Panel Fixed Effects

Prof. Mike Aguilar

https:/www.linkedin.com/in/mike-aguilar-econ

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment
  - With two-way FE, identification uses within-unit variation net of unit and time effects

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment
  - With two-way FE, identification uses within-unit variation net of unit and time effects
- Under these conditions, we do not need to observe the omitted confounders themselves

Housekeeping

- Throughout the slide deck you will see "Q", which indicates a question to you, the reader
- You will also see "A", which indicates the associated answer
- It is generally most efficient to learn this material through active participation. Whenever you encounter a "Q", be sure to try and develop your answer before turning the page to the provided "A" answer
- I recommend reading through these slides before engaging in associated coding exercises
- This slide deck is part of a broader curriculum in causal estimation and inference
- Suggested pre-requisite: foundations of regression, omitted variable bias

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Agile's Productivity Impact



Case Details

Does an Agile method improve productivity?

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an “Agile Project Management” method rather than a traditional linear method.

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years
- Each team can voluntarily decide to adopt an Agile method at any time

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years
- Each team can voluntarily decide to adopt an Agile method at any time
- Team's bonuses are determined by productivity

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years
- Each team can voluntarily decide to adopt an Agile method at any time
- Team's bonuses are determined by productivity
- Productivity Y_{it} is measured for each team/quarter

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years
- Each team can voluntarily decide to adopt an Agile method at any time
- Team's bonuses are determined by productivity
- Productivity Y_{it} is measured for each team/quarter
- We have a dummy variable indicating if Agile was used during that quarter; $D_{it} = 1$ if Agile, 0 otherwise.

Case Details

Does an Agile method improve productivity?

Background

- Software engineering teams often use an "Agile Project Management" method rather than a traditional linear method.
- Advocates of Agile purport enhanced productivity relative to a traditional method.

Data

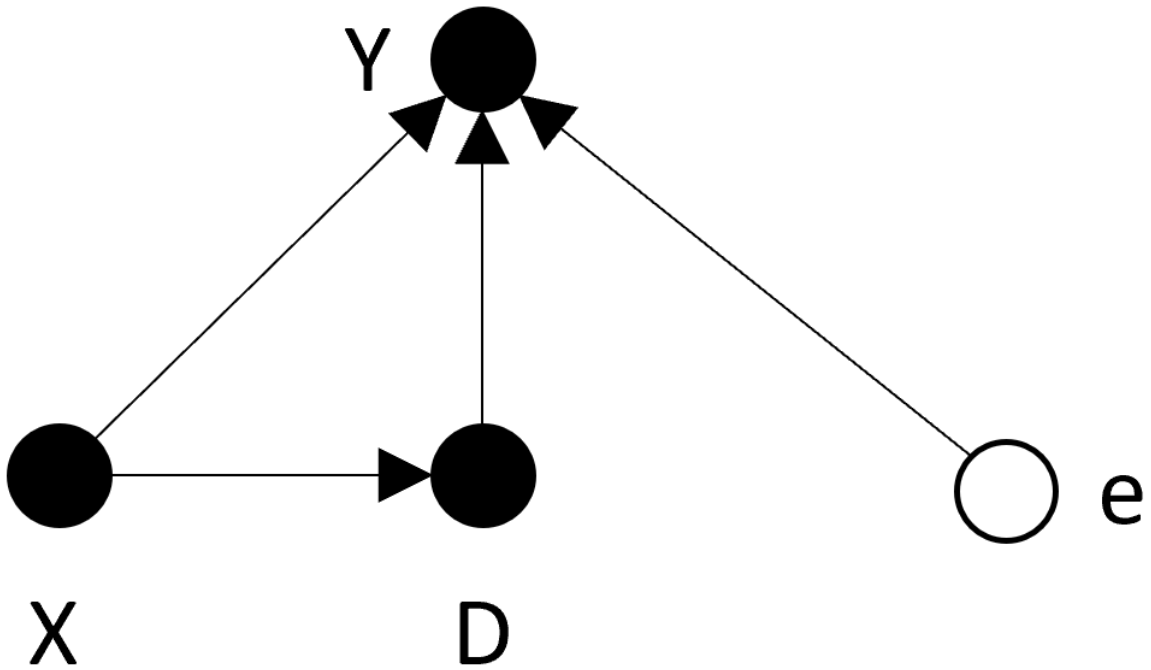
- Unit of measurement is an engineering team (i), for which we have 3 dozen teams
- Each team is tracked every quarter (t) for 2 years
- Each team can voluntarily decide to adopt an Agile method at any time
- Team's bonuses are determined by productivity
- Productivity Y_{it} is measured for each team/quarter
- We have a dummy variable indicating if Agile was used during that quarter; $D_{it} = 1$ if Agile, 0 otherwise.
- Several other covariates are observed for each team through time X_{it}

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

Causal DAG

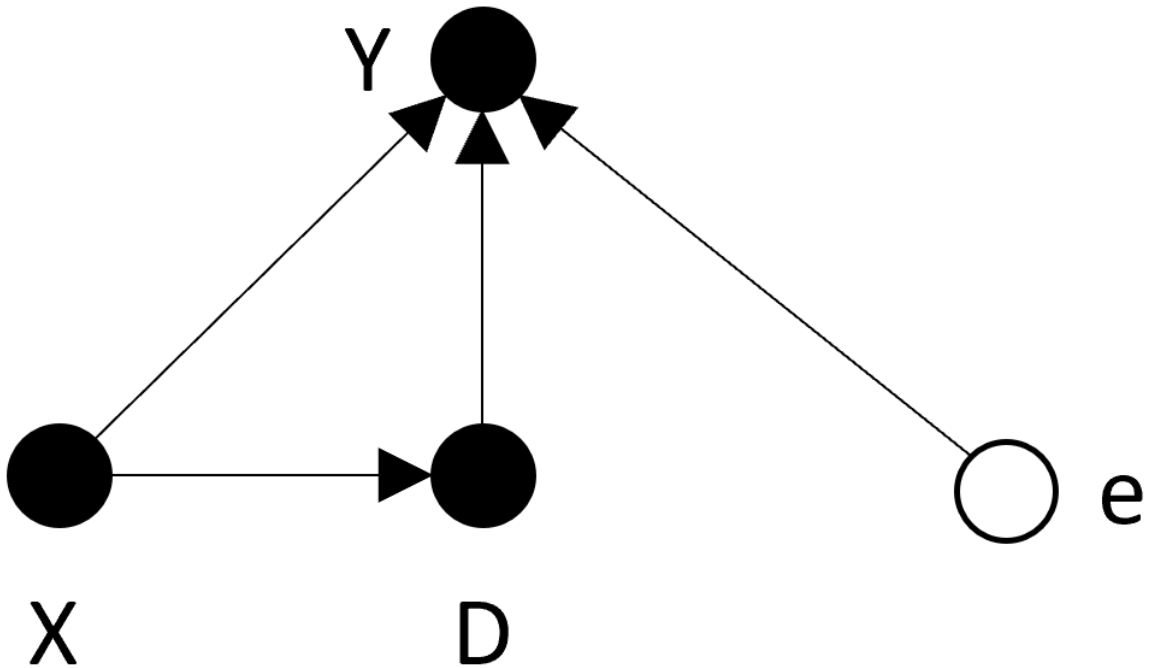
Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?



A:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

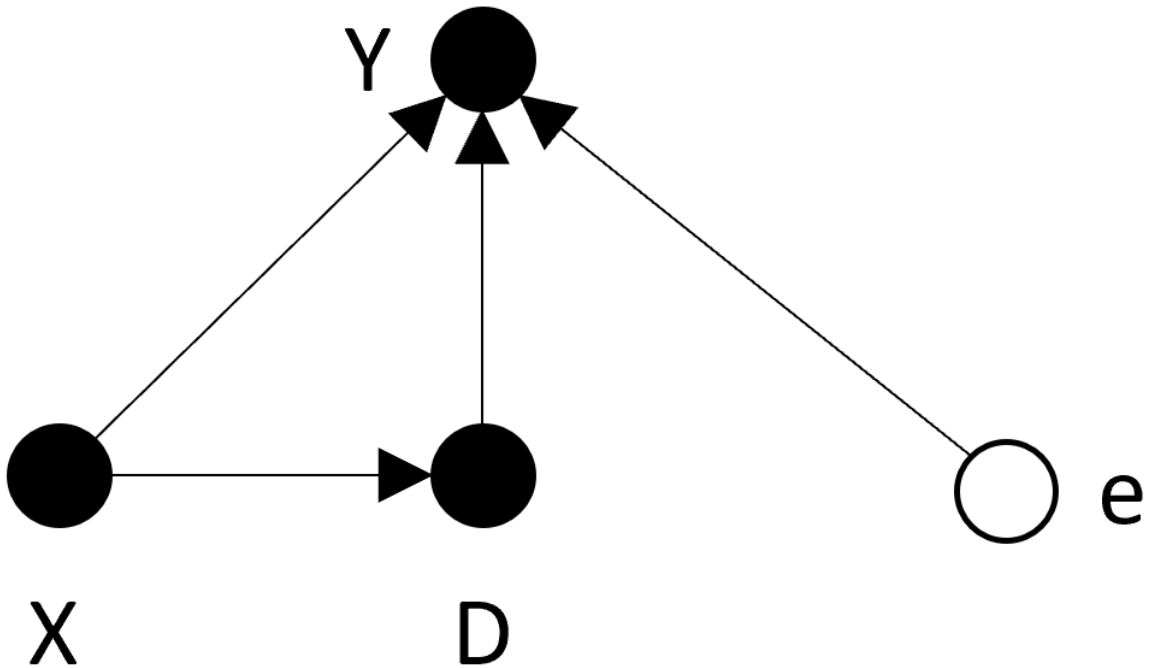


A:

- Treatment:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

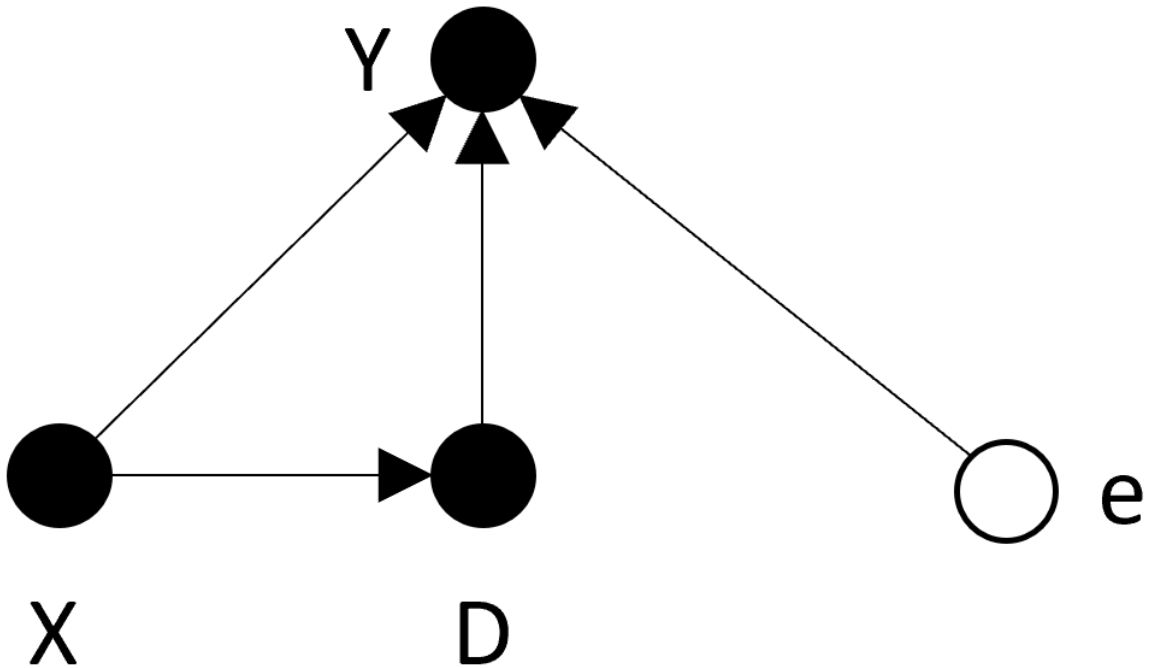


A:

- Treatment: D_{it} (use Agile)

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

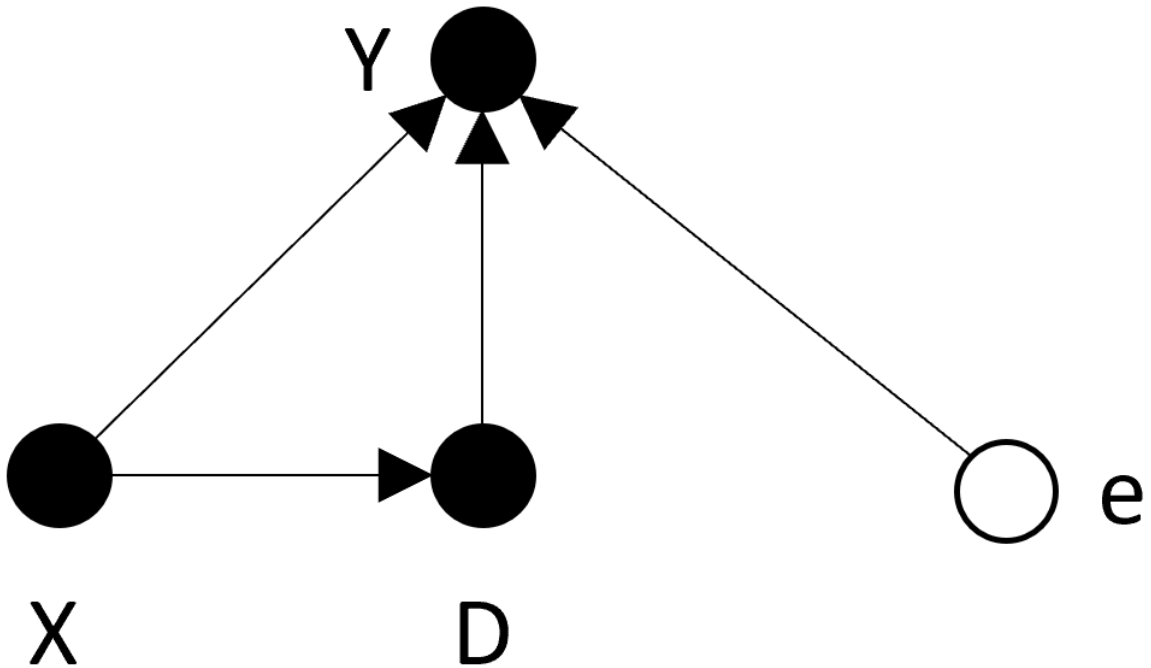


A:

- Treatment: D_{it} (use Agile)
- Outcome:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

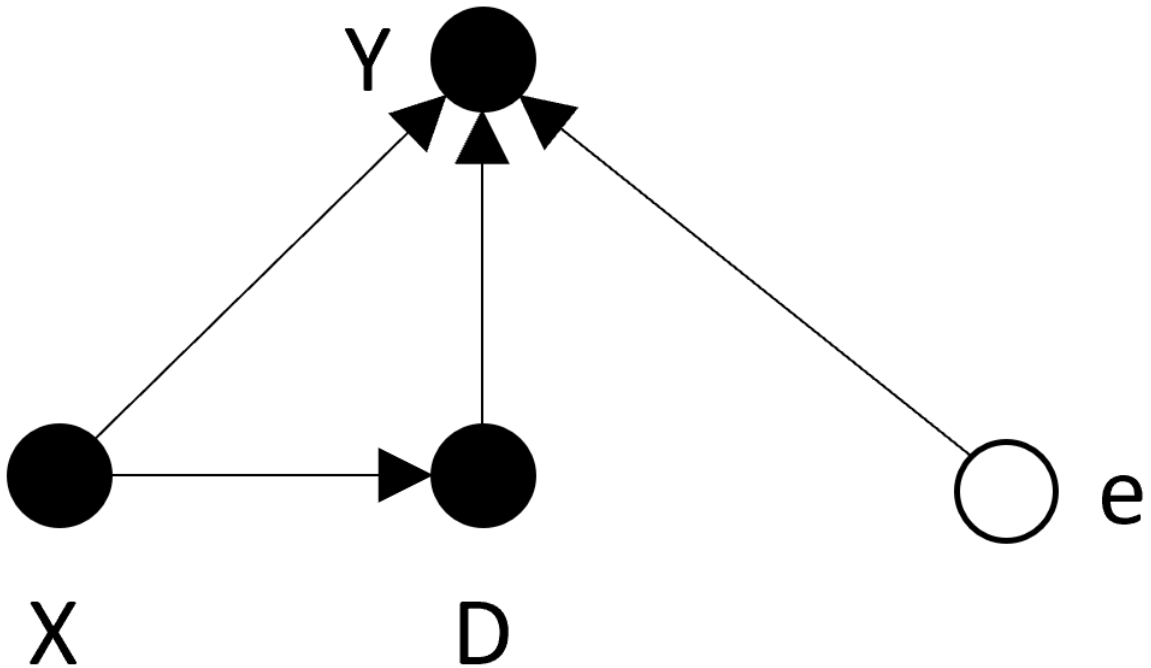


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

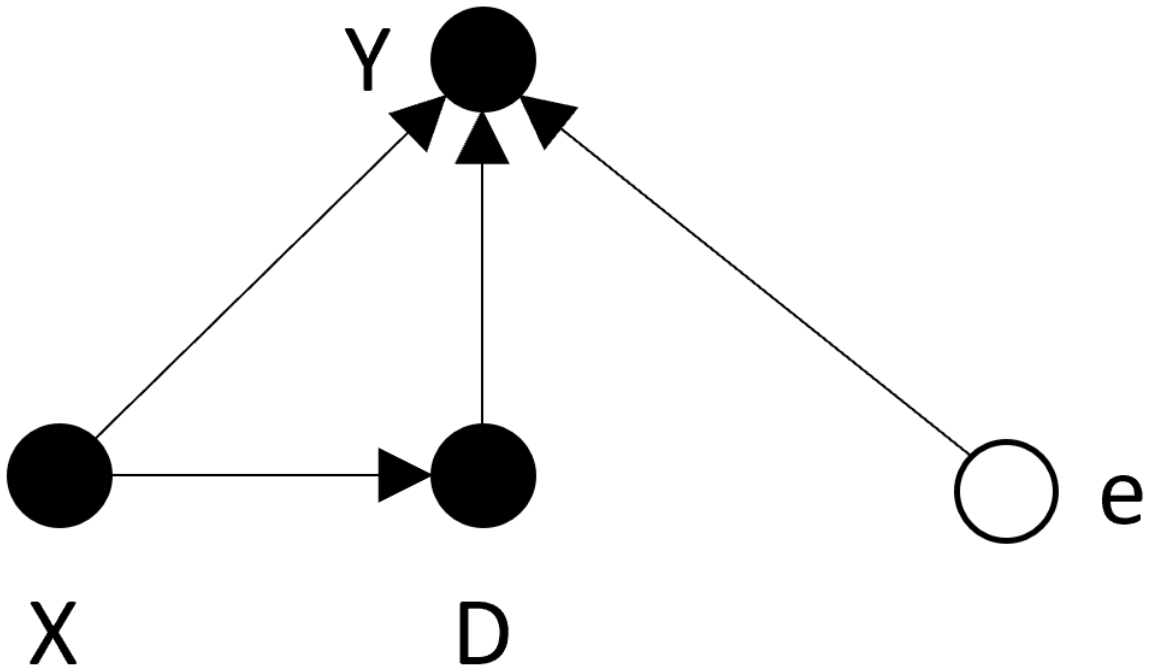


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

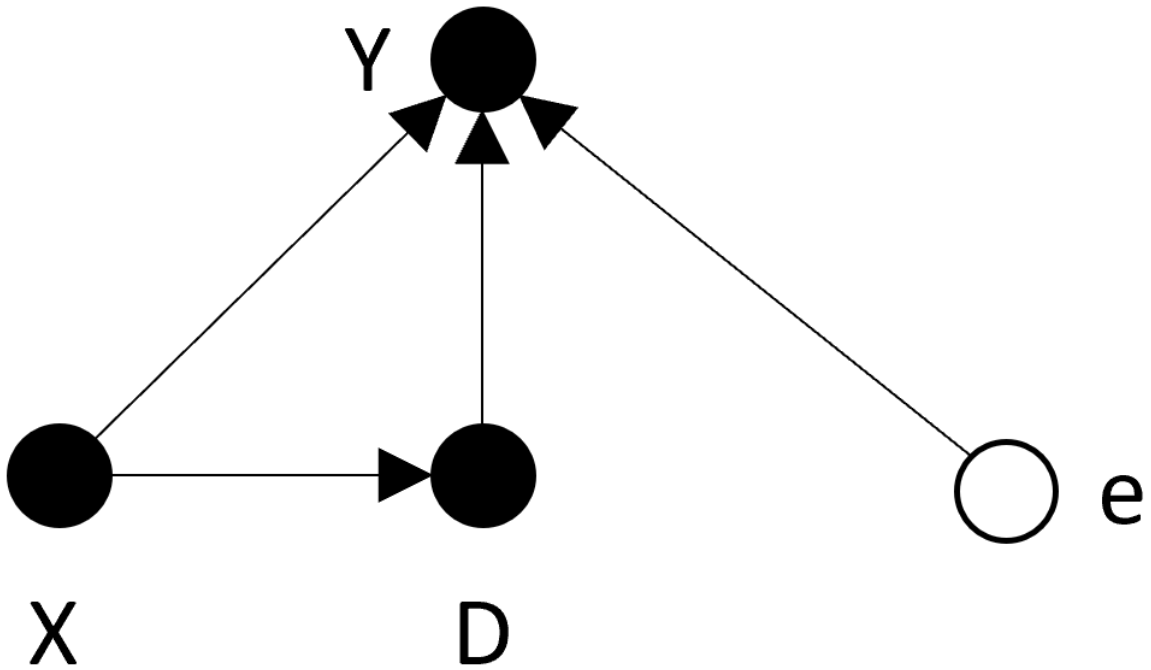


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect: Impact of Agile on Productivity

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

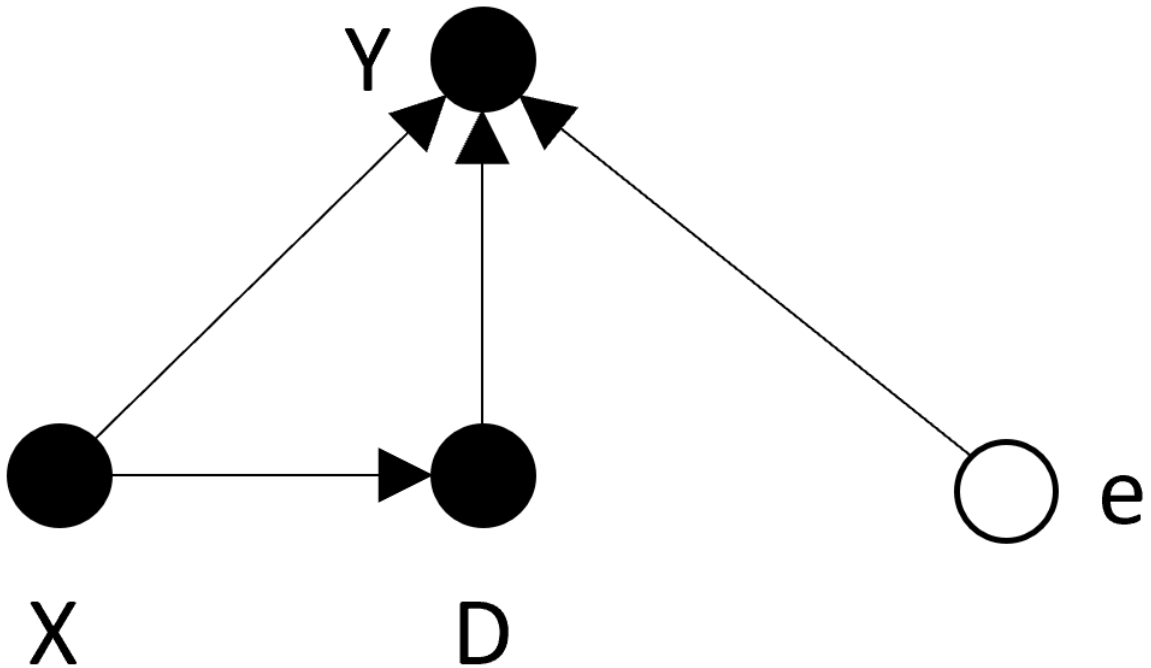


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect: Impact of Agile on Productivity
- Treated Group:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

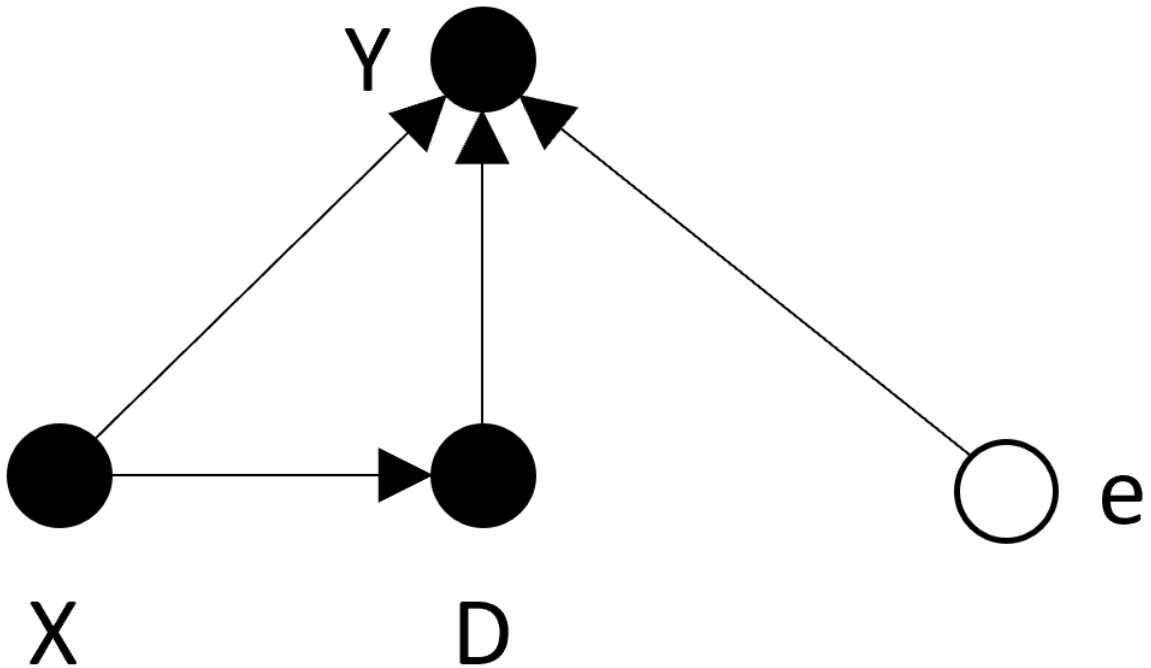


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect: Impact of Agile on Productivity
- Treated Group: Agile users in t

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

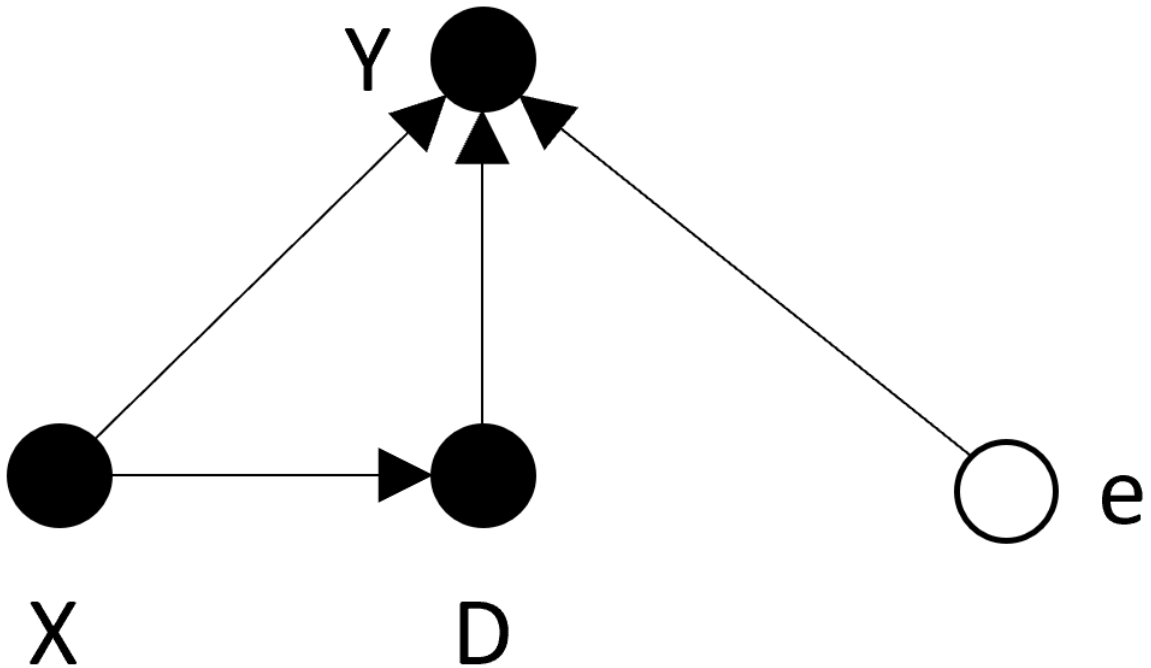


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect: Impact of Agile on Productivity
- Treated Group: Agile users in t
- Control Group:

Causal DAG

Q: Can you draw a DAG? Outcome, Treatment, Treatment Effect, Treated/Control?

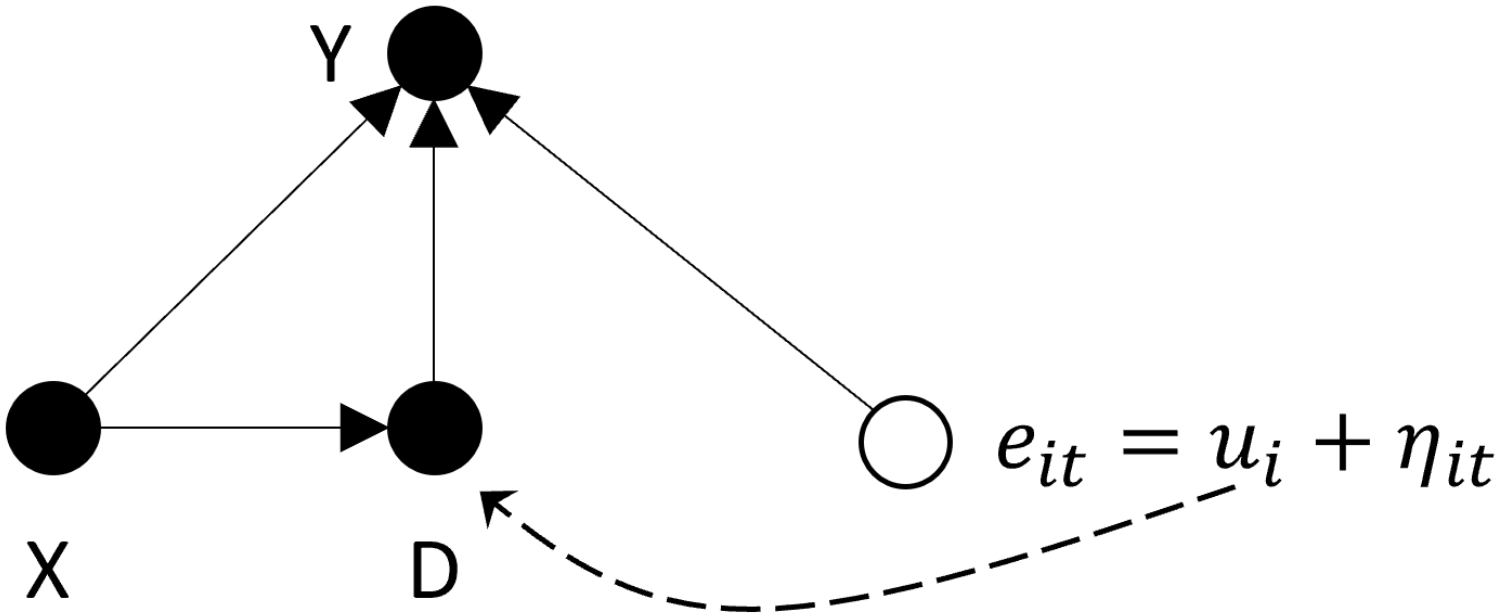


A:

- Treatment: D_{it} (use Agile)
- Outcome: Y_{it} (Productivity)
- Treatment Effect: Impact of Agile on Productivity
- Treated Group: Agile users in t
- Control Group: non-Agile users in t

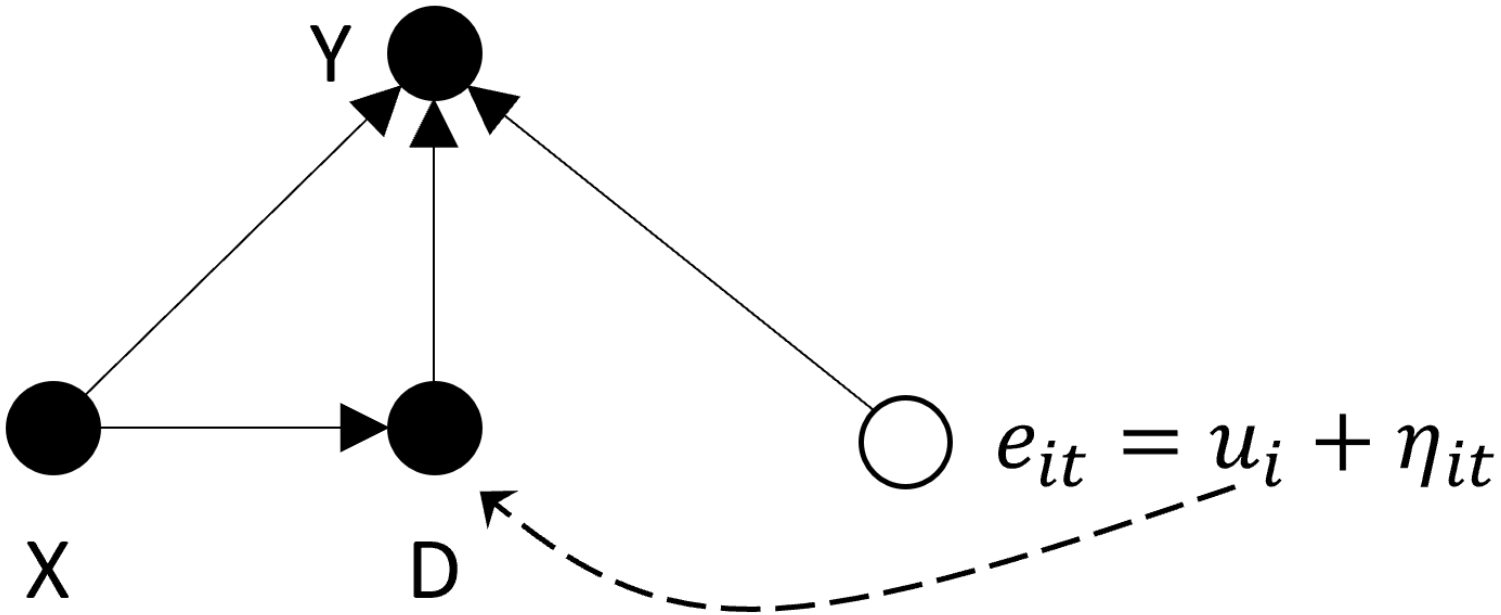
Understanding the Errors

Q: Can you explain the composite error term in our case?



Understanding the Errors

Q: Can you explain the composite error term in our case?

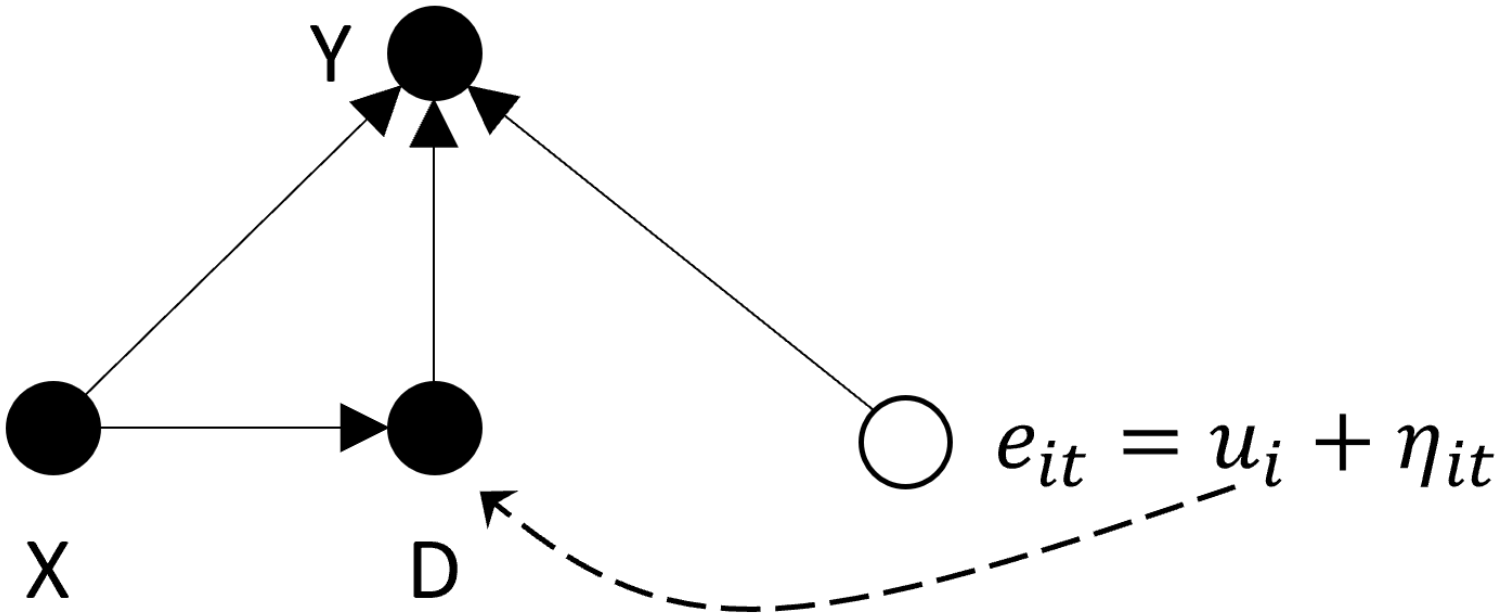


A:

- The error e_{it} consists of two parts

Understanding the Errors

Q: Can you explain the composite error term in our case?

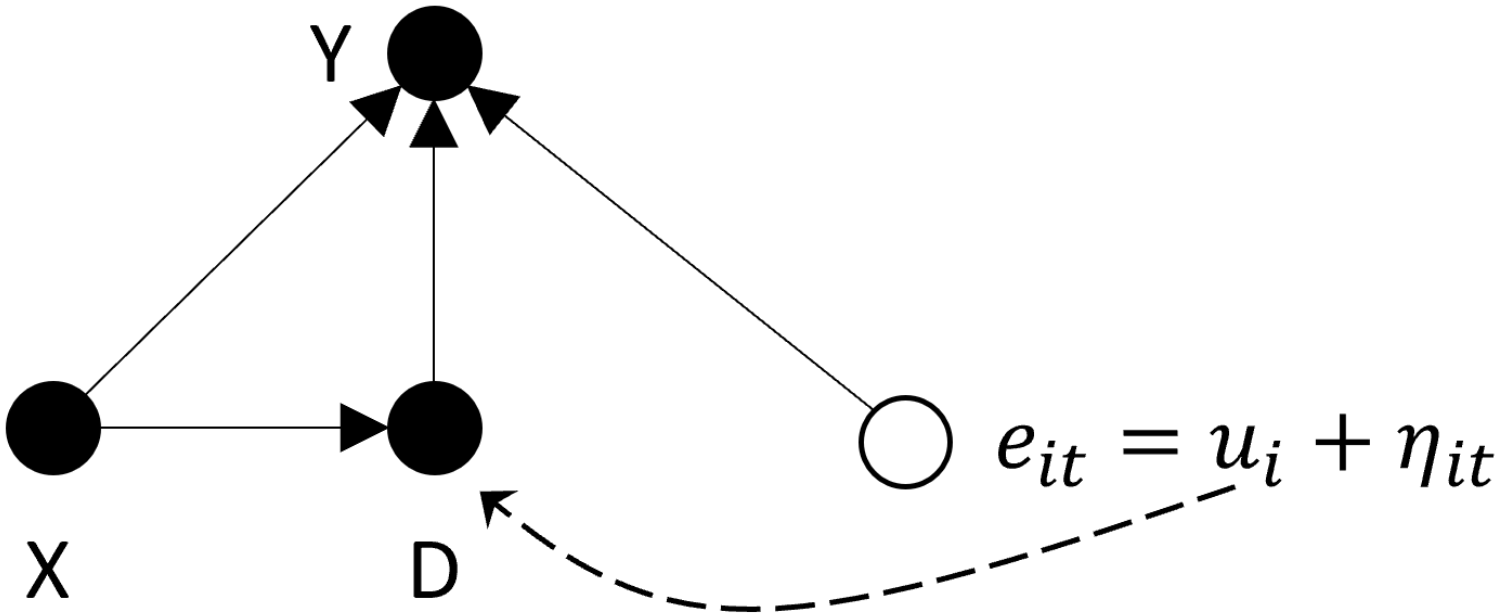


A:

- The error e_{it} consists of two parts
- u_i: Time invariant, but varies with team (i); e.g. innate ability of team

Understanding the Errors

Q: Can you explain the composite error term in our case?

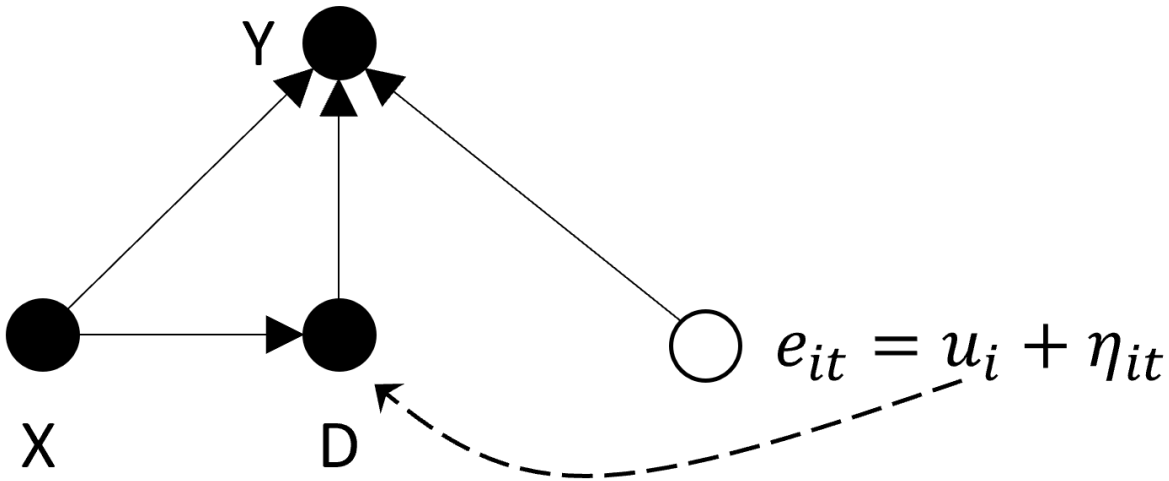


A:

- The error e_{it} consists of two parts
- u_i: Time invariant, but varies with team (i); e.g. innate ability of team
- η_{it}: Time and team varying; e.g. supervisor, client, number of projects assigned

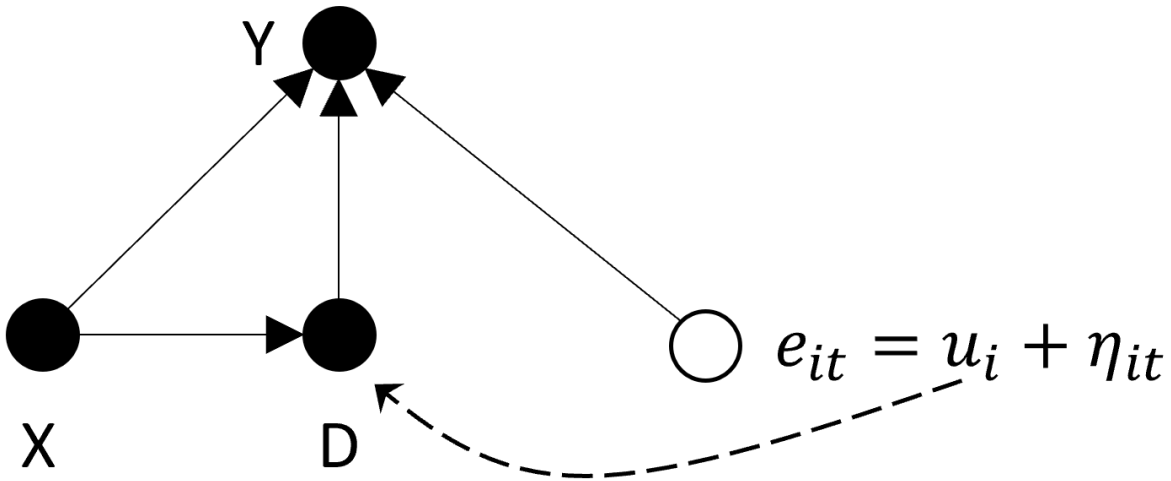
Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?



Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?

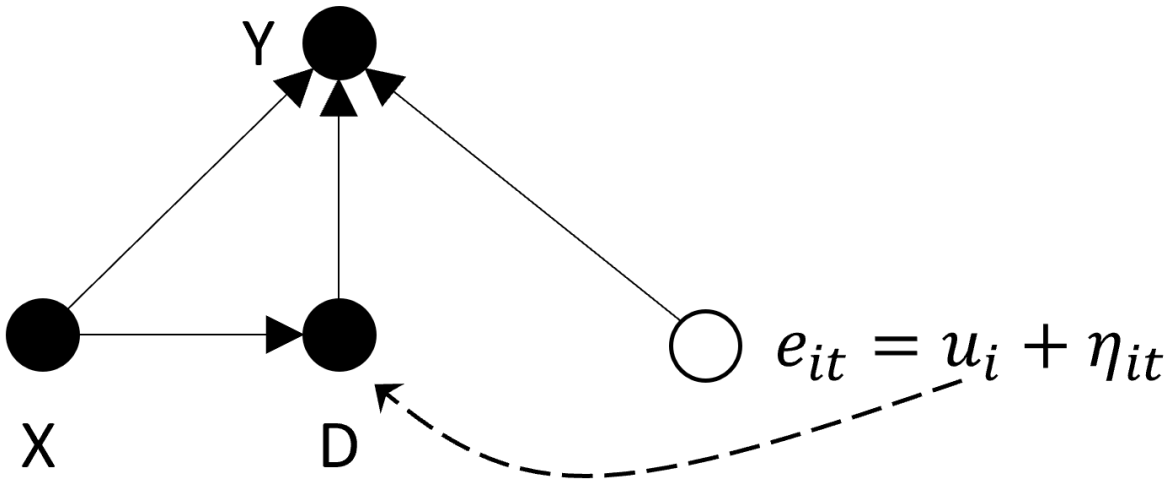


A:

- Consider estimating $Y = \alpha + \beta D + \gamma X + e$

Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?

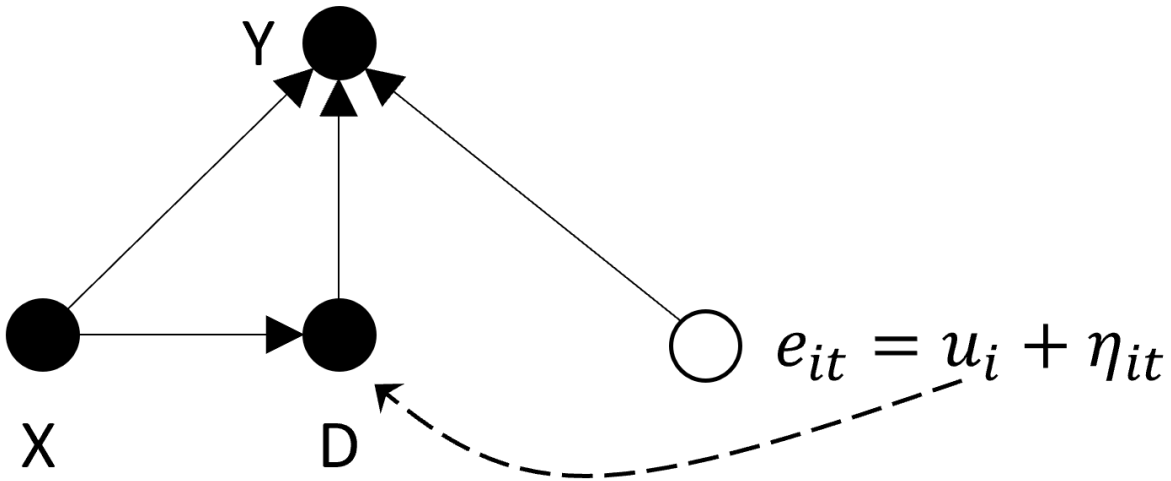


A:

- Consider estimating $Y = \alpha + \beta D + \gamma X + e$
- The DAG suggests $D \not\perp e$ because D is related to u

Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?

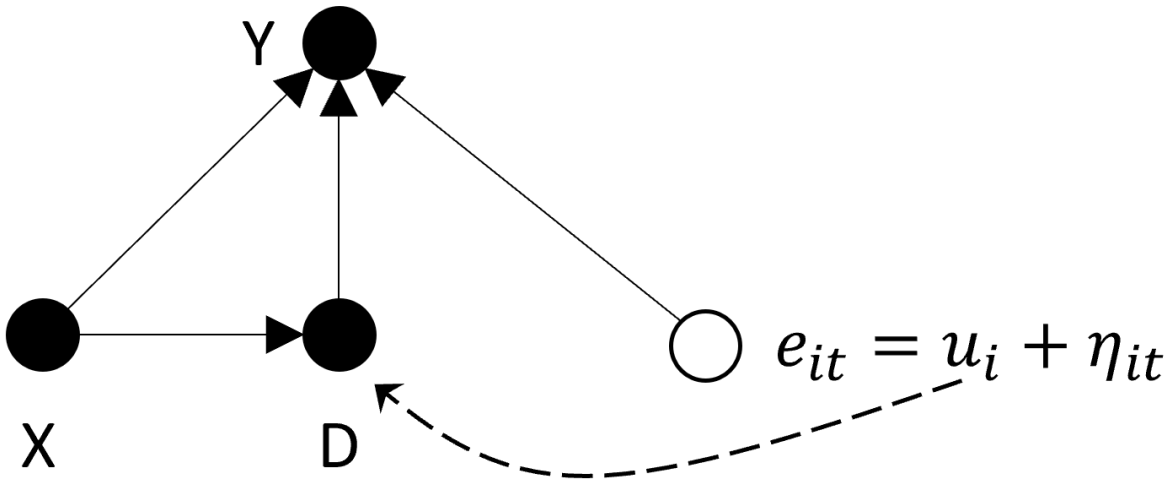


A:

- Consider estimating $Y = \alpha + \beta D + \gamma X + e$
- The DAG suggests $D \not\perp e$ because D is related to u
- This might arise for many reasons. Suppose the teams with the greatest innate ability (u) are the ones who feel most comfortable switching to Agile (D).

Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?

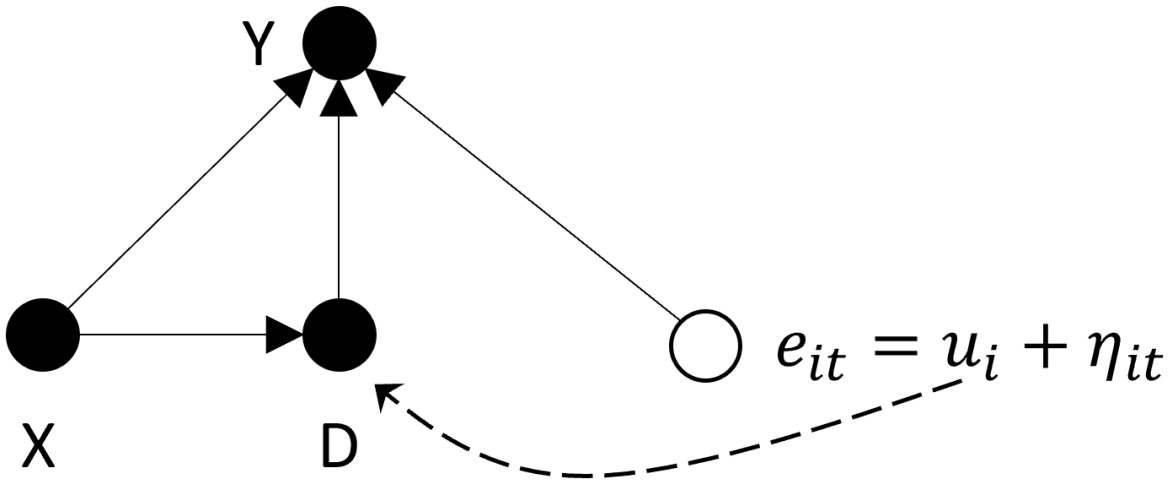


A:

- Consider estimating $Y = \alpha + \beta D + \gamma X + e$
- The DAG suggests $D \not\perp e$ because D is related to u
- This might arise for many reasons. Suppose the teams with the greatest innate ability (u) are the ones who feel most comfortable switching to Agile (D).
- If that is the case, the sign of the bias would be

Finding the OVB

Q: Can you explain the potential for OVB in this case? Can you guess a sign of the bias?



A:

- Consider estimating $Y = \alpha + \beta D + \gamma X + e$
- The DAG suggests $D \not\perp e$ because D is related to u
- This might arise for many reasons. Suppose the teams with the greatest innate ability (u) are the ones who feel most comfortable switching to Agile (D).
- If that is the case, the sign of the bias would be positive since the productivity of the treated group (Agile users) would be higher than that of the control group (non Agile users).

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Panel Data

Q: What is Panel Data?

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods
- Typically index entities with $i = 1, 2, \cdots, N$

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods
- Typically index entities with $i = 1, 2, \cdots, N$
- Typically index time periods with $t = 1, 2, \cdots, T$

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods
- Typically index entities with $i = 1, 2, \cdots, N$
- Typically index time periods with $t = 1, 2, \cdots, T$
- Panel: "same people over time"

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods
- Typically index entities with $i = 1, 2, \cdots, N$
- Typically index time periods with $t = 1, 2, \cdots, T$
- Panel: "same people over time"
- Pool Cross Section: "potentially different people over time"

Panel Data

Q: What is Panel Data?

- Often referred to as Longitudinal Data
- Set of observations collected for the same "N" entities (i.e. people, families, firms, etc..) for at least two periods
- Typically index entities with $i = 1, 2, \cdots, N$
- Typically index time periods with $t = 1, 2, \cdots, T$
- Panel: "same people over time"
- Pool Cross Section: "potentially different people over time"
- Colloquially, the term "panel" is sometimes used to refer to either type.

Panel vs Pooled

Panel vs Pooled

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Panel vs Pooled

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

Panel vs Pooled

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

- We can think of Panel Data as a special case of Pooled data

Panel vs Pooled

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

- We can think of Panel Data as a special case of Pooled data
- These terms (pooled and panel) have similar, but distinct meanings in data type vs model type. e.g. We can estimate a pooled OLS model on panel data

Balanced vs Unbalanced

Unbalanced Panel

	Firm 1	Firm 2	Firm 3
1990	X	X	30
1991	X	75	32
1992	125	76	29
1993	205	84	56

Balanced vs Unbalanced

Unbalanced Panel

	Firm 1	Firm 2	Firm 3
1990	X	X	30
1991	X	75	32
1992	125	76	29
1993	205	84	56

Balanced Panel

	Firm 1	Firm 2	Firm 3
~~1990~~	~~X~~	~~X~~	~~30~~
~~1991~~	~~X~~	~~75~~	~~32~~
1992	125	76	29
1993	205	84	56

Balanced vs Unbalanced

Unbalanced Panel

	Firm 1	Firm 2	Firm 3
1990	X	X	30
1991	X	75	32
1992	125	76	29
1993	205	84	56

Balanced Panel

	Firm 1	Firm 2	Firm 3
~~1990~~	~~X~~	~~X~~	~~30~~
~~1991~~	~~X~~	~~75~~	~~32~~
1992	125	76	29
1993	205	84	56

Our focus will be on balanced panels, but the causal estimation techniques we employ are generally available for either type of data.

Panel $\neq$ Wide Format

Don't confuse Panel for wide vs long format of a cross sectional dataset

Panel $\neq$ Wide Format

Don't confuse Panel for wide vs long format of a cross sectional dataset

Firm ID	Revenues	Costs
A	100	20
B	104	18

Table: Wide Format

Panel $\neq$ Wide Format

Don't confuse Panel for wide vs long format of a cross sectional dataset

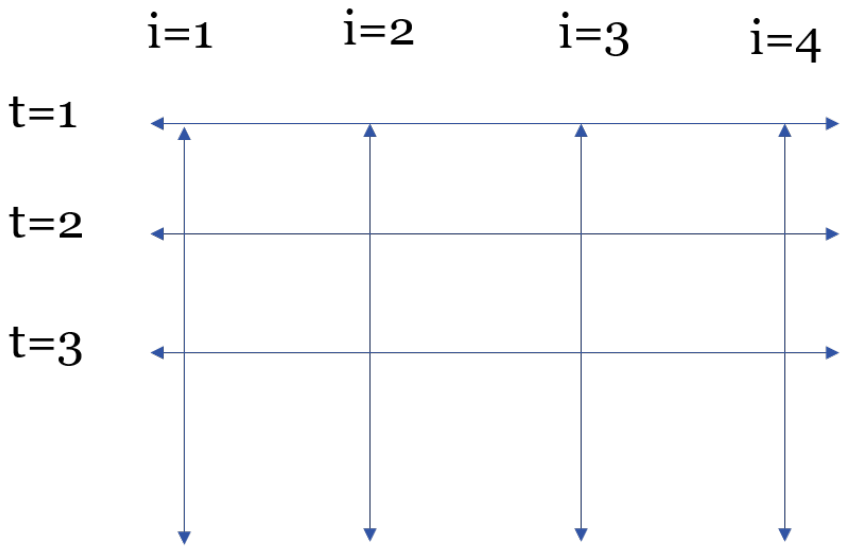
Firm ID	Revenues	Costs
A	100	20
B	104	18

Table: Wide Format

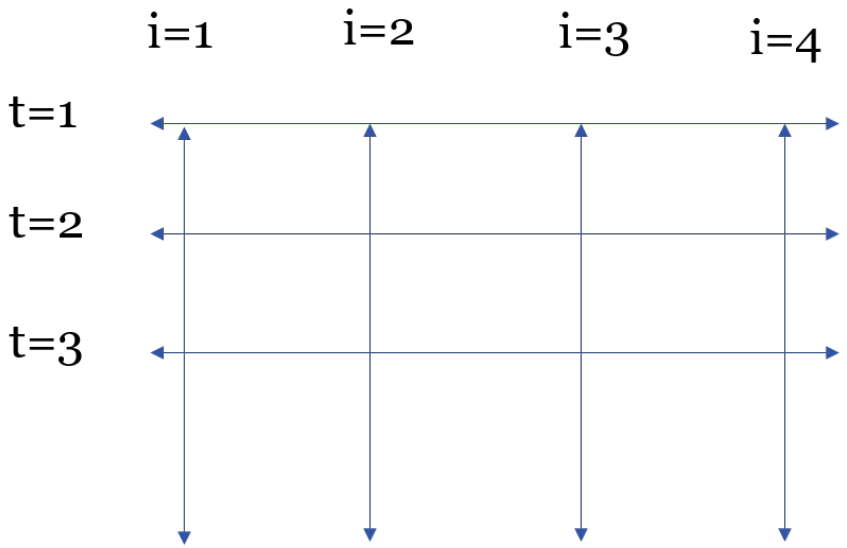
Firm ID	Field	Value
A	Revenue	100
A	Costs	20
B	Revenue	104
B	Costs	18

Table: Long Format

Advantage of Panels

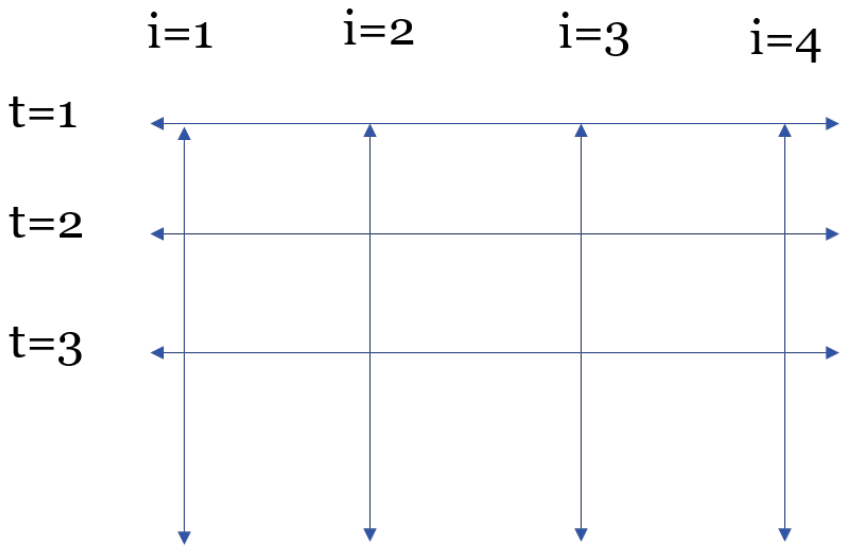


Advantage of Panels



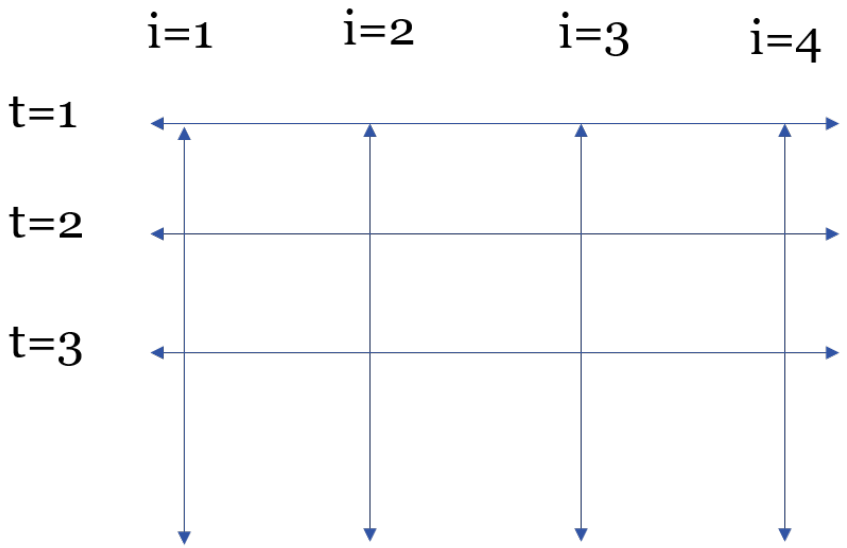
- A single cross sectional regression may be plagued by OVB

Advantage of Panels



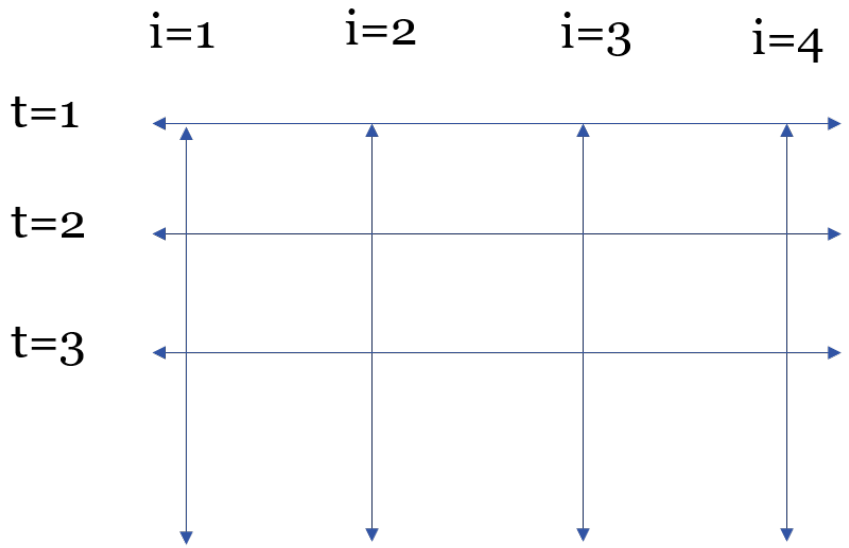
- A single cross sectional regression may be plagued by OVB
  e.g. We might be missing innate ability

Advantage of Panels



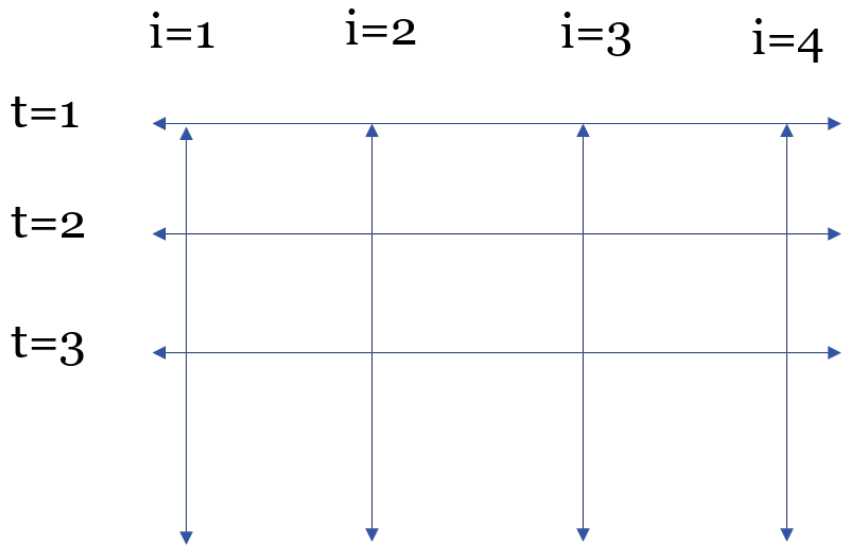
- A single cross sectional regression may be plagued by OVB
  e.g. We might be missing innate ability
- A single time series regression likely won't allow the treatment to vary

Advantage of Panels



- A single cross sectional regression may be plagued by OVB
  - e.g. We might be missing innate ability
- A single time series regression likely won't allow the treatment to vary
  - e.g. Having data on a single team through time may not provide variation in use of Agile methods (treatment)

Advantage of Panels



- A single cross sectional regression may be plagued by OVB
  - e.g. We might be missing innate ability
- A single time series regression likely won't allow the treatment to vary
  - e.g. Having data on a single team through time may not provide variation in use of Agile methods (treatment)
- But a panel/pooled dataset may be able to elicit the treatment effect while controlling for OVB

Disadvantages of Panels

Disadvantages of Panels

- Operations: panel data is costly to collect (true for any longitudinal data) since we need to track individuals through time.

Disadvantages of Panels

- Operations: panel data is costly to collect (true for any longitudinal data) since we need to track individuals through time.
- Strategy: longitudinal panel data is costly because we need to wait until sufficient observations are available before we can make an informed decision.

Panel doesn't need time

Although we usually encounter panels over time (i.e. longitudinal), that isn't essential

Panel doesn't need time

Although we usually encounter panels over time (i.e. longitudinal), that isn't essential

Year	Firm#	Region	Sales
1990	A	NorthEast	100
1990	A	West	102
1990	B	NorthEast	98
1990	B	West	103

Table: Panel dataset varying by Firm i and Region j, not time

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Estimating One Cross Section at a Time

What if we estimate one cross section at a time?

Estimating One Cross Section at a Time

What if we estimate one cross section at a time?

For each t we could estimate

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

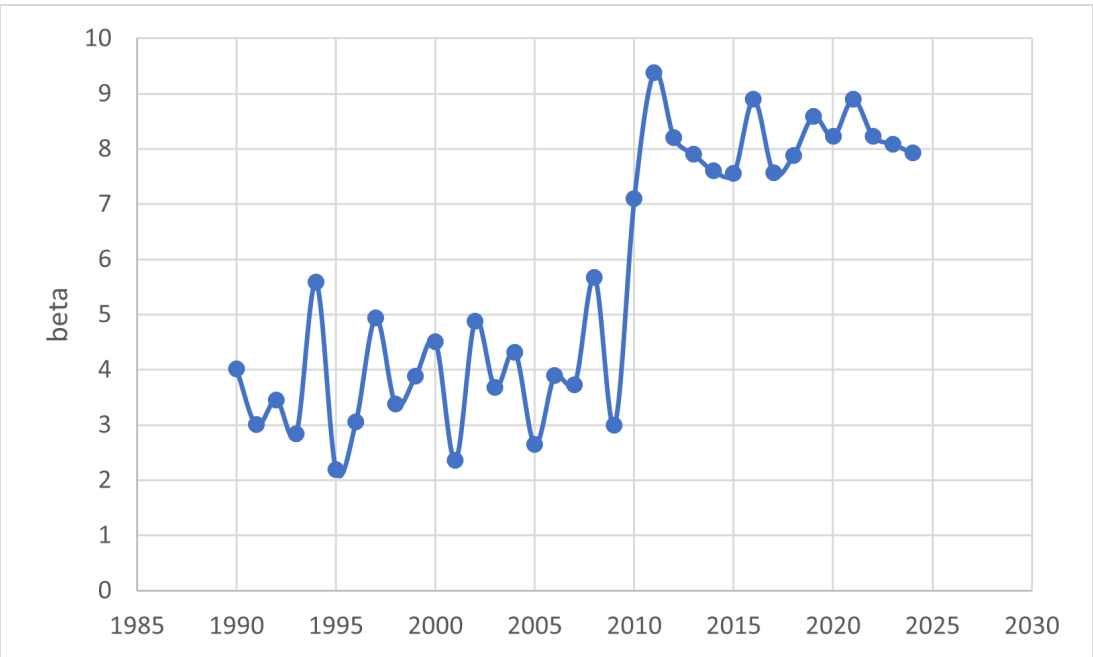
Estimating One Cross Section at a Time

What if we estimate one cross section at a time?

For each t we could estimate

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

This would produce a sequence of β's



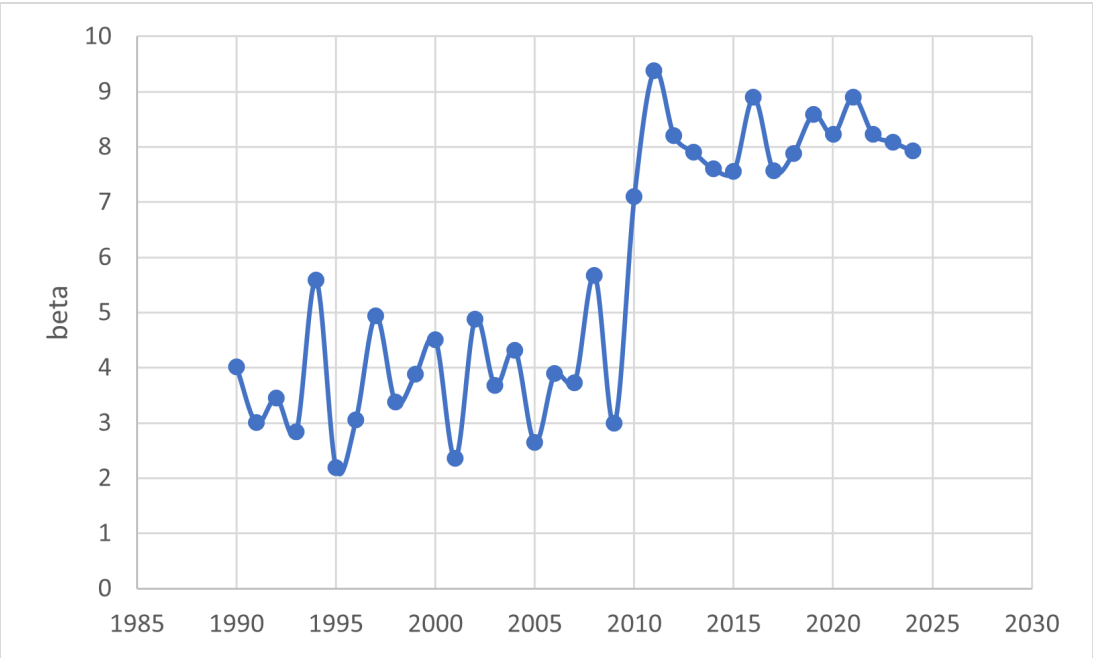
Estimating One Cross Section at a Time

What if we estimate one cross section at a time?

For each t we could estimate

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

This would produce a sequence of β's



Q: What's an advantage of this approach?

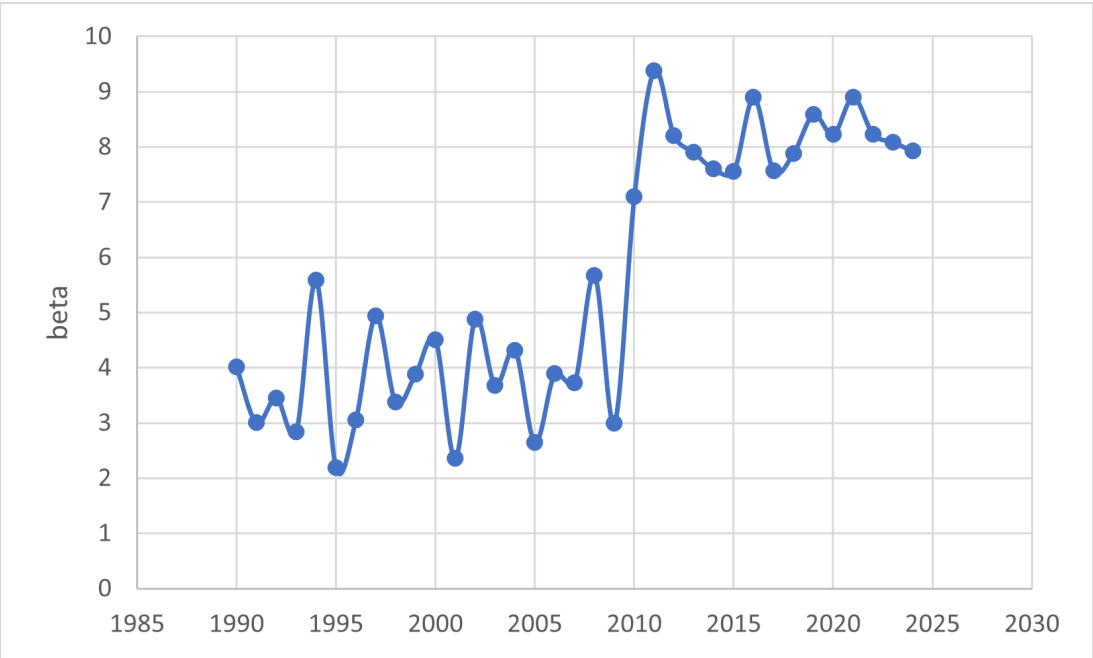
Estimating One Cross Section at a Time

What if we estimate one cross section at a time?

For each t we could estimate

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

This would produce a sequence of β's



Q: What's an advantage of this approach?

A: We might be able to detect time varying treatment effects.

Disadvantages of Individual Cross Sections

Q: What are some disadvantages of this approach?

Disadvantages of Individual Cross Sections

Q: What are some disadvantages of this approach?

A:

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

1. The estimates most likely have different number of observations given the nature of the pooled dataset, which gives different confidence in each estimate

Disadvantages of Individual Cross Sections

Q: What are some disadvantages of this approach?

A:

$$Y_i = \alpha + \beta D_i + \gamma X_i + e_i$$

1. The estimates most likely have different number of observations given the nature of the pooled dataset, which gives different confidence in each estimate
2. More importantly, these estimates would be biased since the OVB remains in each cross section (e.g. innate ability differs across teams within a given point of time)

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Pooled OLS

Recall the difference between Pooled and Panel data

Pooled OLS

Recall the difference between Pooled and Panel data

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled OLS

Recall the difference between Pooled and Panel data

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

Pooled OLS

Recall the difference between Pooled and Panel data

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

- Running OLS on a pooled cross section (i.e. Pooled OLS) is implemented identically to running OLS on a single cross section

Pooled OLS

Recall the difference between Pooled and Panel data

Panel

t	i	X
1990	A	5
1990	B	15
1991	A	6
1991	B	19

Pooled Cross Sectional

t	i	X
1990	A	5
1990	B	15
1991	C	12
1991	D	6

- Running OLS on a pooled cross section (i.e. Pooled OLS) is implemented identically to running OLS on a single cross section
- You can even run Pooled OLS on a true panel dataset if you (accidentally?) ignore the longitudinal aspects.

Pooled OLS for TE

Q: What's wrong with estimating this model by OLS?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

Pooled OLS for TE

Q: What's wrong with estimating this model by OLS?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

A:

- Once again, $D_{it} \not\perp e_{it}$ because u_i is related to treatment D_{it}

Pooled OLS for TE

Q: What's wrong with estimating this model by OLS?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

A:

- Once again, $D_{it} \not\perp e_{it}$ because u_i is related to treatment D_{it}
- An OVB exists

Pooled OLS for TE

Q: What's wrong with estimating this model by OLS?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

A:

- Once again, $D_{it} \not\perp e_{it}$ because u_i is related to treatment D_{it}
- An OVB exists
- Pooled OLS on independent cross-sections cannot uncover an unbiased estimate of the treatment effect
- We need a panel dataset that tracks the same units through time

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \ \ t = 1, 2$$

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \ \ t = 1, 2$$

We can decompose the error as

$$e_{it} = u_i + \eta_{it}$$

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \ \ t = 1, 2$$

We can decompose the error as

$$e_{it} = u_i + \eta_{it}$$

- e_{it} : Composite error

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \;\; t = 1, 2$$

We can decompose the error as

$$e_{it} = u_i + \eta_{it}$$

- e_{it} : Composite error
- u_i : Entity-specific fixed effect; unobserved individual heterogeneity

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \ \ t = 1, 2$$

We can decompose the error as

$$e_{it} = u_i + \eta_{it}$$

- e_{it} : Composite error
- u_i : Entity-specific fixed effect; unobserved individual heterogeneity
- η_{it} : Idiosyncratic error that may vary through time

2-Period Panel Model

Decomposing the Error

Given our panel data model

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}, \;\; t = 1, 2$$

We can decompose the error as

$$e_{it} = u_i + \eta_{it}$$

- e_{it} : Composite error
- u_i : Entity-specific fixed effect; unobserved individual heterogeneity
- η_{it} : Idiosyncratic error that may vary through time
- Note: We could include a "τ_t" in the composite error to accommodate for time fixed effects

OVB

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$
$$e_{it} = u_i + \eta_{it}$$

The violation of $D_{it} \perp e_{it}$ due to $Cov(D_{it}, u_i) \neq 0$ leads to an OVB, which is sometimes called "heterogeneity bias".

OVB

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$
$$e_{it} = u_i + \eta_{it}$$

The violation of $D_{it} \perp e_{it}$ due to $Cov(D_{it}, u_i) \neq 0$ leads to an OVB, which is sometimes called "heterogeneity bias".

Q: What does $Cov(D_{it}, u_i) \neq 0$ mean in our Agile case study?

OVB

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$
$$e_{it} = u_i + \eta_{it}$$

The violation of $D_{it} \perp e_{it}$ due to $Cov(D_{it}, u_i) \neq 0$ leads to an OVB, which is sometimes called "heterogeneity bias".

Q: What does $Cov(D_{it}, u_i) \neq 0$ mean in our Agile case study?
A: Innate ability of a team is related to its choice of using Agile.

OVB

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$
$$e_{it} = u_i + \eta_{it}$$

The violation of $D_{it} \perp e_{it}$ due to $Cov(D_{it}, u_i) \neq 0$ leads to an OVB, which is sometimes called "heterogeneity bias".

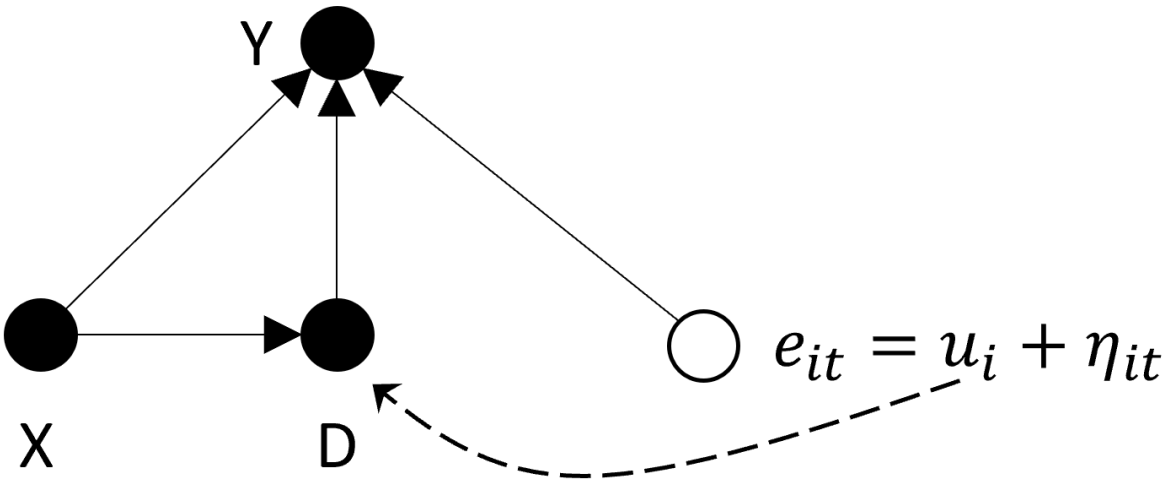
Q: What does $Cov(D_{it}, u_i) \neq 0$ mean in our Agile case study?
A: Innate ability of a team is related to its choice of using Agile.

Understanding this OVB is the key to overcoming the problem with our panel methods.

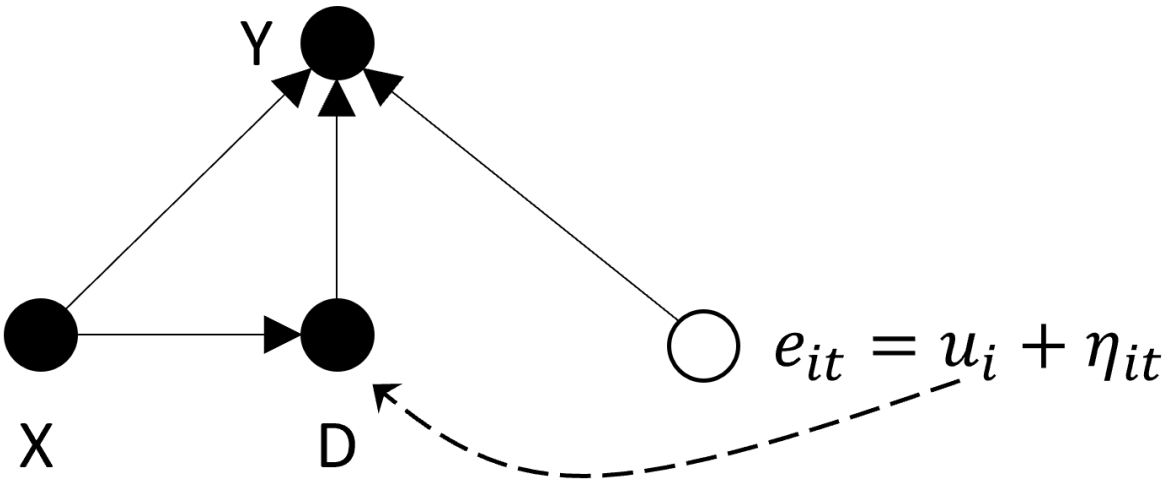
Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Entity Specific Dummies

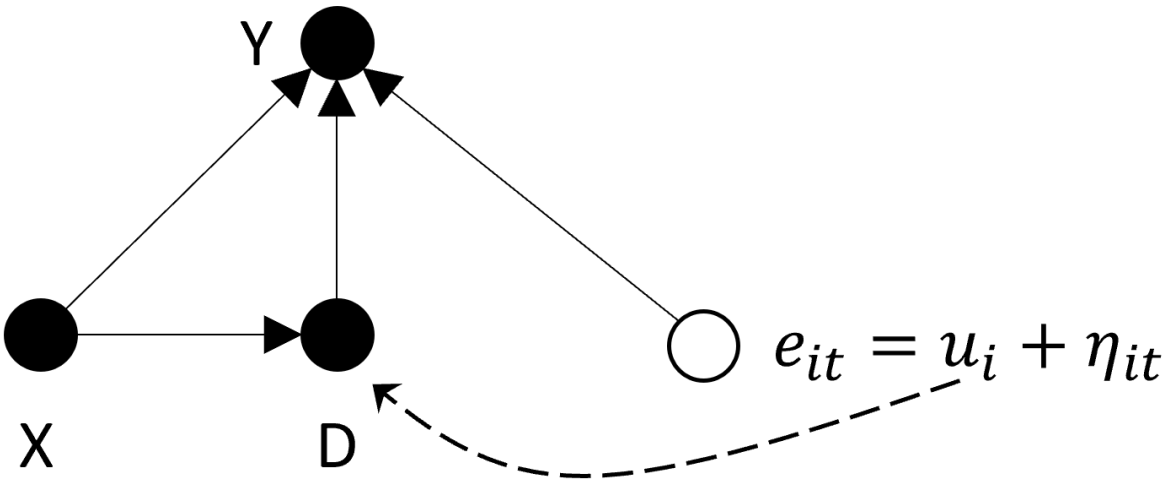


Entity Specific Dummies



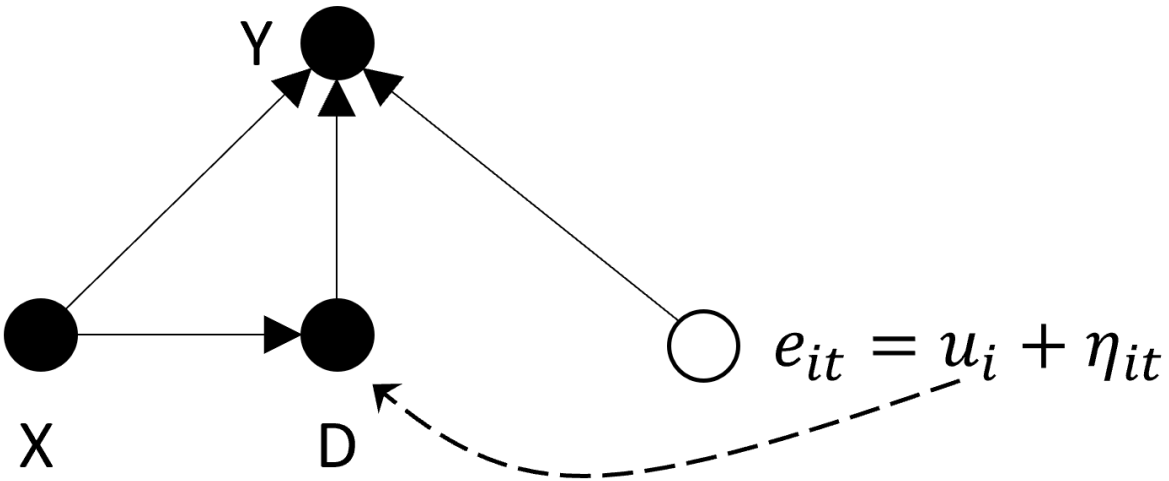
- Recall that the DAG suggested something was "special" about each team (e.g. innate ability).

Entity Specific Dummies



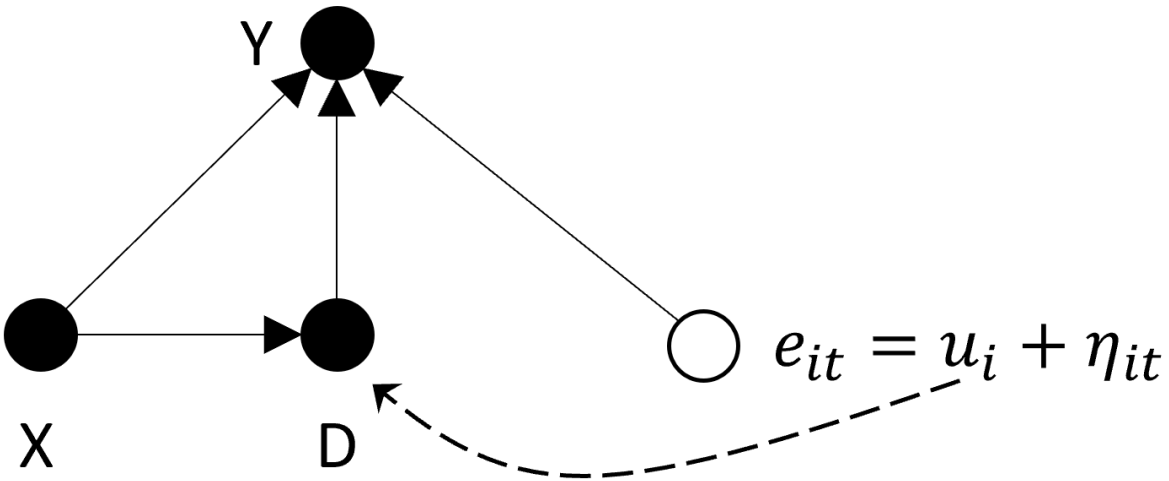
- Recall that the DAG suggested something was "special" about each team (e.g. innate ability).
- We can't measure innate ability, and so we can't include it directly in the regression.

Entity Specific Dummies



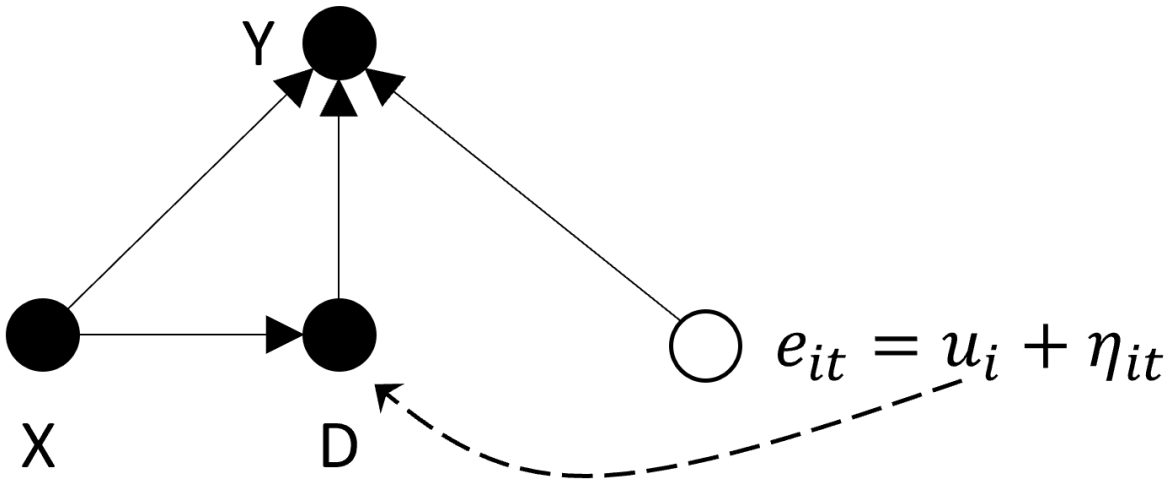
- Recall that the DAG suggested something was "special" about each team (e.g. innate ability).
- We can't measure innate ability, and so we can't include it directly in the regression.
- But innate ability doesn't change through time (u_i has no time subscript); i.e. it is fixed.

Entity Specific Dummies



- Recall that the DAG suggested something was "special" about each team (e.g. innate ability).
- We can't measure innate ability, and so we can't include it directly in the regression.
- But innate ability doesn't change through time (u_i has no time subscript); i.e. it is fixed.
- So we can include a dummy variable for each team.

Entity Specific Dummies



- Recall that the DAG suggested something was "special" about each team (e.g. innate ability).
- We can't measure innate ability, and so we can't include it directly in the regression.
- But innate ability doesn't change through time (u_i has no time subscript); i.e. it is fixed.
- So we can include a dummy variable for each team.
- That dummy variable would represent team $i's$ innate ability (as well as anything else that is fixed and specific to team i)

Least Squares Dummy Variable Regression

Suppose that instead of an unobserved fixed effect in the error term (u_i)

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

Least Squares Dummy Variable Regression

Suppose that instead of an unobserved fixed effect in the error term (u_i)

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \ t = 1, 2, ..., T$$

we explicitly model the team specific fixed effects. Define the dummy variable $Team_i = 1$ for team i and 0 otherwise. We then have

Least Squares Dummy Variable Regression

Suppose that instead of an unobserved fixed effect in the error term (u_i)

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

we explicitly model the team specific fixed effects. Define the dummy variable $Team_i = 1$ for team i and 0 otherwise. We then have

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

Note: We exclude one of the teams to avoid the dummy variable trap.

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

- $H_0 : Team_1 = 0,\ Team_2 = 0,\ \ldots,\ Team_{N-1} = 0$ vs H_A at least one of these restrictions fails (i.e. unit fixed effects are present).

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

- $H_0 : Team_1 = 0,\ Team_2 = 0,\ \ldots,\ Team_{N-1} = 0$ vs H_A at least one of these restrictions fails (i.e. unit fixed effects are present).
  - Note: This is equivalent to considering the base model $Y_{it} = \alpha_i + \beta D_{it} + \gamma X_{it} + \eta_{it}$ and testing $H_0 : \alpha_1 = \cdots = \alpha_N$

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

- $H_0 : Team_1 = 0,\ Team_2 = 0,\ \ldots,\ Team_{N-1} = 0$ vs H_A at least one of these restrictions fails (i.e. unit fixed effects are present).
  - Note: This is equivalent to considering the base model $Y_{it} = \alpha_i + \beta D_{it} + \gamma X_{it} + \eta_{it}$ and testing $H_0 : \alpha_1 = \cdots = \alpha_N$
- Estimate $Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$ and gather restricted residual sum of squares SSR_R

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

- $H_0 : Team_1 = 0,\ Team_2 = 0,\ \ldots,\ Team_{N-1} = 0$ vs H_A at least one of these restrictions fails (i.e. unit fixed effects are present).
  - Note: This is equivalent to considering the base model $Y_{it} = \alpha_i + \beta D_{it} + \gamma X_{it} + \eta_{it}$ and testing $H_0 : \alpha_1 = \cdots = \alpha_N$
- Estimate $Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$ and gather restricted residual sum of squares SSR_R
- Estimate the LSDV model and gather the unrestricted residual sum of squares SSR_U

Testing for Fixed Effects

We can test for the presence of these entity specific fixed effects via a joint F test ala Chow.

- $H_0 : Team_1 = 0,\ Team_2 = 0,\ \ldots,\ Team_{N-1} = 0$ vs H_A at least one of these restrictions fails (i.e. unit fixed effects are present).
  - Note: This is equivalent to considering the base model $Y_{it} = \alpha_i + \beta D_{it} + \gamma X_{it} + \eta_{it}$ and testing $H_0 : \alpha_1 = \cdots = \alpha_N$
- Estimate $Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$ and gather restricted residual sum of squares SSR_R
- Estimate the LSDV model and gather the unrestricted residual sum of squares SSR_U
- Evaluate $F^{stat} = \frac{(SSR_R - SSR_U)/(N-1)}{SSR_U/(NT-N-K)}$, where $i = 1, .., N,\ t = 1, ..., T$, and K is the number of regressors in the unobserved effects model (i.e. 2 in this case)

Shortcoming of LSDV

Q: What's a shortcoming of the LSDV approach?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i \, Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

Shortcoming of LSDV

Q: What's a shortcoming of the LSDV approach?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

A:

- There are $N \times T$ observations

Shortcoming of LSDV

Q: What's a shortcoming of the LSDV approach?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i \, Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

A:

- There are $N \times T$ observations
- By including a separate parameter for each entity we are adding

Shortcoming of LSDV

Q: What's a shortcoming of the LSDV approach?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i \, Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

A:

- There are $N \times T$ observations
- By including a separate parameter for each entity we are adding $N - 1$ of the δ parameters, which

Shortcoming of LSDV

Q: What's a shortcoming of the LSDV approach?

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \sum_{i=1}^{N-1} \delta_i Team_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

A:

- There are $N \times T$ observations
- By including a separate parameter for each entity we are adding $N - 1$ of the δ parameters, which reduces degrees of freedom

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias
- Thankfully, we don't need data on innate ability

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias
- Thankfully, we don't need data on innate ability
- Notice that innate ability is fixed over time and varies across teams

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias
- Thankfully, we don't need data on innate ability
- Notice that innate ability is fixed over time and varies across teams
- Moreover, the ID number is fixed over time and varies across teams

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias
- Thankfully, we don't need data on innate ability
- Notice that innate ability is fixed over time and varies across teams
- Moreover, the ID number is fixed over time and varies across teams
- So, within an FE we don't need the omitted variable (ability), we just need

FE Intuition

Q: What's the intuition behind an FE solution to OVB?

A:

- We don't have data on each team's innate ability
- This omission is generating a bias
- Thankfully, we don't need data on innate ability
- Notice that innate ability is fixed over time and varies across teams
- Moreover, the ID number is fixed over time and varies across teams
- So, within an FE we don't need the omitted variable (ability), we just need each team's ID number

More FE Intuition

If you are attempting to track your weight gain/loss over time, you don't need an accurate scale.



More FE Intuition

If you are attempting to track your weight gain/loss over time, you don't need an accurate scale.

You only need a scale that is consistently wrong.



More FE Intuition

If you are attempting to track your weight gain/loss over time, you don't need an accurate scale.

You only need a scale that is consistently wrong.

Likewise, with FE in our case study, we know our model is wrong because we don't have ability, but we don't need that information as long as natural ability is consistently wrong (i.e. fixed) over time.



Time De-Meaning

Consider the model with decomposed error term

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Time De-Meaning

Consider the model with decomposed error term

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Take averages over time for each team i

$$\bar{Y}_i = \beta \bar{D}_i + \gamma \bar{X}_i + u_i + \bar{\eta}_i$$

where $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$, etc..

Time De-Meaning

Consider the model with decomposed error term

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Take averages over time for each team i

$$\bar{Y}_i = \beta \bar{D}_i + \gamma \bar{X}_i + u_i + \bar{\eta}_i$$

where $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$, etc..

- Q: Why don't we have a bar on u?

Time De-Meaning

Consider the model with decomposed error term

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Take averages over time for each team i

$$\bar{Y}_i = \beta \bar{D}_i + \gamma \bar{X}_i + u_i + \bar{\eta}_i$$

where $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$, etc..

- Q: Why don't we have a bar on u?
- A: u_i does not vary over time, so there is nothing to average.

Time De-Meaning

Consider the model with decomposed error term

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Take averages over time for each team i

$$\bar{Y}_i = \beta \bar{D}_i + \gamma \bar{X}_i + u_i + \bar{\eta}_i$$

where $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$, etc..

- Q: Why don't we have a bar on u?
- A: u_i does not vary over time, so there is nothing to average.

Notice that we don't include an intercept in this model. If we did, that is called the "between" estimator, which is biased with $Cov(u_i, \{D, X\}) \neq 0$, which is the situation we are trying to overcome.

Eliminating the Fixed Effect u_i

By de-meaning the initial model we get

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + \gamma(X_{it} - \bar{X}_i) + \eta_{it} - \bar{\eta}_i$$

Eliminating the Fixed Effect u_i

By de-meaning the initial model we get

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + \gamma(X_{it} - \bar{X}_i) + \eta_{it} - \bar{\eta}_i$$

Q: What parameter/variable is missing?

Eliminating the Fixed Effect u_i

By de-meaning the initial model we get

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + \gamma(X_{it} - \bar{X}_i) + \eta_{it} - \bar{\eta}_i$$

Q: What parameter/variable is missing?
A: They key here is that the team specific fixed effect u_i is gone.

Eliminating the Fixed Effect u_i

By de-meaning the initial model we get

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + \gamma(X_{it} - \bar{X}_i) + \eta_{it} - \bar{\eta}_i$$

Q: What parameter/variable is missing?
A: They key here is that the team specific fixed effect u_i is gone.
Q: Why is that good news?

Eliminating the Fixed Effect u_i

By de-meaning the initial model we get

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + \gamma(X_{it} - \bar{X}_i) + \eta_{it} - \bar{\eta}_i$$

Q: What parameter/variable is missing?
A: They key here is that the team specific fixed effect u_i is gone.
Q: Why is that good news?
A: u_i was the portion of the error term that generated the OVB. The time de-meaning washed it away.

FE Model

If we define $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.. then we can write

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

FE Model

If we define $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.. then we can write

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

- We can now estimate via pooled OLS

FE Model

If we define $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.. then we can write

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

- We can now estimate via pooled OLS
- The u_i is referred to as the entity-specific fixed effect

FE Model

If we define $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.. then we can write

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

- We can now estimate via pooled OLS
- The u_i is referred to as the entity-specific fixed effect
- The estimators $\hat{\beta}$ and $\hat{\gamma}$ are sometimes referred to as Fixed Effects estimators since they leverage the fact that u_i is *fixed* over time

FE Model

If we define $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc.. then we can write

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

- We can now estimate via pooled OLS
- The u_i is referred to as the entity-specific fixed effect
- The estimators $\hat{\beta}$ and $\hat{\gamma}$ are sometimes referred to as Fixed Effects estimators since they leverage the fact that u_i is *fixed* over time
- This time de-meaning is also referred to as a Within Transformation since it makes a transformation *within* each cross sectional observation

Common temporal trends

Q: Suppose productivity Y is rising over time for all teams due to new tech (e.g. AI tools, etc..). Why wouldn't our (one way) fixed effects approach mitigate the source of the bias?

Common temporal trends

Q: Suppose productivity Y is rising over time for all teams due to new tech (e.g. AI tools, etc..). Why wouldn't our (one way) fixed effects approach mitigate the source of the bias?

A:

- Our one way fixed effects approach removes things that are fixed through time but vary in the cross section

Common temporal trends

Q: Suppose productivity Y is rising over time for all teams due to new tech (e.g. AI tools, etc..). Why wouldn't our (one way) fixed effects approach mitigate the source of the bias?

A:

- Our one way fixed effects approach removes things that are fixed through time but vary in the cross section
- It does not remove things that are common across teams but vary through time (e.g. rising productivity trend)

Common temporal trends

Q: Suppose productivity Y is rising over time for all teams due to new tech (e.g. AI tools, etc..). Why wouldn't our (one way) fixed effects approach mitigate the source of the bias?

A:

- Our one way fixed effects approach removes things that are fixed through time but vary in the cross section
- It does not remove things that are common across teams but vary through time (e.g. rising productivity trend)
- The solution is Two Way Fixed Effects

Two Way Fixed Effects

One Way Fixed Effects

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Two Way Fixed Effects

One Way Fixed Effects

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \;\; t = 1, 2$$

Two Way Fixed Effects

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \tau_t + \eta_{it}, \;\; t = 1, 2$$

Two Way Fixed Effects

One Way Fixed Effects

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}, \ \ t = 1, 2$$

Two Way Fixed Effects

$$Y_{it} = \beta D_{it} + \gamma X_{it} + u_i + \tau_t + \eta_{it}, \ \ t = 1, 2$$

- u_i : team specific, time invariant
- τ_t : time specific, team invariant

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?
- A: Use of Agile (D) is not related to other omitted drivers of Productivity that may vary over time and team (e.g. # of team members, etc..)

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?
- A: Use of Agile (D) is not related to other omitted drivers of Productivity that may vary over time and team (e.g. # of team members, etc..)
- Q: So what type of endogeneity is permitted?

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?
- A: Use of Agile (D) is not related to other omitted drivers of Productivity that may vary over time and team (e.g. # of team members, etc..)
- Q: So what type of endogeneity is permitted?
- A: D_{it} can be related to u_i

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?
- A: Use of Agile (D) is not related to other omitted drivers of Productivity that may vary over time and team (e.g. # of team members, etc..)
- Q: So what type of endogeneity is permitted?
- A: D_{it} can be related to u_i
- Q: What does that mean in our case?

Permitting some Endogeneity

- Notice that we don't require $D_{it} \perp e_{it}$, where $e_{it} = u_i + \eta_{it}$
- We only require $D_{it} \perp \eta_{it}$.
- Q: What does this mean in our case?
- A: Use of Agile (D) is not related to other omitted drivers of Productivity that may vary over time and team (e.g. # of team members, etc..)
- Q: So what type of endogeneity is permitted?
- A: D_{it} can be related to u_i
- Q: What does that mean in our case?
- A: The team's natural ability (u_i) can be related to whether they choose to use Agile (D_{it})

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1 - \theta$ never use Agile. What is the impact on our ATE estimate?

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

If our teams never change treatment status over the time span observed, then for each team i, $D_{i,t=1} = D_{i,t=2} = \cdots = \bar{D}_i$, and so $\tilde{D}_i =$

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

If our teams never change treatment status over the time span observed, then for each team i, $D_{i,t=1} = D_{i,t=2} = \cdots = \bar{D}_i$, and so $\tilde{D}_i = 0 \ \forall i$

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

If our teams never change treatment status over the time span observed, then for each team i, $D_{i,t=1} = D_{i,t=2} = \cdots = \bar{D}_i$, and so $\tilde{D}_i = 0 \ \forall i$
The time de-meaned version is

$$\tilde{Y}_{it} = \gamma \tilde{X}_{it} + \tilde{\eta}_{it}$$

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

If our teams never change treatment status over the time span observed, then for each team i, $D_{i,t=1} = D_{i,t=2} = \cdots = \bar{D}_i$, and so $\tilde{D}_i = 0 \ \forall i$
The time de-meaned version is

$$\tilde{Y}_{it} = \gamma \tilde{X}_{it} + \tilde{\eta}_{it}$$

There is no treatment variable, and hence we can't estimate a treatment effect.

Change in Treatment Effect

Q: Suppose some fraction θ of our teams always use Agile and $1-\theta$ never use Agile. What is the impact on our ATE estimate?

A: Our model is

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + u_i + \eta_{it}$$

If our teams never change treatment status over the time span observed, then for each team i, $D_{i,t=1} = D_{i,t=2} = \cdots = \bar{D}_i$, and so $\tilde{D}_i = 0 \ \forall i$
The time de-meaned version is

$$\tilde{Y}_{it} = \gamma \tilde{X}_{it} + \tilde{\eta}_{it}$$

There is no treatment variable, and hence we can't estimate a treatment effect.
Fixed Effects estimation for ATE requires that at least some of our observations have variation in treatment status over time in order to eliminate the OVB

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Using Pooled Cross Sections

Despite not having a causal interpretation, Pooled Cross Sectional models can still be useful in other settings

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

Some (non causal) questions you can address with Pooled Cross sections:

- What is our prediction of a team's productivity after controlling Agile D use and other factors X?

Using Pooled Cross Sections

Despite not having a causal interpretation, Pooled Cross Sectional models can still be useful in other settings

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + e_{it}$$

Some (non causal) questions you can address with Pooled Cross sections:

- What is our prediction of a team's productivity after controlling Agile D use and other factors X?
- If we include time variables, we can ask: Does productivity vary over time once we control for Agile use and other factors?
- etc...

Pooled OLS Assumptions

Q: What assumption does Pooled OLS impose on the Treatment Effect?

Pooled OLS Assumptions

Q: What assumption does Pooled OLS impose on the Treatment Effect?

A: Simple pooling implies

$$Y_{i,1990} = \beta_0 + \beta_{1990} D_{i,1990} + \cdots + e_{i,1990}$$
$$Y_{j,1991} = \beta_0 + \beta_{1990} D_{j,1991} + \cdots + e_{j,1991}$$

Pooled OLS Assumptions

Q: What assumption does Pooled OLS impose on the Treatment Effect?

A: Simple pooling implies

$$Y_{i,1990} = \beta_0 + \beta_{1990} D_{i,1990} + \cdots + e_{i,1990}$$
$$Y_{j,1991} = \beta_0 + \beta_{1990} D_{j,1991} + \cdots + e_{j,1991}$$

But the underlying process may look like

$$Y_{i,1990} = \beta_0 + \beta_{1990} D_{i,1990} + \cdots + e_{i,1990}$$
$$Y_{j,1991} = \beta_0 + \beta_{1991} D_{j,1991} + \cdots + e_{j,1991}$$

Pooled OLS Assumptions

Q: What assumption does Pooled OLS impose on the Treatment Effect?

A: Simple pooling implies

$$Y_{i,1990} = \beta_0 + \beta_{1990} D_{i,1990} + \cdots + e_{i,1990}$$
$$Y_{j,1991} = \beta_0 + \beta_{1990} D_{j,1991} + \cdots + e_{j,1991}$$

But the underlying process may look like

$$Y_{i,1990} = \beta_0 + \beta_{1990} D_{i,1990} + \cdots + e_{i,1990}$$
$$Y_{j,1991} = \beta_0 + \beta_{1991} D_{j,1991} + \cdots + e_{j,1991}$$

Standard Pooled OLS (and Panel methods more generally) assumes time invariant treatment effects.

Should we pool?

A Test for Parameter Stability

By pooling our data we are assuming that the coefficients for $t = 1$ match those for $t = 2$ (i.e. time invariant). We can use a Chow test to determine if this is wise. For simplicity, we will assume $T = 2$ and no X confounders; the method scales easily.

Should we pool?

A Test for Parameter Stability

By pooling our data we are assuming that the coefficients for $t = 1$ match those for $t = 2$ (i.e. time invariant). We can use a Chow test to determine if this is wise. For simplicity, we will assume $T = 2$ and no X confounders; the method scales easily.

1. Estimate the restricted model on the pooled data OLS:
   $Y_{it} = \alpha_0 + \alpha_1 D_{t=1} + \beta_{it} D_{it} + \gamma D_{t=1} \times D_{it} + e_{it} \rightarrow R^2 = SSR_r$

Should we pool?

A Test for Parameter Stability

By pooling our data we are assuming that the coefficients for $t = 1$ match those for $t = 2$ (i.e. time invariant). We can use a Chow test to determine if this is wise. For simplicity, we will assume $T = 2$ and no X confounders; the method scales easily.

1. Estimate the restricted model on the pooled data OLS: $Y_{it} = \alpha_0 + \alpha_1 D_{t=1} + \beta_{it} D_{it} + \gamma D_{t=1} \times D_{it} + e_{it} \rightarrow R^2 = SSR_r$
2. Estimate the unrestricted model for each time period via OLS: $Y_{i1} = \alpha_1 + \beta_1 D_{i1} + e_{i1} \rightarrow R^2 = SSR_1$. Compute SSR_2 analogously. Construct $SSR_u = SSR_1 + SSR_2$.

Should we pool?

A Test for Parameter Stability

By pooling our data we are assuming that the coefficients for $t = 1$ match those for $t = 2$ (i.e. time invariant). We can use a Chow test to determine if this is wise. For simplicity, we will assume $T = 2$ and no X confounders; the method scales easily.

1. Estimate the restricted model on the pooled data OLS: $Y_{it} = \alpha_0 + \alpha_1 D_{t=1} + \beta_{it} D_{it} + \gamma D_{t=1} \times D_{it} + e_{it} \rightarrow R^2 = SSR_r$
2. Estimate the unrestricted model for each time period via OLS: $Y_{i1} = \alpha_1 + \beta_1 D_{i1} + e_{i1} \rightarrow R^2 = SSR_1$. Compute SSR_2 analogously. Construct $SSR_u = SSR_1 + SSR_2$.
3. We are testing if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. The F-statistic would be $[\frac{SSR_r - SSR_u}{SSR_u} \times \frac{N-2(2)}{2+1}]$. Heteroscedasticity robust versions are available.

Should we pool?

A Test for Parameter Stability

By pooling our data we are assuming that the coefficients for $t = 1$ match those for $t = 2$ (i.e. time invariant). We can use a Chow test to determine if this is wise. For simplicity, we will assume $T = 2$ and no X confounders; the method scales easily.

1. Estimate the restricted model on the pooled data OLS: $Y_{it} = \alpha_0 + \alpha_1 D_{t=1} + \beta_{it} D_{it} + \gamma D_{t=1} \times D_{it} + e_{it} \rightarrow R^2 = SSR_r$
2. Estimate the unrestricted model for each time period via OLS: $Y_{i1} = \alpha_1 + \beta_1 D_{i1} + e_{i1} \rightarrow R^2 = SSR_1$. Compute SSR_2 analogously. Construct $SSR_u = SSR_1 + SSR_2$.
3. We are testing if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. The F-statistic would be $[\frac{SSR_r - SSR_u}{SSR_u} \times \frac{N-2(2)}{2+1}]$. Heteroscedasticity robust versions are available.

If reject the null, then the parameters may not be stable through time, implying that pooling the data may not be wise (without adjustments)

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

FE Assumptions

We can estimate the fixed effect model via OLS

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

assuming the typical assumptions of cross sectional regression hold. (Let's focus on D for now, but the following holds for all regressors) Most importantly,

$$E[\eta_{it} | D_{is}, u_i] = 0$$

for all $s = 1, ..., T$.

FE Assumptions

We can estimate the fixed effect model via OLS

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

assuming the typical assumptions of cross sectional regression hold. (Let's focus on D for now, but the following holds for all regressors) Most importantly,

$$E[\eta_{it} | D_{is}, u_i] = 0$$

for all $s = 1, ..., T$. Q: Can you interpret the following implication?

$$Cov(\eta_{i2}, D_{i2}) = Cov(\eta_{i2}, D_{i1}) = Cov(\eta_{i1}, D_{i2}) = Cov(\eta_{i1}, D_{i1}) = 0$$

FE Assumptions

We can estimate the fixed effect model via OLS

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

assuming the typical assumptions of cross sectional regression hold. (Let's focus on D for now, but the following holds for all regressors) Most importantly,

$$E[\eta_{it} | D_{is}, u_i] = 0$$

for all $s = 1, ..., T$. Q: Can you interpret the following implication?

$$Cov(\eta_{i2}, D_{i2}) = Cov(\eta_{i2}, D_{i1}) = Cov(\eta_{i1}, D_{i2}) = Cov(\eta_{i1}, D_{i1}) = 0$$

A: The error is unrelated to the regressor(s) in *all* time periods (Strict Exogeneity).

Goodness of Fit

There are several ways to compute goodness of fit measures within a fixed effects model.

Goodness of Fit

There are several ways to compute goodness of fit measures within a fixed effects model.
A typical implementation is via the Within Transformation

$$\tilde{Y}_{it} = \beta \tilde{D}_{it} + \gamma \tilde{X}_{it} + \tilde{\eta}_{it}, \;\; t = 1, ..., T$$

and we generate

$$R^2 = 1 - \frac{\sum_{it} \hat{\tilde{\eta}}_{it}^2}{\sum_{it} (\tilde{Y}_{it} - \bar{\tilde{Y}}_{it})^2}$$

Outline

1. Motivation - Case Study Details
2. Panel Data
3. Pooled OLS
   - OLS on Separate Cross Sections through Time
   - OLS on Pooled Cross Sectional Data
4. Panel Models
   - OVB
   - LSDV
   - Fixed Effects (FE)
5. Appendix
   - Non Causal uses of Pooled OLS
   - FE Miscellanea
   - Random Effects

Defining a Random Effect

Consider a general Fixed Effect framework (treatment can be any of the X's)

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

Defining a Random Effect

Consider a general Fixed Effect framework (treatment can be any of the X's)

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

We used Time Averaging to eliminate u_i because we were concerned it was correlated with X_{it}

Defining a Random Effect

Consider a general Fixed Effect framework (treatment can be any of the X's)

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

We used Time Averaging to eliminate u_i because we were concerned it was correlated with X_{it}
If u_i is NOT correlated with X_{it} eliminating it might impact our standard errors. Note that this obviates our original need for fixed effects to overcome OVB and estimate the treatment effect. So, RE is a general tool.

Defining a Random Effect

Consider a general Fixed Effect framework (treatment can be any of the X's)

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + u_i + \eta_{it}, \;\; t = 1, 2, ..., T$$

We used Time Averaging to eliminate u_i because we were concerned it was correlated with X_{it}
If u_i is NOT correlated with X_{it} eliminating it might impact our standard errors. Note that this obviates our original need for fixed effects to overcome OVB and estimate the treatment effect. So, RE is a general tool.
In a Random Effects model we assume

$$Cov(X_{itj}, u_i) = 0, \;\; t = 1, 2, ...T; j = 1, 2, ..., K$$

Estimating a Random Effects Model

We can rewrite the model as

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + v_{it}, \;\; t = 1, 2, ..., T$$

where $v_{it} = u_i + \eta_{it}$.

Estimating a Random Effects Model

We can rewrite the model as

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \cdots + \beta_K X_{itK} + v_{it}, \ \ t = 1, 2, ..., T$$

where $v_{it} = u_i + \eta_{it}$.
We can estimate this via typical cross sectional or pooled OLS methods and generate unbiased estimates. However,

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}, \ \ t \neq s$$

If this (serial) correlation among the errors exists, it can impact our standard errors.

Fixing the Standard Errors

For balanced panel with N larger than T we can conduct a quasi-time demeaning

$$Y_{it} - \lambda \bar{Y}_i = \beta_0(1-\lambda) + \beta_1(X_{it1} - \lambda \bar{X}_{i1}) + \cdots + \beta_K(X_{itK} - \bar{X}_{iK}) + (v_{it} - \lambda \bar{v}_i)$$

where $\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$.

Fixing the Standard Errors

For balanced panel with N larger than T we can conduct a quasi-time demeaning

$$Y_{it} - \lambda \bar{Y}_i = \beta_0(1-\lambda) + \beta_1(X_{it1} - \lambda \bar{X}_{i1}) + \cdots + \beta_K(X_{itK} - \bar{X}_{iK}) + (v_{it} - \lambda \bar{v}_i)$$

where $\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$. A common approach to estimate λ:

$$\hat{\sigma}_a^2 = \frac{1}{[NT(T-1)/2 - (k+1)]} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it}\hat{v}_{is}$$

where $\hat{v}$ is derived via pooled OLS. Then, $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

We can see that RE $\Rightarrow$ Pooled OLS as $\lambda \Rightarrow 0$

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

We can see that RE $\Rightarrow$ Pooled OLS as $\lambda \Rightarrow 0$

- This is the case when a_i is unimportant (i.e. σ_a is small relative to σ_u)

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

We can see that RE $\Rightarrow$ FE as $\lambda \Rightarrow 1$

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

We can see that RE $\Rightarrow$ FE as $\lambda \Rightarrow 1$

- If $Cov(X_{itj}, a_i) \neq 0$, then RE estimators are biased

Nice feature of Random Effects

Consider λ

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$$

We can see that RE $\Rightarrow$ FE as $\lambda \Rightarrow 1$

- If $Cov(X_{itj}, a_i) \neq 0$, then RE estimators are biased
- Notice $v_{it} - \lambda\bar{v}_i = (1-\lambda)a_i + u_{it} - \lambda\bar{u}_i$
- Think of $1-\lambda$ as the weight of a in the error, which impacts the bias
- If λ is close to 0, there is high weight and potentially high bias
- If λ is close to 1, then weight and potential bias shrink
- This is sensible since FE is fine under this setting and RE devolves to FE when $\lambda = 1$

Summary of Panel Methods

- Panel methods provide opportunity to explore causality in other settings.
- $T = 2$ — First Difference is attractive for its simplicity
- $N > T$ — Fixed Effects are preferred to First Difference. Can accommodate entity or time fixed effects
- Fixed Effects generally preferred to Random Effects for causality because permits arbitrary correlation between X_{it} and a_i. Random Effects may be useful in settings where the key explanatory variable is constant over time.

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment
  - With two-way FE, identification uses within-unit variation net of unit and time effects

Panel FE — Summary

- Fixed Effects (FE) mitigate omitted variable bias when unobserved confounders have the right structure
  - Panel data: the same units i observed over time
  - Omitted factors vary across i but are fixed over time
  - Identification requires within-unit variation in the treatment
  - With two-way FE, identification uses within-unit variation net of unit and time effects
- Under these conditions, we do not need to observe the omitted confounders themselves