Module 9: Panel Data

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

Where are we? Where are we going?

- Where we have found good controls:

Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control

Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates

Where are we? Where are we going?

- Where we have found good controls:
    - Units randomized to receive control
    - Units with similar values of covariates
    - Units with opposite value of some instrument

Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates
  - Units with opposite value of some instrument
  - At a discontinuity in treatment assignment

Where are we? Where are we going?

- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates
  - Units with opposite value of some instrument
  - At a discontinuity in treatment assignment
- What if we have repeated measurements of the same units?

Where are we? Where are we going?

- Where we have found good controls:
    - Units randomized to receive control
    - Units with similar values of covariates
    - Units with opposite value of some instrument
    - At a discontinuity in treatment assignment
- What if we have repeated measurements of the same units?
- Now there are two possible sources of variation to exploit:

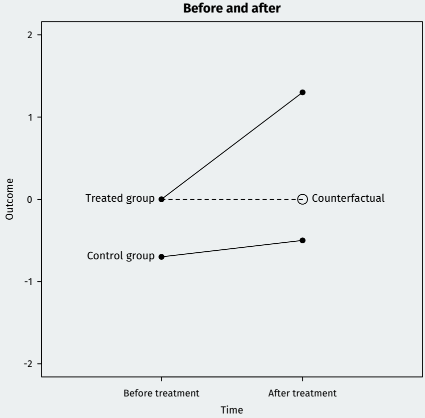
Where are we? Where are we going?

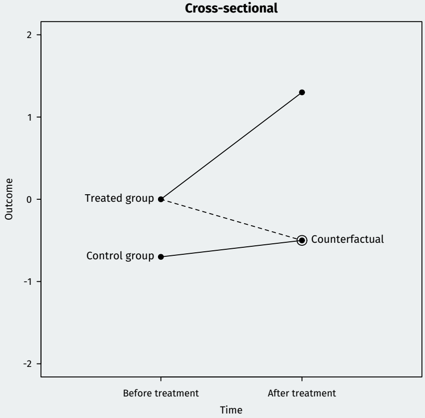
- Where we have found good controls:
  - Units randomized to receive control
  - Units with similar values of covariates
  - Units with opposite value of some instrument
  - At a discontinuity in treatment assignment
- What if we have repeated measurements of the same units?
- Now there are two possible sources of variation to exploit:
  - Exploit **cross-sectional** variation in treatment.

Where are we? Where are we going?

- Where we have found good controls:
    - Units randomized to receive control
    - Units with similar values of covariates
    - Units with opposite value of some instrument
    - At a discontinuity in treatment assignment
- What if we have repeated measurements of the same units?
- Now there are two possible sources of variation to exploit:
    - Exploit **cross-sectional** variation in treatment.
    - Exploit variation in treatment **within a unit over time** (before/after)

Cross-sectional vs before/after





1/ Difference in differences

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from \$4.25 to \$5.05
    - Neighboring PA stays at \$4.25

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
  - Neighboring PA stays at $4.25
  - We observe employment in both states before and after increase

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
    - Neighboring PA stays at $4.25
    - We observe employment in both states before and after increase
- Look at eastern PA and NJ fast food restaurants.

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
    - Neighboring PA stays at $4.25
    - We observe employment in both states before and after increase
- Look at eastern PA and NJ fast food restaurants.
    - Similar prices, wages, products, etc.

Minimum wages (Card & Krueger, 1994)

- Does increasing the minimum wage affect employment?
- Classical economic theory tends to point to negative effects.
- But difficult to randomize changes to the minimum wage.
- In 1992, NJ minimum wage increased from $4.25 to $5.05
    - Neighboring PA stays at $4.25
    - We observe employment in both states before and after increase
- Look at eastern PA and NJ fast food restaurants.
    - Similar prices, wages, products, etc.
    - Most likely to be affected by the change.

Differences-in-differences design

- Basic setup: two groups, two time periods.

Differences-in-differences design

- Basic setup: two groups, two time periods.
    - Pre-period ($t = 0$): neither group is treated.

Differences-in-differences design

- Basic setup: two groups, two time periods.
    - Pre-period ($t = 0$): neither group is treated.
    - Post-period ($t = 1$): one group is treated, other remains untreated.

Differences-in-differences design

- Basic setup: two groups, two time periods.
    - Pre-period ($t = 0$): neither group is treated.
    - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$
  - $G_i = 0$ for those that are always untreated

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$
  - $G_i = 0$ for those that are always untreated
- Treatment status in each period:

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$
  - $G_i = 0$ for those that are always untreated
- Treatment status in each period:
  - No treatment in the first period for either group: $D_{i0} = 0$

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$
  - $G_i = 0$ for those that are always untreated
- Treatment status in each period:
  - No treatment in the first period for either group: $D_{i0} = 0$
  - In treated group, $G_i = 1 \rightsquigarrow D_{i1} = 1$

Differences-in-differences design

- Basic setup: two groups, two time periods.
    - Pre-period ($t = 0$): neither group is treated.
    - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
    - $G_i = 1$ are those that are treated at $t = 1$
    - $G_i = 0$ for those that are always untreated
- Treatment status in each period:
    - No treatment in the first period for either group: $D_{i0} = 0$
    - In treated group, $G_i = 1 \rightsquigarrow D_{i1} = 1$
    - In control group, $G_i = 0 \rightsquigarrow D_{i1} = 0$

Differences-in-differences design

- Basic setup: two groups, two time periods.
  - Pre-period ($t = 0$): neither group is treated.
  - Post-period ($t = 1$): one group is treated, other remains untreated.
- Groups defined by treatment status in post-period:
  - $G_i = 1$ are those that are treated at $t = 1$
  - $G_i = 0$ for those that are always untreated
- Treatment status in each period:
  - No treatment in the first period for either group: $D_{i0} = 0$
  - In treated group, $G_i = 1 \rightsquigarrow D_{i1} = 1$
  - In control group, $G_i = 0 \rightsquigarrow D_{i1} = 0$

	Time period	
	Pre-period $(t = 0)$	Post-period $(t = 1)$
Control group $(G_i = 0)$	$D_{i0} = 0$	$D_{i1} = 0$
Treated group $(G_i = 1)$	$D_{i0} = 0$	$D_{i1} = 1$

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- **Consistency**: $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- **Consistency**: $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$
    - Observe control p.o. for all units in first period: $Y_{i0}(0) = Y_{i0}$

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- **Consistency**: $Y_{it} = D_{it} Y_{it}(1) + (1 - D_{it}) Y_{it}(0)$
    - Observe control p.o. for all units in first period: $Y_{i0}(0) = Y_{i0}$
    - In treated group: $G_i = 1 \rightsquigarrow Y_{i1} = Y_{i1}(1)$

Potential outcomes approach to DID

- $Y_{it}(d)$ is the potential outcome under treatment d at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- **Consistency**: $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$
    - Observe control p.o. for all units in first period: $Y_{i0}(0) = Y_{i0}$
    - In treated group: $G_i = 1 \rightsquigarrow Y_{i1} = Y_{i1}(1)$
    - In control group: $G_i = 0 \rightsquigarrow Y_{i1} = Y_{i1}(0)$

Identification problem

- Average treatment effect on the treated:

Identification problem

- Average treatment effect on the treated:

$$\tau_{ATT} = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1]$$

Identification problem

- Average treatment effect on the treated:

$$\begin{aligned}
\tau_{ATT} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1] \\
&= \mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 1]
\end{aligned}$$

Identification problem

- Average treatment effect on the treated:

$$\begin{aligned}\tau_{ATT} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] \\ &= \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1] \\ &= \underbrace{\mathbb{E}[Y_{i1}|G_i = 1]}_{\text{(a)}} - \underbrace{\mathbb{E}[Y_{i1}(0)|G_i = 1]}_{\text{(b)}}\end{aligned}$$

Identification problem

- Average treatment effect on the treated:

$$\begin{aligned}\tau_{ATT} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1] \\ &= \mathbb{E}[Y_{i1}(1) | G_i = 1] - \mathbb{E}[Y_{i1}(0) | G_i = 1] \\ &= \underbrace{\mathbb{E}[Y_{i1} | G_i = 1]}_{\text{(a)}} - \underbrace{\mathbb{E}[Y_{i1}(0) | G_i = 1]}_{\text{(b)}}\end{aligned}$$

- Part (a) is just a conditional average of observed data $\rightsquigarrow$ identified.

Identification problem

- Average treatment effect on the treated:

$$\begin{aligned}\tau_{ATT} &= \mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] \\ &= \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1] \\ &= \underbrace{\mathbb{E}[Y_{i1}|G_i = 1]}_{\text{(a)}} - \underbrace{\mathbb{E}[Y_{i1}(0)|G_i = 1]}_{\text{(b)}}\end{aligned}$$

- Part (a) is just a conditional average of observed data $\rightsquigarrow$ identified.
- Part (b) is a counterfactual: what would the average outcome in the treated group have been if it have been in control?

Three control strategies

$$\tau_{ATT} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period $(t = 0)$	Post-period $(t = 1)$
Control group $(G_i = 0)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group $(G_i = 1)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

Three control strategies

$$\tau_{ATT} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period $(t = 0)$	Post-period $(t = 1)$
Control group $(G_i = 0)$	$\mathbb{E}[Y_{i0}(0)|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)|G_i = 0]$
Treated group $(G_i = 1)$	$\mathbb{E}[Y_{i0}(0)|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)|G_i = 1]$

- **Cross-sectional design**
  - Assumption: mean independence of treatment

$$\mathbb{E}[Y_{i1}(0)|G_i = 1] = \mathbb{E}[Y_{i1}(0)|G_i = 0]$$

  - Use post-treatment control group:

$$\tau_{ATT} = \mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i1}|G_i = 0]$$

Three control strategies

$$\tau_{ATT} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period $(t = 0)$	Post-period $(t = 1)$
Control group $(G_i = 0)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group $(G_i = 1)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **Before-and-after design**
    - Assumption: no trends

$$\mathbb{E}[Y_{i1}(0)|G_i = 1] = \mathbb{E}[Y_{i0}(0)|G_i = 1]$$

- Use pre-period outcome in treated group:

$$\tau_{ATT} = \mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]$$

Three control strategies

$$\tau_{ATT} = \mathbb{E}[Y_{i1}(1)|G_i = 1] - \mathbb{E}[Y_{i1}(0)|G_i = 1]$$

	Time period	
	Pre-period $(t = 0)$	Post-period $(t = 1)$
Control group $(G_i = 0)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 0]$	$\mathbb{E}[Y_{i1}(0)\|G_i = 0]$
Treated group $(G_i = 1)$	$\mathbb{E}[Y_{i0}(0)\|G_i = 1]$	$\mathbb{E}[Y_{i1}(1)\|G_i = 1]$

- **Difference-in-differences**:
    - Assumption: parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

    - Use pre-period treated outcome plus trend in control group:

$$\begin{aligned}\tau_{ATT} =& (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ & - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])\end{aligned}$$

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
    - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
    - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.
- Why is this weaker than other assumption?

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
  - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.
- Why is this weaker than other assumption?
  - Allows for time-constant unmeasured confounding between Y_{it} and G_i

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0)|G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
    - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.
- Why is this weaker than other assumption?
    - Allows for time-constant unmeasured confounding between Y_{it} and G_i
    - Allows for (common) secular trends in the outcome over time (unlike FE).

Parallel trends

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1]$$

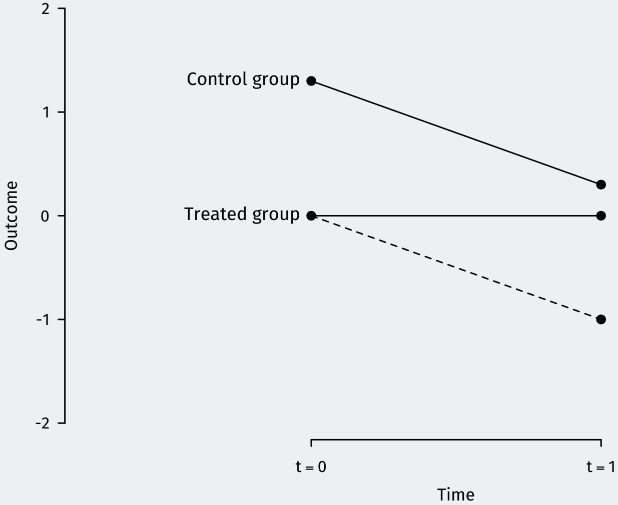
- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
    - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.
- Why is this weaker than other assumption?
    - Allows for time-constant unmeasured confounding between Y_{it} and G_i
    - Allows for (common) secular trends in the outcome over time (unlike FE).
- Not invariant to nonlinear transformations!

Parallel trends

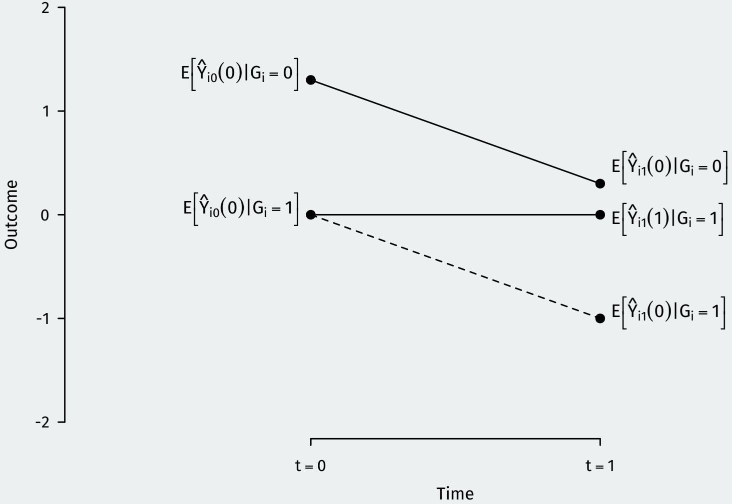
$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1]$$

- Key assumption of differences-in-differences: **parallel trends**
- Interpretation:
  - Secular trend in the control group is a good proxy how the treated group would have changed over time without treatment.
- Why is this weaker than other assumption?
  - Allows for time-constant unmeasured confounding between Y_{it} and G_i
  - Allows for (common) secular trends in the outcome over time (unlike FE).
- Not invariant to nonlinear transformations!
  - Parallel trends for Y_{it} implies non-parallel trends for $\log(Y_{it})$ and vice versa.

Parallel trends in a graph



Parallel trends in a graph



Identification

- Identification result:

$$\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

Identification

- Identification result:

$$\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

- Threat to identification: non-parallel trends

Identification

- Identification result:

$$\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1])$$
$$- (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

- Threat to identification: non-parallel trends
    - **unmeasured time-varying confounding**

Identification

- Identification result:

$$\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

- Threat to identification: non-parallel trends
    - **unmeasured time-varying confounding**
    - **Ashenfelter's dip**: empirical finding that people who enroll in job training programs see their earnings decline prior to that training.

Identification

- Identification result:

$$\tau_{ATT} = (\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]) \\ - (\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0])$$

- Threat to identification: non-parallel trends
  - **unmeasured time-varying confounding**
  - **Ashenfelter's dip**: empirical finding that people who enroll in job training programs see their earnings decline prior to that training.
- Falsification test: check pre-treatment parallel trends.

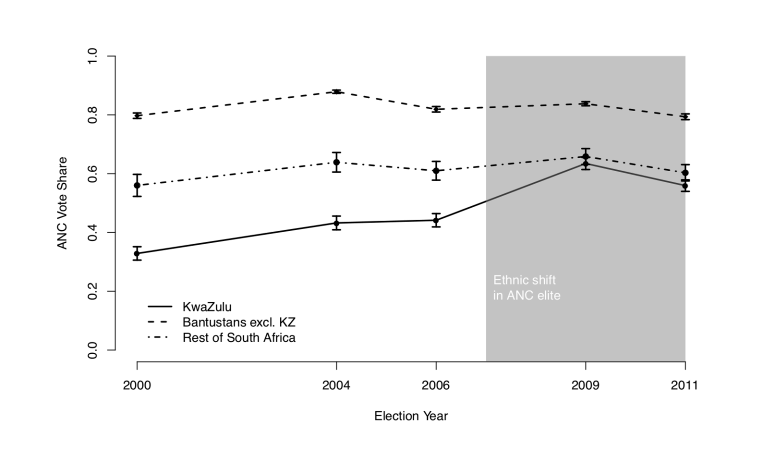
Identification

- Identification result:

$$\begin{aligned}\tau_{ATT} = & \left(\mathbb{E}[Y_{i1}|G_i = 1] - \mathbb{E}[Y_{i0}|G_i = 1]\right) \\ & - \left(\mathbb{E}[Y_{i1}|G_i = 0] - \mathbb{E}[Y_{i0}|G_i = 0]\right)\end{aligned}$$

- Threat to identification: non-parallel trends
  - **unmeasured time-varying confounding**
  - **Ashenfelter's dip**: empirical finding that people who enroll in job training programs see their earnings decline prior to that training.
- Falsification test: check pre-treatment parallel trends.
  - Doesn't imply parallel trends hold for the post-period however!

Checking parallel trends (de Kadt/Larreguy, 2018)



Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\text{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\text{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

  - Standard errors from standard difference in means.

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\text{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

  - Standard errors from standard difference in means.

- Regression implementation:

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\mathrm{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

  - Standard errors from standard difference in means.
- Regression implementation:
  - Regress $\Delta Y_i = Y_{i1} - Y_{i0}$ on G_i.

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\mathrm{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

  - Standard errors from standard difference in means.
- Regression implementation:
  - Regress $\Delta Y_i = Y_{i1} - Y_{i0}$ on G_i.
  - Use (cluster) robust SEs

Estimation

- Estimation with panel data:

$$\widehat{\tau}_{\mathrm{ATT}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i \left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in treated group}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i)\left\{Y_{i1} - Y_{i0}\right\}}_{\text{average trend in the control group}}$$

  - Standard errors from standard difference in means.
- Regression implementation:
  - Regress $\Delta Y_i = Y_{i1} - Y_{i0}$ on G_i.
  - Use (cluster) robust SEs
- Also possible to use DID on repeated cross sections.

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - Be sure to cluster by unit (or level of treatment assignment)

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - Be sure to cluster by unit (or level of treatment assignment)
- Coefficient on D_{it} equivalent to DID estimation.

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - Be sure to cluster by unit (or level of treatment assignment)

- Coefficient on D_{it} equivalent to DID estimation.

- Only holds for the 2 group, 2 period case!

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - Be sure to cluster by unit (or level of treatment assignment)
- Coefficient on D_{it} equivalent to DID estimation.
- Only holds for the 2 group, 2 period case!
  - Large new literature on interpretation of TWFE in more general cases.

DID and linear two-way fixed effects

- Linear two-way (group and time) fixed effect model:

$$Y_{it} = \alpha + \gamma G_i + \beta t + \tau D_{it} + \varepsilon_{it}$$

  - Fixed effect for group and time.
  - Be sure to cluster by unit (or level of treatment assignment)
- Coefficient on D_{it} equivalent to DID estimation.
- Only holds for the 2 group, 2 period case!
  - Large new literature on interpretation of TWFE in more general cases.
  - Basically, TWFE is an odd weighted average of DID effects with sometimes negative weights.

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.
  - Benefit over parallel trends: it is scale-free.

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.
  - Benefit over parallel trends: it is scale-free.
  - Equivalent to parallel trends if $\mathbb{E}[Y_{i0} \mid G_i = 1] = \mathbb{E}[Y_{i0} \mid G_i = 0]$

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.
  - Benefit over parallel trends: it is scale-free.
  - Equivalent to parallel trends if $\mathbb{E}[Y_{i0} \mid G_i = 1] = \mathbb{E}[Y_{i0} \mid G_i = 0]$
- Different ideas about why there is imbalance on the LDV:

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.
  - Benefit over parallel trends: it is scale-free.
  - Equivalent to parallel trends if $\mathbb{E}[Y_{i0} \mid G_i = 1] = \mathbb{E}[Y_{i0} \mid G_i = 0]$
- Different ideas about why there is imbalance on the LDV:
  - DID: time-constant unmeasured confounder creates imbalance.

DID vs lagged dependent variable

- Alternative identification assumption:

$$Y_{i1}(0) \perp\!\!\!\perp G_i \mid Y_{i0}$$

  - Doesn't imply and isn't implied by parallel trends.
  - Benefit over parallel trends: it is scale-free.
  - Equivalent to parallel trends if $\mathbb{E}[Y_{i0} \mid G_i = 1] = \mathbb{E}[Y_{i0} \mid G_i = 0]$
- Different ideas about why there is imbalance on the LDV:
  - DID: time-constant unmeasured confounder creates imbalance.
  - LDV: previous outcome directly affects treatment assignment.

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i1}}_{\text{difference in post period}}$$

$$- \hat{\rho}_{LDV} \underbrace{\left\{ \frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i0} \right\}}_{\text{difference in pre period}}$$

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i1}}_{\text{difference in post period}}$$

$$- \hat{\rho}_{LDV} \underbrace{\left\{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i0}\right\}}_{\text{difference in pre period}}$$

- If $\hat{\rho}_{LDV} = 1$ then $\widehat{\tau}_{DID} = \widehat{\tau}_{LDV}$ and if $0 \leq \hat{\rho}_{LDV} < 1$:

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i1}}_{\text{difference in post period}}$$

$$- \hat{\rho}_{LDV} \underbrace{\left\{ \frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i0} \right\}}_{\text{difference in pre period}}$$

- If $\hat{\rho}_{LDV} = 1$ then $\widehat{\tau}_{DID} = \widehat{\tau}_{LDV}$ and if $0 \leq \hat{\rho}_{LDV} < 1$:
    - If $G_i = 1$ has higher baseline outcomes $\rightsquigarrow \widehat{\tau}_{LDV} > \widehat{\tau}_{DID}$.

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i1}}_{\text{difference in post period}}$$

$$- \hat{\rho}_{LDV} \underbrace{\left\{ \frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i0} \right\}}_{\text{difference in pre period}}$$

- If $\hat{\rho}_{LDV} = 1$ then $\widehat{\tau}_{DID} = \widehat{\tau}_{LDV}$ and if $0 \leq \hat{\rho}_{LDV} < 1$:
  - If $G_i = 1$ has higher baseline outcomes $\leadsto \widehat{\tau}_{LDV} > \widehat{\tau}_{DID}$.
  - If $G_i = 1$ has lower baseline outcomes $\leadsto \widehat{\tau}_{DID} > \widehat{\tau}_{LDV}$.

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i1}}_{\text{difference in post period}}$$

$$- \hat{\rho}_{LDV} \underbrace{\left\{ \frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1 - G_i) Y_{i0} \right\}}_{\text{difference in pre period}}$$

- If $\hat{\rho}_{LDV} = 1$ then $\widehat{\tau}_{DID} = \widehat{\tau}_{LDV}$ and if $0 \leq \hat{\rho}_{LDV} < 1$:
  - If $G_i = 1$ has higher baseline outcomes $\rightsquigarrow \widehat{\tau}_{LDV} > \widehat{\tau}_{DID}$.
  - If $G_i = 1$ has lower baseline outcomes $\rightsquigarrow \widehat{\tau}_{DID} > \widehat{\tau}_{LDV}$.
- Bracketing relationship: if you willing to assume parallel trends or LDV,

$$\mathbb{E}[\widehat{\tau}_{LDV}] \geq \tau_{\text{att}} \geq \mathbb{E}[\widehat{\tau}_{DID}]$$

DID/LDV bracketing

- Estimator: estimate CEF $\mathbb{E}[Y_{i1} \mid Y_{i0}, G_i] = \alpha + \rho Y_{i0} + \tau G_i$

$$\widehat{\tau}_{LDV} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i1} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i1}}_{\text{difference in post period}}$$
$$- \hat{\rho}_{LDV} \underbrace{\left\{\frac{1}{n_1}\sum_{i=1}^{n} G_i Y_{i0} - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)Y_{i0}\right\}}_{\text{difference in pre period}}$$

- If $\hat{\rho}_{LDV} = 1$ then $\widehat{\tau}_{DID} = \widehat{\tau}_{LDV}$ and if $0 \leq \hat{\rho}_{LDV} < 1$:
  - If $G_i = 1$ has higher baseline outcomes $\rightsquigarrow \widehat{\tau}_{LDV} > \widehat{\tau}_{DID}$.
  - If $G_i = 1$ has lower baseline outcomes $\rightsquigarrow \widehat{\tau}_{DID} > \widehat{\tau}_{LDV}$.
- Bracketing relationship: if you willing to assume parallel trends or LDV,

$$\mathbb{E}[\widehat{\tau}_{LDV}] \geq \tau_{\text{att}} \geq \mathbb{E}[\widehat{\tau}_{DID}]$$

  - Holds nonparametrically as well.

Nonparametric identification

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?

Nonparametric identification

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

$$E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 0]$$

Nonparametric identification

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

$$E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 0]$$

- What does this assumption say? It says that the potential trend under control is the same for the control and treated groups, conditional on covariates.

Nonparametric identification

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

$$E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 0]$$

- What does this assumption say? It says that the potential trend under control is the same for the control and treated groups, conditional on covariates.
    - Units that are similar at baseline will follow similar paths under no treatment.

Nonparametric identification

- Up until now, we assumed unconditional parallel trends. What if this doesn't hold?
- Alternative identification: **conditional parallel trends**

$$E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid \mathbf{X}_i, G_i = 0]$$

- What does this assumption say? It says that the potential trend under control is the same for the control and treated groups, conditional on covariates.
    - Units that are similar at baseline will follow similar paths under no treatment.
- **Matching**: conduct DID analysis on units with similar values of $\mathbf{X}_i$

Semiparametric estimation with repeated outcomes

- How to estimate regression DID without strong linearity assumptions?

Semiparametric estimation with repeated outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

Semiparametric estimation with repeated outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

- Reweights control group to have the same distribution of $\mathbf{X}_i$ as treated group.

Semiparametric estimation with repeated outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

- Reweights control group to have the same distribution of $\mathbf{X}_i$ as treated group.
- Have to estimate the **propensity score** $\mathbb{P}(G_i = 1 \mid \mathbf{X}_i)$

Semiparametric estimation with repeated outcomes

- How to estimate regression DID without strong linearity assumptions?
- Abadie (2005) derives **weighting estimators** in this setting:

$$\mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1] = \mathbb{E}\left[\frac{(Y_{i1} - Y_{i0})}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}{1 - \mathbb{P}(G_i = 1 \mid \mathbf{X}_i)}\right]$$

- Reweights control group to have the same distribution of $\mathbf{X}_i$ as treated group.
- Have to estimate the **propensity score** $\mathbb{P}(G_i = 1 \mid \mathbf{X}_i)$
    - Possible model misspecification!

2/ Fixed effects

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.

Basic idea of fixed effects

- “One way” fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
  - Arbitrary treatment timing, covariates, etc.
  - Units: $i = 1, \ldots, n$
  - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
  - Arbitrary treatment timing, covariates, etc.
  - Units: $i = 1, \dots, n$
  - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
  - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$
  - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
  - Arbitrary treatment timing, covariates, etc.
  - Units: $i = 1, \ldots, n$
  - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
  - History of a variable: $\overline{D}_{it} = (D_{i1}, \ldots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i' \beta + \varepsilon_{it}$
  - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
  - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \ldots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \ldots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i' \beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!
    - Imai and Kim (2019, AJPS) give clarification on these identification issues.

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \not\rightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!
    - Imai and Kim (2019, AJPS) give clarification on these identification issues.
- Implicit assumption of **no carryover**? $Y_{it}(d_1, \dots, d_t) = Y_{it}(d_t)$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \dots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i' \beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!
    - Imai and Kim (2019, AJPS) give clarification on these identification issues.
- Implicit assumption of **no carryover**? $Y_{it}(d_1, \dots, d_t) = Y_{it}(d_t)$
    - More a choice of estimand: focuses on **contemporaneous** effect.

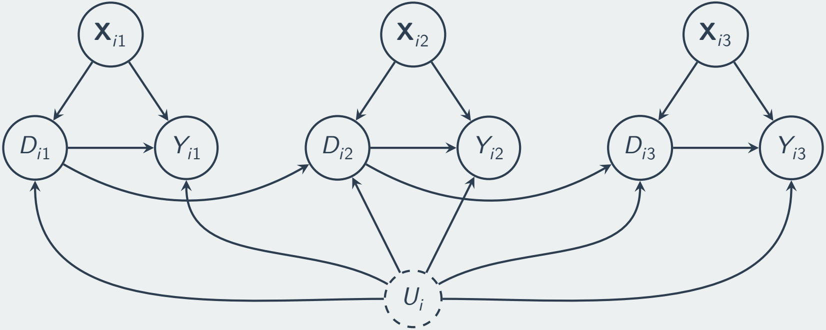
Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
  - Arbitrary treatment timing, covariates, etc.
  - Units: $i = 1, \dots, n$
  - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
  - History of a variable: $\overline{D}_{it} = (D_{i1}, \dots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$
  - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
  - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
  - $\rightsquigarrow$ LDV cannot be a confounder!
  - Imai and Kim (2019, AJPS) give clarification on these identification issues.
- Implicit assumption of **no carryover**? $Y_{it}(d_1, \dots, d_t) = Y_{it}(d_t)$
  - More a choice of estimand: focuses on **contemporaneous** effect.
  - Treatment history follows observed path through $t-1$:
    $Y_{it}(d_t) = Y_{it}(D_{i1}, \dots, D_{i,t-1}, d_t)$

Basic idea of fixed effects

- "One way" fixed effects generalizes the before/after design.
    - Arbitrary treatment timing, covariates, etc.
    - Units: $i = 1, \ldots, n$
    - Causal ordering with time: covariates $\mathbf{X}_{it}$, treatment D_{it}, outcome Y_{it}
    - History of a variable: $\overline{D}_{it} = (D_{i1}, \ldots, D_{it})$ and $\overline{D}_i \equiv \overline{D}_{iT}$
- Linear fixed effects model: $Y_{it} = \alpha_i + \tau D_{it} + \mathbf{X}_i'\beta + \varepsilon_{it}$
    - Key assumption: **strict exogeneity** $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_i, \overline{D}_i, \alpha_i] = 0$
    - Implies **no feedback** between outcome and treatment ($Y_{it} \nrightarrow D_{i,t+1}$)
    - $\rightsquigarrow$ LDV cannot be a confounder!
    - Imai and Kim (2019, AJPS) give clarification on these identification issues.
- Implicit assumption of **no carryover**? $Y_{it}(d_1, \ldots, d_t) = Y_{it}(d_t)$
    - More a choice of estimand: focuses on **contemporaneous** effect.
    - Treatment history follows observed path through $t-1$: $Y_{it}(d_t) = Y_{it}(D_{i1}, \ldots, D_{i,t-1}, d_t)$
    - $\rightsquigarrow$ lags of treatments become part of time-varying confounders.

Strict exogeneity DAG



Strict exogeneity implied by strict ignorability $Y_{it}(d) \perp\!\!\!\perp \overline{D}_i \mid \overline{X}_i, U_i$

FE estimation

- With linear models, two transformations can purge the fixed effects.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1}\sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau\ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}' \beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}'\beta + \tau \Delta D_{it} + \Delta \varepsilon_{it}$$

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}'\beta + \tau \Delta D_{it} + \Delta \varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.

- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}'\beta + \tau \Delta D_{it} + \Delta\varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$
  - Both consistent under strict exogeneity.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1}\sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau\ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta\mathbf{X}_{it}'\beta + \tau\Delta D_{it} + \Delta\varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$
  - Both consistent under strict exogeneity.
  - FE more efficient if original errors, ε_{it}, are serially uncorrelated.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

  - Time-demeaning Y_{it} purges the time constant fixed effect.
  - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}'\beta + \tau \Delta D_{it} + \Delta\varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$
  - Both consistent under strict exogeneity.
  - FE more efficient if original errors, ε_{it}, are serially uncorrelated.
  - FD more efficient if differences, $\Delta\varepsilon_{it}$, are serially uncorrelated.

FE estimation

- With linear models, two transformations can purge the fixed effects.
- **Within/FE transformation**: $\ddot{Z}_{it} = Z_{it} - T^{-1} \sum_{s=1}^{T} Z_{is}$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

    - Time-demeaning Y_{it} purges the time constant fixed effect.
    - But they retain the same coefficients as the original model.
- **First differences**: $\Delta Z_{it} = Z_{it} - Z_{i,t-1}$

$$\Delta Y_{it} = \Delta \mathbf{X}_{it}'\beta + \tau \Delta D_{it} + \Delta\varepsilon_{it}$$

- Estimation: pooled OLS of either specification, $\widehat{\tau}_{\text{fe}}$, $\widehat{\tau}_{\text{fd}}$
    - Both consistent under strict exogeneity.
    - FE more efficient if original errors, ε_{it}, are serially uncorrelated.
    - FD more efficient if differences, $\Delta\varepsilon_{it}$, are serially uncorrelated.
    - Latter allows for substantial serial dependence in the original errors.

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\arg\max_{\alpha,\beta,\tau,\gamma} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}'_{it}\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n
  - Computationally inefficient for large n (number of dummies grows with n)

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n
  - Computationally inefficient for large n (number of dummies grows with n)

- Best practice: cluster variances at the unit level.

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n
  - Computationally inefficient for large n (number of dummies grows with n)

- Best practice: cluster variances at the unit level.

  - With CR variance estimators, LSDV "double counts" degrees of freedom

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n
  - Computationally inefficient for large n (number of dummies grows with n)

- Best practice: cluster variances at the unit level.

  - With CR variance estimators, LSDV “double counts” degrees of freedom
  - Better to use within estimator in that case.

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

  - **Least squares dummy variable** estimator reasonable for moderate n
  - Computationally inefficient for large n (number of dummies grows with n)
- Best practice: cluster variances at the unit level.
  - With CR variance estimators, LSDV "double counts" degrees of freedom
  - Better to use within estimator in that case.
- Best choice: use canned packages.

Estimation notes

- Within estimator can be implemented by adding unit dummy variables.

$$\underset{\alpha,\beta,\tau,\gamma}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \alpha - \mathbf{X}_{it}'\beta - \tau D_{it} - \sum_{k=2}^{n} \gamma_k \mathbb{1}(i = k) \right)^2$$

    - **Least squares dummy variable** estimator reasonable for moderate n
    - Computationally inefficient for large n (number of dummies grows with n)

- Best practice: cluster variances at the unit level.
    - With CR variance estimators, LSDV "double counts" degrees of freedom
    - Better to use within estimator in that case.

- Best choice: use canned packages.
    - `{fixest}` in R, `-reghdfe-` in Stata

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
  - OLS typically biased because nonconstant effects induce correlation between treatment and error.

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
  - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
  - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

  - S_i^2 is the within-unit treatment variance.

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
  - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

  - S_i^2 is the within-unit treatment variance.
  - Units with even treatment/control split upweighted.

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
    - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

    - S_i^2 is the within-unit treatment variance.
    - Units with even treatment/control split upweighted.
- Imai, Kim & Wang (2019, AJPS): use a matching to target the ATE.

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
    - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \xrightarrow{p} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

    - S_i^2 is the within-unit treatment variance.
    - Units with even treatment/control split upweighted.
- Imai, Kim & Wang (2019, AJPS): use a matching to target the ATE.
    - Match treated and control periods within units (also weakens linearity).

Non-constant treatment effects

- LFE models assume constant treatment effects. What happens if not?
    - OLS typically biased because nonconstant effects induce correlation between treatment and error.
- With no covariates and no only treated/control units:

$$\widehat{\tau}_{\text{fe}} \overset{p}{\to} \frac{\mathbb{E}\left[\left(\frac{\sum_t D_{it} Y_{it}}{\sum_t D_{it}} - \frac{\sum_t (1-D_{it}) Y_{it}}{\sum_t (1-D_{it})}\right) S_i^2\right]}{\mathbb{E}[S_i^2]} \neq \tau$$

    - S_i^2 is the within-unit treatment variance.
    - Units with even treatment/control split upweighted.
- Imai, Kim & Wang (2019, AJPS): use a matching to target the ATE.
    - Match treated and control periods within units (also weakens linearity).
    - `{PanelMatch}` R package.

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

    - Allow Y_{it} to be related to future $D_{i,t+s}$

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

    - Allow Y_{it} to be related to future $D_{i,t+s}$
- This implies **sequential exogeneity** of the errors: $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{it}, \alpha_i] = 0$.

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

    - Allow Y_{it} to be related to future $D_{i,t+s}$
- This implies **sequential exogeneity** of the errors: $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{it}, \alpha_i] = 0$.
- Estimation to these **dynamic panel models**:

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

    - Allow Y_{it} to be related to future $D_{i,t+s}$
- This implies **sequential exogeneity** of the errors: $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{it}, \alpha_i] = 0$.
- Estimation to these **dynamic panel models**:
    - instrumental variables (Arellano and Bond) using lagged difference and levels as instruments (only valid for linear models).

Strict vs. sequential exogeneity/ignorability

- Strict exogeneity/ignorability is **very strong**.
    - Remember: rules out all outcome-treatment feedback.
- Weaker assumption: **Sequential ignorability**:

$$Y_{it}(d) \perp\!\!\!\perp D_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{i,t-1}, \alpha_i$$

    - Allow Y_{it} to be related to future $D_{i,t+s}$
- This implies **sequential exogeneity** of the errors: $\mathbb{E}[\varepsilon_{it} \mid \overline{\mathbf{X}}_{it}, \overline{D}_{it}, \alpha_i] = 0$.
- Estimation to these **dynamic panel models**:
    - instrumental variables (Arellano and Bond) using lagged difference and levels as instruments (only valid for linear models).
    - bias correction: estimate the bias and subtract it off (valid for nonlinear models too).

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
    - Complicated by the effect of treatment on time-varying confounders.

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
  - Complicated by the effect of treatment on time-varying confounders.
  - Pathways involving $\mathbf{X}_{it}(d_{t-1})$ difficult to identify.

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
    - Complicated by the effect of treatment on time-varying confounders.
    - Pathways involving $\mathbf{X}_{it}(d_{t-1})$ difficult to identify.
- Possible approach: **propensity score FEs** (Blackwell & Yamauchi, 2021)

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
  - Complicated by the effect of treatment on time-varying confounders.
  - Pathways involving $\mathbf{X}_{it}(d_{t-1})$ difficult to identify.
- Possible approach: **propensity score FEs** (Blackwell & Yamauchi, 2021)
  - Include unit dummies in propensity score model.

Effect of lagged treatments

- Focused on the contemporaneous effect of D_{it}.
- What about treatment histories $Y_{it}(d_{t-1}, d_t)$?
- Very difficult, if not impossible with fixed effects models.
  - Complicated by the effect of treatment on time-varying confounders.
  - Pathways involving $\mathbf{X}_{it}(d_{t-1})$ difficult to identify.
- Possible approach: **propensity score FEs** (Blackwell & Yamauchi, 2021)
  - Include unit dummies in propensity score model.
  - Bias from incidental parameters, but disappears as $T \to \infty$