Gov 2002 - Causal Inference IV: Repeated Measurements

Matthew Blackwell   Arthur Spirling

October 23rd, 2014

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization
  - Instrumental variables: randomized instrument

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization
  - Instrumental variables: randomized instrument
  - RD: discontinuity at the threshold

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization
  - Instrumental variables: randomized instrument
  - RD: discontinuity at the threshold
- Today we're going to look to another possible source of variation: repeated measurements on the same unit over time.

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization
  - Instrumental variables: randomized instrument
  - RD: discontinuity at the threshold
- Today we're going to look to another possible source of variation: repeated measurements on the same unit over time.
- What if selection on the observables doesn't hold, but do have repeated measurements. Can we use this to identify and estimate effects?

Repeated measurements

- For every week on causal inference, we have identified a different source of exogenous variation:
  - Selection on the observables: as-if randomization
  - Instrumental variables: randomized instrument
  - RD: discontinuity at the threshold
- Today we're going to look to another possible source of variation: repeated measurements on the same unit over time.
- What if selection on the observables doesn't hold, but do have repeated measurements. Can we use this to identify and estimate effects?
- Message: simply having panel data does not identify an effect, but it does allow us to rely on different identifying assumptions.

Basic Idea

- The basic idea is that ignorability doesn't hold, conditional on the observed covariates, $Y_{it}(a) \not\perp\!\!\!\perp A_{it}|X_{it}$, but ignorability might hold conditional on some unobserved, time-constant, variable:
$$Y_{it}(a) \perp\!\!\!\perp A_{it}|X_{it}, U_i.$$

Basic Idea

- The basic idea is that ignorability doesn't hold, conditional on the observed covariates, $Y_{it}(a) \not\perp\!\!\!\perp A_{it}|X_{it}$, but ignorability might hold conditional on some unobserved, time-constant, variable:
$$Y_{it}(a) \perp\!\!\!\perp A_{it}|X_{it}, U_i.$$
- Within units, effects are identified.

Basic Idea

- The basic idea is that ignorability doesn't hold, conditional on the observed covariates, $Y_{it}(a) \not\perp\!\!\!\perp A_{it}|X_{it}$, but ignorability might hold conditional on some unobserved, time-constant, variable:
$$Y_{it}(a) \perp\!\!\!\perp A_{it}|X_{it}, U_i.$$
- Within units, effects are identified.
- This is because, even if U_i is unobserved, it is held constant within a unit.

Basic Idea

- The basic idea is that ignorability doesn't hold, conditional on the observed covariates, $Y_{it}(a) \not\perp\!\!\!\perp A_{it}|X_{it}$, but ignorability might hold conditional on some unobserved, time-constant, variable:
$$Y_{it}(a) \perp\!\!\!\perp A_{it}|X_{it}, U_i.$$
- Within units, effects are identified.
- This is because, even if U_i is unobserved, it is held constant within a unit.
- Thus, by performing analyses within the units, we can control for this unobserved heterogeneity.

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).
- **TSCS data**: high T, low medium N.

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).
- **TSCS data**: high T, low medium N.
  - U.S. states over time

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).
- **TSCS data**: high T, low medium N.
  - U.S. states over time
  - Western European countries over time.

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).
- **TSCS data**: high T, low medium N.
  - U.S. states over time
  - Western European countries over time.
- For the most part, the issues of causality are the same for these two types of data, so I will refer to them both as panel data.

Note about terminology

- Generally, we talk about *panel data* and *time-series cross-sectional data* in political science.
- **Panel data**: small T, large N
  - The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections).
- **TSCS data**: high T, low medium N.
  - U.S. states over time
  - Western European countries over time.
- For the most part, the issues of causality are the same for these two types of data, so I will refer to them both as panel data.
- But estimation is a different issue. Different estimators work differently under either data types.

Fixed effects estimators

Notation

- Units $i = 1, \ldots, N$

Notation

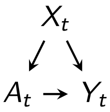
- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,

Notation

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, A_{it} are the outcome and treatment for unit i in period t
  We have a set of covariates in each period, as well,

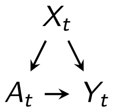
Notation

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, A_{it} are the outcome and treatment for unit i in period t
  We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to A_{it}.



Notation

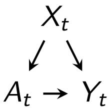
- Units $i = 1, \dots, N$
- Time periods $t = 1, \dots, T$ with $T \geq 2$,
- Y_{it}, A_{it} are the outcome and treatment for unit i in period t
  We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to A_{it}.



- U_i = unobserved, time-invariant unit effects (causally prior to everything)

Notation

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, A_{it} are the outcome and treatment for unit i in period t
  We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to A_{it}.



- U_i = unobserved, time-invariant unit effects (causally prior to everything)
- History of some variable: $\underline{A}_{it} = (A_1, \ldots, A_t)$.

Basic linear fixed-effects model

- The typical way that we write a fixed effect model is as a linear regression:

$$Y_{it} = X_{it}'\beta + \tau A_{it} + U_i + \varepsilon_{it}$$

Basic linear fixed-effects model

- The typical way that we write a fixed effect model is as a linear regression:
$$Y_{it} = X_{it}'\beta + \tau A_{it} + U_i + \varepsilon_{it}$$
- Here we focus on contemporaneous effects, $\tau \cdot A_{it}$

Basic linear fixed-effects model

- The typical way that we write a fixed effect model is as a linear regression:
$$Y_{it} = X_{it}'\beta + \tau A_{it} + U_i + \varepsilon_{it}$$
- Here we focus on contemporaneous effects, $\tau \cdot A_{it}$
- ... but could have something more complicated:
$= \tau_1 A_{it} + \tau_2 A_{i,t-1} \rightarrow$ much harder

Basic linear fixed-effects model

- The typical way that we write a fixed effect model is as a linear regression:
$$Y_{it} = X_{it}'\beta + \tau A_{it} + U_i + \varepsilon_{it}$$
- Here we focus on contemporaneous effects, $\tau \cdot A_{it}$
- ... but could have something more complicated: $= \tau_1 A_{it} + \tau_2 A_{i,t-1} \rightarrow$ much harder
- Key assumptions will be on the relationship between U_i and ε_{it}.

Basic linear fixed-effects model

- The typical way that we write a fixed effect model is as a linear regression:
$$Y_{it} = X_{it}'\beta + \tau A_{it} + U_i + \varepsilon_{it}$$
- Here we focus on contemporaneous effects, $\tau \cdot A_{it}$
- …but could have something more complicated: $= \tau_1 A_{it} + \tau_2 A_{i,t-1} \rightarrow$ much harder
- Key assumptions will be on the relationship between U_i and ε_{it}.
- With no lagged dependent variables in X_{it}, we usually rely on what is called a **strict exogeneity** assumption:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This combining this with the above regression, we get the following conditional expectation function:

$$E[Y_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = X_{it}'\beta + \tau A_{it} + U_i$$

- This combining this with the above regression, we get the following conditional expectation function:

$$E[Y_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = X_{it}'\beta + \tau A_{it} + U_i$$

- Need to fix one of the unobserved unit effects at $U_1 = 0$ (or fix the mean at 0), $U_2, \ldots, U_N$ are parameters/constants.

Fixed-effects within estimator

- Define the "within" estimator:

$$(Y_{it} - \overline{Y}_i) = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

Fixed-effects within estimator

- Define the "within" estimator:

$$(Y_{it} - \overline{Y}_i) = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

- Here, let $\overline{Y}_i$ be the unit averages. Note that:

$$\overline{Y}_i = \overline{X}_i'\beta + \tau\overline{A}_i + U_i + \overline{\varepsilon}_i$$

Fixed-effects within estimator

- Define the "within" estimator:

$$(Y_{it} - \overline{Y}_i) = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

- Here, let $\overline{Y}_i$ be the unit averages. Note that:

$$\overline{Y}_i = \overline{X}_i'\beta + \tau\overline{A}_i + U_i + \overline{\varepsilon}_i$$

- Logic: since the unobserved effect is constant over time, subtracting off the mean also subtracts that unobserved effect:

$$U_i - \frac{1}{T}\sum_{t=1}^{T} U_i = U_i - U_i = 0$$

Fixed-effects within estimator

- Define the "within" estimator:

$$(Y_{it} - \overline{Y}_i) = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

- Here, let $\overline{Y}_i$ be the unit averages. Note that:

$$\overline{Y}_i = \overline{X}_i'\beta + \tau\overline{A}_i + U_i + \overline{\varepsilon}_i$$

- Logic: since the unobserved effect is constant over time, subtracting off the mean also subtracts that unobserved effect:

$$U_i - \frac{1}{T}\sum_{t=1}^{T} U_i = U_i - U_i = 0$$

- This also demonstrates why the assumption of the fixed effects being time-constant is so important.

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.
- Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors:

$$E[Y_{it} - \overline{Y}_i|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i)$$

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.
- Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors:

$$E[Y_{it} - \overline{Y}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i)$$

- OLS and go!

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.
- Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors:

$$E[Y_{it} - \overline{Y}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i)$$

- OLS and go!

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.
- Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors:

$$E[Y_{it} - \overline{Y}_i | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i)$$

- OLS and go! Hidden assumption?

Fixed-effects within estimator

- Informal proof. We have strict exogeneity:

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This implies exogeneity of the mean-differenced errors:

$$E[\varepsilon_{it} - \overline{\varepsilon}_i|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- Mean-differenced errors are uncorrelated with the treatment or regressors from *any* time period.
- Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors:

$$E[Y_{it} - \overline{Y}_i|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = (X_{it} - \overline{X}_i)'\beta + \tau(A_{it} - \overline{A}_i)$$

- OLS and go! Hidden assumption?
- Full rank: $\text{rank}[E[(X_{it} - \overline{X}_i)'(X_{it} - \overline{X}_i)]] = K$

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.
- For some variable, Z_{it}, let

$$\Delta Z_{it} = Z_{it} - Z_{i,t-1}$$

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.
- For some variable, Z_{it}, let

$$\Delta Z_{it} = Z_{it} - Z_{i,t-1}$$

- The **first difference model** is the following:

$$\Delta Y_{it} = \Delta X_{it}' \beta + \tau \Delta A_{it} + \Delta \varepsilon_{it}$$

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.
- For some variable, Z_{it}, let

$$\Delta Z_{it} = Z_{it} - Z_{i,t-1}$$

- The **first difference model** is the following:

$$\Delta Y_{it} = \Delta X_{it}'\beta + \tau \Delta A_{it} + \Delta \varepsilon_{it}$$

- This follows from the fact that $\Delta U_i = 0$

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.
- For some variable, Z_{it}, let

$$\Delta Z_{it} = Z_{it} - Z_{i,t-1}$$

- The **first difference model** is the following:

$$\Delta Y_{it} = \Delta X_{it}'\beta + \tau \Delta A_{it} + \Delta \varepsilon_{it}$$

- This follows from the fact that $\Delta U_i = 0$
- By the same logic as above, we get the following:

$$E[\Delta Y_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = \Delta X_{it}'\beta + \tau \Delta A_{it}$$

First differences

- Because the U_i are time-fixed, first-differences are an alternative to mean-differences.
- For some variable, Z_{it}, let

$$\Delta Z_{it} = Z_{it} - Z_{i,t-1}$$

- The **first difference model** is the following:

$$\Delta Y_{it} = \Delta X_{it}'\beta + \tau \Delta A_{it} + \Delta \varepsilon_{it}$$

- This follows from the fact that $\Delta U_i = 0$
- By the same logic as above, we get the following:

$$E[\Delta Y_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = \Delta X_{it}'\beta + \tau \Delta A_{it}$$

- More efficient than regular FE when there is serial correlation exists in the errors.

Random effects

- With fixed effects, we made an assumption that

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0]$$

Random effects

- With fixed effects, we made an assumption that

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0]$$

- "Random effects" models make an additional assumption:

$$E[U_i|\underline{X}_{iT}, \underline{A}_{iT}] = E[U_i] = 0$$

Random effects

- With fixed effects, we made an assumption that

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0]$$

- "Random effects" models make an additional assumption:

$$E[U_i|\underline{X}_{iT}, \underline{A}_{iT}] = E[U_i] = 0$$

- That is, random effects models make an additional assumption about the distribution of unobserved effects.

Random effects

- With fixed effects, we made an assumption that

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0]$$

- "Random effects" models make an additional assumption:

$$E[U_i|\underline{X}_{iT}, \underline{A}_{iT}] = E[U_i] = 0$$

- That is, random effects models make an additional assumption about the distribution of unobserved effects.
- Unit-level effects are uncorrelated with treatment and covariates.

Random effects

- With fixed effects, we made an assumption that

$$E[\varepsilon_{it}|\underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0]$$

- "Random effects" models make an additional assumption:

$$E[U_i|\underline{X}_{iT}, \underline{A}_{iT}] = E[U_i] = 0$$

- That is, random effects models make an additional assumption about the distribution of unobserved effects.
- Unit-level effects are uncorrelated with treatment and covariates.
- **Important**: implies that ignorability holds without conditioning on U_i, so this is **not** helping us identify causal effects beyond typical regressions.

Why random effects?

- So why do people use random effects? Standard errors!

Why random effects?

- So why do people use random effects? Standard errors!
- Under the RE assumption, we have the following:

$$Y_{it} = X_{it}'\beta + \tau A_{it} + \nu_i$$

where $\nu_i = U_i + \varepsilon_{it}$.

Why random effects?

- So why do people use random effects? Standard errors!
- Under the RE assumption, we have the following:

$$Y_{it} = X_{it}'\beta + \tau A_{it} + \nu_i$$

where $\nu_i = U_i + \varepsilon_{it}$.

- Now, notice that

$$\text{cov}[Y_{i1}, Y_{i2}|\underline{X}_{it}, \underline{A}_{it}] = \sigma_u^2$$

where σ_u^2 is the variance of the U_i.

Why random effects?

- So why do people use random effects? Standard errors!
- Under the RE assumption, we have the following:

$$Y_{it} = X_{it}'\beta + \tau A_{it} + \nu_i$$

  where $\nu_i = U_i + \varepsilon_{it}$.
- Now, notice that

$$\text{cov}[Y_{i1}, Y_{i2} | \underline{X}_{it}, \underline{A}_{it}] = \sigma_u^2$$

  where σ_u^2 is the variance of the U_i.
- This violates the assumption of no autocorrelation for OLS. What's the problem with this?

Why random effects?

- So why do people use random effects? Standard errors!
- Under the RE assumption, we have the following:

$$Y_{it} = X_{it}'\beta + \tau A_{it} + \nu_i$$

  where $\nu_i = U_i + \varepsilon_{it}$.
- Now, notice that

$$\text{cov}[Y_{i1}, Y_{i2} | \underline{X}_{it}, \underline{A}_{it}] = \sigma_u^2$$

  where σ_u^2 is the variance of the U_i.
- This violates the assumption of no autocorrelation for OLS. What's the problem with this?
- Random effects models gets us consistent standard error estimates.

Lagged dependent variables

- Strict exogeneity is quite strong.

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence
  - any other time-varying covariate

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence
  - any other time-varying covariate
- Rules out lagged dependent variables. Why? Informally:

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence
  - any other time-varying covariate
- Rules out lagged dependent variables. Why? Informally:
  - Let $Y_{i,t-1} \in X_{it}$.

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence
  - any other time-varying covariate
- Rules out lagged dependent variables. Why? Informally:
  - Let $Y_{i,t-1} \in X_{it}$.
  - By strict exogeneity, ε_{it} uncorrelated with $X_{i,t+1}$

Lagged dependent variables

- Strict exogeneity is quite strong.
- Example: economic interdependence between countries ($A_{it} = 1$ if county-dyad i is interdependent in period t) and conflict severity (Y_{it}) between countries.
- Strict exogeneity assumption implies shocks to conflict severity at t uncorrelated with:
  - future values of conflict severity
  - economic interdendence
  - any other time-varying covariate
- Rules out lagged dependent variables. Why? Informally:
  - Let $Y_{i,t-1} \in X_{it}$.
  - By strict exogeneity, ε_{it} uncorrelated with $X_{i,t+1}$
  - Implies that ε_{it} uncorrelated with Y_{it}, but this can't be since it is the error for that variable!

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it}|\underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it}-Y_{i,t-1}) = \beta(Y_{i,t-1}-Y_{i,t-2})+\tau(A_{it}-A_{i,t-1})+(\varepsilon_{it}-\varepsilon_{i,t-1})$$

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.
- This is sometimes called a **dynamic panel model**, where we can't rely on the exogeneity assumption alone. We need an instrument.

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.
- This is sometimes called a **dynamic panel model**, where we can't rely on the exogeneity assumption alone. We need an instrument.
- Conveniently, $Y_{i,t-2}$ is a good instrument:

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.
- This is sometimes called a **dynamic panel model**, where we can't rely on the exogeneity assumption alone. We need an instrument.
- Conveniently, $Y_{i,t-2}$ is a good instrument:
    - correlated with $(Y_{i,t-1} - Y_{i,t-2})$

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it}|\underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.
- This is sometimes called a **dynamic panel model**, where we can't rely on the exogeneity assumption alone. We need an instrument.
- Conveniently, $Y_{i,t-2}$ is a good instrument:
  - correlated with $(Y_{i,t-1} - Y_{i,t-2})$
  - unrelated to the error $(\varepsilon_{it} - \varepsilon_{i,t-1})$

LDVs and exogeneity

- Weaker assumption: errors uncorrelated with past values, which we call **sequential exogeneity**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- But this doesn't identify our parameters. To see this, imagine the LDV is the only covariate and we're using first differences:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, $Y_{i,t-1}$ is correlated with the $\varepsilon_{i,t-1}$.
- This is sometimes called a **dynamic panel model**, where we can't rely on the exogeneity assumption alone. We need an instrument.
- Conveniently, $Y_{i,t-2}$ is a good instrument:
  - correlated with $(Y_{i,t-1} - Y_{i,t-2})$
  - unrelated to the error $(\varepsilon_{it} - \varepsilon_{i,t-1})$
- Dynamic panel literature full of examples of how to use different IV approaches

Heterogeneous treatment effects

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.
- $\tau_{it} = Y_{it}(1) - Y_{it}(0)$

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.
- $\tau_{it} = Y_{it}(1) - Y_{it}(0)$
- Consistency:

$$Y_{it} = A_{it} Y_{it}(1) + (1 - A_{it}) Y_{it}(0)$$

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.
- $\tau_{it} = Y_{it}(1) - Y_{it}(0)$
- Consistency:

$$Y_{it} = A_{it} Y_{it}(1) + (1 - A_{it}) Y_{it}(0)$$

- Which implies the following:

$$Y_{it} = E[Y_{it}(0)|\underline{X}_{iT}, U_i] + A_{it}\tau_{it} + Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]$$

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.
- $\tau_{it} = Y_{it}(1) - Y_{it}(0)$
- Consistency:

$$Y_{it} = A_{it} Y_{it}(1) + (1 - A_{it}) Y_{it}(0)$$

- Which implies the following:

$$Y_{it} = E[Y_{it}(0) | \underline{X}_{iT}, U_i] + A_{it}\tau_{it} + Y_{it}(0) - E[Y_{it}(0) | \underline{X}_{iT}, U_i]$$

- Here, $\underline{X}_{it}$ might include lags of A_{it} as well.

Potential outcomes in the general setting

- Let $Y_{it}(1)$ be the potential outcome when a person gets treatment at time t. Similarly for $Y_{it}(0)$.
- $\tau_{it} = Y_{it}(1) - Y_{it}(0)$
- Consistency:

$$Y_{it} = A_{it} Y_{it}(1) + (1 - A_{it}) Y_{it}(0)$$

- Which implies the following:

$$Y_{it} = E[Y_{it}(0)|\underline{X}_{iT}, U_i] + A_{it}\tau_{it} + Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]$$

- Here, $\underline{X}_{it}$ might include lags of A_{it} as well.
- Get us even closer, note we can do the following:

$$A_i\tau_{it} = A_{it}\tau_c + A_{it}(\tau_{it} - \tau_c)$$

where $\tau_c = E[\tau_{it}]$ is the ATE.

Putting it all together

- Finally, let's assume that the mean of the potential outcome under control is linear:

$$E[Y_{it}(0)|\underline{X}_{iT}, U_i] = X_{ij}\beta_c + U_i$$

Putting it all together

- Finally, let's assume that the mean of the potential outcome under control is linear:

$$E[Y_{it}(0)|\underline{X}_{iT}, U_i] = X_{ij}\beta_c + U_i$$

- Combing these facts, we get:

$$Y_{it} = X_{ij}\beta_c + U_i + A_{it}\tau_c + \eta_{it}$$

Putting it all together

- Finally, let's assume that the mean of the potential outcome under control is linear:

$$E[Y_{it}(0)|\underline{X}_{iT}, U_i] = X_{ij}\beta_c + U_i$$

- Combing these facts, we get:

$$Y_{it} = X_{ij}\beta_c + U_i + A_{it}\tau_c + \eta_{it}$$

- Where the combined error is:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

Putting it all together

- Finally, let's assume that the mean of the potential outcome under control is linear:

$$E[Y_{it}(0)|\underline{X}_{iT}, U_i] = X_{ij}\beta_c + U_i$$

- Combing these facts, we get:

$$Y_{it} = X_{ij}\beta_c + U_i + A_{it}\tau_c + \eta_{it}$$

- Where the combined error is:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- When will $\tau_c = \tau$ from the fixed effects regression models above?

Assumptions

- Let's make a **strict ignorability** assumption conditional on U_i:

$$Y_{it}(a) \perp\!\!\!\perp A_{it} | \{\underline{X}_{iT} \setminus A_{it}\}, U_i$$

Assumptions

- Let's make a **strict ignorability** assumption conditional on U_i:

$$Y_{it}(a) \perp\!\!\!\perp A_{it} | \{\underline{X}_{iT} \setminus A_{it}\}, U_i$$

- Very strong assumption: basically that A_{it} is randomized within unit-periods, with no feedback.

Assumptions

- Let's make a **strict ignorability** assumption conditional on U_i:

$$Y_{it}(a) \perp\!\!\!\perp A_{it} | \{\underline{X}_{iT} \setminus A_{it}\}, U_i$$

- Very strong assumption: basically that A_{it} is randomized within unit-periods, with no feedback.
- Does this imply the strict exogeneity assumption on our combined error?

$$E[\eta_{it} | \underline{X}_{iT}, U_i, A_{it}] = 0$$

Assumptions

- Let's make a **strict ignorability** assumption conditional on U_i:

$$Y_{it}(a) \perp\!\!\!\perp A_{it} | \{\underline{X}_{iT} \setminus A_{it}\}, U_i$$

- Very strong assumption: basically that A_{it} is randomized within unit-periods, with no feedback.
- Does this imply the strict exogeneity assumption on our combined error?

$$E[\eta_{it} | \underline{X}_{iT}, U_i, A_{it}] = 0$$

- Let's find out!

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Let's take the "typical errors" first:

$$E\left[\left(Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]\right)|\underline{X}_{iT}, U_i, A_{it}\right]$$

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Let's take the "typical errors" first:

$$E\left[\left(Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]\right)|\underline{X}_{iT}, U_i, A_{it}\right]$$

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Let's take the "typical errors" first:

$$E\left[(Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i])\,|\underline{X}_{iT}, U_i, A_{it}\right]$$
$$= E[Y_{it}(0)|\underline{X}_{iT}, U_i, A_{it}] - E[Y_{it}(0)|\underline{X}_{iT}, U_i]$$

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Let's take the "typical errors" first:

$$\begin{aligned} & E\left[\left(Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]\right)|\underline{X}_{iT}, U_i, A_{it}\right] \\ & = E[Y_{it}(0)|\underline{X}_{iT}, U_i, A_{it}] - E[Y_{it}(0)|\underline{X}_{iT}, U_i] \\ & = E[Y_{it}(0)|\underline{X}_{iT}, U_i] - E[Y_{it}(0)|\underline{X}_{iT}, U_i] \end{aligned}$$

Errors

- Remember the decomposition:

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Let's take the "typical errors" first:

$$\begin{aligned}
&E\left[\left(Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]\right)|\underline{X}_{iT}, U_i, A_{it}\right] \\
&= E[Y_{it}(0)|\underline{X}_{iT}, U_i, A_{it}] - E[Y_{it}(0)|\underline{X}_{iT}, U_i] \\
&= E[Y_{it}(0)|\underline{X}_{iT}, U_i] - E[Y_{it}(0)|\underline{X}_{iT}, U_i] \\
&= 0
\end{aligned}$$

- Great, the "typical errors" are 0 on average under our strict ignorability assumption.

Non-constant effects errors

- What about the non-constant effects part?

$$E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] = A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}])$$

Non-constant effects errors

- What about the non-constant effects part?

$$E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] = A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}])$$

Non-constant effects errors

- What about the non-constant effects part?

$$\begin{aligned} E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] &= A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}]) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i, A_{it}] - \tau_c) \end{aligned}$$

Non-constant effects errors

- What about the non-constant effects part?

$$\begin{aligned} E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] &= A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}]) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i, {\color{red}A_{it}}] - \tau_c) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i] - \tau_c) \end{aligned}$$

Non-constant effects errors

- What about the non-constant effects part?

$$\begin{aligned} E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] &= A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}]) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i, A_{it}] - \tau_c) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i] - \tau_c) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i] - E[\tau_{it}]) \end{aligned}$$

Non-constant effects errors

- What about the non-constant effects part?

$$\begin{aligned} E[A_{it}(\tau_{it} - \tau_c)|\underline{X}_{iT}, U_i, A_{it}] &= A_{it}(E[\tau_{it} - \tau_c|\underline{X}_{iT}, U_i, A_{it}]) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i, A_{it}] - \tau_c) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i] - \tau_c) \\ &= A_{it}(E[\tau_{it}|\underline{X}_{iT}, U_i] - E[\tau_{it}]) \end{aligned}$$

- Thus, we can see that the combined error will only satisfy the strict exogeneity assumption of fixed effects when

$$E[\tau_{it}|\underline{X}_{iT}, U_i] = E[\tau_{it}]$$

Non-constant effects errors

- What about the non-constant effects part?

$$\begin{aligned} E[A_{it}(\tau_{it} - \tau_c) | \underline{X}_{iT}, U_i, A_{it}] &= A_{it}(E[\tau_{it} - \tau_c | \underline{X}_{iT}, U_i, A_{it}]) \\ &= A_{it}(E[\tau_{it} | \underline{X}_{iT}, U_i, A_{it}] - \tau_c) \\ &= A_{it}(E[\tau_{it} | \underline{X}_{iT}, U_i] - \tau_c) \\ &= A_{it}(E[\tau_{it} | \underline{X}_{iT}, U_i] - E[\tau_{it}]) \end{aligned}$$

- Thus, we can see that the combined error will only satisfy the strict exogeneity assumption of fixed effects when

$$E[\tau_{it} | \underline{X}_{iT}, U_i] = E[\tau_{it}]$$

- This is when the treatment effects are independent of the unit effects and the covariates.

Regression bias?

- We've seen this before: it's a general problem with regression and varying treatment effects.

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

Regression bias?

- We've seen this before: it's a general problem with regression and varying treatment effects.

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Generally the issue here is that non-constant effects induce correlation between the treatment and the error term.

Regression bias?

- We've seen this before: it's a general problem with regression and varying treatment effects.

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Generally the issue here is that non-constant effects induce correlation between the treatment and the error term.
- Distinct from confounding bias since we could, in principle, estimate $E[\tau_{it}|\underline{X}_{iT}, U_i]$ to then calculate $E[\tau_{it}]$

Regression bias?

- We've seen this before: it's a general problem with regression and varying treatment effects.

$$\eta_{it} = \underbrace{A_{it}(\tau_{it} - \tau_c)}_{\text{non-constant effects}} + \underbrace{Y_{it}(0) - E[Y_{it}(0)|\underline{X}_{iT}, U_i]}_{\text{typical errors}}$$

- Generally the issue here is that non-constant effects induce correlation between the treatment and the error term.
- Distinct from confounding bias since we could, in principle, estimate $E[\tau_{it}|\underline{X}_{iT}, U_i]$ to then calculate $E[\tau_{it}]$
- Overall ATE still nonparametrically identified, even if the FE regression doesn't estimate it.

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:
  - A_{it} affects Y_{it} which then affects $A_{i,t+1}$

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:
  - A_{it} affects Y_{it} which then affects $A_{i,t+1}$
  - Basically, any feedback between treatment and the outcome

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:
  - A_{it} affects Y_{it} which then affects $A_{i,t+1}$
  - Basically, any feedback between treatment and the outcome
- Can we weaken this? Yes! **Sequential ignorability**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0$$

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:
  - A_{it} affects Y_{it} which then affects $A_{i,t+1}$
  - Basically, any feedback between treatment and the outcome
- Can we weaken this? Yes! **Sequential ignorability**:

$$E[\varepsilon_{it} | \underline{X}_{it}, \underline{A}_{it}, U_i] = 0$$

- Note here that the we only condition up to t so that the errors can be correlated with future $A_{i,t+1}$ and so on.

Strict exogeneity/ignorability

- We've needed strict exogeneity to derive many of these results, but this is very restrictive.
- Rules out the following:
  - A_{it} affects Y_{it} which then affects $A_{i,t+1}$
  - Basically, any feedback between treatment and the outcome
- Can we weaken this? Yes! **Sequential ignorability**:

$$E[\varepsilon_{it}|\underline{X}_{it}, \underline{A}_{it}, U_i] = 0$$

- Note here that the we only condition up to t so that the errors can be correlated with future $A_{i,t+1}$ and so on.
- Estimation here gets more difficult (see Wooldridge, 2002, 11.2)

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?
- Very difficult, if not impossible with fixed effects models.

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?
- Very difficult, if not impossible with fixed effects models.
- Why?

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?
- Very difficult, if not impossible with fixed effects models.
- Why?
  - For cumulative effects, we need to consider the effects of treatment on time-varying confounders, $X_{it}(\underline{a}_{i,t-1})$.

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?
- Very difficult, if not impossible with fixed effects models.
- Why?
  - For cumulative effects, we need to consider the effects of treatment on time-varying confounders, $X_{it}(\underline{a}_{i,t-1})$.
  - Those pathways might be hard to identify

Contemporaneous vs Cumulative effects

- Another assumption we've been making is that there is only a contemporaneous effect: τA_{it}.
- Implicitly or explicitly fixing the past history of the treatment.
- What if we want to estimate the cumulative effects?
- Very difficult, if not impossible with fixed effects models.
- Why?
  - For cumulative effects, we need to consider the effects of treatment on time-varying confounders, $X_{it}(\underline{a}_{i,t-1})$.
  - Those pathways might be hard to identify
- We would just need more assumptions

Basic differences-in-differences model

Setup

- Basic setup: two groups, two time periods.

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome
- Differences in differences (diff-in-diff, DID, DD): difference between $t = 1$ to $t = 2$ changes in treatment and control groups

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome
- Differences in differences (diff-in-diff, DID, DD): difference between $t = 1$ to $t = 2$ changes in treatment and control groups
  - Resolution: changes in the control group identifies the secular trend

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome
- Differences in differences (diff-in-diff, DID, DD): difference between $t = 1$ to $t = 2$ changes in treatment and control groups
  - Resolution: changes in the control group identifies the secular trend
- Examples:

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome
- Differences in differences (diff-in-diff, DID, DD): difference between $t = 1$ to $t = 2$ changes in treatment and control groups
  - Resolution: changes in the control group identifies the secular trend
- Examples:
  - Minimum wage changes in NJ with PA as control (Card and Krueger)

Setup

- Basic setup: two groups, two time periods.
- At $t = 0$, neither group is treated and in period $t = 1$, one (and only one) of the groups is treated.
- Differences: changes in treated group from $t = 0$ to $t = 1$
  - Problem: Might be secular changes in the outcome
- Differences in differences (diff-in-diff, DID, DD): difference between $t = 1$ to $t = 2$ changes in treatment and control groups
  - Resolution: changes in the control group identifies the secular trend
- Examples:
  - Minimum wage changes in NJ with PA as control (Card and Krueger)
  - Effect of artillery shelling on insurgent attacks (Lyall)

Constant effects linear DID model

- Focus on fixed effect setup with two periods, pretreatment ($t = 0$) and posttreatment ($t = 1$)

Constant effects linear DID model

- Focus on fixed effect setup with two periods, pretreatment ($t = 0$) and posttreatment ($t = 1$)
- By design, we have $A_{i0} = 0$ for all i

Constant effects linear DID model

- Focus on fixed effect setup with two periods, pretreatment ($t = 0$) and posttreatment ($t = 1$)
- By design, we have $A_{i0} = 0$ for all i
- Start with linear models, then move to potential outcomes

Constant effects linear DID model

- Focus on fixed effect setup with two periods, pretreatment ($t = 0$) and posttreatment ($t = 1$)
- By design, we have $A_{i0} = 0$ for all i
- Start with linear models, then move to potential outcomes
- The specific model we will assume is this:

$$Y_{it} = \delta_t + \tau A_{it} + \alpha_i + \eta_{it}$$

Constant effects linear DID model

- Focus on fixed effect setup with two periods, pretreatment ($t = 0$) and posttreatment ($t = 1$)
- By design, we have $A_{i0} = 0$ for all i
- Start with linear models, then move to potential outcomes
- The specific model we will assume is this:

$$Y_{it} = \delta_t + \tau A_{it} + \alpha_i + \eta_{it}$$

- Here we have a period effect, δ_t and a unit effect α_i, and a transitory shock, η_{it}, which has mean zero.

Common trends

- Without further assumptions, τ not identified because A_{i1} might be correlated with shocks.

$$\Pr[A_{i1} = 1 | \eta_{i0}, \eta_{i1}] = \Pr[A_{i1} = 1]$$

Common trends

- Without further assumptions, τ not identified because A_{i1} might be correlated with shocks.

$$\Pr[A_{i1} = 1 | \eta_{i0}, \eta_{i1}] = \Pr[A_{i1} = 1]$$

- Errors are also independent of the treatment, so that $\eta_{i1} - \eta_{i0}$ is independent of A_{it}.

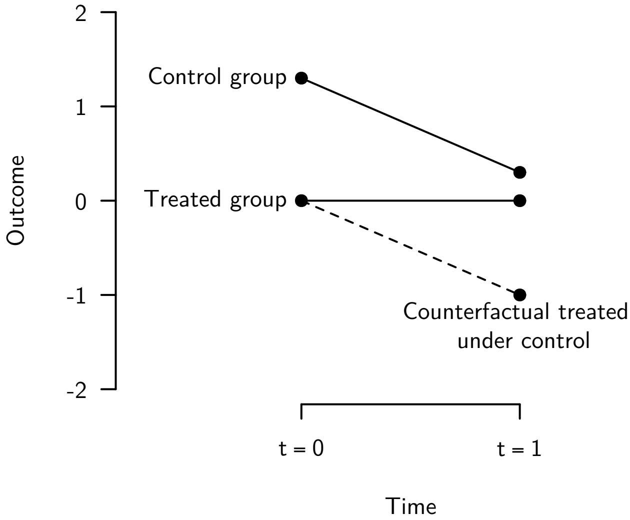
Common trends

- Without further assumptions, τ not identified because A_{i1} might be correlated with shocks.

$$\Pr[A_{i1} = 1 | \eta_{i0}, \eta_{i1}] = \Pr[A_{i1} = 1]$$

- Errors are also independent of the treatment, so that $\eta_{i1} - \eta_{i0}$ is independent of A_{it}.
- Specifically, this means treated and control groups have the same trends in the error (on average)

Common trends in a graph



Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\delta = (\delta_1 - \delta_0) \qquad \text{(time trend)}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\delta = (\delta_1 - \delta_0) \qquad \text{(time trend)}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\delta = (\delta_1 - \delta_0) \qquad \text{(time trend)}$$
$$\mu = E[\alpha_i | A_{i1} = 0] + \delta_0 \qquad \text{(control start)}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\delta = (\delta_1 - \delta_0) \quad \text{(time trend)}$$
$$\mu = E[\alpha_i | A_{i1} = 0] + \delta_0 \quad \text{(control start)}$$
$$\gamma = E[\alpha_i | A_{i1} = 1] - E[\alpha_i | A_{i1} = 0] \quad \text{(baseline difference)}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\begin{aligned}
\delta &= (\delta_1 - \delta_0) && \text{(time trend)} \\
\mu &= E[\alpha_i | A_{i1} = 0] + \delta_0 && \text{(control start)} \\
\gamma &= E[\alpha_i | A_{i1} = 1] - E[\alpha_i | A_{i1} = 0] && \text{(baseline difference)} \\
\varepsilon_{it} &= \alpha_i - E[\alpha_i | A_{i1}] + \eta_{it} && \text{(new error)}
\end{aligned}$$

Identification

- With this assumption, we can rewrite the above model as the following:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- The parameters are the following:

$$\begin{aligned}
\delta &= (\delta_1 - \delta_0) && \text{(time trend)} \\
\mu &= E[\alpha_i | A_{i1} = 0] + \delta_0 && \text{(control start)} \\
\gamma &= E[\alpha_i | A_{i1} = 1] - E[\alpha_i | A_{i1} = 0] && \text{(baseline difference)} \\
\varepsilon_{it} &= \alpha_i - E[\alpha_i | A_{i1}] + \eta_{it} && \text{(new error)}
\end{aligned}$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$E[\varepsilon_{it}|A_{i1}, A_{i0}] = E[\varepsilon_{it}|A_{i1}]$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$E[\varepsilon_{it}|A_{i1}, A_{i0}] = E[\varepsilon_{it}|A_{i1}]$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$\begin{aligned} E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\ &= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \end{aligned}$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$
\begin{aligned}
E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\
&= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \\
&= E[\alpha_i|A_{i1}] - E[E[\alpha_i|A_{i1}]|A_{i1}] + E[\eta_{it}|A_{i1}]
\end{aligned}
$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$
\begin{aligned}
E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\
&= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \\
&= E[\alpha_i|A_{i1}] - E[E[\alpha_i|A_{i1}]|A_{i1}] + E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}|A_{i1}]
\end{aligned}
$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$
\begin{aligned}
E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\
&= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \\
&= E[\alpha_i|A_{i1}] - E[E[\alpha_i|A_{i1}]|A_{i1}] + E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}] = 0
\end{aligned}
$$

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$
\begin{aligned}
E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\
&= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \\
&= E[\alpha_i|A_{i1}] - E[E[\alpha_i|A_{i1}]|A_{i1}] + E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}] = 0
\end{aligned}
$$

- Note that we have this even though we have made no assumptions on the distribution of the unit-specific effects and their relation to the treatment.

- Using the above assumption, we can show that the treatment is independent of the error in this model:

$$
\begin{aligned}
E[\varepsilon_{it}|A_{i1}, A_{i0}] &= E[\varepsilon_{it}|A_{i1}] \\
&= E[(\alpha_i - E[\alpha_i|A_{i1}] + \eta_{it})|A_{i1}] \\
&= E[\alpha_i|A_{i1}] - E[E[\alpha_i|A_{i1}]|A_{i1}] + E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}|A_{i1}] \\
&= E[\eta_{it}] = 0
\end{aligned}
$$

- Note that we have this even though we have made no assumptions on the distribution of the unit-specific effects and their relation to the treatment.
- Just assumed that control and treatment have the same average secular trends

- Now, we can investigate how two differences here. First, the time trend for the untreated:

$$E[Y_{i1}|A_{i1} = 0] - E[Y_{i0}|A_{i1} = 0] = \delta$$

- Now, we can investigate how two differences here. First, the time trend for the untreated:

$$E[Y_{i1}|A_{i1}=0]-E[Y_{i0}|A_{i1}=0]=\delta$$

- And now the trend for the treated group:

$$E[Y_{i1}|A_{i1}=1]-E[Y_{i0}|A_{i1}=1]=\delta+\tau$$

- Now, we can investigate how two differences here. First, the time trend for the untreated:

$$E[Y_{i1}|A_{i1} = 0] - E[Y_{i0}|A_{i1} = 0] = \delta$$

- And now the trend for the treated group:

$$E[Y_{i1}|A_{i1} = 1] - E[Y_{i0}|A_{i1} = 1] = \delta + \tau$$

- Therefore:

$$\begin{aligned} &(E[Y_{i1}|A_{i1} = 1] - E[Y_{i0}|A_{i1} = 1]) \\ -&(E[Y_{i1}|A_{i1} = 0] - E[Y_{i0}|A_{i1} = 0]) = \tau \end{aligned}$$

- Now, we can investigate how two differences here. First, the time trend for the untreated:

$$E[Y_{i1}|A_{i1} = 0] - E[Y_{i0}|A_{i1} = 0] = \delta$$

- And now the trend for the treated group:

$$E[Y_{i1}|A_{i1} = 1] - E[Y_{i0}|A_{i1} = 1] = \delta + \tau$$

- Therefore:

$$\begin{gathered}(E[Y_{i1}|A_{i1} = 1] - E[Y_{i0}|A_{i1} = 1]) \\ -(E[Y_{i1}|A_{i1} = 0] - E[Y_{i0}|A_{i1} = 0]) = \tau\end{gathered}$$

- This motivates the differences-in-differences estimator as the difference between these two differences. We can estimate each of these CEFs from the data and compute their sample versions to get an estimate of τ.

Estimation

- For the two period, binary treatment case, a regression of the outcome on time (pre-treatment, post-treatment), treated group, and their interaction can estimate τ:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{i1} \cdot t + \varepsilon_{it}$$

Estimation

- For the two period, binary treatment case, a regression of the outcome on time (pre-treatment, post-treatment), treated group, and their interaction can estimate τ:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{i1} \cdot t + \varepsilon_{it}$$

- $\widehat{\tau}$ would be the coefficient on the interaction between time and the treatment.

Estimation

- For the two period, binary treatment case, a regression of the outcome on time (pre-treatment, post-treatment), treated group, and their interaction can estimate τ:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{i1} \cdot t + \varepsilon_{it}$$

- $\widehat{\tau}$ would be the coefficient on the interaction between time and the treatment.
- This is for two cross-sections.

Estimation

- For the two period, binary treatment case, a regression of the outcome on time (pre-treatment, post-treatment), treated group, and their interaction can estimate τ:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{i1} \cdot t + \varepsilon_{it}$$

- $\widehat{\tau}$ would be the coefficient on the interaction between time and the treatment.
- This is for two cross-sections.
- If we have panel data, then we can estimate this in a different, more direct way. Note that:

$$\tau = E[Y_{i1} - Y_{i0} | A_{i1} = 1] - E[Y_{i1} - Y_{i0} | A_{i1} = 0]$$

Estimation

- For the two period, binary treatment case, a regression of the outcome on time (pre-treatment, post-treatment), treated group, and their interaction can estimate τ:

$$Y_{it} = \mu + \delta \cdot t + \gamma A_{i1} + \tau A_{i1} \cdot t + \varepsilon_{it}$$

- $\widehat{\tau}$ would be the coefficient on the interaction between time and the treatment.
- This is for two cross-sections.
- If we have panel data, then we can estimate this in a different, more direct way. Note that:

$$\tau = E[Y_{i1} - Y_{i0} | A_{i1} = 1] - E[Y_{i1} - Y_{i0} | A_{i1} = 0]$$

- Thus, in the panel data case, we can estimate the effect by regressing the change for each unit, $Y_{i1} - Y_{i0}$, on the treatment.

Threats to identification

- Obviously, the treatment needs to be independent of the idiosyncratic shocks so that the variation of the outcome is the same for the treated and control groups, but this might not be plausible.

Threats to identification

- Obviously, the treatment needs to be independent of the idiosyncratic shocks so that the variation of the outcome is the same for the treated and control groups, but this might not be plausible.
- Ashenfelter's dip: which is a empirical finding that people who enroll in job training programs see their earnings decline prior to that training.

Threats to identification

- Obviously, the treatment needs to be independent of the idiosyncratic shocks so that the variation of the outcome is the same for the treated and control groups, but this might not be plausible.
- Ashenfelter's dip: which is a empirical finding that people who enroll in job training programs see their earnings decline prior to that training.
- In the Lyall paper, it might be the case that insurgent attacks might be falling in places where there is shelling because rebels attacked in those areas and have moved on.

Threats to identification

- Obviously, the treatment needs to be independent of the idiosyncratic shocks so that the variation of the outcome is the same for the treated and control groups, but this might not be plausible.
- Ashenfelter's dip: which is a empirical finding that people who enroll in job training programs see their earnings decline prior to that training.
- In the Lyall paper, it might be the case that insurgent attacks might be falling in places where there is shelling because rebels attacked in those areas and have moved on.
- Thus, the independence of the treatment and idiosyncratic shocks might only hold conditional on covariates.

Regression DD

- Sometimes key assumption only holds conditional on X_i.

Regression DD

- Sometimes key assumption only holds conditional on X_i.
- Can incorporate those with a regression DID, which includes covariates in a linear, additive manner:

$$Y_{it} = \mu + X_i' \beta_t + \delta t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

Regression DD

- Sometimes key assumption only holds conditional on X_i.
- Can incorporate those with a regression DID, which includes covariates in a linear, additive manner:

$$Y_{it} = \mu + X_i'\beta_t + \delta t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- If we have repeated observations, we can take the differences between $t = 0$ and $t = 1$:

$$Y_{i1} - Y_{i0} = \delta + X_i'\beta + \tau(A_{i1} - A_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

Regression DD

- Sometimes key assumption only holds conditional on X_i.
- Can incorporate those with a regression DID, which includes covariates in a linear, additive manner:

$$Y_{it} = \mu + X_i'\beta_t + \delta t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- If we have repeated observations, we can take the differences between $t = 0$ and $t = 1$:

$$Y_{i1} - Y_{i0} = \delta + X_i'\beta + \tau(A_{i1} - A_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

- Here, we have $\beta = \beta_1 - \beta_0$. Further note that because everyone is untreated in the first period, $A_{i1} - A_{i0} = A_{i1}$.

Regression DD

- Sometimes key assumption only holds conditional on X_i.
- Can incorporate those with a regression DID, which includes covariates in a linear, additive manner:

$$Y_{it} = \mu + X_i'\beta_t + \delta t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- If we have repeated observations, we can take the differences between $t = 0$ and $t = 1$:

$$Y_{i1} - Y_{i0} = \delta + X_i'\beta + \tau(A_{i1} - A_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

- Here, we have $\beta = \beta_1 - \beta_0$. Further note that because everyone is untreated in the first period, $A_{i1} - A_{i0} = A_{i1}$.
- As usual, for panel data, regress changes on treatment.

Regression DD

- Sometimes key assumption only holds conditional on X_i.
- Can incorporate those with a regression DID, which includes covariates in a linear, additive manner:

$$Y_{it} = \mu + X_i'\beta_t + \delta t + \gamma A_{i1} + \tau A_{it} + \varepsilon_{it}$$

- If we have repeated observations, we can take the differences between $t = 0$ and $t = 1$:

$$Y_{i1} - Y_{i0} = \delta + X_i'\beta + \tau(A_{i1} - A_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

- Here, we have $\beta = \beta_1 - \beta_0$. Further note that because everyone is untreated in the first period, $A_{i1} - A_{i0} = A_{i1}$.
- As usual, for panel data, regress changes on treatment.
- This approach depends on constant effects and linearity in X_i. Can we generalize?

Robustness checks

- Lags and Leads

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.
  - Include lags $A_{i,t-1}, A_{i,t-1}$ etc and leads $A_{i,t+1}, A_{i,t+2}$ in the model and see if pattern holds.

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.
  - Include lags $A_{i,t-1}, A_{i,t-1}$ etc and leads $A_{i,t+1}, A_{i,t+2}$ in the model and see if pattern holds.
- Time trends

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.
  - Include lags $A_{i,t-1}, A_{i,t-1}$ etc and leads $A_{i,t+1}, A_{i,t+2}$ in the model and see if pattern holds.
- Time trends
- With $T > 2$, we can add unit-specific linear trends to the regression DID model:

$$Y_{it} = \delta_t + \tau A_{it} + \alpha_{0i} + \alpha_{1i} \cdot t + \eta_{it}$$

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.
  - Include lags $A_{i,t-1}, A_{i,t-1}$ etc and leads $A_{i,t+1}, A_{i,t+2}$ in the model and see if pattern holds.
- Time trends
- With $T > 2$, we can add unit-specific linear trends to the regression DID model:

$$Y_{it} = \delta_t + \tau A_{it} + \alpha_{0i} + \alpha_{1i} \cdot t + \eta_{it}$$

  - Helps detect if there really are varying trends, if estimated from pre-treatment data.

Robustness checks

- Lags and Leads
  - if A_{it} causes Y_{it}, and not the other way around, then current and lagged values of A_{it} should have an effect on Y_{it}, but future values of A_{it} should not.
  - Include lags $A_{i,t-1}, A_{i,t-1}$ etc and leads $A_{i,t+1}, A_{i,t+2}$ in the model and see if pattern holds.
- Time trends
- With $T > 2$, we can add unit-specific linear trends to the regression DID model:

$$Y_{it} = \delta_t + \tau A_{it} + \alpha_{0i} + \alpha_{1i} \cdot t + \eta_{it}$$

  - Helps detect if there really are varying trends, if estimated from pre-treatment data.
  - Synthetic control matching leverages this type of idea

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Because no one is treated at time $t = 0$, we have

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Because no one is treated at time $t = 0$, we have
  - $A_{i1} = A_i$

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Because no one is treated at time $t = 0$, we have
  - $A_{i1} = A_i$
  - $Y_{i0}(0) = Y_{i0}$

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Because no one is treated at time $t = 0$, we have
  - $A_{i1} = A_i$
  - $Y_{i0}(0) = Y_{i0}$
  - $Y_{i1} = A_i Y_{i1}(1) + (1 - A_i) Y_{i1}(0)$

Potential outcomes approach to DID

- Parametric models aren't cool. You what's cool? Nonparametric identification.
- Let $Y_{it}(a)$ be the potential outcome under treatment a at time t.
- Again, the individual causal effect is just $Y_{it}(1) - Y_{it}(0)$.
- Because no one is treated at time $t = 0$, we have
  - $A_{i1} = A_i$
  - $Y_{i0}(0) = Y_{i0}$
  - $Y_{i1} = A_i Y_{i1}(1) + (1 - A_i) Y_{i1}(0)$
- We'll focus on two estimands, the ATT,

$$\tau_{ATT} = E[Y_{it}(1) - Y_{it}(0) | A_i = 1]$$

and the conditional ATT:

$$\tau_{ATT}(x) = E[Y_{it}(1) - Y_{it}(0) | X_i = x, A_i = 1]$$

.

Nonparametric identification

- Let's make the crucial identifying assumption of a DID model:

$$E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]$$

Nonparametric identification

- Let's make the crucial identifying assumption of a DID model:

$$E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]$$

- 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

- Let's make the crucial identifying assumption of a DID model:

$$E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]$$

- 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.
- Just parallel trends in terms of potential outcomes.

Nonparametric identification

- Let's make the crucial identifying assumption of a DID model:

$$E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1] = E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]$$

- 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.
- Just parallel trends in terms of potential outcomes.
- Note that, if the two groups have the same mean potential outcome under control in the first period,

$$E[Y_{i0}(0)|X_i, A_i = 1] = E[Y_{i0}(0)|X_i, A_i = 0]$$

then this assumption just becomes regular ignorability:

$$E[Y_{i1}(1)|X_i, A_i = 1] = E[Y_{i1}(1)|X_i, A_i = 0]$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0) | X_i, A_i = 1] \\
= & E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0) | X_i, A_i = 1]
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
&E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=&E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1]
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
= & E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
= & (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1])
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=& E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1]) \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0])
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=& E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1]) \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]) \\
=& (E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0)|X_i, A_i = 0] - E[Y_{i0}(0)|X_i, A_i = 0])
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=& E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1]) \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]) \\
=& (E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0)|X_i, A_i = 0] - E[Y_{i0}(0)|X_i, A_i = 0]) \\
=& \underbrace{(E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1])}_{\text{differences for } A_i=1} - \underbrace{(E[Y_{i1}|X_i, A_i = 0] - E[Y_{i0}|X_i, A_i = 0])}_{\text{differences for } A_i=0}
\end{aligned}
$$

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=& E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1]) \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]) \\
=& (E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0)|X_i, A_i = 0] - E[Y_{i0}(0)|X_i, A_i = 0]) \\
=& \underbrace{(E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1])}_{\text{differences for } A_i=1} - \underbrace{(E[Y_{i1}|X_i, A_i = 0] - E[Y_{i0}|X_i, A_i = 0])}_{\text{differences for } A_i=0}
\end{aligned}
$$

- Very similar to results above.

- We can show that this is the key assumption for identifying the DID approach:

$$
\begin{aligned}
& E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] \\
=& E[Y_{i1}(1) - Y_{i0}(0) + Y_{i0}(0) - Y_{i1}(0)|X_i, A_i = 1] \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}(0)|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 1]) \\
=& (E[Y_{i1}(1)|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]) \\
=& (E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1]) - (E[Y_{i1}(0)|X_i, A_i = 0] - E[Y_{i0}(0)|X_i, A_i = 0]) \\
=& \underbrace{(E[Y_{i1}|X_i, A_i = 1] - E[Y_{i0}|X_i, A_i = 1])}_{\text{differences for } A_i=1} - \underbrace{(E[Y_{i1}|X_i, A_i = 0] - E[Y_{i0}|X_i, A_i = 0])}_{\text{differences for } A_i=0}
\end{aligned}
$$

- Very similar to results above.
- Each CEF could be estimated nonparametrically, but we would run into the curse of dimensionality if X_i is complicated

Nonparametric DID notes

- Note what is powerful here: no ignorability assumption.

Nonparametric DID notes

- Note what is powerful here: no ignorability assumption.
- Relies only on parallel trends assumption.

Nonparametric DID notes

- Note what is powerful here: no ignorability assumption.
- Relies only on parallel trends assumption.
- Parallel trends might be more plausible when treatment and control are similar

Nonparametric DID notes

- Note what is powerful here: no ignorability assumption.
- Relies only on parallel trends assumption.
- Parallel trends might be more plausible when treatment and control are similar
- Also, parallel trends assumption may not hold for transformations of the data.

Nonparametric DID notes

- Note what is powerful here: no ignorability assumption.
- Relies only on parallel trends assumption.
- Parallel trends might be more plausible when treatment and control are similar
- Also, parallel trends assumption may not hold for transformations of the data.
- Nonparametrics will hard with moderately-sized covariate space

Semiparametric estimation with repeated outcomes

- Abadie (2005) on how to use weighting estimators to help with estimation.

Semiparametric estimation with repeated outcomes

- Abadie (2005) on how to use weighting estimators to help with estimation.
- Basically, we are going to weight the treated and control groups so that they are balanced on the covariates.

Semiparametric estimation with repeated outcomes

- Abadie (2005) on how to use weighting estimators to help with estimation.
- Basically, we are going to weight the treated and control groups so that they are balanced on the covariates.
- Recover the "hidden" balanced experiment.

Semiparametric estimation with repeated outcomes

- Abadie (2005) on how to use weighting estimators to help with estimation.
- Basically, we are going to weight the treated and control groups so that they are balanced on the covariates.
- Recover the "hidden" balanced experiment.

Semiparametric estimation with repeated outcomes

- Abadie (2005) on how to use weighting estimators to help with estimation.
- Basically, we are going to weight the treated and control groups so that they are balanced on the covariates.
- Recover the "hidden" balanced experiment.

$$E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] = E\left[\frac{A_i(Y_{i1} - Y_{i0})}{\Pr[A_i = 1|X_i]} - \frac{(1 - A_i)(Y_{i1} - Y_{i0})}{1 - \Pr[A_i = 1|X_i]}\middle| X_i\right]$$

- The tradeoff here is that we have to estimate the propensity score to estimate these weights for each unit:

$$\rho_0(A_i, X_i) = \frac{A_i - \Pr[A_i = 1|X_i]}{\Pr[A_i = 1|X_i](1 - \Pr[A_i = 1|X_i])}$$

How weighting works

- Intuitively, weighting is very similar to survey weights.

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.
- We want to weight the "sample" to be similar to/representative of that balanced "population"

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.
- We want to weight the "sample" to be similar to/representative of that balanced "population"
- This is a general technique, not limited to DID.

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.
- We want to weight the "sample" to be similar to/representative of that balanced "population"
- This is a general technique, not limited to DID.
- Weights for treated units are: $\frac{1}{\Pr[A_i=1|X_i]}$

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.
- We want to weight the "sample" to be similar to/representative of that balanced "population"
- This is a general technique, not limited to DID.
- Weights for treated units are: $\frac{1}{\Pr[A_i=1|X_i]}$
- Weights for control units are: $\frac{1}{\Pr[A_i=0|X_i]}$

How weighting works

- Intuitively, weighting is very similar to survey weights.
- If the data were perfectly balanced on the covariates, we wouldn't need to control for them.
- We want to weight the "sample" to be similar to/representative of that balanced "population"
- This is a general technique, not limited to DID.
- Weights for treated units are: $\frac{1}{\Pr[A_i=1|X_i]}$
- Weights for control units are: $\frac{1}{\Pr[A_i=0|X_i]}$
- Sometimes called "inverse probability of treatment weighting" (IPTW)

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1] \Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0] \Pr[A_i = 0|X_i]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1] \Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0] \Pr[A_i = 0|X_i]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1] \Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0] \Pr[A_i = 0|X_i] \\
= \quad & E\left[\frac{1}{\Pr[A_i = 1|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 1\right] \Pr[A_i = 1|X_i] \\
& + E\left[\frac{1}{\Pr[A_i = 0|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 0\right] \Pr[A_i = 0|X_i]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1]\,\Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0]\,\Pr[A_i = 0|X_i] \\
= \quad & E\left[\frac{1}{\Pr[A_i = 1|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 1\right]\Pr[A_i = 1|X_i] \\
& + E\left[\frac{1}{\Pr[A_i = 0|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 0\right]\Pr[A_i = 0|X_i] \\
= & E[Y_{i1} - Y_{i0}|X_i, A_i = 1] - E[Y_{i1} - Y_{i0}|X_i, A_i = 0]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1]\Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0]\Pr[A_i = 0|X_i] \\
= \quad & E\left[\frac{1}{\Pr[A_i = 1|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 1\right]\Pr[A_i = 1|X_i] \\
& + E\left[\frac{1}{\Pr[A_i = 0|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 0\right]\Pr[A_i = 0|X_i] \\
= & E[Y_{i1} - Y_{i0}|X_i, A_i = 1] - E[Y_{i1} - Y_{i0}|X_i, A_i = 0] \\
= & E[Y_{i1}(1) - Y_{i0}(1)|X_i, A_i = 1] - E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1}-Y_{i0})|X_i] = & \quad E[\rho_0(A_i,X_i)(Y_{i1}-Y_{i0})|X_i,A_i=1]\Pr[A_i=1|X_i] \\
& + E[\rho_0(A_i,X_i)(Y_{i1}-Y_{i0})|X_i,A_i=0]\Pr[A_i=0|X_i] \\
= & \quad E\left[\frac{1}{\Pr[A_i=1|X_i]}(Y_{i1}-Y_{i0})\middle|X_i,A_i=1\right]\Pr[A_i=1|X_i] \\
& + E\left[\frac{1}{\Pr[A_i=0|X_i]}(Y_{i1}-Y_{i0})\middle|X_i,A_i=0\right]\Pr[A_i=0|X_i] \\
= & E[Y_{i1}-Y_{i0}|X_i,A_i=1]-E[Y_{i1}-Y_{i0}|X_i,A_i=0] \\
= & E[Y_{i1}(1)-Y_{i0}(1)|X_i,A_i=1]-E[Y_{i1}(0)-Y_{i0}(0)|X_i,A_i=0] \\
= & E[Y_{i1}(1)-Y_{i1}(0)|X_i,A_i=1]-E[Y_{i0}(1)-Y_{i0}(0)|X_i,A_i=1]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = &\quad E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1]\Pr[A_i = 1|X_i] \\
&+ E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0]\Pr[A_i = 0|X_i] \\
= &\quad E\left[\frac{1}{\Pr[A_i = 1|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 1\right]\Pr[A_i = 1|X_i] \\
&+ E\left[\frac{1}{\Pr[A_i = 0|X_i]}(Y_{i1} - Y_{i0})\middle|X_i, A_i = 0\right]\Pr[A_i = 0|X_i] \\
= &E[Y_{i1} - Y_{i0}|X_i, A_i = 1] - E[Y_{i1} - Y_{i0}|X_i, A_i = 0] \\
= &E[Y_{i1}(1) - Y_{i0}(1)|X_i, A_i = 1] - E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0] \\
= &E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] - E[Y_{i0}(1) - Y_{i0}(0)|X_i, A_i = 1] \\
= &E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] - E[Y_{i0}(0) - Y_{i0}(0)|X_i, A_i = 1]
\end{aligned}
$$

Proof

- The proof is actually quite straightforward:

$$
\begin{aligned}
E[\rho_0(Y_{i1} - Y_{i0})|X_i] = \quad & E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 1] \Pr[A_i = 1|X_i] \\
& + E[\rho_0(A_i, X_i)(Y_{i1} - Y_{i0})|X_i, A_i = 0] \Pr[A_i = 0|X_i] \\
= \quad & E\left[\frac{1}{\Pr[A_i = 1|X_i]}(Y_{i1} - Y_{i0})\middle| X_i, A_i = 1\right] \Pr[A_i = 1|X_i] \\
& + E\left[\frac{1}{\Pr[A_i = 0|X_i]}(Y_{i1} - Y_{i0})\middle| X_i, A_i = 0\right] \Pr[A_i = 0|X_i] \\
= & E[Y_{i1} - Y_{i0}|X_i, A_i = 1] - E[Y_{i1} - Y_{i0}|X_i, A_i = 0] \\
= & E[Y_{i1}(1) - Y_{i0}(1)|X_i, A_i = 1] - E[Y_{i1}(0) - Y_{i0}(0)|X_i, A_i = 0] \\
= & E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] - E[Y_{i0}(1) - Y_{i0}(0)|X_i, A_i = 1] \\
= & E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1] - E[Y_{i0}(0) - Y_{i0}(0)|X_i, A_i = 1] \\
= & E[Y_{i1}(1) - Y_{i1}(0)|X_i, A_i = 1]
\end{aligned}
$$

Readings

Readings