Frontier Topics in Empirical Economics: Week 8
Causal Inference with Panel Data I

Zibin Huang [1]

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

November 14, 2025

Fixed Effect: Panel Data

- In the previous lectures, we mostly consider 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 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 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 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 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_i, military service status D_{it}, ability A_i, covariates X_{it}
- Assume constant TE and linear CEF, we have:

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

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

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

Fixed Effect: FE Settings

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

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

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

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

Fixed Effect: FE Settings

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

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

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

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

Fixed Effect: FE Settings

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

$$Y_{it} = \alpha + \rho D_{it} + A_i'\gamma + X_{it}'\beta + \epsilon_{it} \tag{1}$$

$$E[Y_{it}|A_i, X_{it}, D_{it}] = \alpha + \rho D_{it} + A_i'\gamma + X_{it}'\beta \tag{2}$$

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

Fixed Effect: FE Settings

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

$$Y_{it} = \alpha + \rho D_{it} + A_i' \gamma + X_{it}' \beta + \epsilon_{it} \tag{1}$$

$$E[Y_{it} | A_i, X_{it}, D_{it}] = \alpha + \rho D_{it} + A_i' \gamma + X_{it}' \beta \tag{2}$$

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

Fixed Effect: FE Settings

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

$$Y_{it} = \alpha + \rho D_{it} + A_i' \gamma + X_{it}' \beta + \epsilon_{it} \tag{1}$$

$$E[Y_{it} | A_i, X_{it}, D_{it}] = \alpha + \rho D_{it} + A_i' \gamma + X_{it}' \beta \tag{2}$$

- A_i is the unobserved confounding factor, $\epsilon_{it} \perp\!\!\!\perp D_{it} | A_i, X_{it}$
- How to estimate ρ? 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

$$\bar{Y}_{it} = \alpha + \rho \bar{D}_{it} + A_i^{'}\gamma + \bar{X}_{it}^{'}\beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

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

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

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

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

$$\bar{Y}_{it} = \alpha + \rho\bar{D}_{it} + A_i^{'}\gamma + \bar{X}_{it}^{'}\beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

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

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

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

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

$$\bar{Y}_{it} = \alpha + \rho\bar{D}_{it} + A_i'\gamma + \bar{X}_{it}'\beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

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

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

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

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

$$\bar{Y}_{it} = \alpha + \rho \bar{D}_{it} + A_i' \gamma + \bar{X}_{it}' \beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

$$\begin{aligned} Y_{it} - \bar{Y}_{it} &= \alpha - \alpha + \rho(D_{it} - \bar{D}_{it}) + A_i'\gamma - A_i'\gamma + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (3) \\ &= \rho(D_{it} - \bar{D}_{it}) + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (4) \end{aligned}$$

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

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

$$\bar{Y}_{it} = \alpha + \rho \bar{D}_{it} + A_i' \gamma + \bar{X}_{it}' \beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

$$\begin{aligned} Y_{it} - \bar{Y}_{it} &= \alpha - \alpha + \rho(D_{it} - \bar{D}_{it}) + A_i'\gamma - A_i'\gamma + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (3) \\ &= \rho(D_{it} - \bar{D}_{it}) + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (4) \end{aligned}$$

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

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

$$\bar{Y}_{it} = \alpha + \rho \bar{D}_{it} + A_i' \gamma + \bar{X}_{it}' \beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

$$\begin{aligned} Y_{it} - \bar{Y}_{it} &= \alpha - \alpha + \rho(D_{it} - \bar{D}_{it}) + A_i'\gamma - A_i'\gamma + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (3) \\ &= \rho(D_{it} - \bar{D}_{it}) + (X_{it}' - \bar{X}_{it}')\beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (4) \end{aligned}$$

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

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

$$\bar{Y}_{it} = \alpha + \rho \bar{D}_{it} + A_i' \gamma + \bar{X}_{it}' \beta + \bar{\epsilon}_{it}$$

- Step 2: Subtract the mean from the original regression

$$\begin{aligned} Y_{it} - \bar{Y}_{it} &= \alpha - \alpha + \rho(D_{it} - \bar{D}_{it}) + A_i' \gamma - A_i' \gamma + (X_{it}' - \bar{X}_{it}') \beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (3) \\ &= \rho(D_{it} - \bar{D}_{it}) + (X_{it}' - \bar{X}_{it}') \beta + (\epsilon_{it} - \bar{\epsilon}_{it}) && (4) \end{aligned}$$

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

Fixed Effect: Dummy Estimator

Method 2: Dummy Estimator

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

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

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

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$Y_{it} = (\alpha + A_i^{'}\gamma) + \rho D_{it} + X_{it}^{'}\beta + \epsilon_{it}$$

$$Y_{it} = \alpha_i + \rho D_{it} + X_{it}^{'}\beta + \epsilon_{it} \tag{5}$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$Y_{it} = (\alpha + A_i'\gamma) + \rho D_{it} + X_{it}'\beta + \epsilon_{it}$$
$$Y_{it} = \alpha_i + \rho D_{it} + X_{it}'\beta + \epsilon_{it} \quad (5)$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$Y_{it} = (\alpha + A_i'\gamma) + \rho D_{it} + X_{it}'\beta + \epsilon_{it}$$
$$Y_{it} = \alpha_i + \rho D_{it} + X_{it}'\beta + \epsilon_{it} \quad (5)$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$Y_{it} = (\alpha + A_i' \gamma) + \rho D_{it} + X_{it}' \beta + \epsilon_{it}$$
$$Y_{it} = \alpha_i + \rho D_{it} + X_{it}' \beta + \epsilon_{it} \quad (5)$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$
\begin{aligned}
Y_{it} &= (\alpha + A_i' \gamma) + \rho D_{it} + X_{it}' \beta + \epsilon_{it} \\
Y_{it} &= \alpha_i + \rho D_{it} + X_{it}' \beta + \epsilon_{it} \qquad (5)
\end{aligned}
$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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

$$\begin{aligned} Y_{it} &= (\alpha + A_i'\gamma) + \rho D_{it} + X_{it}'\beta + \epsilon_{it} \\ Y_{it} &= \alpha_i + \rho D_{it} + X_{it}'\beta + \epsilon_{it} \end{aligned} \tag{5}$$

- Unobserved A_i is absorbed in dummy α_i
- Just run regression (5) and get ρ
- 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
- Substracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \tag{6}$$

- 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
- Substracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}' \beta + \Delta \epsilon_{it} \tag{6}$$

- 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
- Substracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}^{l} \beta + \Delta \epsilon_{it} \tag{6}$$

- 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
- Substracting regression in t by $t - 1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}^{\prime} \beta + \Delta \epsilon_{it} \tag{6}$$

- 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
- Substracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}^{\prime} \beta + \Delta \epsilon_{it} \tag{6}$$

- 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
- Substracting regression in t by $t-1$, we have:

$$\Delta Y_{it} = \rho \Delta D_{it} + \Delta X_{it}^{\prime} \beta + \Delta \epsilon_{it} \tag{6}$$

- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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 $T > 2$
- 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

DID: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}'\beta + \lambda_t + \alpha_i + \epsilon_{it} \tag{7}$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{\prime}\beta + \lambda_t + \alpha_i + \epsilon_{it} \tag{7}$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{\prime}\beta + \lambda_t + \alpha_i + \epsilon_{it} \tag{7}$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{\prime}\beta + \lambda_t + \alpha_i + \epsilon_{it} \tag{7}$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{I}\beta + \lambda_t + \alpha_i + \epsilon_{it} \quad (7)$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{\prime}\beta + \lambda_t + \alpha_i + \epsilon_{it} \quad (7)$$

- 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: TWFE

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

$$Y_{it} = \rho D_{it} + X_{it}^{\prime}\beta + \lambda_t + \alpha_i + \epsilon_{it} \tag{7}$$

- 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

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

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

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

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

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, $D_{st} = NJ_s d_t$, 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, $D_{st} = NJ_s d_t$, 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, $D_{st} = NJ_s d_t$, 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, $D_{st} = NJ_s d_t$, 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, $D_{st} = NJ_s d_t$, 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

- 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

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t \tag{8}$$

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

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

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t \tag{8}$$

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

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

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t \tag{8}$$

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

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

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t \quad (8)$$

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

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

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t \tag{8}$$

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

DID: Identification

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

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \epsilon_{ist} \tag{9}$$

- First difference: For same state, dif across time

$$E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb} \tag{10}$$
$$E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb] = \lambda_{Nov} - \lambda_{Feb} + \delta \tag{11}$$

- Second difference: Difference in trends across states

$$(11) - (10) = \delta \tag{12}$$

DID: Identification

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

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \epsilon_{ist} \tag{9}$$

- First difference: For same state, dif across time

$$E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb} \tag{10}$$
$$E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb] = \lambda_{Nov} - \lambda_{Feb} + \delta \tag{11}$$

- Second difference: Difference in trends across states

$$(11) - (10) = \delta \tag{12}$$

DID: Identification

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

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \epsilon_{ist} \tag{9}$$

- First difference: For same state, dif across time

$$E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb} \tag{10}$$
$$E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb] = \lambda_{Nov} - \lambda_{Feb} + \delta \tag{11}$$

- Second difference: Difference in trends across states

$$(11) - (10) = \delta \tag{12}$$

DID: Identification

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

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \epsilon_{ist} \tag{9}$$

- First difference: For same state, dif across time

$$E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb} \tag{10}$$
$$E[Y_{ist}|s = NJ, t = Nov] - E[Y_{ist}|s = NJ, t = Feb] = \lambda_{Nov} - \lambda_{Feb} + \delta \tag{11}$$

- Second difference: Difference in trends across states

$$(11) - (10) = \delta \tag{12}$$

DID: Identification

We are taking untreated group as the control!

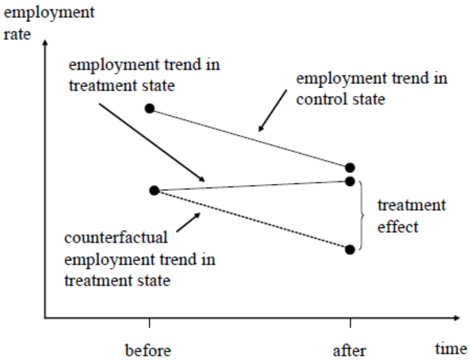


Figure 5.2.1: Causal effects in the differences-in-differences model

DID: Test of Parallel Trend

- After the implementation of the policy at t_0, we can no longer observe Y_{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

- 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)

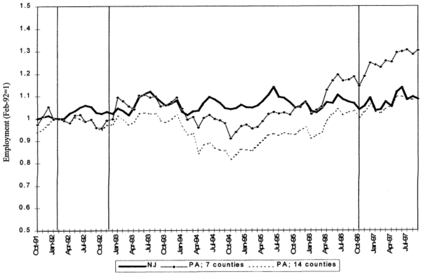


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

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)

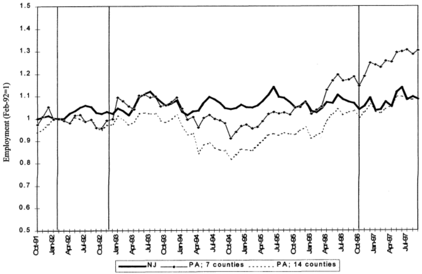


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

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)

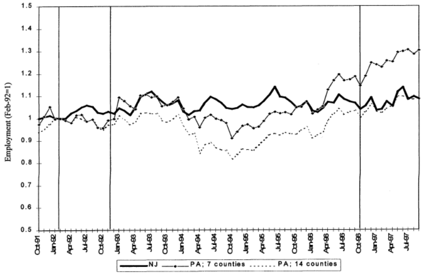


Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

DID: Test of Parallel Trend

- What about this?

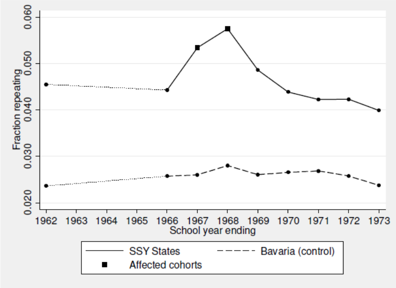


Figure 5.2.3: Average rates of grade repetition in second grade for treatment and control schools in Germany (from Pischke 2007). The data span a period before and after a change in term length for students outside of Bavaria.

DID: Test of Parallel Trend

- What about this?

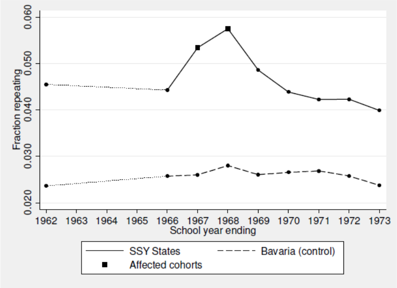


Figure 5.2.3: Average rates of grade repetition in second grade for treatment and control schools in Germany (from Pischke 2007). The data span a period before and after a change in term length for students outside of Bavaria.

DID: Test of Parallel Trend

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

- δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

- δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

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

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

DID: Test of Parallel Trend

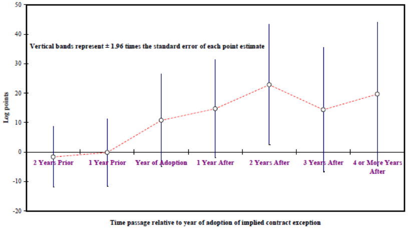
2. Event Study Regression
   - If we have data from -T to T', and the policy D_{it} is implemented at $t = 0$
   - Let D_s be the dummy of whether in the treated group
   - Run the following regression

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau \neq -1} 1(t = \tau)\delta_\tau D_s + \epsilon_{ist} \tag{13}$$

   - δ_τ 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 δ_τ

DID: Test of Parallel Trend

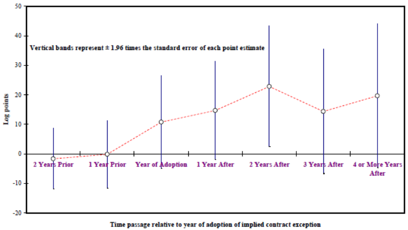
- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of δ for each time period and have:



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

DID: Test of Parallel Trend

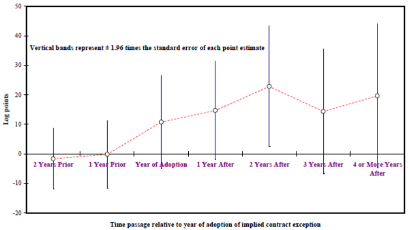
- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of δ for each time period and have:



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

DID: Test of Parallel Trend

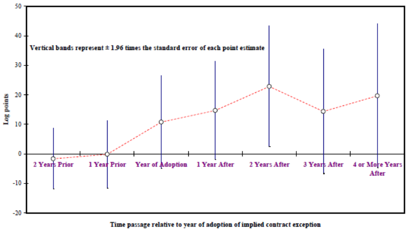
- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of δ for each time period and have:



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

DID: Test of Parallel Trend

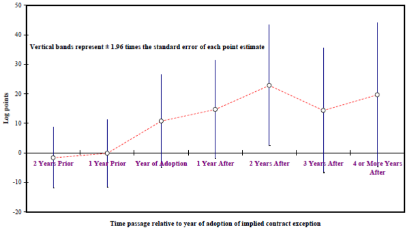
- Usually $t = -1$ (just before the policy) is omitted as the baseline
- Then we draw the changes of δ for each time period and have:



- Points before $t = 0$ are not significant $\Rightarrow$ 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 δ for each time period and have:



- Points before $t = 0$ are not significant $\Rightarrow$ 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 $\Leftarrow$ 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 $\Leftarrow$ 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 $\Leftarrow$ 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 $\Longleftarrow$ 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 $\Leftarrow$ 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 i from family j at time t:

$$wage_{ijt} = \beta_0 + \beta_1 exp_{ijt} + \epsilon_{ijt} \tag{14}$$

Identification Source

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

$$wage_{ijt} = \beta_0 + \beta_1 exp_{ijt} + \epsilon_{ijt} \tag{14}$$

Identification Source

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

$$wage_{ijt} = \beta_0 + \beta_1 exp_{ijt} + \epsilon_{ijt} \tag{14}$$

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 ($i \times t$ 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 ($i \times t$ 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 ($i \times t$ 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 ($i \times t$ 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 pre-trend testing
  - Synthetic Control Method: When you do not have parallel trend
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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
  - Triple differences (DDD)
  - 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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

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: $\delta_{1t} - \delta_{0t} = \gamma t$
   - 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!
   - Why? It goes back to the nature of the statistical test

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1-\beta$
- Power means the power to reject a false H0

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1 - \beta$
- Power means the power to reject a false H0

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1 - \beta$
- Power means the power to reject a false H0

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1 - \beta$
- Power means the power to reject a false H0

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1 - \beta$
- Power means the power to reject a false H0

Type I and Type II Error

Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

- Type I error: H0 is true but we reject it, α
- Type II error: H0 is false but we do not reject it, β
- Significance level: Probability of commiting Type I error, α
- Power: Probability of rejecting H0 if it is false, $1 - \beta$
- Power means the power to reject a false H0

Type I and Type II Error

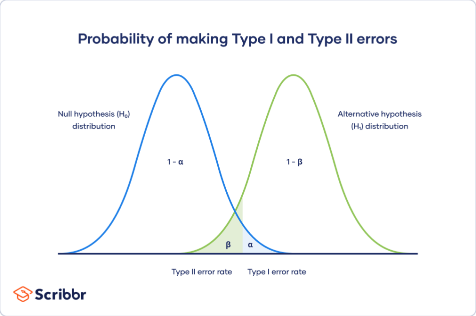
Null hypothesis is ...	True	False
Rejected	Type I error False positive Probability = α	Correct decision True positive Probability = 1 - β
Not rejected	Correct decision True negative Probability = 1 - α	Type II error False negative Probability = β

Scribbr

Pre-trend Testing: New Development

Tradeoff!!!

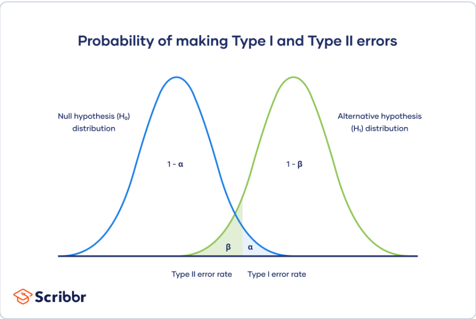
- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger α; Go right, you have larger β



Pre-trend Testing: New Development

Tradeoff!!!

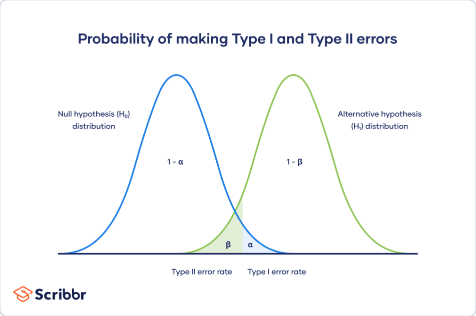
- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger α; Go right, you have larger β



Pre-trend Testing: New Development

Tradeoff!!!

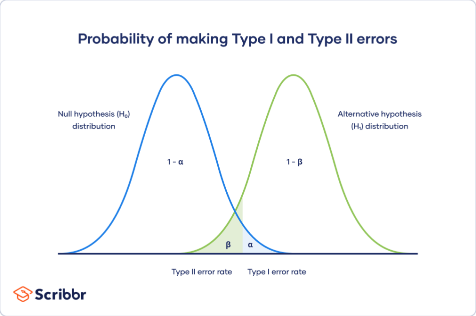
- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger α; Go right, you have larger β



Pre-trend Testing: New Development

Tradeoff!!!

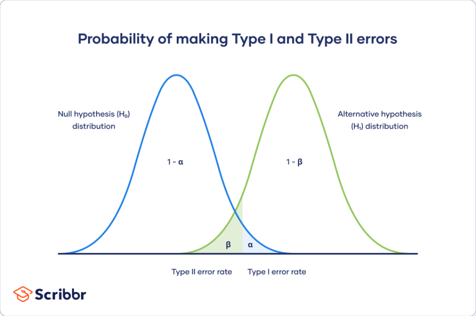
- Now you have to choose a threshold critical value to make your rejection decision
- Go left, you have larger α; Go right, you have larger β



Pre-trend Testing: New Development

Tradeoff!!!

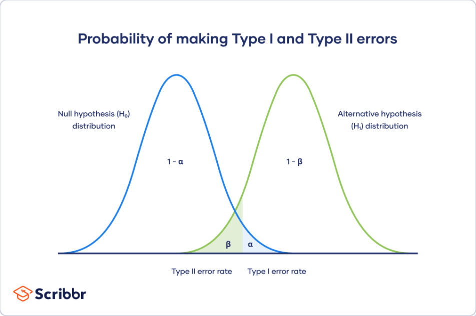
- You can decrease T1ER by decreasing $\alpha \implies$ increasing β
- You can decrease T2ER by decreasing $\beta \implies$ increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative



Pre-trend Testing: New Development

Tradeoff!!!

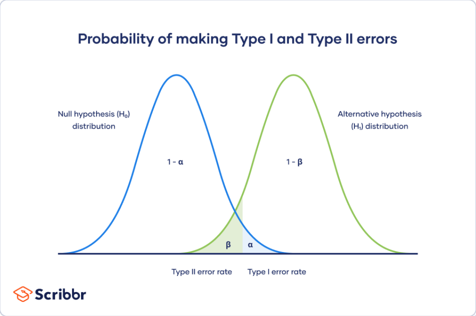
- You can decrease T1ER by decreasing $\alpha \implies$ increasing β
- You can decrease T2ER by decreasing $\beta \implies$ increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative



Pre-trend Testing: New Development

Tradeoff!!!

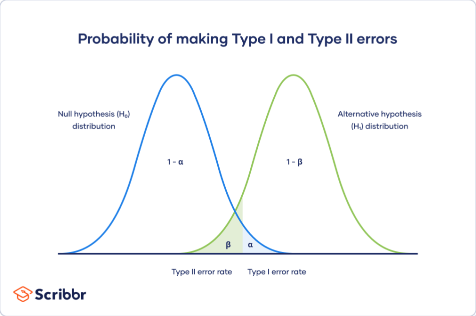
- You can decrease T1ER by decreasing $\alpha \Rightarrow$ increasing β
- You can decrease T2ER by decreasing $\beta \Rightarrow$ increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative



Pre-trend Testing: New Development

Tradeoff!!!

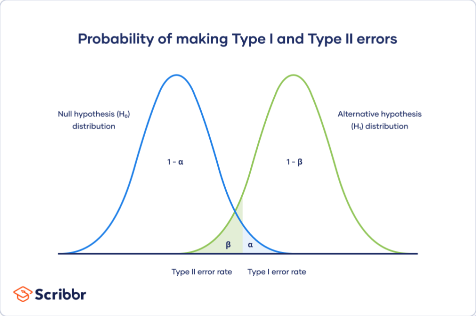
- You can decrease T1ER by decreasing $\alpha \Rightarrow$ increasing β
- You can decrease T2ER by decreasing $\beta \Rightarrow$ increasing α
- If you want H_0 to be rejected less easily, you have to tolerate large probability to have false negative



Pre-trend Testing: New Development

Tradeoff!!!

- You can decrease T1ER by decreasing $\alpha \Rightarrow$ increasing β
- You can decrease T2ER by decreasing $\beta \Rightarrow$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\implies$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\Longrightarrow$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\Longrightarrow$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\Longrightarrow$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\implies$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\Rightarrow$ 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 H0
- Minimize Type I error probability α to be smaller than some level (10%, 5%, 1%)
- It then leads to large β! $\Longrightarrow$ 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

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

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! $\Longrightarrow$ 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! $\Rightarrow$ 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! $\Longrightarrow$ 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! $\Longrightarrow$ 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! $\Longrightarrow$ 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! $\Longrightarrow$ 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 unaffected units often provides a more appropriate comparison than any single unaffected 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 unaffected units often provides a more appropriate comparison than any single unaffected 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 unaffected units often provides a more appropriate comparison than any single unaffected 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

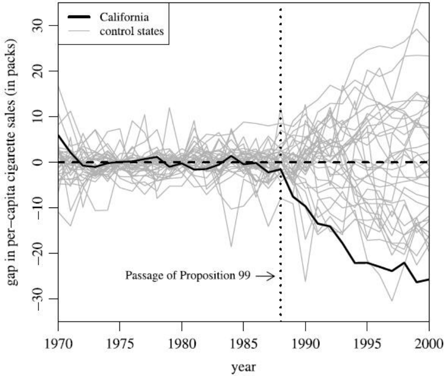


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

Synthetic Control: Main Idea

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

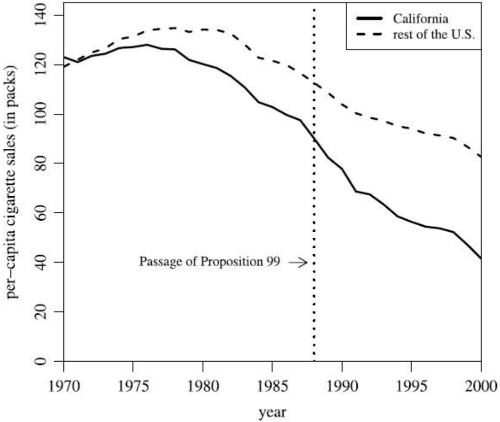


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Synthetic Control: Main Idea

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

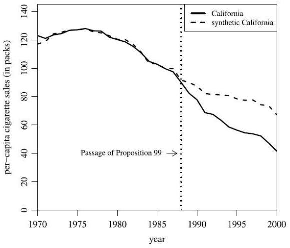


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

(a)

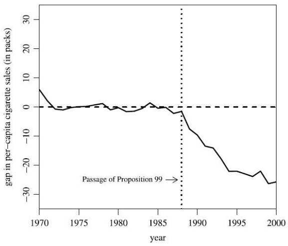


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

(b)

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, \ldots, J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, \ldots, J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Suppose we have $j = 1, 2, ..., J + 1$ units (provinces, cities...), spanning T periods
- T_0 is the treatment starting period, $j = 1$ is the treated unit
- We call $j = 2, 3, ..., J + 1$ 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 Y
- X_{kj} are unaffected by treatment

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^I, Y_{jt}^N
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- Define potential outcome: Y_{jt}^{I}, Y_{jt}^{N}
- Treatment effect of interest: $\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$ for $t > T_0$
- 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:

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

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

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

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

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

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

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

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

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

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

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

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

$$||X_1 - X_0 W|| = (\sum_{h=1}^{k} v_h (X_{h1} - w_2 X_{h2} - ... - w_{J+1} X_{hJ+1})^2)^{1/2} \tag{15}$$

- 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 v and weights w

Synthetic Control: Settings

- How to define the weights?
- Assume that we have $X = (X_1, X_2, ..., X_h, ..., X_k)$
- We minimize the following:

$$||X_1 - X_0 W|| = \left(\sum_{h=1}^{k} v_h (X_{h1} - w_2 X_{h2} - ... - w_{J+1} X_{hJ+1})^2\right)^{1/2} \tag{15}$$

- 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 v 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 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 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 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

TABLE 1
ECONOMIC GROWTH PREDICTOR MEANS BEFORE THE GERMAN REUNIFICATION

	West Germany (1)	Synthetic West Germany (2)	OECD average (3)	Austria (nearest neighbor) (4)
GDP per capita	15,808.9	15,802.2	13,669.4	14,817.0
Trade openness	56.8	56.9	59.8	74.6
Inflation rate	2.6	3.5	7.6	3.5
Industry share	34.5	34.4	33.8	35.5
Schooling	55.5	55.2	38.7	60.9
Investment rate	27.0	27.0	25.9	26.6

Synthetic Control: An Example

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

TABLE 2
SYNTHETIC CONTROL WEIGHTS FOR WEST GERMANY

Australia	—
Austria	0.42
Belgium	—
Denmark	—
France	—
Greece	—
Italy	—
Japan	0.16
Netherlands	0.09
New Zealand	—
Norway	—
Portugal	—
Spain	—
Switzerland	0.11
United Kingdom	—
United States	0.22

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \tilde{w}_2(V)Y_{2t} - ... - \tilde{w}_{J+1}(V)Y_{J+1t}\right)^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \tilde{w}_2(V)Y_{2t} - \ldots - \tilde{w}_{J+1}(V)Y_{J+1t}\right)^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} (Y_{1t} - \tilde{w}_2(V)Y_{2t} - ... - \tilde{w}_{J+1}(V)Y_{J+1t})^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} \left(Y_{1t} - \tilde{w}_2(V)Y_{2t} - \ldots - \tilde{w}_{J+1}(V)Y_{J+1t}\right)^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} (Y_{1t} - \tilde{w}_2(V)Y_{2t} - \ldots - \tilde{w}_{J+1}(V)Y_{J+1t})^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} (Y_{1t} - \tilde{w}_2(V)Y_{2t} - ... - \tilde{w}_{J+1}(V)Y_{J+1t})^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: Procedure

How to determine v?

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

$$\sum_{t=t_0+1}^{T_0} (Y_{1t} - \tilde{w}_2(V)Y_{2t} - ... - \tilde{w}_{J+1}(V)Y_{J+1t})^2$$

- Step 4: Using V^* and validation data to calculate W by $\tilde{w}(v)$
- V becomes a tuning parameter

Synthetic Control: An Example

- Synthetic West Germany

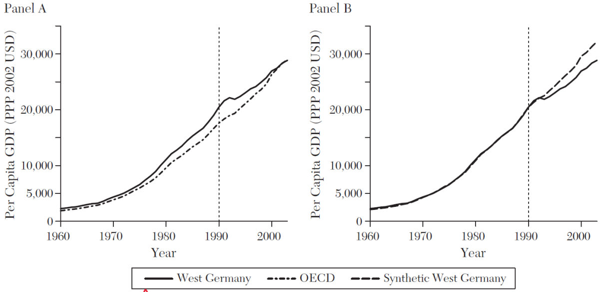


Figure 1. Synthetic Control Estimation in the German Reunification Example

Notes: Panel A compares the evolution of per capita GDP in West Germany to the evolution of per capita GDP for a simple average of OECD countries. In panel B the comparison is with a synthetic control calculated in the manner explained in subsection 3.2. See Abadie, Diamond, and Hainmueller (2015) for details.

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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 in X! (but pre-treatment outcomes can)
- There can be bias if there are unobserved endogenous characteristics that cannot be matched
- When the difference $X_1 - X_0 W^*$ 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 $[0, 1]$

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD)

- Another method to solve the unparallel trend is triple differences (DDD)
- Assume that in Card and Krueger (1994), PA and NJ can have different time trends in the average wage due to some common shock
- However, minimum wage only affects cashier at McDonald's, but not teachers
- The wage gap between cashiers and teachers have the same trend in PA and NJ
- So the only treated group is cashier in NJ
- The control group includes cashiers and teachers in PA, and teachers in NJ

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \quad (16)$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

Triple Differences (DDD): 3WFE

- Denote i as individual, g as occupation group ($g = 1$ if cashier), s as state ($s = 1$ if NJ), t as time ($t = 1$ if Nov)
- We can extend the TWFE to have

$$Y_{it} = \rho D_{it} + \lambda_{g,t} + \delta_{s,t} + \alpha_i + \epsilon_{it} \tag{16}$$

- This is called Three-way Fixed Effect Model (3WFE)
- We control for i-level (thus, g, s-level), g, t-level, and s, t-level FE
- $D_{it} = 1$ only if $t = 1, g = 1, s = 1$
- It can be easily understand as a difference between two DIDs

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t}, \delta_{s,t}, \alpha_i, D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s=0, g=0, t=1] - E[Y_{ist}|s=0, g=0, t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$
$$E[Y_{ist}|s=0, g=1, t=1] - E[Y_{ist}|s=0, g=1, t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$
$$E[Y_{ist}|s=1, g=0, t=1] - E[Y_{ist}|s=1, g=0, t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$
$$E[Y_{ist}|s=1, g=1, t=1] - E[Y_{ist}|s=1, g=1, t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$
$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t}, \delta_{s,t}, \alpha_i, D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s=0,g=0,t=1] - E[Y_{ist}|s=0,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$
$$E[Y_{ist}|s=0,g=1,t=1] - E[Y_{ist}|s=0,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$
$$E[Y_{ist}|s=1,g=0,t=1] - E[Y_{ist}|s=1,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$
$$E[Y_{ist}|s=1,g=1,t=1] - E[Y_{ist}|s=1,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$
$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t}, \delta_{s,t}, \alpha_i, D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s = 0, g = 0, t = 1] - E[Y_{ist}|s = 0, g = 0, t = 0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$

$$E[Y_{ist}|s = 0, g = 1, t = 1] - E[Y_{ist}|s = 0, g = 1, t = 0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$

$$E[Y_{ist}|s = 1, g = 0, t = 1] - E[Y_{ist}|s = 1, g = 0, t = 0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$

$$E[Y_{ist}|s = 1, g = 1, t = 1] - E[Y_{ist}|s = 1, g = 1, t = 0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$

$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t}, \delta_{s,t}, \alpha_i, D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s=0,g=0,t=1] - E[Y_{ist}|s=0,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$

$$E[Y_{ist}|s=0,g=1,t=1] - E[Y_{ist}|s=0,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$

$$E[Y_{ist}|s=1,g=0,t=1] - E[Y_{ist}|s=1,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$

$$E[Y_{ist}|s=1,g=1,t=1] - E[Y_{ist}|s=1,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$

$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t}, \delta_{s,t}, \alpha_i, D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s=0, g=0, t=1] - E[Y_{ist}|s=0, g=0, t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$
$$E[Y_{ist}|s=0, g=1, t=1] - E[Y_{ist}|s=0, g=1, t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$
$$E[Y_{ist}|s=1, g=0, t=1] - E[Y_{ist}|s=1, g=0, t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$
$$E[Y_{ist}|s=1, g=1, t=1] - E[Y_{ist}|s=1, g=1, t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$
$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

DDD: Identification

- Here, the extended parallel trend assumption is that $E[\epsilon_{it}|\lambda_{g,t},\delta_{s,t},\alpha_i,D_{it}] = 0$
- With the extended parallel trend assumption, we can identify the policy effect δ:
- First difference: For same state, same occupation, dif across time (α_i is canceled)

$$E[Y_{ist}|s=0,g=0,t=1] - E[Y_{ist}|s=0,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{0,1} - \delta_{0,0} \quad (17)$$
$$E[Y_{ist}|s=0,g=1,t=1] - E[Y_{ist}|s=0,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{0,1} - \delta_{0,0} \quad (18)$$
$$E[Y_{ist}|s=1,g=0,t=1] - E[Y_{ist}|s=1,g=0,t=0] = \lambda_{0,1} - \lambda_{0,0} + \delta_{1,1} - \delta_{1,0} \quad (19)$$
$$E[Y_{ist}|s=1,g=1,t=1] - E[Y_{ist}|s=1,g=1,t=0] = \lambda_{1,1} - \lambda_{1,0} + \delta_{1,1} - \delta_{1,0} + \rho \quad (20)$$

- Second difference: For same state, dif in trends across occupation

$$(18) - (17) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) \quad (21)$$
$$(20) - (19) = \lambda_{1,1} - \lambda_{1,0} - (\lambda_{0,1} - \lambda_{0,0}) + \rho \quad (22)$$

- Third difference: Difference across state

$$(22) - (21) = \rho \quad (23)$$

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 figure
  - Run event study
- However, statistical test minimizes T1ER, which inflates 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 inflates 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 inflates 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 inflates 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 inflates 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 inflates 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 inflates 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 inflates 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 characteristics
  - Weights for each donor unit
- Then you create a control group to mimic the behavior of the treated group
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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
- Or you can use DDD if there is a further layer of difference which gives you the exogeneity

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.