{v}

Synthetic Control: Procedure

How to determine \hat{w} ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1; \dots; t_0$; and Validation: $t = t_0 + 1; \dots; T_0$
- Step 2: For every possible \hat{w} , calculate \hat{w} by minimizing equation (15) using training data $\{w, V\}$
- Step 3: Find the best \hat{w} that minimizes MSPE using validation data:

$$\hat{w} = \underset{w}{\text{argmin}} \sum_{t=t_0+1}^{T_0} (Y_{1t} - w_1 V_t - Y_{2t} - \dots - w_{J-1} V_t - Y_{J-1t})^2$$

- Step 4: Using \hat{w} and validation data to calculate \hat{w}

Synthetic Control: Procedure

How to determine \hat{w} ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1; \dots; t_0$; and Validation: $t = t_0 + 1; \dots; T_0$
- Step 2: For every possible \hat{w} , calculate \hat{w} by minimizing equation (15) using training data $\mathbf{w}^T \mathbf{V}$
- Step 3: Find the best \hat{w} that minimizes MSPE using validation data:

$$\hat{w} = \underset{t = t_0 + 1}{\text{arg min}} \sum_{t=t_0+1}^{T_0} (Y_{1t} - \hat{w}_1 - \hat{w}_2 V_{2t} - \dots - \hat{w}_{J-1} V_{J-1t})^2$$

- Step 4: Using \hat{w} and validation data to calculate \hat{w}

Synthetic Control: Procedure

How to determine \tilde{w} ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1; \dots; t_0$; and Validation: $t = t_0 + 1; \dots; T_0$
- Step 2: For every possible \tilde{w} , calculate \tilde{w} by minimizing equation (15) using training data $\tilde{w} = \tilde{w}_V$
- Step 3: Find the best \tilde{w}^* that minimizes MSPE using validation data:

$$\tilde{w}^* = \underset{\tilde{w}}{\text{arg min}} \sum_{t=t_0+1}^{T_0} (Y_{1t} - \tilde{w}_1 - \tilde{w}_2 V - \dots - \tilde{w}_{J-1} V)^2$$

- Step 4: Using \tilde{w}^* and validation data to calculate \tilde{w}

Synthetic Control: Procedure

How to determine v ?

- Step 1: Divide all pre-treatment sample into 2 parts
Training: $t = 1; \dots; t_0$; and Validation: $t = t_0 + 1; \dots; T_0$
- Step 2: For every possible v , calculate w by minimizing equation (15) using training data $w = V$
- Step 3: Find the best v^* that minimizes MSPE using validation data:

$$= \sum_{t=t_0+1}^{T_0} (Y_{1t} - w_1 V - Y_{2t} - \dots - w_{J-1} V - Y_{J-1t})^2$$

- Step 4: Using v^* and validation data to calculate w

Synthetic Control: An Example

- Synthetic West Germany

Synthetic Control

- Homework: Explain the reason why we split the data into two parts. No math!

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_{it} (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0W$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_{1t} (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_{1t} - X_0W^*$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within $[0, 1]$

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_1 (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0W^*$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used ~~X₁~~ (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0W^*$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_1 (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 \tilde{W}$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_1 (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 \tilde{W}$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
 - So, choose the donor pool judiciously
 - All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_1 (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 \tilde{W}$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within 0; 1

Synthetic Control: Conclusion

Several things you should remember for synthetic control

- Post-treatment outcomes cannot be used X_1 (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 \tilde{W}$ is large (pre-treatment fit is bad), do not use synthetic control
- If the size of the donor pool is too large, there can be over-fitting
- So, choose the donor pool judiciously
- All weights must be within $[0, 1]$

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- When we have data across time for different units, we have panel data
- FE, dummy, and FD regressions can cancel out time-invariant confounders
- FE and dummy regressions are identical
- FE and FD are identical in 2-period cases, but different for more than 2
- FE is more efficient if errors are not correlated; FE is more efficient if errors are random walks

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a graph
 - Run event study
- However, statistical test minimizes T1ER, which in ates T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a figure
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a figure
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a picture
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a figure
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a picture
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a picture
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- The main assumption for DID is parallel trend
- Traditionally, we validate parallel trend assumption in two steps:
 - Draw a picture
 - Run event study
- However, statistical test minimizes T1ER, which ignores T2ER
- Be careful using them!
- Check the power and validate your assumption using economic knowledge

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each control unit
 - Weights for each covariate
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

Conclusion

- When it is hard to find a good control group with parallel trend
- You can create one using synthetic control method
- You assign two sets of weights and taking the weighted average
 - Weights for each characteristics
 - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group

References

- Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59 (2):391–425.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490):493–505.
- . 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59 (2):495–510.
- 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.
- Rambachan, Ashesh and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* :rdad018.
- Roth, Jonathan. 2022. "Pretest with Caution: Event-study Estimates after Testing for Parallel Trends." *American Economic Review: Insights* 4 (3):305–322.