Difference in Differences

Prof. Mike Aguilar

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

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Difference in Differences (DiD) - Summary

- Natural Experiment

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data
- Treated observations switch from untreated to treated state around some exogenous policy event

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data
- Treated observations switch from untreated to treated state around some exogenous policy event
- Identification (i.e. estimation of the treatment effect) rests upon a common trends assumption; i.e. if the event never happened the difference in treated and control groups would remain unchanged.

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: panel data, fixed effects

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Case Study

Does Remote Work Impact Productivity?



Background

- Remote work boomed in the aftermath of Covid-19.

Background

- Remote work boomed in the aftermath of Covid-19.
- Proponents believe that remote work enhances productivity through various channels:

Background

- Remote work boomed in the aftermath of Covid-19.
- Proponents believe that remote work enhances productivity through various channels: more working hours due to shorter commute times, effective pay raises due to less commuting costs, increased morale, etc..

Background

- Remote work boomed in the aftermath of Covid-19.
- Proponents believe that remote work enhances productivity through various channels: more working hours due to shorter commute times, effective pay raises due to less commuting costs, increased morale, etc..
- Opponents believe that remote work reduces productivity through various channels:

Background

- Remote work boomed in the aftermath of Covid-19.
- Proponents believe that remote work enhances productivity through various channels: more working hours due to shorter commute times, effective pay raises due to less commuting costs, increased morale, etc..
- Opponents believe that remote work reduces productivity through various channels: distractions in the home workspace, conflicting personal obligations during working hours, etc..

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

- Unit of measurement is

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

- Unit of measurement is a call-center-worker i

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

- Unit of measurement is a call-center-worker *i*
- Productivity measured in various ways (e.g.

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

- Unit of measurement is a call-center-worker i
- Productivity measured in various ways (e.g. number of calls completed per day, customer satisfaction ratings, employee reviews, etc..)

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

- Unit of measurement is a call-center-worker i
- Productivity measured in various ways (e.g. number of calls completed per day, customer satisfaction ratings, employee reviews, etc..)
- Myriad worker characteristics available, such as their pre-COVID office/remote status

Study Details

Background

- A large call-center operator permitted employees to choose working in-office or remotely.
- Amid COVID, the operator went fully remote on 15Mar2020.
- Further details in Emanuel and Harrington, FRBNY Working Paper No.1061 May 2023

Data

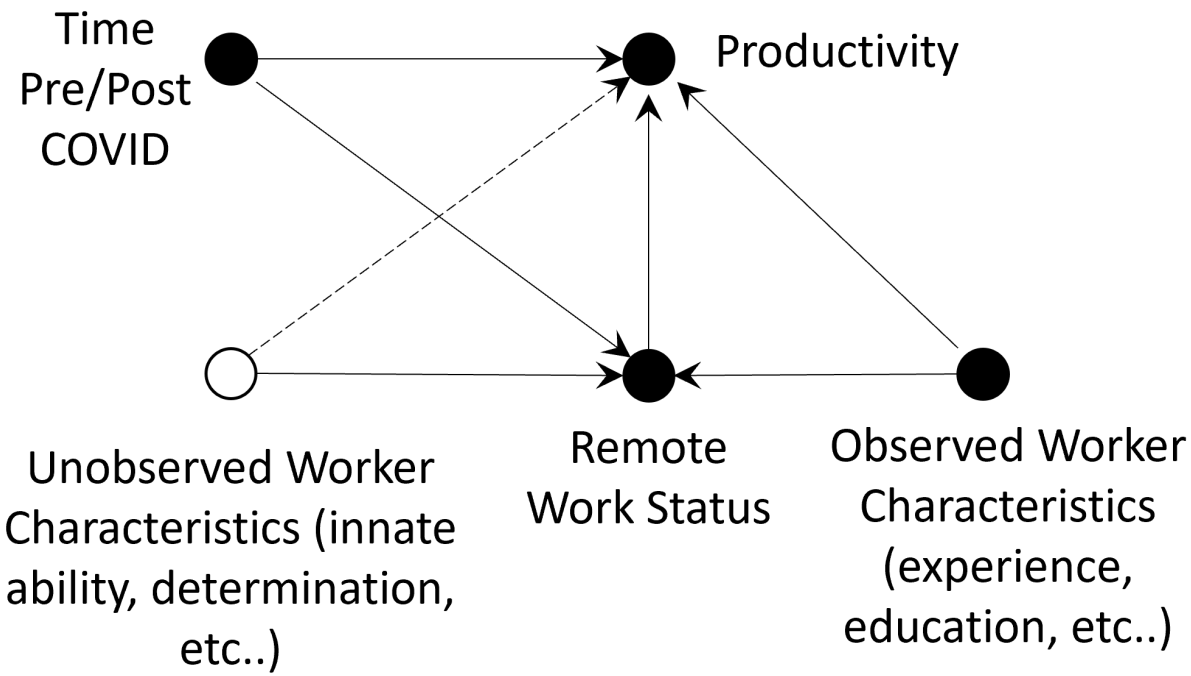
- Unit of measurement is a call-center-worker *i*
- Productivity measured in various ways (e.g. number of calls completed per day, customer satisfaction ratings, employee reviews, etc..)
- Myriad worker characteristics available, such as their pre-COVID office/remote status
- Data measured daily Jan2019-Oct2021

Identify the Causal Path

Q: Can you draw a DAG and identify key causal variables?

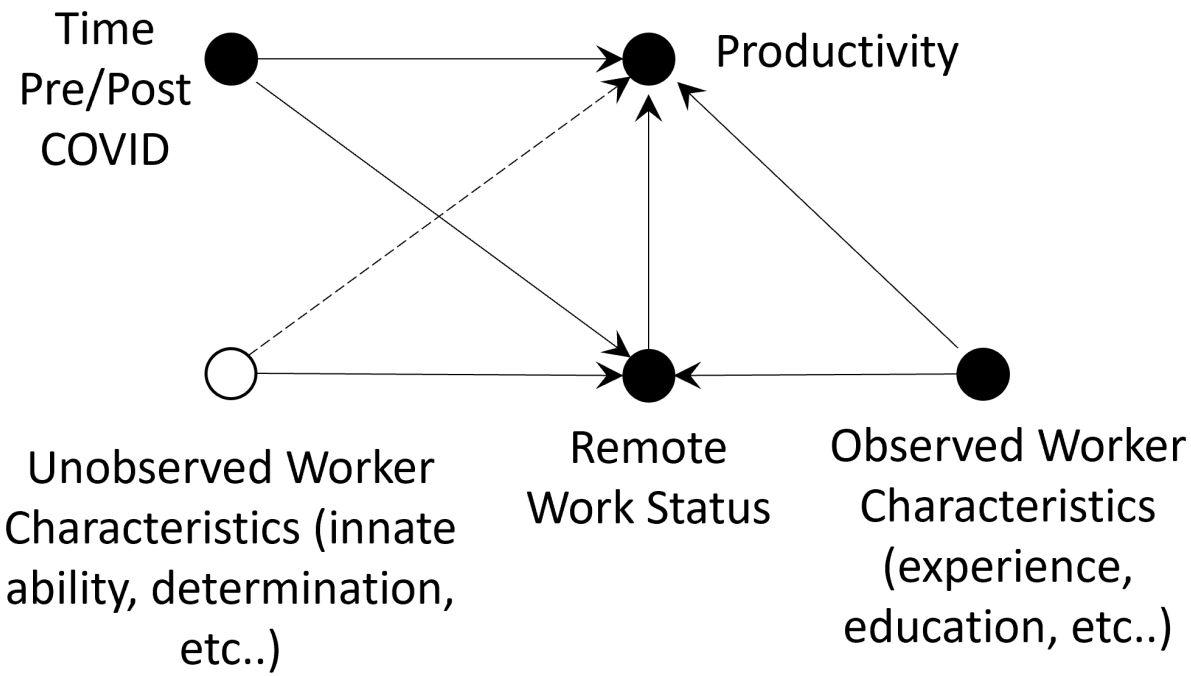
Identify the Causal Path

Q: Can you draw a DAG and identify key causal variables?



Identify the Causal Path

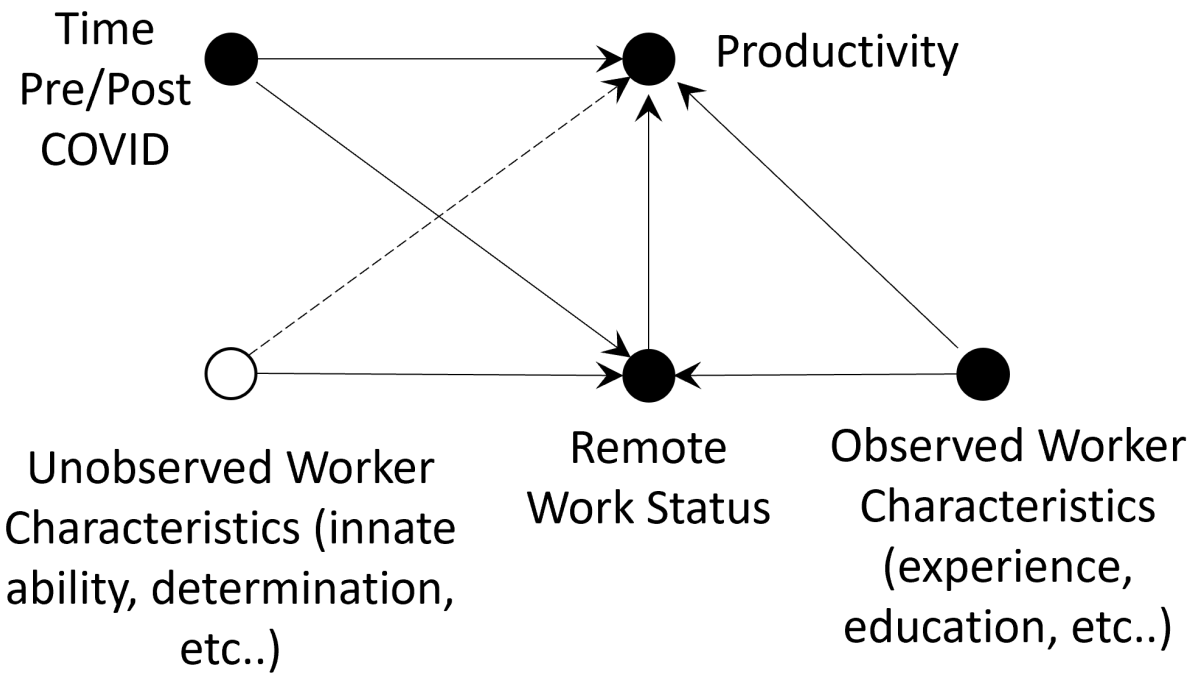
Q: Can you draw a DAG and identify key causal variables?



- Outcome:

Identify the Causal Path

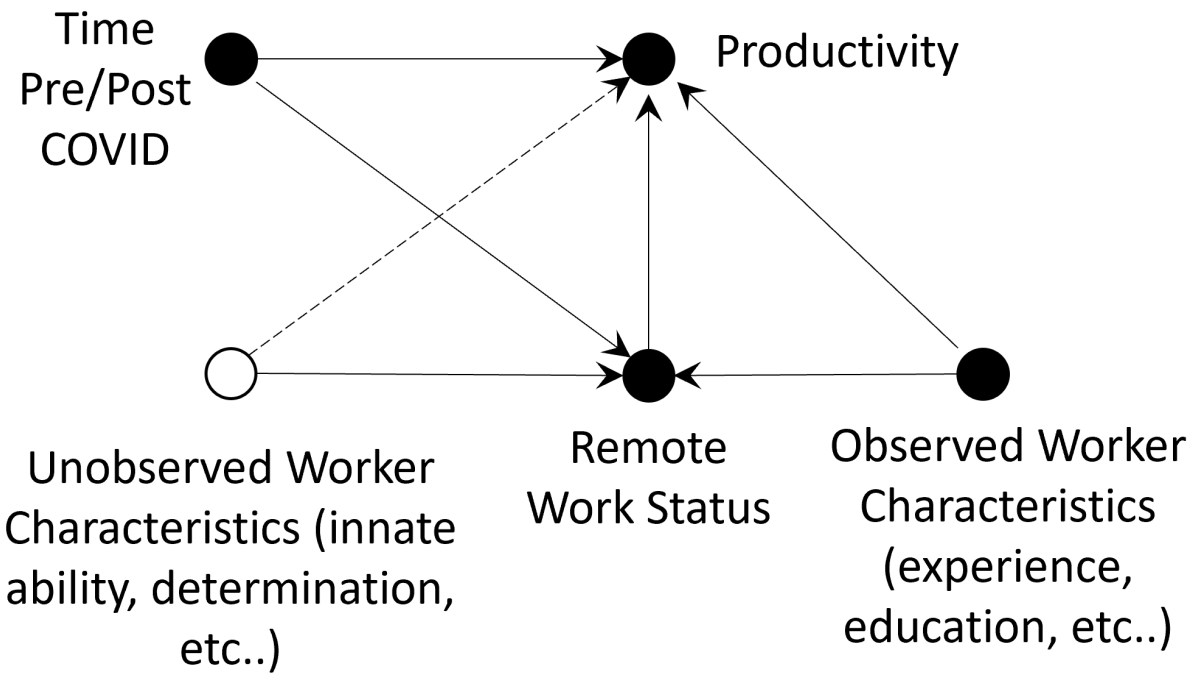
Q: Can you draw a DAG and identify key causal variables?



- Outcome: Productivity
- Treatment:

Identify the Causal Path

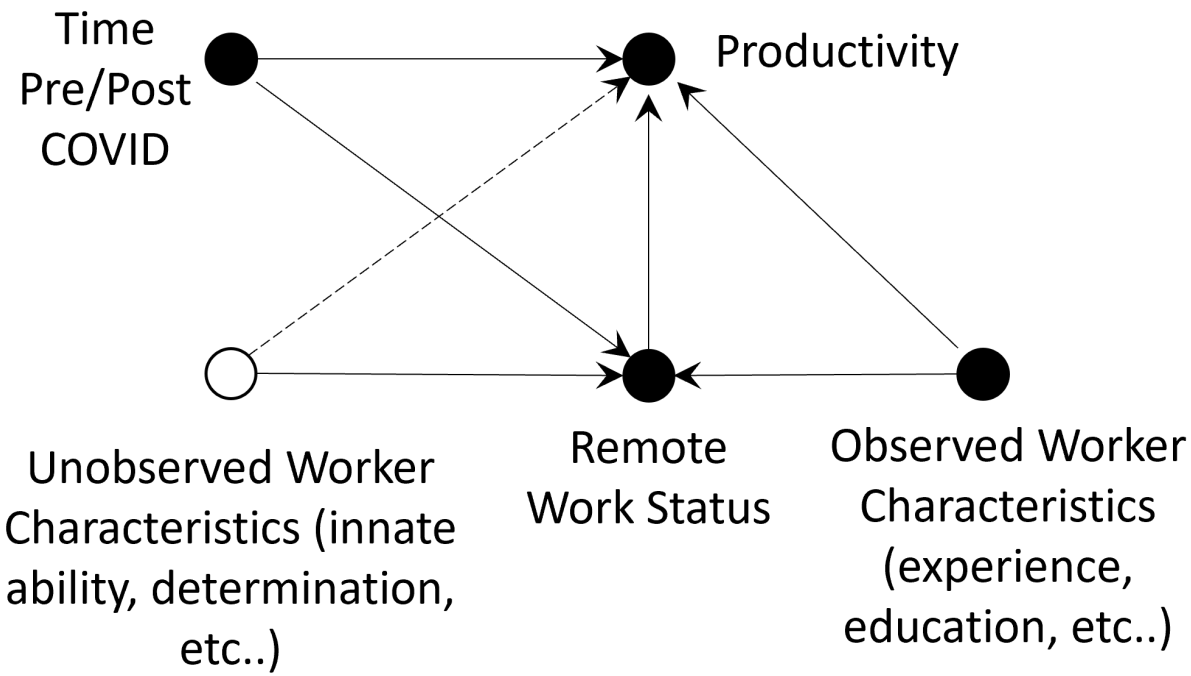
Q: Can you draw a DAG and identify key causal variables?



- Outcome: Productivity
- Treatment: Remote work status
- Treatment Effect:

Identify the Causal Path

Q: Can you draw a DAG and identify key causal variables?



- Outcome: Productivity
- Treatment: Remote work status
- Treatment Effect: Impact of remote work on productivity

Note: The dashed line has no special meaning in this case

Natural Experiment

Q: Why is this a natural experiment?

Natural Experiment

Q: Why is this a natural experiment?

- Controlled Experiment: The analyst knows and can control every aspect of the data generating process. The analyst can change any variable within the model to evoke a treatment effect. (Very unlikely to see)

Natural Experiment

Q: Why is this a natural experiment?

- Controlled Experiment: The analyst knows and can control every aspect of the data generating process. The analyst can change any variable within the model to evoke a treatment effect. (Very unlikely to see)
- Random Experiment: The analyst randomly selects individuals into treated and control group (i.e. RCT).

Natural Experiment

Q: Why is this a natural experiment?

- Controlled Experiment: The analyst knows and can control every aspect of the data generating process. The analyst can change any variable within the model to evoke a treatment effect. (Very unlikely to see)
- Random Experiment: The analyst randomly selects individuals into treated and control group (i.e. RCT).
- Natural Experiment: An exogenous policy intervention separates individuals into treated and control groups.

Natural Experiment

Q: Why is this a natural experiment?

- Controlled Experiment: The analyst knows and can control every aspect of the data generating process. The analyst can change any variable within the model to evoke a treatment effect. (Very unlikely to see)
- Random Experiment: The analyst randomly selects individuals into treated and control group (i.e. RCT).
- Natural Experiment: An exogenous policy intervention separates individuals into treated and control groups.
- Observational Study: The treated and control individuals are observed without any intervention from the analyst.

Natural Experiment

Q: Why is this a natural experiment?

- Controlled Experiment: The analyst knows and can control every aspect of the data generating process. The analyst can change any variable within the model to evoke a treatment effect. (Very unlikely to see)
- Random Experiment: The analyst randomly selects individuals into treated and control group (i.e. RCT).
- Natural Experiment: An exogenous policy intervention separates individuals into treated and control groups.
- Observational Study: The treated and control individuals are observed without any intervention from the analyst.

A: The exogenous nature of the corporate policy change is the key. The call-center employees did not make the policy change to go fully remote nor did we (the analysts). Of course, this identification assumption needs to be probed.

DiD

In the following we will motivate the Difference in Differences (DiD) estimator from several perspectives.

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Potential Outcomes Framework

Standard potential outcomes notation is complicated by the presence of 2 groups of workers (treat, control), 2 potential outcomes (work remotely or in-office), and 2 time periods (2019, 2020).

Potential Outcomes Framework

Standard potential outcomes notation is complicated by the presence of 2 groups of workers (treat, control), 2 potential outcomes (work remotely or in-office), and 2 time periods (2019, 2020).

We could define $D_i = 1$ if individual i worked in office in 2019 and switched to remote in 2020, and $D_t = 1$ if the year is 2019.

Potential Outcomes Framework

Standard potential outcomes notation is complicated by the presence of 2 groups of workers (treat, control), 2 potential outcomes (work remotely or in-office), and 2 time periods (2019, 2020).

We could define $D_i = 1$ if individual i worked in office in 2019 and switched to remote in 2020, and $D_t = 1$ if the year is 2019.

Then $E[Y_{1,i}|D_i = 1, D_t = 1]$ captures the expected productivity for employees that chose to work in the office in 2019, in the year 2019, under the potential state of working remotely.

Potential Outcomes Framework

Standard potential outcomes notation is complicated by the presence of 2 groups of workers (treat, control), 2 potential outcomes (work remotely or in-office), and 2 time periods (2019, 2020).

We could define $D_i = 1$ if individual i worked in office in 2019 and switched to remote in 2020, and $D_t = 1$ if the year is 2019.

Then $E[Y_{1,i}|D_i = 1, D_t = 1]$ captures the expected productivity for employees that chose to work in the office in 2019, in the year 2019, under the potential state of working remotely.

We could replicate this notation for various combinations of D_i and D_t, but that is cumbersome.

Simplifying Notation

To simplify notation, let's define the following two groups.

Simplifying Notation

To simplify notation, let's define the following two groups.

- Treated aka Switchers (worked in-office in 2019 then switched to remote in 2020)

Simplifying Notation

To simplify notation, let's define the following two groups.

- Treated aka Switchers (worked in-office in 2019 then switched to remote in 2020)
- Control aka Stayers (worked remotely in 2019 and 2020)

Simplifying Notation

To simplify notation, let's define the following two groups.

- Treated aka Switchers (worked in-office in 2019 then switched to remote in 2020)
- Control aka Stayers (worked remotely in 2019 and 2020)

We can now rewrite the potential outcome as

$$E[Y_{1,i}|D_i = 1, D_t = 1] = Y^{Switcher}_{Remote,2019}$$

Simplifying Notation

To simplify notation, let's define the following two groups.

- Treated aka Switchers (worked in-office in 2019 then switched to remote in 2020)
- Control aka Stayers (worked remotely in 2019 and 2020)

We can now rewrite the potential outcome as

$$E[Y_{1,i}|D_i = 1, D_t = 1] = Y^{Switcher}_{Remote,2019}$$

This is the expected productivity of a switcher working remotely in 2019.

Simplifying Notation

To simplify notation, let's define the following two groups.

- Treated aka Switchers (worked in-office in 2019 then switched to remote in 2020)
- Control aka Stayers (worked remotely in 2019 and 2020)

We can now rewrite the potential outcome as

$$E[Y_{1,i}|D_i = 1, D_t = 1] = Y^{Switcher}_{Remote,2019}$$

This is the expected productivity of a switcher working remotely in 2019. Let's apply this notation by exploring how to structure the DiD estimator.

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias.

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2020}] =$

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2020}] = \text{ATT}$

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}] =$

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}] = \text{Selection Bias}$

Using 2020

Q: Can we compare the Switchers and Stayers in 2020 for TE?

A: No. Using only 2020 opens us up to a selection bias. Leveraging the observed outcomes in 2020

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}] = \text{Selection Bias}$

where in this case the Selection Bias captures inherent productivity differences between the Switchers and Stayers measured in 2020.

Using 2020

Signing the bias

We identified the selection bias as

$$E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

Using 2020

Signing the bias

We identified the selection bias as

$$E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

Q: Do you suspect a positive or negative selection bias?

We identified the selection bias as

$$E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

Q: Do you suspect a positive or negative selection bias?
A: Either is possible. An argument can be made for positive. e.g. The selection bias here compares the Switchers and Stayers in their preferred states (Office and Remote, respectively). Perhaps the Stayers initially chose Remote work because they were less productive due to family situations that required Remote work, but which impeded their productivity relative to the Switchers.

Using 2020

Signing the bias

We identified the selection bias as

$$E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

Q: Do you suspect a positive or negative selection bias?
A: Either is possible. An argument can be made for positive. e.g. The selection bias here compares the Switchers and Stayers in their preferred states (Office and Remote, respectively). Perhaps the Stayers initially chose Remote work because they were less productive due to family situations that required Remote work, but which impeded their productivity relative to the Switchers.

Q: Based on the suspected sign of the bias, would the (biased) estimates that use 2020 data overstate or understate work from home productivity?

Using 2020

Signing the bias

We identified the selection bias as

$$E[Y^{Switchers}_{Office,2020}] - E[Y^{Stayers}_{Remote,2020}]$$

Q: Do you suspect a positive or negative selection bias?
A: Either is possible. An argument can be made for positive. e.g. The selection bias here compares the Switchers and Stayers in their preferred states (Office and Remote, respectively). Perhaps the Stayers initially chose Remote work because they were less productive due to family situations that required Remote work, but which impeded their productivity relative to the Switchers.

Q: Based on the suspected sign of the bias, would the (biased) estimates that use 2020 data overstate or understate work from home productivity?
A: If the bias is positive, then the estimate based only on 2020 data would be overstating the treatment effect.

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Office,2019}] - E[Y^{Switchers}_{Remote,2019}] =$

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Office,2019}] - E[Y^{Switchers}_{Remote,2019}] =$ "ATT"

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Office,2019}] - E[Y^{Switchers}_{Remote,2019}] =$ "ATT"
- $E[Y^{Switchers}_{Remote,2019}] - E[Y^{Stayers}_{Remote,2019}] =$

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Office,2019}] - E[Y^{Switchers}_{Remote,2019}] =$ "ATT"
- $E[Y^{Switchers}_{Remote,2019}] - E[Y^{Stayers}_{Remote,2019}] =$ Selection Bias

Using 2019

Q: Can we compare the Switchers and Stayers in 2019 for TE?

A: No. Again, there is a selection bias.

Leveraging the observed outcomes in 2019

$$E[Y^{Switchers}_{Office,2019}] - E[Y^{Stayers}_{Remote,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switchers}_{Office,2019}] - E[Y^{Switchers}_{Remote,2019}] =$ "ATT"
- $E[Y^{Switchers}_{Remote,2019}] - E[Y^{Stayers}_{Remote,2019}] =$ Selection Bias

Notice that the ATT is of the opposite sign as intended (Office - Remote) instead of (Remote - Office).

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] =$

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] = \text{ATT}$

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] =$

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = \Delta E[Y^{Switcher}_{Office}] =$

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = \Delta E[Y^{Switcher}_{Office}] = \text{Organic Growth}$

Using 2019 and 2020 for Treated

Q: Can we compare the switchers in 2019 and 2020 TE?

A: No, but we are getting close.

Leveraging what we observe in each year for the Switchers

$$E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2019}]$$

is comprised of two unobserved parts

- $E[Y^{Switcher}_{Remote,2020}] - E[Y^{Switcher}_{Office,2020}] = \text{ATT}$
- $E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = \Delta E[Y^{Switcher}_{Office}] = \text{Organic Growth}$

In this case the "Organic Growth" captures the expected change in productivity between 2019 and 2020 among the Switchers if they had stayed in the office. This could be influenced by COVID and other forces.

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

So the probability limit of $\hat{\beta}$ is

$$plim\, \hat{\beta} = \beta + \frac{Cov(D_{it} - \bar{D}_i,\ \tau_t - \bar{\tau})}{Var(D_{it} - \bar{D}_i)} + \frac{Cov(D_{it} - \bar{D}_i,\ \eta_{it} - \bar{\eta}_i)}{Var(D_{it} - \bar{D}_i)}$$

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

So the probability limit of $\hat{\beta}$ is

$$plim\, \hat{\beta} = \beta + \frac{Cov(D_{it} - \bar{D}_i,\ \tau_t - \bar{\tau})}{Var(D_{it} - \bar{D}_i)} + \frac{Cov(D_{it} - \bar{D}_i,\ \eta_{it} - \bar{\eta}_i)}{Var(D_{it} - \bar{D}_i)}$$

- We allow $Cov(D_{it}, u_i) \neq 0$ (FE removes u_i)

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

So the probability limit of $\hat{\beta}$ is

$$plim\,\hat{\beta} = \beta + \frac{Cov(D_{it} - \bar{D}_i,\ \tau_t - \bar{\tau})}{Var(D_{it} - \bar{D}_i)} + \frac{Cov(D_{it} - \bar{D}_i,\ \eta_{it} - \bar{\eta}_i)}{Var(D_{it} - \bar{D}_i)}$$

- We allow $Cov(D_{it}, u_i) \neq 0$ (FE removes u_i)
- If D_{it} is correlated with $\tau_t \Rightarrow$ common time-shock bias

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

So the probability limit of $\hat{\beta}$ is

$$plim\,\hat{\beta} = \beta + \frac{Cov(D_{it} - \bar{D}_i,\ \tau_t - \bar{\tau})}{Var(D_{it} - \bar{D}_i)} + \frac{Cov(D_{it} - \bar{D}_i,\ \eta_{it} - \bar{\eta}_i)}{Var(D_{it} - \bar{D}_i)}$$

- We allow $Cov(D_{it}, u_i) \neq 0$ (FE removes u_i)
- If D_{it} is correlated with $\tau_t \Rightarrow$ common time-shock bias
- If D_{it} is correlated with $\eta_{it} \Rightarrow$ differential-trend bias

An FE Solution?

Q: Can we use one-way unit-FE to resolve the problem?

Suppose the true DGP is $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
But we estimate a one-way FE model:

$$Y_{it} = \beta D_{it} + u_i + \varepsilon_{it} \quad \text{where } \varepsilon_{it} = \tau_t + \eta_{it}$$

After applying the Within (time-demeaning) transformation:

$$Y_{it} - \bar{Y}_i = \beta(D_{it} - \bar{D}_i) + (\tau_t - \bar{\tau}) + (\eta_{it} - \bar{\eta}_i)$$

So the probability limit of $\hat{\beta}$ is

$$plim\,\hat{\beta} = \beta + \frac{Cov(D_{it} - \bar{D}_i,\ \tau_t - \bar{\tau})}{Var(D_{it} - \bar{D}_i)} + \frac{Cov(D_{it} - \bar{D}_i,\ \eta_{it} - \bar{\eta}_i)}{Var(D_{it} - \bar{D}_i)}$$

- We allow $Cov(D_{it}, u_i) \neq 0$ (FE removes u_i)
- If D_{it} is correlated with $\tau_t \Rightarrow$ common time-shock bias
- If D_{it} is correlated with $\eta_{it} \Rightarrow$ differential-trend bias

One-way FE removes selection on u_i, but not time-related sources of bias.

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Our estimate of β is: (technically, this is the plim of $\hat{\beta}$)

$$\hat{\beta} = \beta + \frac{Cov(\tilde{D}_{it},\ \tilde{\eta}_{it})}{Var(\tilde{D}_{it})}$$

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Our estimate of β is: (technically, this is the plim of $\hat{\beta}$)

$$\hat{\beta} = \beta + \frac{Cov(\tilde{D}_{it},\ \tilde{\eta}_{it})}{Var(\tilde{D}_{it})}$$

- α_i removes time-invariant unit heterogeneity.

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Our estimate of β is: (technically, this is the plim of $\hat{\beta}$)

$$\hat{\beta} = \beta + \frac{Cov(\tilde{D}_{it},\ \tilde{\eta}_{it})}{Var(\tilde{D}_{it})}$$

- α_i removes time-invariant unit heterogeneity.
- γ_t removes common time shocks (the τ_t term).

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Our estimate of β is: (technically, this is the plim of $\hat{\beta}$)

$$\hat{\beta} = \beta + \frac{Cov(\tilde{D}_{it},\ \tilde{\eta}_{it})}{Var(\tilde{D}_{it})}$$

- α_i removes time-invariant unit heterogeneity.
- γ_t removes common time shocks (the τ_t term).
- Remaining bias requires $Cov(\tilde{D}_{it},\ \tilde{\eta}_{it}) \neq 0$.

Can we use Two-Way FE as a Solution?

Q: Does adding time fixed effects remove the bias?

Start from the same DGP $Y_{it} = \beta D_{it} + u_i + \tau_t + \eta_{it}$
Estimate the two-way FE model (include time effects):

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \nu_{it}$$

Define the double-demeaned (within-by-time) variables

$$\tilde{D}_{it} \equiv D_{it} - \bar{D}_i - \bar{D}_t + \bar{D}, \qquad \tilde{\eta}_{it} \equiv \eta_{it} - \bar{\eta}_i - \bar{\eta}_t + \bar{\eta}.$$

Our estimate of β is: (technically, this is the plim of $\hat{\beta}$)

$$\hat{\beta} = \beta + \frac{Cov(\tilde{D}_{it},\ \tilde{\eta}_{it})}{Var(\tilde{D}_{it})}$$

- α_i removes time-invariant unit heterogeneity.
- γ_t removes common time shocks (the τ_t term).
- Remaining bias requires $Cov(\tilde{D}_{it},\ \tilde{\eta}_{it}) \neq 0$.

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

We can estimate the TE by

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

We can estimate the TE by

- comparing the Switchers' growth over time with the Stayers' growth over time (i.e. DiD)...

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

We can estimate the TE by

- comparing the Switchers' growth over time with the Stayers' growth over time (i.e. DiD)...
- ...if we assume Common (aka Parallel) Trends

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

We can estimate the TE by

- comparing the Switchers' growth over time with the Stayers' growth over time (i.e. DiD)...
- ...if we assume Common (aka Parallel) Trends , which we'll see as $Cov(\tilde{D}_{it},\ \tilde{\eta}_{it}) = 0$

Difference in Differences

We can't estimate TE by

- comparing the individuals before treatment (2019)
- comparing the individuals after treatment (2020)
- computing the growth of one individual before and after treatment
- one-way (but Two Way FE was close)

We can estimate the TE by

- comparing the Switchers' growth over time with the Stayers' growth over time (i.e. DiD)...
- ...if we assume Common (aka Parallel) Trends , which we'll see as $Cov(\tilde{D}_{it},\ \tilde{\eta}_{it}) = 0$
- In this sense, DiD is applying Two-Way FE in a natural experiment under the Common (aka Parallel) Trends assumption

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

$$E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = E[Y^{Stayer}_{Remote,2020}] - E[Y^{Stayer}_{Remote,2019}]$$

$$\Delta E[Y^{Switcher}_{Office}] = \Delta E[Y^{Stayer}_{Remote}]$$

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

$$E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = E[Y^{Stayer}_{Remote,2020}] - E[Y^{Stayer}_{Remote,2019}]$$

$$\Delta E[Y^{Switcher}_{Office}] = \Delta E[Y^{Stayer}_{Remote}]$$

This is NOT saying productivity is the same for Switchers and Stayers in any year or in any state of nature.

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

$$E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = E[Y^{Stayer}_{Remote,2020}] - E[Y^{Stayer}_{Remote,2019}]$$

$$\Delta E[Y^{Switcher}_{Office}] = \Delta E[Y^{Stayer}_{Remote}]$$

This is NOT saying productivity is the same for Switchers and Stayers in any year or in any state of nature.
This is NOT saying that productivity growth is the same for Switchers and Stayers.

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

$$E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = E[Y^{Stayer}_{Remote,2020}] - E[Y^{Stayer}_{Remote,2019}]$$

$$\Delta E[Y^{Switcher}_{Office}] = \Delta E[Y^{Stayer}_{Remote}]$$

This is NOT saying productivity is the same for Switchers and Stayers in any year or in any state of nature.
This is NOT saying that productivity growth is the same for Switchers and Stayers.
The Parallel Trends Assumption says that productivity growth of Stayers matches the productivity growth of Switchers in the potential states where those Switchers remained in the office.

Parallel Trends Assumption

Assume the Organic Growth of the Stayers is the same as the Switchers

$$E[Y^{Switcher}_{Office,2020}] - E[Y^{Switcher}_{Office,2019}] = E[Y^{Stayer}_{Remote,2020}] - E[Y^{Stayer}_{Remote,2019}]$$

$$\Delta E[Y^{Switcher}_{Office}] = \Delta E[Y^{Stayer}_{Remote}]$$

This is NOT saying productivity is the same for Switchers and Stayers in any year or in any state of nature.

This is NOT saying that productivity growth is the same for Switchers and Stayers.

The Parallel Trends Assumption says that productivity growth of Stayers matches the productivity growth of Switchers in the potential states where those Switchers remained in the office.

This is a questionable assumption that is difficult to test directly.

See appendix for a derivation of ATT by leveraging the Common trends

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

A Simplified Dataset

Stacia=Stayer (Control); Sweeney=Switcher (Treated)



Allowed to keep working remotely



2019



Forced to leave the office and go remote

2020



2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher			
Stayer			
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic		
Stayer			
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	
Stayer			
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer			
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic		
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	Time
Switcher-Stayer			

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	Time
Switcher-Stayer	Diff.Intrins.		

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	Time
Switcher-Stayer	Diff.Intrins.	Diff.Intrins. + Treatment	

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	Time
Switcher-Stayer	Diff.Intrins.	Diff.Intrins. + Treatment	Treatment

2x2 DiD

2 groups (Treated=Switcher, Control=Stayer), over 2 periods (2020/2019)

Productivity	2019	2020	Difference
Switcher	Intrinsic	Intrinsic + Time + Treatment	Time + Treatment
Stayer	Intrinsic	Intrinsic + Time	Time
Switcher-Stayer	Diff.Intrins.	Diff.Intrins. + Treatment	Treatment

By taking the difference in the differences, we can isolate a treatment effect.

Computing ATE in 2 ways

$\widehat{ATE}$

Computing ATE in 2 ways

$$\widehat{ATE} = (Y_{2020}^{Switcher} - Y_{2019}^{Switcher}) -$$

Computing ATE in 2 ways

$$\widehat{ATE} = (Y_{2020}^{Switcher} - Y_{2019}^{Switcher}) - (Y_{2020}^{Stayer} - Y_{2019}^{Stayer})$$

Computing ATE in 2 ways

$$\begin{aligned}\widehat{ATE} &= (Y_{2020}^{Switcher} - Y_{2019}^{Switcher}) - (Y_{2020}^{Stayer} - Y_{2019}^{Stayer}) \\ &= (Y_{2020}^{Switcher} - Y_{2020}^{Stayer}) -\end{aligned}$$

Computing ATE in 2 ways

$$\begin{aligned}\widehat{ATE} &= (Y_{2020}^{Switcher} - Y_{2019}^{Switcher}) - (Y_{2020}^{Stayer} - Y_{2019}^{Stayer}) \\ &= (Y_{2020}^{Switcher} - Y_{2020}^{Stayer}) - (Y_{2019}^{Switcher} - Y_{2019}^{Stayer})\end{aligned}$$

Technical note: DiD provides ATE under full independence of errors and regressors (see Blundell and MacCurdy 2000). However, we can relax full independence to lack of correlation and impose constant treatment effects across all individuals. In which case we produce the ATT (see Athey and Imbens 2006 pg436).

Visualizing the 2x2 DiD

There are several ways to visualize the 2x2 DiD

Visualizing the 2x2 DiD

There are several ways to visualize the 2x2 DiD

1. Difference in the growth rates over time of Switchers and Stayers

Visualizing the 2x2 DiD

There are several ways to visualize the 2x2 DiD

1. Difference in the growth rates over time of Switchers and Stayers
2. Difference in Switchers and Stayers in 2020 and 2019

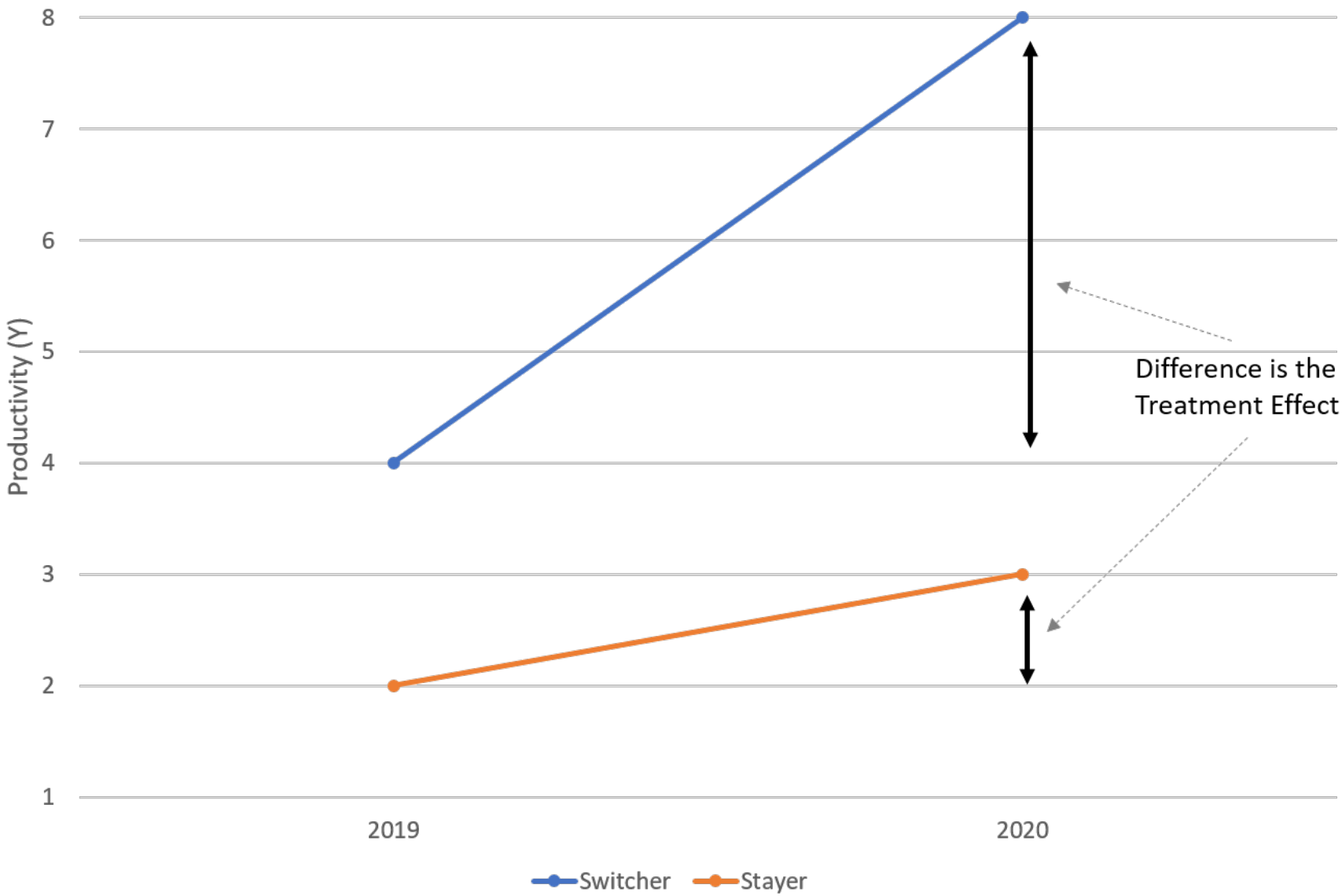
Visualizing the 2x2 DiD

There are several ways to visualize the 2x2 DiD

1. Difference in the growth rates over time of Switchers and Stayers
2. Difference in Switchers and Stayers in 2020 and 2019
3. Difference in Switcher Productivity in 2020 and potential outcome in 2019

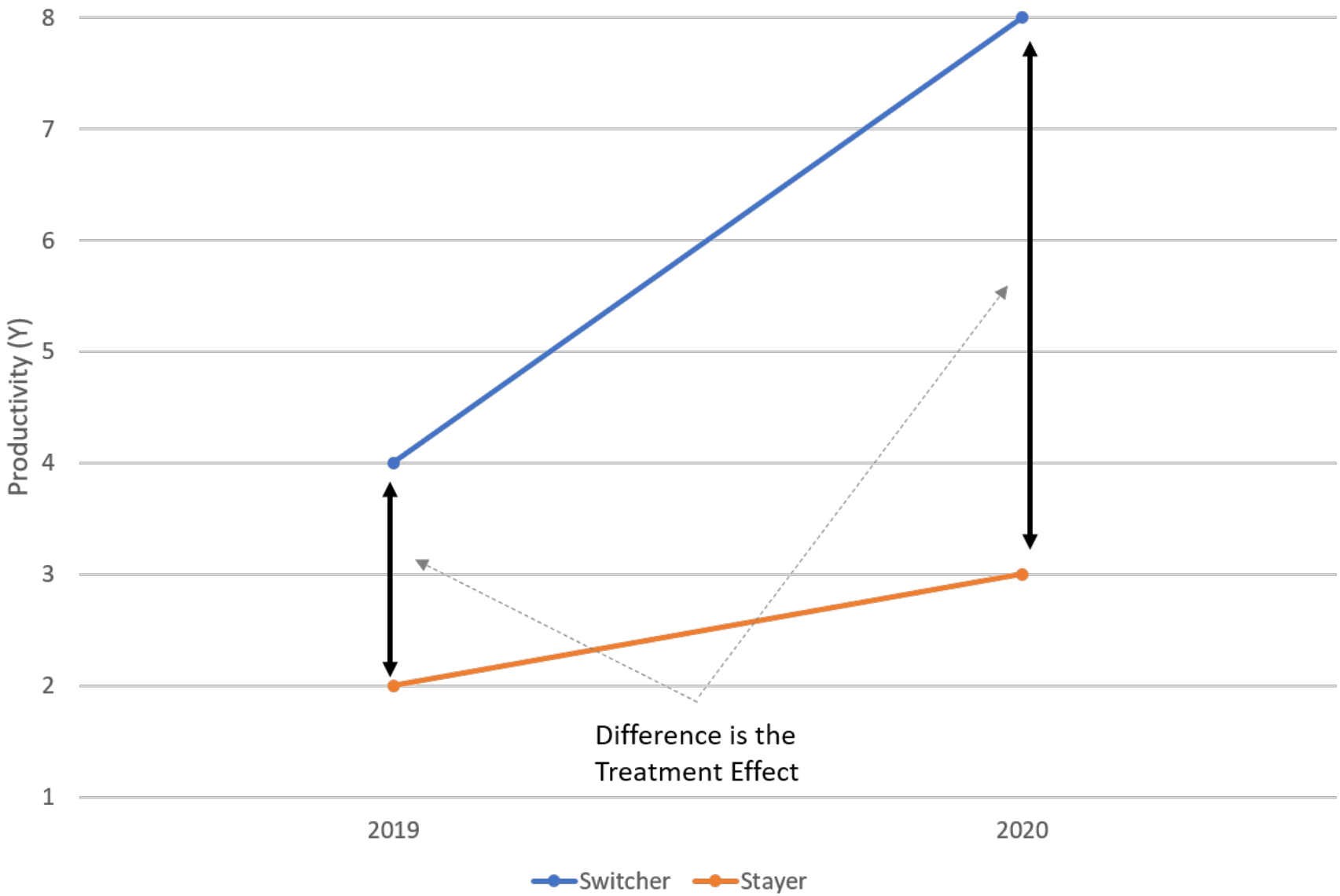
Visualizing the 2x2 DiD

Difference in the growth rates over time of Switchers and Stayers



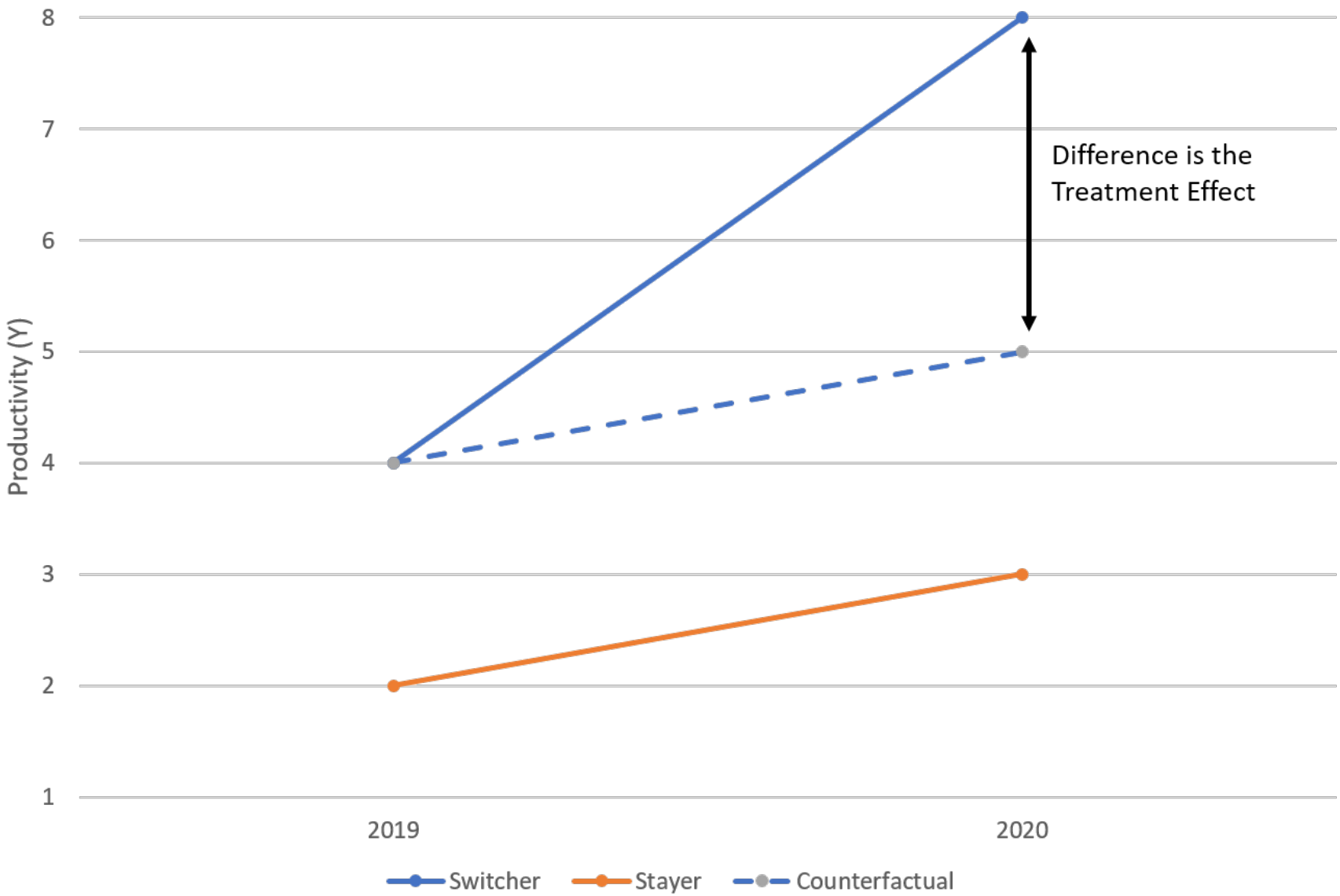
Visualizing the 2x2 DiD

Difference in Switchers and Stayers in 2020 and 2019



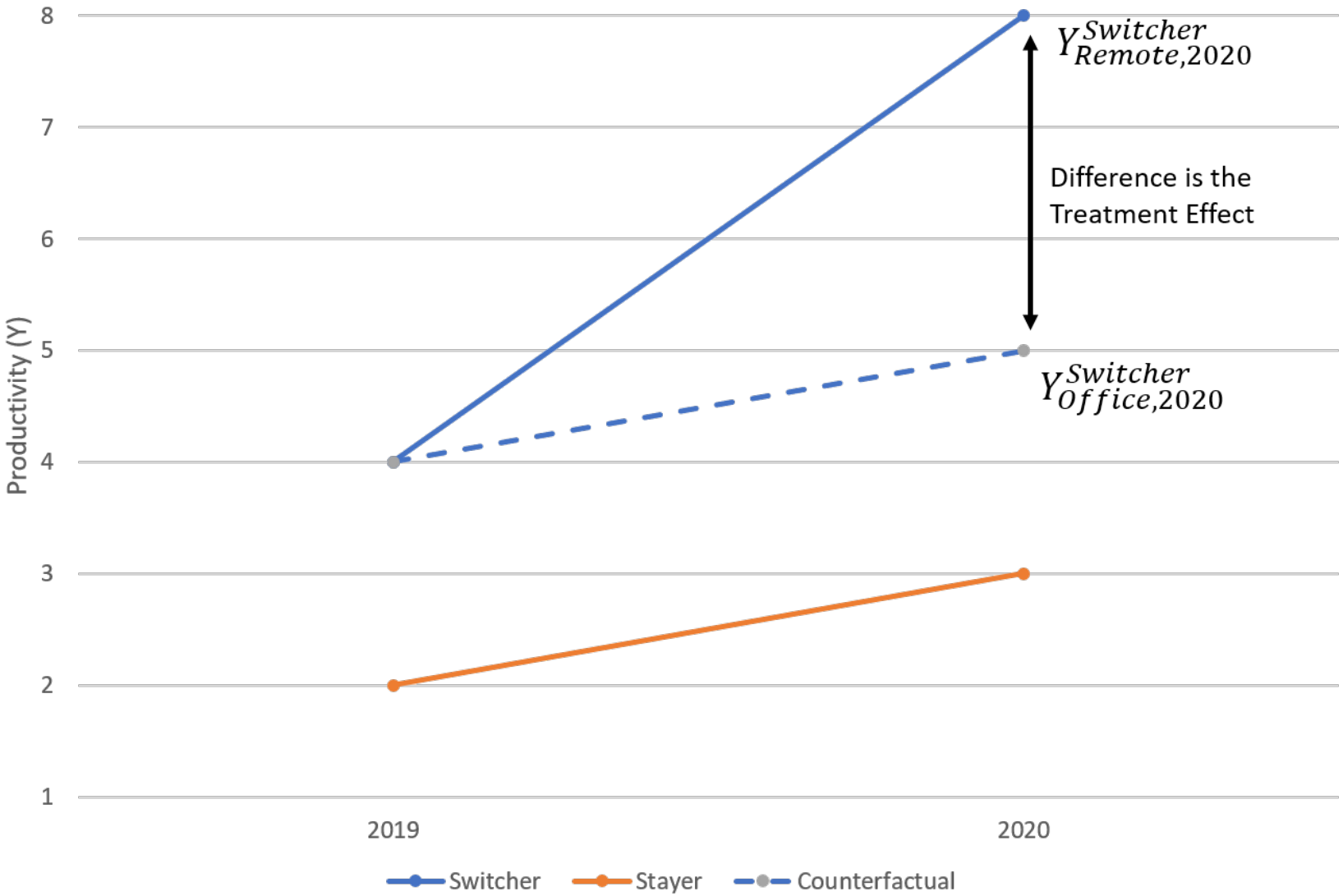
Visualizing the 2x2 DiD

Difference in Observed Productivity in 2020 and potential outcome in 2020



Visualizing the 2x2 DiD

Estimate TE via Sweeney's productivity in 2020 minus what Sweeney's productivity would have been had he followed his relationship with Stacia and worked in the office



Pros and Cons of 2x2 DiD

Q: What are the Pros of 2x2 DiD? Cons?

Pros and Cons of 2x2 DiD

Q: What are the Pros of 2x2 DiD? Cons?

Pros

- Very intuitive
- Very simple

Pros and Cons of 2x2 DiD

Q: What are the Pros of 2x2 DiD? Cons?

Pros

- Very intuitive
- Very simple

Cons

- No control variables
- Small sample
  - Lack of statistical confidence
  - No inference
  - Can't probe parallel trends assumption

Pros and Cons of 2x2 DiD

Q: What are the Pros of 2x2 DiD? Cons?

Pros

- Very intuitive
- Very simple

Cons

- No control variables
- Small sample
  - Lack of statistical confidence
  - No inference
  - Can't probe parallel trends assumption

We can overcome these weaknesses if we have more time and more people. But first we need to recast this simple 2x2 case as a regression.

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

A 2x2 Regression

In order to accommodate more time and more people we recast as regression

Define $Treat_i = 1$ if in the treated group (i.e. Sweeney the Switcher), and 0 if in the control group (i.e. Stacia the Stayer).

A 2x2 Regression

In order to accommodate more time and more people we recast as regression

Define $Treat_i = 1$ if in the treated group (i.e. Sweeney the Switcher), and 0 if in the control group (i.e. Stacia the Stayer).
Define $Post_t = 1$ if after treatment (i.e. 2020), and 0 if before treatment (i.e. 2019).

A 2x2 Regression

In order to accommodate more time and more people we recast as regression

Define $Treat_i = 1$ if in the treated group (i.e. Sweeney the Switcher), and 0 if in the control group (i.e. Stacia the Stayer).
Define $Post_t = 1$ if after treatment (i.e. 2020), and 0 if before treatment (i.e. 2019).

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$E[Y_{it}|Treat = 1, Post = 1] =$$
$$E[Y_{it}|Treat = 1, Post = 0] =$$
$$E[Y_{it}|Treat = 0, Post = 1] =$$
$$E[Y_{it}|Treat = 0, Post = 0] =$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned}
E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\
E[Y_{it}|Treat = 1, Post = 0] =& \\
E[Y_{it}|Treat = 0, Post = 1] =& \\
E[Y_{it}|Treat = 0, Post = 0] =&
\end{aligned}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned}
E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\
E[Y_{it}|Treat = 1, Post = 0] =& \alpha + \beta \\
E[Y_{it}|Treat = 0, Post = 1] =& \\
E[Y_{it}|Treat = 0, Post = 0] =&
\end{aligned}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned}
E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\
E[Y_{it}|Treat = 1, Post = 0] =& \alpha + \beta \\
E[Y_{it}|Treat = 0, Post = 1] =& \alpha + \gamma \\
E[Y_{it}|Treat = 0, Post = 0] =&
\end{aligned}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned}
E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\
E[Y_{it}|Treat = 1, Post = 0] =& \alpha + \beta \\
E[Y_{it}|Treat = 0, Post = 1] =& \alpha + \gamma \\
E[Y_{it}|Treat = 0, Post = 0] =& \alpha
\end{aligned}$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned} E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\ E[Y_{it}|Treat = 1, Post = 0] =& \alpha + \beta \\ E[Y_{it}|Treat = 0, Post = 1] =& \alpha + \gamma \\ E[Y_{it}|Treat = 0, Post = 0] =& \alpha \end{aligned}$$

$$ATE = \Delta E[Y_{it}|Treated = 1] - \Delta E[Y_{it}|Treated = 0] = (\gamma + \delta) - \gamma = \delta$$

Interpreting the DiD Regression

Q: Which parameter is the TE estimate?

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times Post_t + e_{it}$$

A:

$$\begin{aligned} E[Y_{it}|Treat = 1, Post = 1] =& \alpha + \beta + \gamma + \delta \\ E[Y_{it}|Treat = 1, Post = 0] =& \alpha + \beta \\ E[Y_{it}|Treat = 0, Post = 1] =& \alpha + \gamma \\ E[Y_{it}|Treat = 0, Post = 0] =& \alpha \end{aligned}$$

$$ATE = \Delta E[Y_{it}|Treated = 1] - \Delta E[Y_{it}|Treated = 0] = (\gamma + \delta) - \gamma = \delta$$

$$\widehat{ATE} = \Delta Y^{Switcher} - \Delta Y^{Stayer} = \hat{\delta}$$

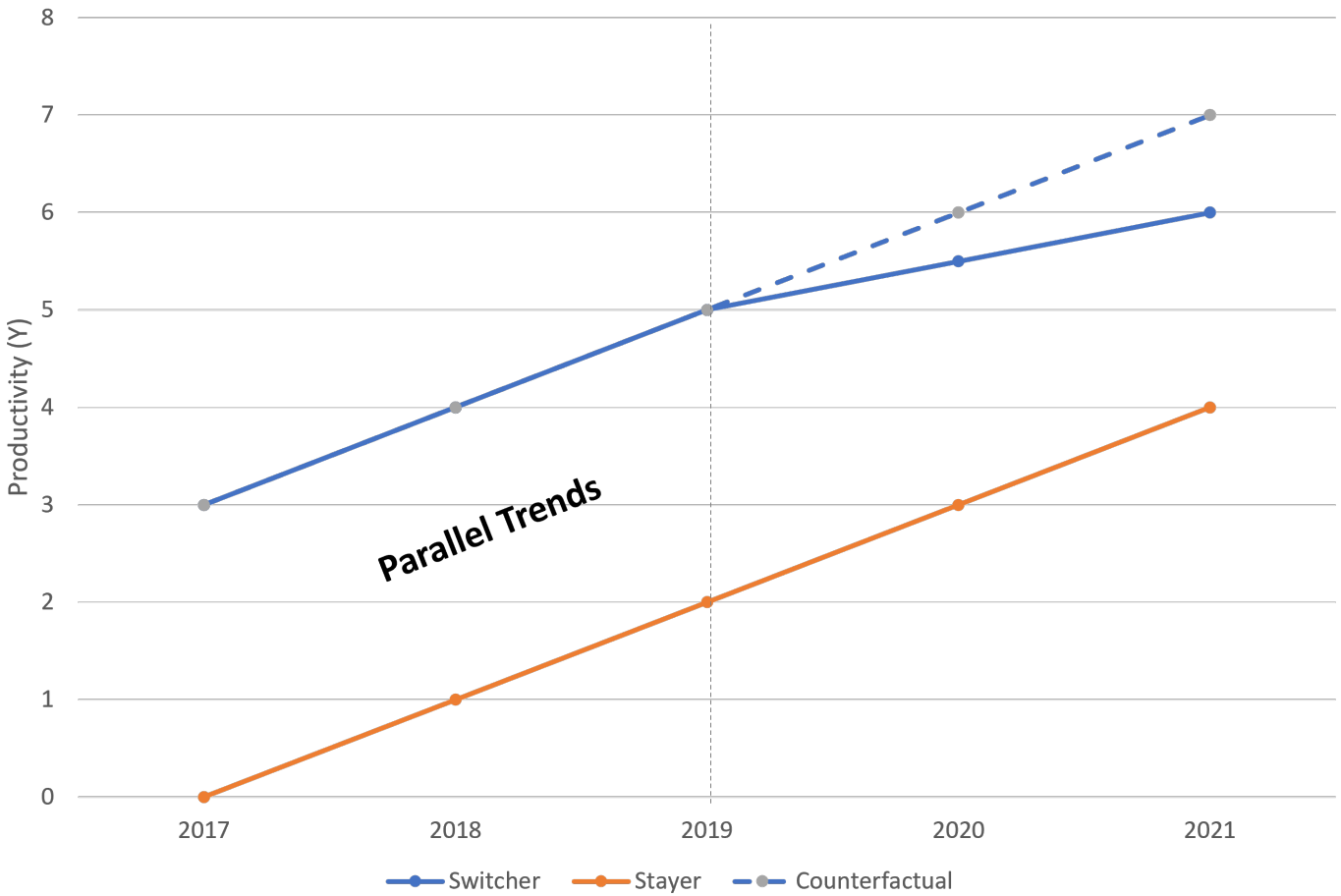
Extending over time

Extending data to cover more time is straightforward

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2018	0	1	0
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	1	1	1
Stacia (Stayer)	2018	0	0	0
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	1	0	0

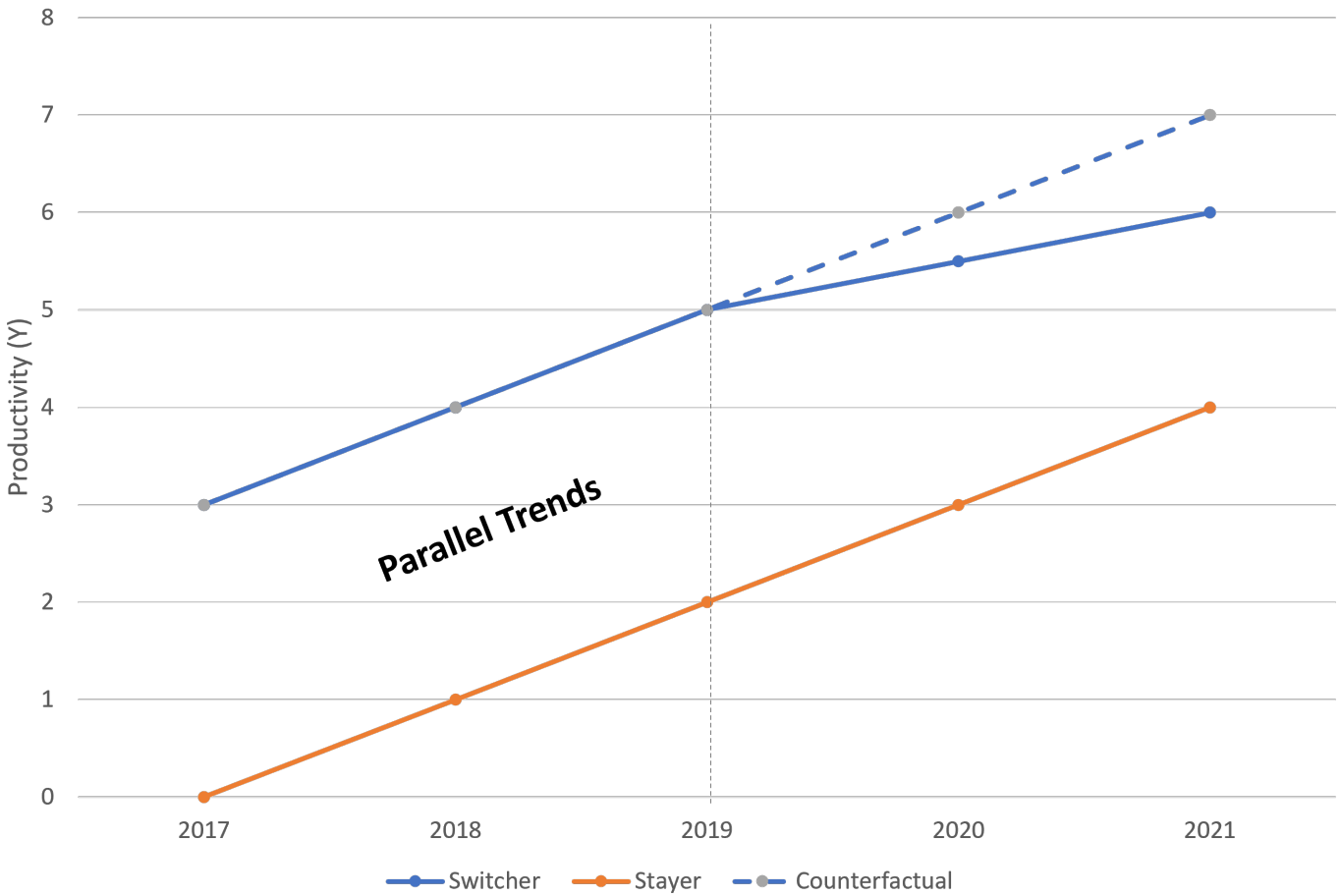
Probing Common Trends

More data pre-intervention permits visual inspection of parallel trends assumption



Probing Common Trends

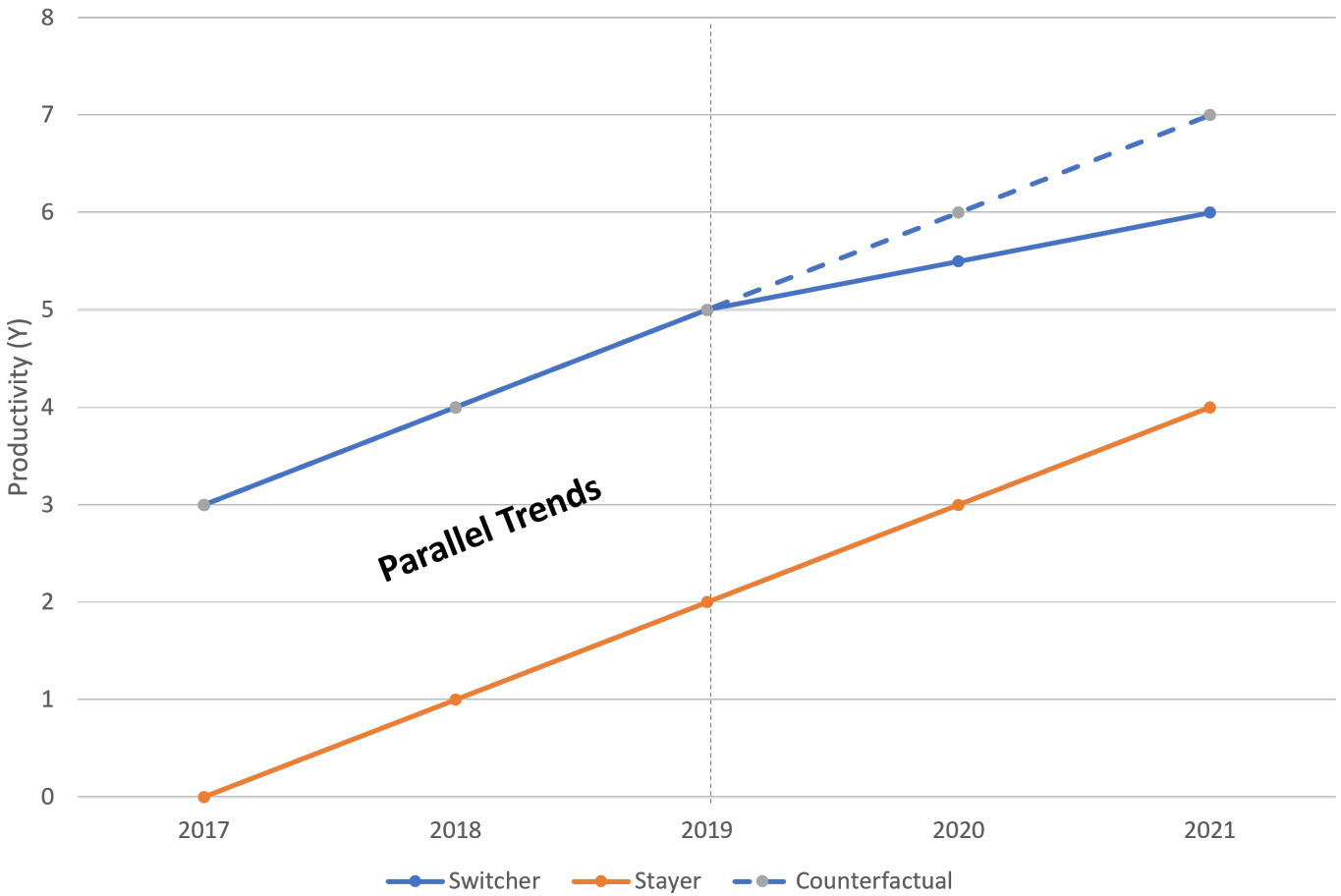
More data pre-intervention permits visual inspection of parallel trends assumption



Q: Why is this often-used heuristic dangerous?

Probing Common Trends

More data pre-intervention permits visual inspection of parallel trends assumption



Q: Why is this often-used heuristic dangerous?

A: It uses only the pre-treatment periods, whereas the Parallel Trends assumption is about post treatment

Extending to more individuals

Extending data to cover more employees requires care

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2018	0	1	0
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	1	1	1
Stacia (Stayer)	2018	0	0	0
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	1	0	0
Swane (Switcher)	2018	0	1	0
⋮	⋮	⋮	⋮	⋮

Extending to more individuals

Extending data to cover more employees requires care

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2018	0	1	0
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	1	1	1
Stacia (Stayer)	2018	0	0	0
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	1	0	0
Swane (Switcher)	2018	0	1	0
⋮	⋮	⋮	⋮	⋮

We can recast ATE as averages: In this case $\widehat{ATE} = (\bar{Y}_{2020}^{Switchers} - \bar{Y}_{2020}^{Stayers}) - (\bar{Y}_{2019}^{Switchers} - \bar{Y}_{2019}^{Stayers})$, which equals δ from the regression setting.

Extending to more individuals

Extending data to cover more employees requires care

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2018	0	1	0
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	1	1	1
Stacia (Stayer)	2018	0	0	0
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	1	0	0
Swane (Switcher)	2018	0	1	0
⋮	⋮	⋮	⋮	⋮

We can recast ATE as averages: In this case $\widehat{ATE} = (\bar{Y}_{2020}^{Switchers} - \bar{Y}_{2020}^{Stayers}) - (\bar{Y}_{2019}^{Switchers} - \bar{Y}_{2019}^{Stayers})$, which equals δ from the regression setting.

Caution: Including more employees into this setup requires that the treatment was imposed at the same time for everyone.

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

A:

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	?	1	?
Sweeney (Switcher)	2021	1	1	1
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	?	0	?
Stacia (Stayer)	2021	1	0	0
Swane (Switcher)	2019	0	1	0
Swane (Switcher)	2020	?	1	?
Swane (Switcher)	2021	1	1	0

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

A:

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	?	1	?
Sweeney (Switcher)	2021	1	1	1
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	?	0	?
Stacia (Stayer)	2021	1	0	0
Swane (Switcher)	2019	0	1	0
Swane (Switcher)	2020	?	1	?
Swane (Switcher)	2021	1	1	0

- $Post_t$ is supposed to vary over time, not across individuals

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

A:

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	?	1	?
Sweeney (Switcher)	2021	1	1	1
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	?	0	?
Stacia (Stayer)	2021	1	0	0
Swane (Switcher)	2019	0	1	0
Swane (Switcher)	2020	?	1	?
Swane (Switcher)	2021	1	1	0

- $Post_t$ is supposed to vary over time, not across individuals
- For Stacia and Sweeney $Post_{2020} = 1$, $Post_{2021} = 1$

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

A:

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	?	1	?
Sweeney (Switcher)	2021	1	1	1
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	?	0	?
Stacia (Stayer)	2021	1	0	0
Swane (Switcher)	2019	0	1	0
Swane (Switcher)	2020	?	1	?
Swane (Switcher)	2021	1	1	0

- $Post_t$ is supposed to vary over time, not across individuals
- For Stacia and Sweeney $Post_{2020} = 1$, $Post_{2021} = 1$
- But, Swane didn't switch to remote until 2021, so for him we should have $Post_{2020} = 0$ and $Post_{2021} = 1$

Different Treatment Times

Q: What is the problem if Swane switched in 2021 (e.g. Staggered policy implementation)

A:

Employee	Year	Post	Treat	Post x Treat
Sweeney (Switcher)	2019	0	1	0
Sweeney (Switcher)	2020	?	1	?
Sweeney (Switcher)	2021	1	1	1
Stacia (Stayer)	2019	0	0	0
Stacia (Stayer)	2020	?	0	?
Stacia (Stayer)	2021	1	0	0
Swane (Switcher)	2019	0	1	0
Swane (Switcher)	2020	?	1	?
Swane (Switcher)	2021	1	1	0

- $Post_t$ is supposed to vary over time, not across individuals
- For Stacia and Sweeney $Post_{2020} = 1$, $Post_{2021} = 1$
- But, Swane didn't switch to remote until 2021, so for him we should have $Post_{2020} = 0$ and $Post_{2021} = 1$
- $Post_t$ is ill defined since it varies across individuals. We need to re-engineer the data.

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Restructuring the dataset

Employee	Year	Post	Treat	PostxTreat
Sweeney	2019	0	1	0
Sweeney	2020	?	1	?
Sweeney	2021	1	1	1
Stacia	2019	0	0	0
Stacia	2020	?	0	?
Stacia	2021	1	0	0
Swane	2019	0	1	0
Swane	2020	?	1	?
Swane	2021	1	1	0

Restructuring the dataset

Define

- Dummy variables for each employee

Employee	Year	Post	Treat	PostxTreat
Sweeney	2019	0	1	0
Sweeney	2020	?	1	?
Sweeney	2021	1	1	1
Stacia	2019	0	0	0
Stacia	2020	?	0	?
Stacia	2021	1	0	0
Swane	2019	0	1	0
Swane	2020	?	1	?
Swane	2021	1	1	0

Restructuring the dataset

Define

- Dummy variables for each employee
- Dummy variable for each time period

Employee	Year	Post	Treat	PostxTreat
Sweeney	2019	0	1	0
Sweeney	2020	?	1	?
Sweeney	2021	1	1	1
Stacia	2019	0	0	0
Stacia	2020	?	0	?
Stacia	2021	1	0	0
Swane	2019	0	1	0
Swane	2020	?	1	?
Swane	2021	1	1	0

Restructuring the dataset

Define

- Dummy variables for each employee
- Dummy variable for each time period
- Dummy $D_{it}^{Treat} = 1$ if employee i at time t is treated ("forced" to work remotely) and 0 otherwise.

Employee	Year	Post	Treat	PostxTreat
Sweeney	2019	0	1	0
Sweeney	2020	?	1	?
Sweeney	2021	1	1	1
Stacia	2019	0	0	0
Stacia	2020	?	0	?
Stacia	2021	1	0	0
Swane	2019	0	1	0
Swane	2020	?	1	?
Swane	2021	1	1	0

Restructuring the dataset

Define

- Dummy variables for each employee
- Dummy variable for each time period
- Dummy $D_{it}^{Treat} = 1$ if employee i at time t is treated ("forced" to work remotely) and 0 otherwise.

Employee	Year	Post	Treat	PostxTreat	D(2019)	D(2020)	D(2021)	D(Sweeney)	D(Stacia)	D(Swane)	D(Treat)
Sweeney	2019	0	1	0	1	0	0	1	0	0	0
Sweeney	2020	?	1	?	0	1	0	1	0	0	1
Sweeney	2021	1	1	1	0	0	1	1	0	0	1
Stacia	2019	0	0	0	0	0	0	0	1	0	0
Stacia	2020	?	0	?	0	1	0	0	1	0	0
Stacia	2021	1	0	0	0	0	1	0	1	0	0
Swane	2019	0	1	0	0	0	0	0	0	1	0
Swane	2020	?	1	?	0	1	0	0	0	1	0
Swane	2021	1	1	0	0	0	1	0	0	1	1

Restructuring the dataset

Define

- Dummy variables for each employee
- Dummy variable for each time period
- Dummy $D_{it}^{Treat} = 1$ if employee i at time t is treated ("forced" to work remotely) and 0 otherwise.

Employee	Year	Post	Treat	PostxTreat	D(2019)	D(2020)	D(2021)	D(Sweeney)	D(Stacia)	D(Swane)	D(Treat)
Sweeney	2019	0	1	0	1	0	0	1	0	0	0
Sweeney	2020	?	1	?	0	1	0	1	0	0	1
Sweeney	2021	1	1	1	0	0	1	1	0	0	1
Stacia	2019	0	0	0	0	0	0	0	1	0	0
Stacia	2020	?	0	?	0	1	0	0	1	0	0
Stacia	2021	1	0	0	0	0	1	0	1	0	0
Swane	2019	0	1	0	0	0	0	0	0	1	0
Swane	2020	?	1	?	0	1	0	0	0	1	0
Swane	2021	1	1	0	0	0	1	0	0	1	1

Notice how we can recreate the original variables (left block) from the expanded structure (right block).

Reframing the model

Our original model

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times After_t + e_{it}$$

Reframing the model

Our original model

$$Y_{it} = \alpha + \beta\, Treat_i + \gamma Post_t + \delta\, Treat_i \times After_t + e_{it}$$

translates to an unobserved fixed effects model with a composite error term

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j\, Time_{jt} + u_i + \eta_{it}$$

which will permit a lot more flexibility. We can estimate this model in several ways.

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Differencing to Eliminate Fixed Effect and OVB

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

For simplicity, consider the case of two time periods (2019,2020) so that $Time_t$ devolves to $Time_t = 1$ for t $= 2020$ and $Time_t = 0$ for 2019.

Differencing to Eliminate Fixed Effect and OVB

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

For simplicity, consider the case of two time periods (2019,2020) so that $Time_t$ devolves to $Time_t = 1$ for t = 2020 and $Time_t = 0$ for 2019.

$$\begin{aligned} Y_{i,2020} &= \alpha + \beta D_{i,2020}^{Treat} + \gamma + u_i + \eta_{i,2020} \\ Y_{i,2019} &= \alpha + \beta D_{i,2019}^{Treat} + u_i + \eta_{i,2019} \end{aligned}$$

Differencing to Eliminate Fixed Effect and OVB

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

For simplicity, consider the case of two time periods (2019,2020) so that $Time_t$ devolves to $Time_t = 1$ for t = 2020 and $Time_t = 0$ for 2019.

$$\begin{aligned} Y_{i,2020} &= \alpha + \beta D_{i,2020}^{Treat} + \gamma + u_i + \eta_{i,2020} \\ Y_{i,2019} &= \alpha + \beta D_{i,2019}^{Treat} + u_i + \eta_{i,2019} \end{aligned}$$

Recall that the fixed effect u_i is the source of the OVB.

Differencing to Eliminate Fixed Effect and OVB

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

For simplicity, consider the case of two time periods (2019,2020) so that $Time_t$ devolves to $Time_t = 1$ for t = 2020 and $Time_t = 0$ for 2019.

$$\begin{aligned} Y_{i,2020} &= \alpha + \beta D_{i,2020}^{Treat} + \gamma + u_i + \eta_{i,2020} \\ Y_{i,2019} &= \alpha + \beta D_{i,2019}^{Treat} + u_i + \eta_{i,2019} \end{aligned}$$

Recall that the fixed effect u_i is the source of the OVB. By computing the (first) difference in these periods we get

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

Differencing to Eliminate Fixed Effect and OVB

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

For simplicity, consider the case of two time periods (2019,2020) so that $Time_t$ devolves to $Time_t = 1$ for t = 2020 and $Time_t = 0$ for 2019.

$$\begin{aligned} Y_{i,2020} &= \alpha + \beta D_{i,2020}^{Treat} + \gamma + u_i + \eta_{i,2020} \\ Y_{i,2019} &= \alpha + \beta D_{i,2019}^{Treat} + u_i + \eta_{i,2019} \end{aligned}$$

Recall that the fixed effect u_i is the source of the OVB. By computing the (first) difference in these periods we get

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

Notice that first differencing removes the fixed effect u_i, and the associated OVB.

Identifying the TE

Q: Which parameter is the TE?

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

Identifying the TE

Q: Which parameter is the TE?

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

A:

$$E[\Delta Y_{it} | \Delta D^{Treat} = 1] =$$
$$E[\Delta Y_{it} | \Delta D^{Treat} = 0] =$$

Identifying the TE

Q: Which parameter is the TE?

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

A:

$$E[\Delta Y_{it} | \Delta D^{Treat} = 1] = \beta + \gamma$$
$$E[\Delta Y_{it} | \Delta D^{Treat} = 0] =$$

Identifying the TE

Q: Which parameter is the TE?

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

A:

$$E[\Delta Y_{it} | \Delta D^{Treat} = 1] = \beta + \gamma$$
$$E[\Delta Y_{it} | \Delta D^{Treat} = 0] = \gamma$$

Identifying the TE

Q: Which parameter is the TE?

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

A:

$$E[\Delta Y_{it} | \Delta D^{Treat} = 1] = \beta + \gamma$$
$$E[\Delta Y_{it} | \Delta D^{Treat} = 0] = \gamma$$

$ATE = \beta$ (i.e. the difference in the differences)

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Q: When might this be violated in our case study?

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Q: When might this be violated in our case study?
A:

- This might be violated if the units that switch treatment also experience changes in unobserved productivity shocks.

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Q: When might this be violated in our case study?
A:

- This might be violated if the units that switch treatment also experience changes in unobserved productivity shocks.
- For example, suppose some of the Switchers also had changes in their home workspace because their children switch between in-person and remote school days.

Exogeneity

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \gamma + \Delta \eta_{it}$$

We can estimate via OLS with typical assumptions holding. Most importantly,

$$Cov(\Delta \eta_{it}, \Delta D_{it}) = 0$$

Q: When might this be violated in our case study?
A:

- This might be violated if the units that switch treatment also experience changes in unobserved productivity shocks.
- For example, suppose some of the Switchers also had changes in their home workspace because their children switch between in-person and remote school days.
- This is effectively the parallel trends assumption in first difference format.

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

$$Cov(D_{it}, u_i) \neq 0$$

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

$$Cov(D_{it}, u_i) \neq 0$$

That is, the regressors can be related to the time invariant component of the errors.

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

$$Cov(D_{it}, u_i) \neq 0$$

That is, the regressors can be related to the time invariant component of the errors.
Q: What does that mean in our case?

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

$$Cov(D_{it}, u_i) \neq 0$$

That is, the regressors can be related to the time invariant component of the errors.
Q: What does that mean in our case?
A: In our case, the First Difference estimator permits the employee effect to be related to productivity, which is essential because

Fixed Effect Endogeneity

Notice that in the First Difference Estimator we can permit the regressor to be related to the errors (in a certain way)

$$Cov(D_{it}, u_i) \neq 0$$

That is, the regressors can be related to the time invariant component of the errors.
Q: What does that mean in our case?
A: In our case, the First Difference estimator permits the employee effect to be related to productivity, which is essential because that was the source of the OVB we suspected via the DAG.

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Least Squares Dummy variable (LSDV)

Rather than estimating via first differences, we can also estimate via traditional FE/Within/LSDV.

Note: FD will provide the same estimates and std.errors as these other methods when T=2, but not when T>2. FD is nice because it's conceptually clear, and has some better bias properties under certain circumstances such as dynamic panels (see Arellano and Bond)

Least Squares Dummy variable (LSDV)

Rather than estimating via first differences, we can also estimate via traditional FE/Within/LSDV.

Note: FD will provide the same estimates and std.errors as these other methods when T=2, but not when T>2. FD is nice because it's conceptually clear, and has some better bias properties under certain circumstances such as dynamic panels (see Arellano and Bond)

Consider again our model with a composite error term, which generates an OVB:

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}$$

LSDV

Q: What is the meaning of δ_k?

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \eta_{it}$$

LSDV

Q: What is the meaning of δ_k?

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \eta_{it}$$

A:

- The δ_k represent the coefficients on the dummies for the employee specific fixed effects
- i.e.

LSDV

Q: What is the meaning of δ_k?

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \eta_{it}$$

A:

- The δ_k represent the coefficients on the dummies for the employee specific fixed effects
- i.e. the part of productivity that is unique to each employee.

LSDV

Q: What is the meaning of δ_k?

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \eta_{it}$$

A:

- The δ_k represent the coefficients on the dummies for the employee specific fixed effects
- i.e. the part of productivity that is unique to each employee.

Q: Which parameter captures the treatment effect?

LSDV

Q: What is the meaning of δ_k?

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \eta_{it}$$

A:

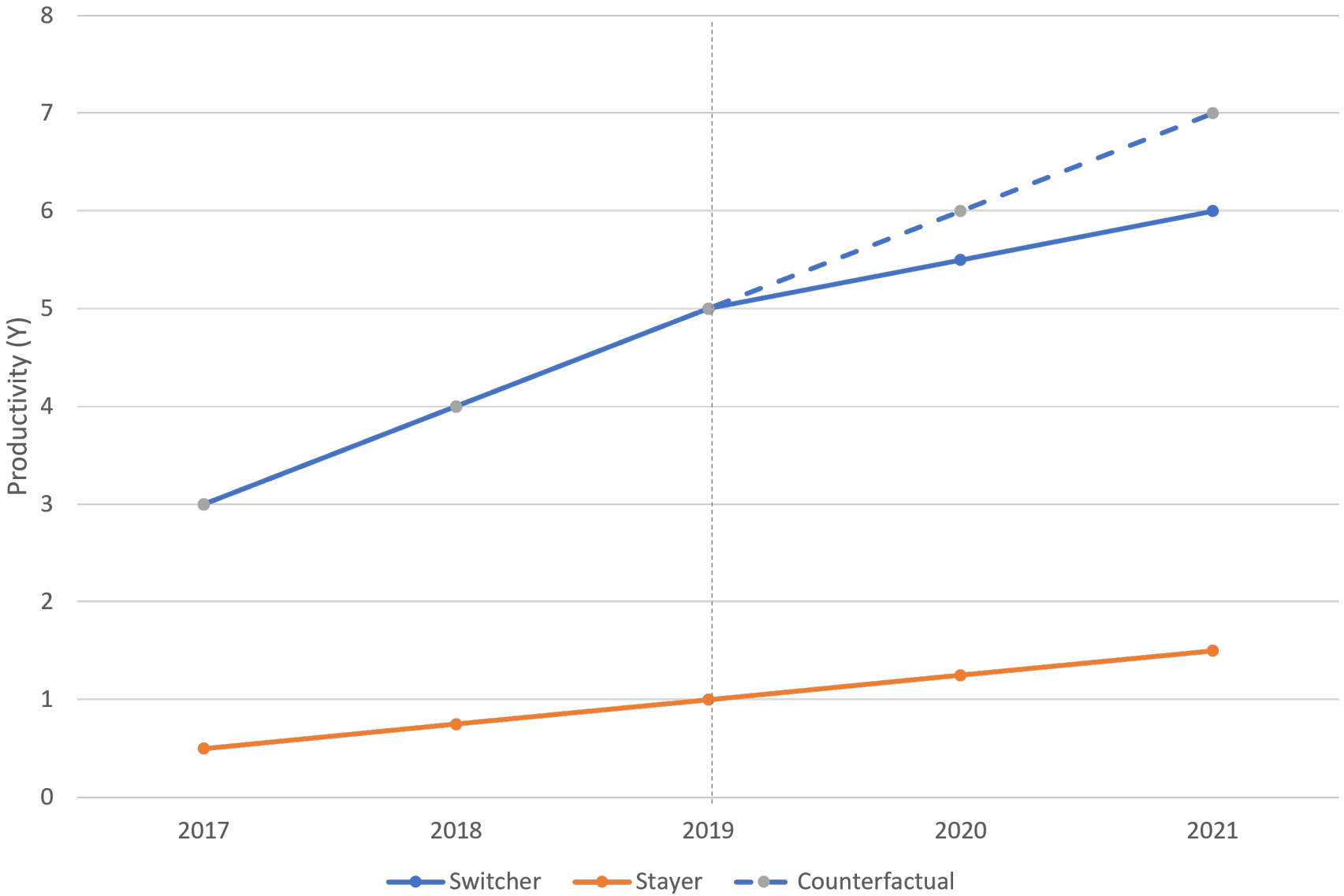
- The δ_k represent the coefficients on the dummies for the employee specific fixed effects
- i.e. the part of productivity that is unique to each employee.

Q: Which parameter captures the treatment effect?

A: β from this LSDV model would capture our treatment effect estimate.

Probing Common Trends Assumption

Visual inspection suggests trends not parallel (aka not "common")



Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Q: What does θ mean?

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Q: What does θ mean?

A:

- The θ parameters capture how employee fixed effect varies over time.

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Q: What does θ mean?

A:

- The θ parameters capture how employee fixed effect varies over time.
- We can run a joint test that all $\theta_l = 0$. Failing to reject

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Q: What does θ mean?

A:

- The θ parameters capture how employee fixed effect varies over time.
- We can run a joint test that all $\theta_l = 0$. Failing to reject supports (but doesn't prove) the parallel trends assumption

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

Q: What does θ mean?

A:

- The θ parameters capture how employee fixed effect varies over time.
- We can run a joint test that all $\theta_l = 0$. Failing to reject supports (but doesn't prove) the parallel trends assumption
- We can also compare the β in this interacted model with those from a model without interactions. If similar, then

Probing Common Trends

An advantage of the LSDV approach (as opposed to the Within transformation) is that we can easily interact the entity fixed effects with time.

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + \sum_{k=1}^{N-1} \delta_k Employee_{ki} + \sum_{l=1}^{N-1} \theta_{li} Employee_{li} \times t + \eta_{it}$$

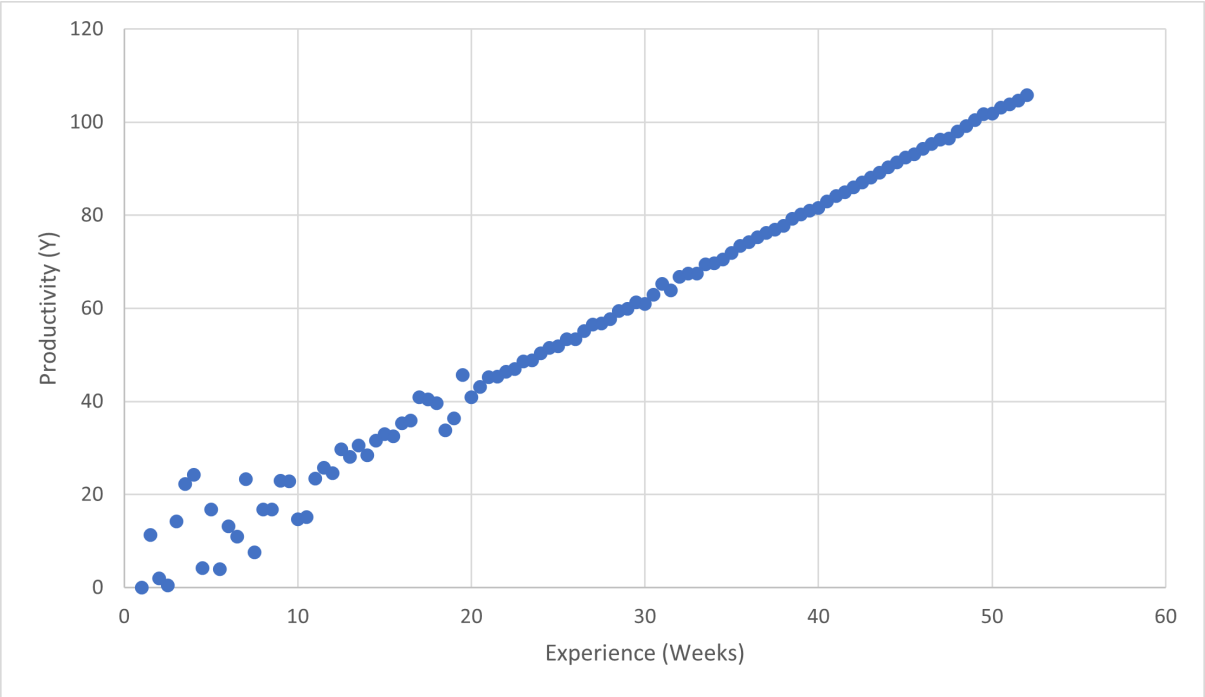
Q: What does θ mean?

A:

- The θ parameters capture how employee fixed effect varies over time.
- We can run a joint test that all $\theta_l = 0$. Failing to reject supports (but doesn't prove) the parallel trends assumption
- We can also compare the β in this interacted model with those from a model without interactions. If similar, then lends credence to the parallel trends assumption.

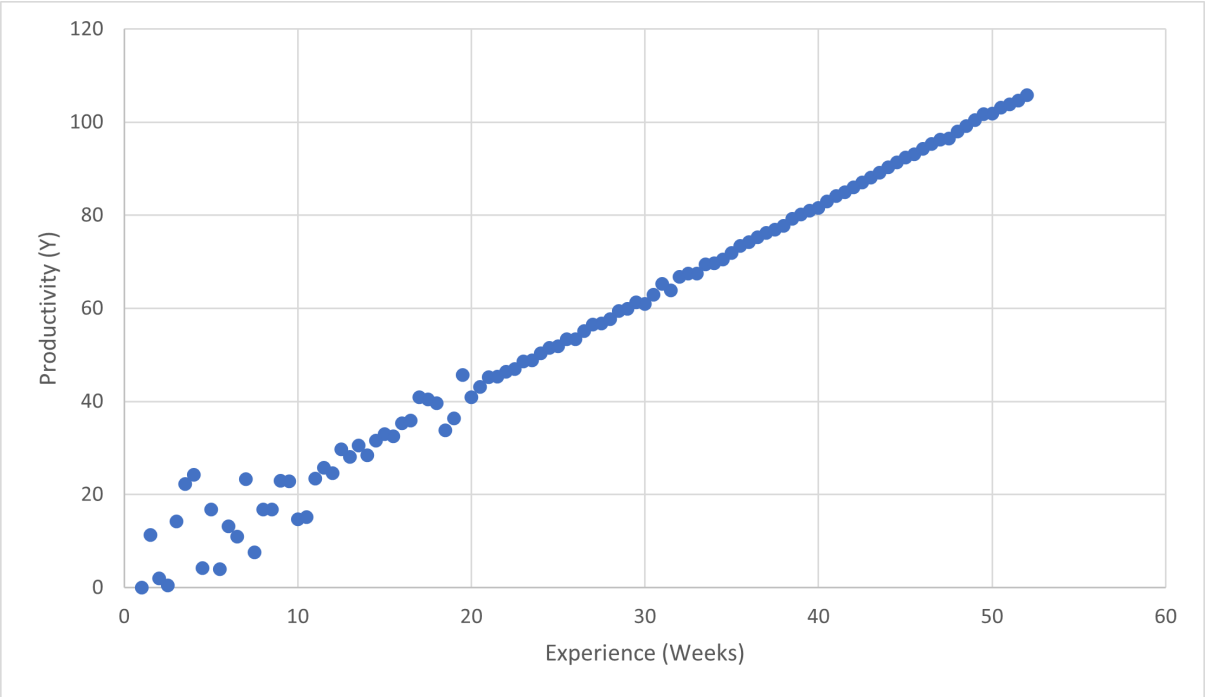
Weighting

Q: What regression "problem" is apparent here?



Weighting

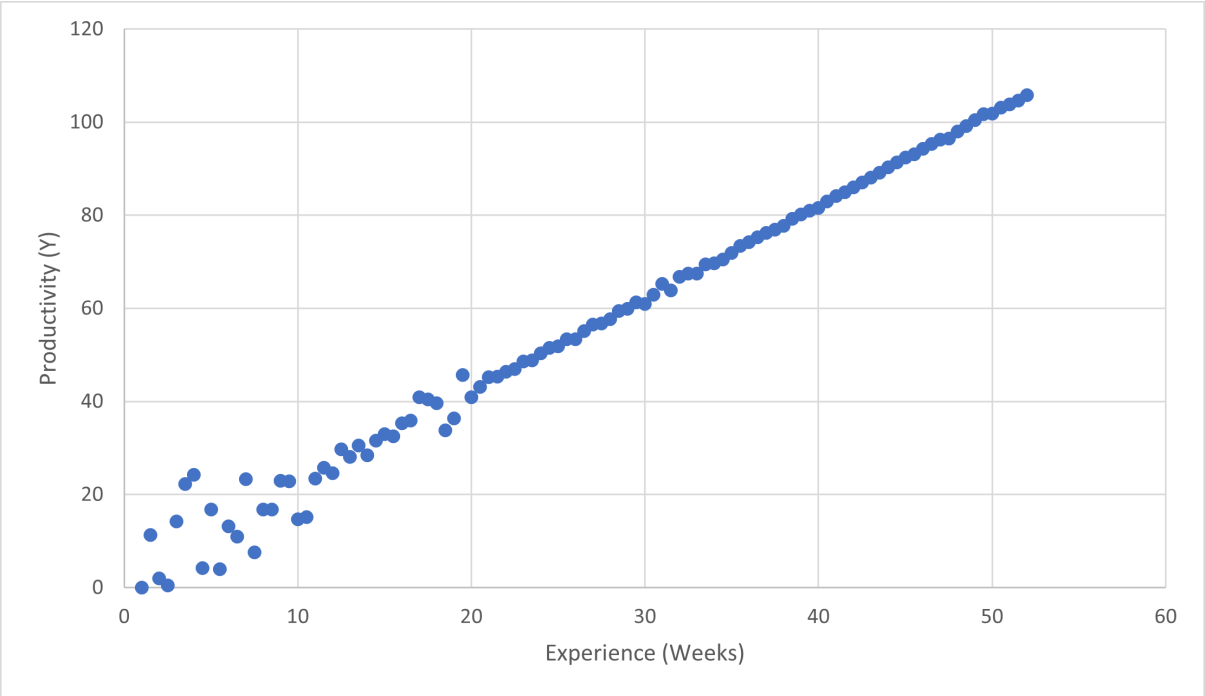
Q: What regression "problem" is apparent here?



- A: Heteroscedasticity appears present. There is more variability in productivity for inexperienced workers

Weighting

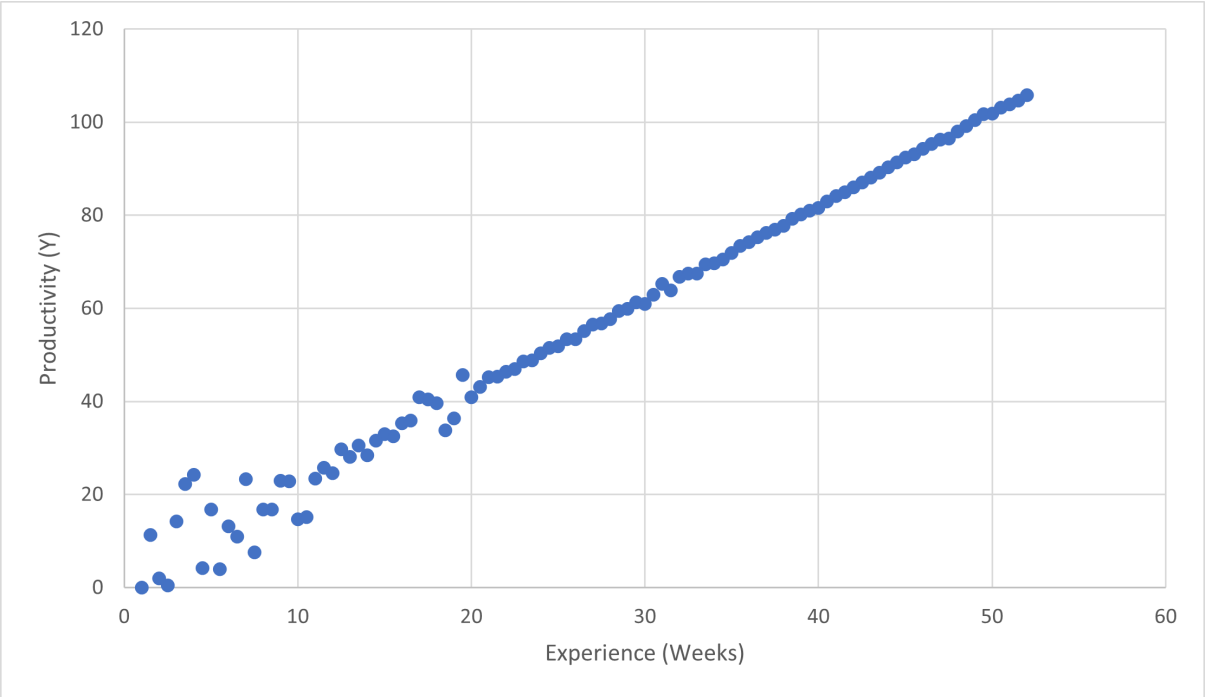
Q: What regression "problem" is apparent here?



- A: Heteroscedasticity appears present. There is more variability in productivity for inexperienced workers
- Our DiD estimates ignore this fact by giving each employee equal weight

Weighting

Q: What regression "problem" is apparent here?



- A: Heteroscedasticity appears present. There is more variability in productivity for inexperienced workers
- Our DiD estimates ignore this fact by giving each employee equal weight
- We can use Weighted Least Squares (WLS) rather than OLS to weight by experience

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Using Parallel Trends to our Advantage

Recall

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] = ATT + \Delta E[Y^{Swticher}_{Office}]$$

Using Parallel Trends to our Advantage

Recall

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] = ATT + \Delta E[Y^{Swticher}_{Office}]$$

We want to isolate ATT. So, subtract the Stayer Organic Growth from both sides

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] - \Delta E[Y^{Stayer}_{Remote}]$$
$$= ATT + \Delta E[Y^{Swticher}_{Office}] - \Delta E[Y^{Stayer}_{Remote}]$$

Using Parallel Trends to our Advantage

Recall

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] = ATT + \Delta E[Y^{Swticher}_{Office}]$$

We want to isolate ATT. So, subtract the Stayer Organic Growth from both sides

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] - \Delta E[Y^{Stayer}_{Remote}] = ATT + \Delta E[Y^{Swticher}_{Office}] - \Delta E[Y^{Stayer}_{Remote}]$$

Under Parallel Trends the organic growth is the same for Switchers and Stayers, so we can isolate the ATT

$$E[Y^{Switchers}_{Remote,2020}] - E[Y^{Switchers}_{Office,2019}] - \Delta E[Y^{Stayer}_{Remote}] = ATT$$

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing.

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta)$

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta) \uparrow \Rightarrow$

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta) \uparrow \Rightarrow t^{stat}$

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta) \uparrow \Rightarrow t^{stat} \downarrow \Rightarrow$ Type 2 Error

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta) \uparrow \Rightarrow t^{stat} \downarrow \Rightarrow$ Type 2 Error $\uparrow$ (i.e. more likely to "Accept" false null) $\Rightarrow$

Sufficient Cross Sectional Variation

Let's expand the FE model slightly in order to illustrate a shortcoming of First Differencing. Consider including a covariate X_{it} which captures employee experience (e.g. number of calls answered).

- Define $\mathbf{X} = \{X, D\}$.
- Even if $\mathbf{X}$ varies in the cross section, $\Delta\mathbf{X}$ may not (e.g. even if the experience varies greatly across teams, the change in experience may not differ much over time).
- Recall $se(\hat{\beta}) = \frac{\hat{\sigma}_e}{SST_{\Delta\mathbf{X}}(1-R^2_{\Delta\mathbf{X}})}$.
- If $SST_{\Delta\mathbf{X}} \downarrow \Rightarrow se(\beta) \uparrow \Rightarrow t^{stat} \downarrow \Rightarrow$ Type 2 Error $\uparrow$ (i.e. more likely to "Accept" false null) $\Rightarrow$ more likely to say that treatment effect (impact of remote work upon productivity) is important when it is not.

First Differences over Multiple Periods

Differencing over multiple periods is an easy extension

Consider

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}, \;\; t = 1, 2, 3$$

Given the standard assumptions (e.g. exogeneity) we can write

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \sum_{t=1}^{T-1} \gamma_t \Delta Time_{jt} + \Delta \eta_{it}, \;\; t = 2, 3$$

First Differences over Multiple Periods

Differencing over multiple periods is an easy extension

Consider

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \sum_{j=1}^{T-1} \gamma_j Time_{jt} + u_i + \eta_{it}, \;\; t = 1, 2, 3$$

Given the standard assumptions (e.g. exogeneity) we can write

$$\Delta Y_{it} = \beta \Delta D_{it}^{Treat} + \sum_{t=1}^{T-1} \gamma_t \Delta Time_{jt} + \Delta \eta_{it}, \;\; t = 2, 3$$

Notice that there is no intercept. This is usually not preferred because it complicates computation of R^2. Instead, a common, but equivalent formulation, is to include an interpret and estimate via Pooled OLS. One "problem" that might arise is serial correlation (beyond our purview).

Testing Common Trends Assumption

A more formal test of common trends is as per Pischke 2005

$$Y_{it} = \sum_{j=-m}^{q} \beta_j D_{it} \times (t = k + j) + u_i + \gamma_t + \eta_{it}$$

where k is the time at which the treatment is being switched on for employee i. We have multiple treatment effects with m leads and q lags.

Testing Common Trends Assumption

A more formal test of common trends is as per Pischke 2005

$$Y_{it} = \sum_{j=-m}^{q} \beta_j D_{it} \times (t = k + j) + u_i + \gamma_t + \eta_{it}$$

where k is the time at which the treatment is being switched on for employee i. We have multiple treatment effects with m leads and q lags.

Testing common trends amounts to testing if $\beta_j = 0 \forall j < 0$

Outline

1. Case Study - Motivation
2. Potential Outcomes
3. 2x2
   - Recast 2x2 as Regression
4. Panel
   - First Differences
   - LSDV
5. Appendix
   - Pooled Cross Section

Pooled Cross Section, not Panel

Suppose our dataset contains a sample of workers in 2019 and a sample of possibly different workers in 2020. We then have a Pooled Cross Section, not a Panel.



DiD with a Pooled Cross Section

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \gamma Time_t + u_{it} + \eta_{it}$$

Notice that the employee fixed effect u_{it} now varies with time since we have different individuals in each year.

DiD with a Pooled Cross Section

$$Y_{it} = \alpha + \beta D_{it}^{Treat} + \gamma Time_t + u_{it} + \eta_{it}$$

Notice that the employee fixed effect u_{it} now varies with time since we have different individuals in each year.
Our differencing trick won't eliminate the employee fixed effect.

Assumptions for DiD with Pooled Cross Sections

Our First Difference estimator requires

$$\bar{u}_{2020}^{Treated} - \bar{u}_{2019}^{Treated} \to 0$$
$$\bar{u}_{2020}^{Control} - \bar{u}_{2019}^{Control} \to 0$$

Assumptions for DiD with Pooled Cross Sections

Our First Difference estimator requires

$$\bar{u}_{2020}^{Treated} - \bar{u}_{2019}^{Treated} \to 0$$
$$\bar{u}_{2020}^{Control} - \bar{u}_{2019}^{Control} \to 0$$

This implies that as our sample size gets large, the average fixed effect for the treated (control) group doesn't change over time.

Assumptions for DiD with Pooled Cross Sections

Our First Difference estimator requires

$$\bar{u}_{2020}^{Treated} - \bar{u}_{2019}^{Treated} \to 0$$
$$\bar{u}_{2020}^{Control} - \bar{u}_{2019}^{Control} \to 0$$

This implies that as our sample size gets large, the average fixed effect for the treated (control) group doesn't change over time.
Q: When might this assumption be met?

Assumptions for DiD with Pooled Cross Sections

Our First Difference estimator requires

$$\bar{u}_{2020}^{Treated} - \bar{u}_{2019}^{Treated} \to 0$$
$$\bar{u}_{2020}^{Control} - \bar{u}_{2019}^{Control} \to 0$$

This implies that as our sample size gets large, the average fixed effect for the treated (control) group doesn't change over time.
Q: When might this assumption be met?

A: We might see this if

Assumptions for DiD with Pooled Cross Sections

Our First Difference estimator requires

$$\bar{u}_{2020}^{Treated} - \bar{u}_{2019}^{Treated} \to 0$$
$$\bar{u}_{2020}^{Control} - \bar{u}_{2019}^{Control} \to 0$$

This implies that as our sample size gets large, the average fixed effect for the treated (control) group doesn't change over time.
Q: When might this assumption be met?

A: We might see this if treatment assignment is random.

Difference in Differences (DiD) - Summary

- Natural Experiment

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data
- Treated observations switch from untreated to treated state around some exogenous policy event

Difference in Differences (DiD) - Summary

- Natural Experiment
- Requires panel/longitudinal data
- Treated observations switch from untreated to treated state around some exogenous policy event
- Identification rests upon a common trends assumption; i.e. if the event never happened the difference in treated and control groups would remain unchanged.