

PUBL0055: Introduction to Quantitative Methods

Lecture 7: Causality III (Observational Data)

Michal Ovádek & Indraneel Sircar

Academic year 2024-25

Outline

Design-based causal inference

Fixed effects models with panel data

Conclusion

Design-based causal inference

Design-based causal inference

- In week 2, we discussed the fundamental problem of causal inference and saw why and how randomised experiments can help us overcome the problem (for groups).
- However, researchers using observational data cannot rely on randomization to solve the confounding problem.
- Last week, we saw that if, *and only if*, we can control for ***all*** potential confounders, we can use multiple linear regression to address the omitted variable bias problem in cross-sectional data (selection on observables)
- Today, we will explore further designs that can help us make causal claims using observational data.

Motivation Part 1

How does raising the minimum wage affect employment?

Is there a *causal* effect of raising the minimum wage on employment? [Card and Krueger \(1994\)](#) explore this question by examining full-time employment in fast-food restaurants in New Jersey (NJ) before and after the minimum wage increased there in April 1992 (from \$4.25 to \$5.05). The authors also collected data on employment in fast-food restaurants in neighbouring eastern Pennsylvania (PA), which did not have an increase in the minimum wage.

- **Y (Outcome):** *fullProp*
 - Proportion of full time employees in a fast-food restaurant
- **X (Treatment):** *increased minimum wage*
 - Increase in minimum wage in NJ fast-food restaurants in April 1992

Causal inference with observational data

Researchers using observational data cannot rely on randomization to solve the selection bias problem so they must employ alternative designs.

A design helps us to estimate *causal* effects, ***if*** it allows us to satisfy the key assumption of “unconfoundedness”:

That treatment and control groups are comparable with respect to everything other than the treatment.

Causal inference with observational data

Researchers using observational data cannot rely on randomization to solve the selection bias problem so they must employ alternative designs.

A design helps us to estimate *causal* effects, **if** it allows us to satisfy the key assumption of “unconfoundedness”:

That treatment and control groups are comparable with respect to everything other than the treatment.

In the minimum wage example:

Question: What is the treatment group?

Causal inference with observational data

Researchers using observational data cannot rely on randomization to solve the selection bias problem so they must employ alternative designs.

A design helps us to estimate *causal* effects, **if** it allows us to satisfy the key assumption of “unconfoundedness”:

That treatment and control groups are comparable with respect to everything other than the treatment.

In the minimum wage example:

Question: What is the treatment group?

Answer: NJ fast-food restaurants after April 1992.

Causal inference with observational data

Researchers using observational data cannot rely on randomization to solve the selection bias problem so they must employ alternative designs.

A design helps us to estimate *causal* effects, **if** it allows us to satisfy the key assumption of “unconfoundedness”:

That treatment and control groups are comparable with respect to everything other than the treatment.

In the minimum wage example:

Question: What is the treatment group?

Answer: NJ fast-food restaurants after April 1992.

Question: How can we find a good comparison group?

Which observations should we compare?

1. **Cross-sectional approach:** We could look at a cross-section of fast-food restaurants in the period after the minimum-wage increase in New Jersey and Pennsylvania.

Which observations should we compare?

1. **Cross-sectional approach:** We could look at a cross-section of fast-food restaurants in the period after the minimum-wage increase in New Jersey and Pennsylvania.

If we observe multiple units, such as restaurants, countries or individuals, at one point in time, we call this a cross-section.

chain	location	fullProp
wendys	PA	0.000
burgerking	PA	0.455
roys	NJ	0.714
burgerking	NJ	0.333

Which observations should we compare?

1. **Cross-sectional approach:** We could look at a cross-section of fast-food restaurants in the period after the minimum-wage increase in New Jersey and Pennsylvania.
 - **Y (Outcome):** *fullProp*
 - Proportion of full time employees in a fast-food restaurant
 - **X (Treatment):** *location*
 - Whether or not the fast-food restaurants was located in NJ after April 1992

Which observations should we compare?

1. **Cross-sectional approach:** We could look at a cross-section of fast-food restaurants in the period after the minimum-wage increase in New Jersey and Pennsylvania.

```
# we could take the difference in the mean proportion of  
# full time employment after the increase  
mean(minwage_after$fullProp[minwage_after$location=="NJ"]) -  
  mean(minwage_after$fullProp[minwage_after$location=="PA"])  
  
## [1] 0.04811886
```

Which observations should we compare?

1. **Cross-sectional approach:** We could look at a cross-section of fast-food restaurants in the period after the minimum-wage increase in New Jersey and Pennsylvania.

```
# we could regress outcome on whether  
# or not fast-food restaurant is in NJ?
```

```
m1 <- lm(fullProp ~ location,  
         data = minwage_after)
```

```
# control for chain
```

```
m2 <- lm(fullProp ~ location + chain,  
         data = minwage_after)
```

```
##  
## =====  
##                Model 1      Model 2  
## -----  
## (Intercept)    0.27 ***    0.31 ***  
##              (0.03)      (0.03)  
## locationNJ     0.05        0.05  
##              (0.03)      (0.03)  
## chainkfc       -0.04  
##              (0.04)  
## chainroys      -0.08 *  
##              (0.03)  
## chainwendys    -0.09 *  
##              (0.04)  
## -----  
## R^2            0.01        0.03  
## Adj. R^2       0.00        0.02  
## Num. obs.     358         358  
## =====  
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of all confounders?

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of all confounders?

Answer: Probably not! In addition to the fast-food chain, there might be state-specific confounders, i.e., factors that affect both the proportion of full time employees and a restaurant's location.

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of all confounders?

Answer: Probably not! In addition to the fast-food chain, there might be state-specific confounders, i.e., factors that affect both the proportion of full time employees and a restaurant's location.

- Or it could be that due to minimum wage law, some NJ fast-food restaurants shut down and moved to PA.)

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of all confounders?

Answer: Probably not! In addition to the fast-food chain, there might be state-specific confounders, i.e., factors that affect both the proportion of full time employees and a restaurant's location.

- Or it could be that due to minimum wage law, some NJ fast-food restaurants shut down and moved to PA.)

→ We need another strategy!

→ We might want to investigate whether NJ fast-food restaurants decreased employment after April 1992 compared to before.

Which observations should we compare?

2. **Before-after approach:** We could compare the same fast-food restaurants in New Jersey before and after the minimum-wage increase.

Which observations should we compare?

- 2. Before-after approach:** We could compare the same fast-food restaurants in New Jersey before and after the minimum-wage increase.
 - **Y (Outcome):** *fullProp*
 - Proportion of full time employees in a fast-food restaurant
 - **X (Treatment):** *time period*
 - Whether or not the proportion of full time employees is measured before or after April 1992

Which observations should we compare?

2. **Before-after approach:** We could compare the same fast-food restaurants in New Jersey before and after the minimum-wage increase.

If we observe data from the same unit at multiple points in time, we call this a time-series.

id	chain	location	time	fullProp
68	roys	NJ	Before	0.500
68	roys	NJ	After	0.714

Which observations should we compare?

2. **Before-after approach:** We could compare the same fast-food restaurants in New Jersey before and after the minimum-wage increase.

If we observe multiple units each at multiple points in in time, we call this **panel data**.

id	chain	location	time	fullProp
68	roys	NJ	Before	0.500
68	roys	NJ	After	0.714
69	kfc	NJ	Before	0.556
69	kfc	NJ	After	0.190

Recap Cross-Sectional Data Notation

- In cross-sectional data, we denote each unit/observation with the subscript i .
- For example, in a bivariate linear regression:¹

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where

- X_i refers to the observation of variable X for unit i
- Y_i refers to the observation of variable Y for unit i
- ϵ_i refers to the error term for unit i

¹Note that β_0 is another (common) way to denote the intercept other than α .

Panel Data Notation

- In panel data, we have *multiple* observations for each unit i .

Panel Data Notation

- In panel data, we have *multiple* observations for each unit i .
- We need additional notation to refer to the time period of each observation, so we use two subscripts instead of one.
 - We still use the subscript i to denote observations for each unit
 - We add subscript t to refer to observations at time period t

Panel Data Notation

- In panel data, we have *multiple* observations for each unit i .
- We need additional notation to refer to the time period of each observation, so we use two subscripts instead of one.
 - We still use the subscript i to denote observations for each unit
 - We add subscript t to refer to observations at time period t

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \epsilon_{i,t}$$

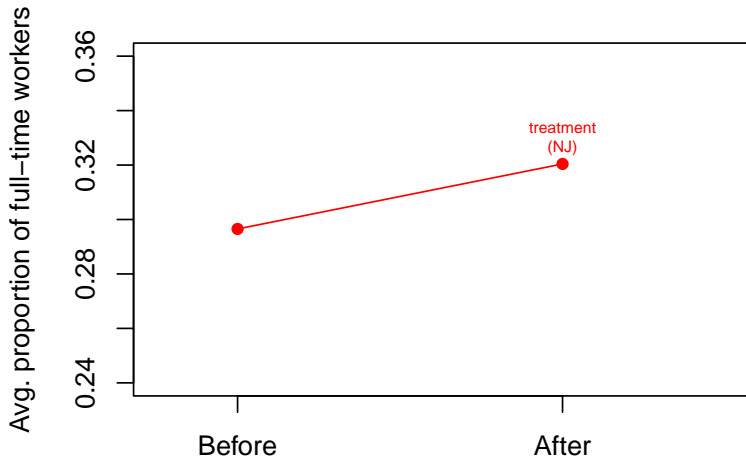
where

- $X_{i,t}$ refers to the observation of variable X for unit i at time period t
- $Y_{i,t}$ refers to the observation of variable Y for unit i at time period t
- $\epsilon_{i,t}$ refers to the error term for unit i at time period t

Before and after designs

- **Data:** Many observations of treated units over time
- **Approach:** Difference in average outcomes before and after treatment
- **Advantage:** Unit-specific confounding factors are held constant

Before and after designs



Before and after designs

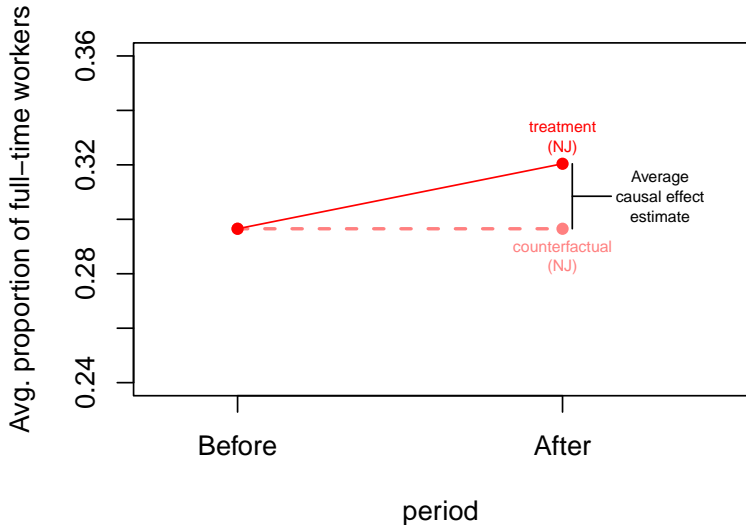
Question: What are the assumptions required to identify the causal effect of the treatment?

Before and after designs

Question: What are the assumptions required to identify the causal effect of the treatment?

Assumption: No time varying confounding, e.g. no time trends

Before and after designs



Before and after designs

Once again we can use the difference in means!

```
# Avg. proportion of full-time employees in NJ before the change  
nj.before <- mean(minwage$fullProp[minwage$time=="Before" &  
                  minwage$location=="NJ"])
```

```
# Avg. proportion of full-time employees in NJ after the change  
nj.after <- mean(minwage$fullProp[minwage$time=="After" &  
                  minwage$location=="NJ"])
```

```
# Estimate using before and after design  
nj.change <- nj.after-nj.before  
nj.change
```

```
## [1] 0.02387474
```


Before and after designs

... or regression. time indicates when employment was measured.

```
before.after.reg <- lm(fullProp~time,  
                        data = minwage[minwage$location=="NJ",])
```

```
before.after.reg
```

```
##
```

```
## Call:
```

```
## lm(formula = fullProp ~ time, data = minwage[minwage$location ==  
##      "NJ", ])
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)      timeAfter
```

```
##      0.29653      0.02387
```

Before and after designs

... or regression. time indicates when employment was measured.

```
before.after.reg <- lm(fullProp~time,
                       data = minwage[minwage$location=="NJ",])
before.after.reg

##
## Call:
## lm(formula = fullProp ~ time, data = minwage[minwage$location ==
##      "NJ", ])
##
## Coefficients:
## (Intercept)      timeAfter
##      0.29653      0.02387
```

We find that the average proportion of full-time fast-food employees in New Jersey increased by 0.02 (2 pp) after the increase in minimum wage.

Before and after design summary

Advantages:

- Accounts for *time invariant* confounders; i.e., holds unit-specific features “constant” by comparing outcomes within units over time
- Only requires data on treated units

Disadvantages:

- Requires assuming that *nothing else* other than the treatment changed for treated units
- Requires data from multiple time points

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of everything except the treatment?

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of everything except the treatment?

Answer: Probably not! There might be time variant confounding.

- For example, maybe employment was getting better over time in New Jersey anyhow?

What about unconfoundedness assumption?

Question: Are the treatment and control groups identical on average in terms of everything except the treatment?

Answer: Probably not! There might be time variant confounding.

- For example, maybe employment was getting better over time in New Jersey anyhow?

→ We need another strategy!

→ We might want to investigate whether fast food restaurants in other states had similar trends in employment.

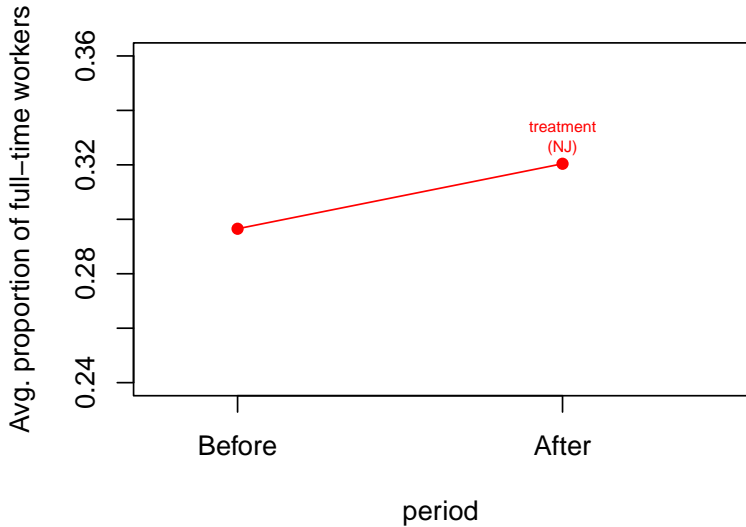
Which observations should we compare?

3. **Difference-in-differences approach:** We could compare the before-after employment difference in NJ fast food restaurants with the before-after employment trend in PA fast food restaurants.

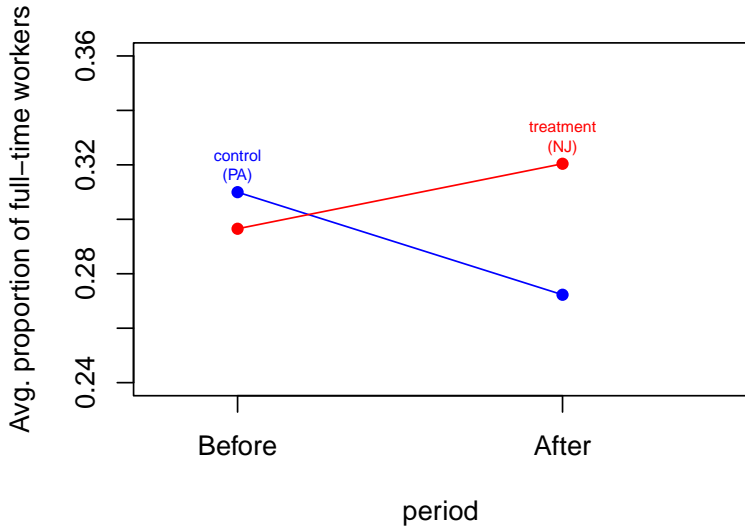
Which observations should we compare?

- 3. Difference-in-differences approach:** We could compare the before-after employment difference in NJ fast food restaurants with the before-after employment trend in PA fast food restaurants.
 - **Data:** Many observations of many units, some of which are treated at some point in time
 - **Approach:** Difference in average outcomes before and after treatment for treated group, compared to same difference for control group
 - **Advantage:** Unit-specific confounders **and** (some) time-variant confounders can be accounted for

Difference-in-differences designs



Difference-in-differences designs



Difference-in-differences designs

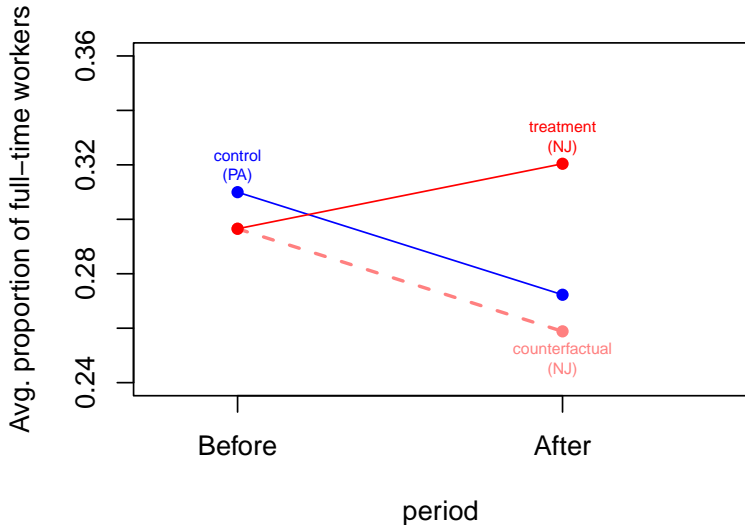
Question: What are the assumptions required to identify the effect of the treatment?

Difference-in-differences designs

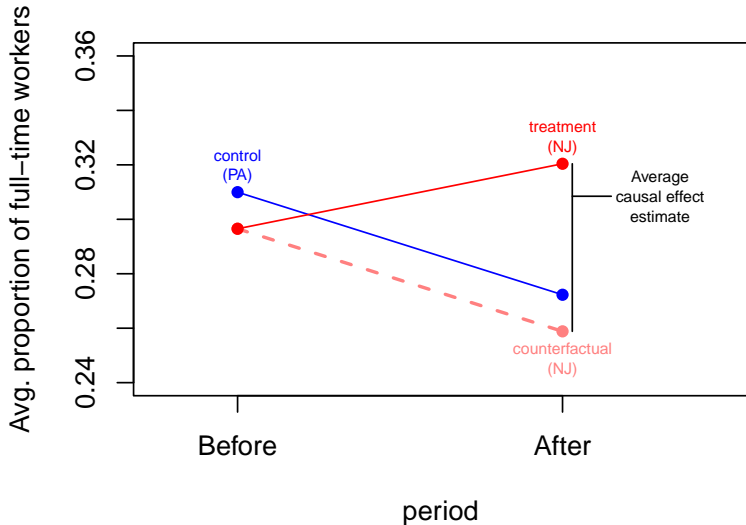
Question: What are the assumptions required to identify the effect of the treatment?

Assumption: No time *and* unit varying confounding/selection bias, i.e. without treatment, treatment group outcome *changes* would have mirrored those in control group (“parallel trends”)

Difference-in-differences designs



Difference-in-differences designs



Difference in differences designs

We can take the difference between two difference in means.

```
# Avg. proportion of full-time employees in PA before the change  
pa.before <- mean(minwage$fullProp[minwage$time=="Before" &  
                    minwage$location=="PA"])
```

```
# Avg. proportion of full-time employees in PA after the change  
pa.after <- mean(minwage$fullProp[minwage$time=="After" &  
                    minwage$location=="PA"])
```

```
# The change in PA (We already calculated the change in NJ)  
pa.change <- pa.after - pa.before
```

```
# Difference-in-differences estimate  
diff.in.diff <- nj.change - pa.change  
diff.in.diff
```

```
## [1] 0.06155831
```

Difference in differences via regression

There are several ways to calculate the difference-in-differences estimate using a regression model.

Difference in differences via regression

There are several ways to calculate the difference-in-differences estimate using a regression model.

We will do it the following way:

- Run a regression for $Y_{i,t}$ where we include observations for each unit in both periods in the data set, and include
 1. a dummy for time period;
 2. a dummy for whether or not the unit receives the treatment at $t1$;
 3. and their interaction.

Difference in differences via regression

There are several ways to calculate the difference-in-differences estimate using a regression model.

We will do it the following way:

- Run a regression for $Y_{i,t}$ where we include observations for each unit in both periods in the data set, and include
 1. a dummy for time period;
 2. a dummy for whether or not the unit receives the treatment at $t1$;
 3. and their interaction.
- In this model, the coefficient on the interaction term is the difference-in-differences estimate.

Difference in differences via regression

```
# Fit a regression with time period, state, and their interaction
```

```
diff.in.diff.reg <- lm(fullProp~location*time, data = minwage)
```

```
diff.in.diff.reg
```

```
##
```

```
## Call:
```

```
## lm(formula = fullProp ~ location * time, data = minwage)
```

```
##
```

```
## Coefficients:
```

```
##          (Intercept)          locationNJ          timeAfter
```

```
##          0.30997          -0.01344          -0.03768
```

```
## locationNJ:timeAfter
```

```
##          0.06156
```

Note that the interaction term estimate of 0.06 is the same as above.

Difference in differences via regression

This gives the following regression equation:

$$\widehat{fullProp}_{i,t} = 0.31 - 0.01 \times location_i - 0.04 \times time_t + 0.06 \times location_i \times time_t$$

Difference in differences via regression

This gives the following regression equation:

$$\widehat{fullProp}_{i,t} = 0.31 - 0.01 \times location_i - 0.04 \times time_t + 0.06 \times location_i \times time_t$$

1. The intercept (0.31) is the average value of fullProp when $location_{NJ} = 0$ and $time_{After} = 0$, i.e. in *Pennsylvania*, *before* treatment occurs.
2. The first coefficient (-0.01) is the average difference in fullProp between fast food restaurants in NJ and PA when $time_{After} = 0$, i.e. *before* treatment occurs.
3. The second coefficient (-0.04) is the average difference in fullProp between fast food restaurants before and after time of treatment when $location_{NJ} = 0$, i.e. in *Pennsylvania*.
4. The third coefficient (0.06) is the average difference in the *before-after treatment change* of fullProp between fast food restaurants in NJ and PA.

Difference in differences via regression

This gives the following regression equation:

$$\widehat{fullProp}_{i,t} = 0.31 - 0.01 \times location_i - 0.04 \times time_t + 0.06 \times location_i \times time_t$$

1. What is the average value of fullProp in *Pennsylvania*, before treatment occurs, i.e. when locationNJ = 0 and timeAfter = 0?
 - $0.31 - 0.01 \times 0 - 0.04 \times 0 + 0.06 \times 0 \times 0 = 0.31$
2. What is the average value of fullProp in *New Jersey*, before treatment occurs, i.e. when locationNJ = 1 and timeAfter = 0?
 - $0.31 - 0.01 \times 1 - 0.04 \times 0 + 0.06 \times 1 \times 0 = 0.3$

Difference in differences via regression

This gives the following regression equation:

$$\widehat{fullProp}_{i,t} = 0.31 - 0.01 \times location_i - 0.04 \times time_t + 0.06 \times location_i \times time_t$$

3. What is the average value of fullProp in *New Jersey, before* treatment occurs, i.e. when locationNJ = 0 and timeAfter = 1?
 - $0.31 - 0.01 \times 0 - 0.04 \times 1 + 0.06 \times 0 \times 1 = 0.27$
4. What is the average value of fullProp in *New Jersey, after* treatment occurs, i.e. when locationNJ = 1 and timeAfter = 1?
 - $0.31 - 0.01 \times 1 - 0.04 \times 1 + 0.06 \times 1 \times 1 = 0.32$

Difference in differences designs

Advantages:

- Holds unit-specific features “constant” by comparing outcomes within units over time
- Removes confounding by shocks that are common to all units (treatment and control)
- Does not assume that outcomes remain constant over time in the absence of treatment

Disadvantages:

- Requires assuming that nothing else changed *only* for treated units
 - E.g. Did other policies change affecting treated units at the same time?
- Requires data from multiple time points and multiple types of units

Summarising our findings

- Using a before and after design, we estimate that the minimum wage bump increased the proportion of full-time employment in fast-food restaurants by 0.02, on average.
- Using a difference-in-differences design, we estimate that the minimum wage bump increased the proportion of full-time employment in fast-food restaurants by 0.06, on average.
- Our estimates suggest that raising the minimum wage in New Jersey tended to *increase* full-time employment in fast-food restaurants.

Fixed effects models with panel data

Which observations should we compare?

We might also be interested in the causal effect of treatments beyond interventions like the minimum wage law that has a clear “before” and “after” the intervention.

Which observations should we compare?

We might also be interested in the causal effect of treatments beyond interventions like the minimum wage law that has a clear “before” and “after” the intervention.

4. **Panel data approach:** We could use fixed effects models to estimate causal effects using panel data.

Motivation Part 2

Is there a “Resource Curse”?

Does the presence of natural resources affect the quality of political institutions? Rentier capitalism on a systemic level may be a curse: if states can extract rents from easily lootable resources instead of taxing their people, they have no incentive to create high quality political institutions – “no representation without taxation”

- **Y (Dependent variable):** Institutional quality index
- **X (Independent variable):** Oil revenue (as a percentage of GDP)

Data source

Observational data

We will study the resource curse with data from 58 countries from 1996 to 2010. Dataset:

- **Y (Outcome):** institutions
 - Institutional Quality index
- **X (Treatment):** oil
 - Oil rents (% of GDP)
- country
- year
- aid - Foreign aid (% of GDP)
- gdp.capita - GDP per capita in US dollars
- polity2 - Polity IV index

“Pooled” data regression

How could we analyse such data using regression?

“Pooled” data regression

How could we analyse such data using regression?

- If we fit a simple or multiple regression model to panel data, *without taking into account the time dimension*, this is called a “pooled model”
 - We are “pooling” all the country-year observations together

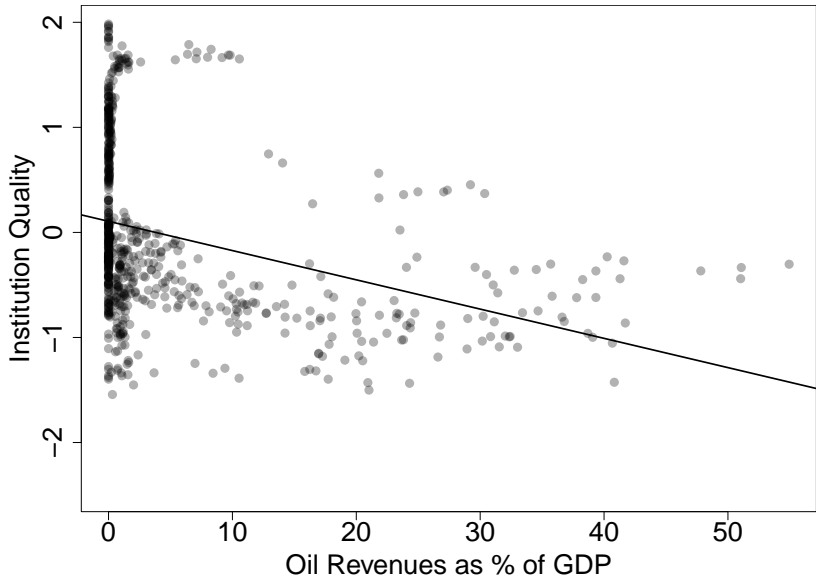
$$\text{institutions}_{i,t} = \beta_0 + \beta_1 \text{oil}_{i,t} + \epsilon_{i,t}$$

“Pooled” data regression

```
pooled_model <- lm(institutions ~ oil,  
                   data = resource_curse)
```

```
##  
## =====  
##                Model 1  
## -----  
## (Intercept)    0.106 **  
##                (0.034)  
## oil            -0.028 ***  
##                (0.003)  
## -----  
## R^2             0.114  
## Adj. R^2       0.113  
## Num. obs.     682  
## =====  
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

“Pooled” data plot



What about unconfoundedness assumption?

Lots of things might “confound” the relationship between oil and institutional quality.

- That is, the observations with high oil revenues might be from country-years that are different (in ways which affect institution quality) from those with low oil revenues.
- We can add control variables to account for all potential confounders, just as before: e.g., per capita income, whether a country is a democracy, foreign aid, etc

“Pooled” data regression with control variables

```
pooled_model_2 <- lm(institutions ~ oil + log(gdp.capita)
                     + polity2 + aid,
                     data = resource_curse)
```

```
##
## =====
##                Model 1      Model 2
## -----
## (Intercept)      0.106 **   -3.356 ***
##                (0.034)     (0.099)
## oil              -0.028 ***  -0.015 ***
##                (0.003)     (0.002)
## log(gdp.capita)                0.403 ***
##                (0.013)
## polity2                0.031 ***
##                (0.003)
## aid                0.012 ***
##                (0.003)
## -----
## R^2                0.114      0.744
## Adj. R^2           0.113      0.742
## Num. obs.          682        672
## =====
```

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil

- Differences in past institutional history
- Economic differences
- Cultural differences
- Etc

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil

- Differences in past institutional history
- Economic differences
- Cultural differences
- Etc

→ We already know how to control for confounding variables we can measure...

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil

- Differences in past institutional history
- Economic differences
- Cultural differences
- Etc

→ We already know how to control for confounding variables we can measure...But what about those we cannot (easily) measure?

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil

- Differences in past institutional history
- Economic differences
- Cultural differences
- Etc

→ We already know how to control for confounding variables we can measure...But what about those we cannot (easily) measure?

→ With panel data we can reduce problems of omitted variable bias in observational data without observing all of the confounders – we can just control for the unit!

Unit fixed effects models

A model with “unit fixed effects” is a model to which we have added dummy variables for every unit (country), except one.

By adding these to our model, we control for all differences between countries that do not change over time.

Unit fixed effects models

A model with “unit fixed effects” is a model to which we have added dummy variables for every unit (country), except one.

By adding these to our model, we control for all differences between countries that do not change over time.

- Recall that a coefficient on a dummy variable shifts the intercept but does not affect the slope
- The intercept shift controls for all differences across countries at one point in time
- It does not, however, control for any *changes* over time
- Unit fixed effects are sometimes also referred to as “individual” fixed effects depending on the context
- Since our cross-sectional units are countries in the resource curse dataset, we will refer to them as “country” fixed effects.

The unit fixed effect model

We could write this up in this way:

$$Y_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \gamma_2 C_{2i} + \gamma_3 C_{3i} + \cdots + \gamma_n C_{ni} + \epsilon_{i,t}$$

where

- $Y_{i,t}$ is institutional quality for country i in year t
- β_0 is the intercept
- $X_{1i,t}$ is oil revenue for country i in year t
- β_1 is the coefficient estimate for oil revenue
- γ_n is the coefficient estimate for the binary variable that equals 1 when country $i = n$.
- C_n is the binary variable that equals 1 when country $i = n$.
- $\epsilon_{i,t}$ is the error term or residual for country i in year t

The unit fixed effect model

However, rather than write out all the dummy variables, we define the unit fixed effects model as follows:

$$Y_{i,t} = \alpha_i + \beta_1 X_{1i,t} + \epsilon_{i,t}$$

where

- $Y_{i,t}$ is institutional quality for country i in year t
- $\alpha_i = \beta_0 + \gamma_i$ **is the intercept for country i**
 - for country 1, the regression equation is $\beta_0 + \beta_1 X_{1i,t}$
 - for country i , it is $\beta_0 + \gamma_i + \beta_1 X_{1i,t}$ – so $\alpha_i = \beta_0 + \gamma_i$
- $X_{1i,t}$ is oil revenue for country i in year t
- β_1 is the coefficient estimate for oil revenue
- $\epsilon_{i,t}$ is the error term or residual for country i in year t

Unit fixed effects = Unit-specific dummies

We have already seen that we can incorporate information about differences between (many) groups of observations by using categorical variables

- In Week 5, the `region` variable was a categorical variable with the German regions
- Instead of 16 regions, we have 58 countries in the resource curse example

Unit fixed effects = Unit-specific dummies

We have already seen that we can incorporate information about differences between (many) groups of observations by using categorical variables

- In Week 5, the `region` variable was a categorical variable with the German regions
- Instead of 16 regions, we have 58 countries in the resource curse example
- We incorporate information about them (unit fixed effects) by including a set of dummy variables for all countries ($n-1$ dummies)

Unit-specific dummies

Example set of dummy variables for a categorical variable with levels of Brazil, Argentina, and Colombia

<i>Country</i>	<i>Year</i>	X_B	X_A	X_C
Brazil	2006	1	0	0
Brazil	2007	1	0	0
Brazil	2008	1	0	0
Argentina	2006	0	1	0
Argentina	2007	0	1	0
Argentina	2008	0	1	0
Colombia	2006	0	0	1
Colombia	2007	0	0	1
Colombia	2008	0	0	1
\vdots	\vdots	\vdots	\vdots	\vdots

Unit fixed effects models in R

It is possible to estimate a fixed effects model using `lm()`

- Simply add the categorical variable to the regression model (as a factor variable):

```
country_fe_model <- lm(institutions ~ oil +  
                        as.factor(country), # the unit FE  
                        data = resource_curse)
```

Unit fixed effects models in R

```
summary(country_fe_model)
```

```
##
```

```
## Call:
```

```
## lm(formula = institutions ~ oil + as.factor(country), data = resource_curse)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.38818 -0.06665 -0.00369  0.06641  0.48773
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value
## (Intercept)  -0.443459   0.034922 -12.698
## oil          -0.002672   0.001487  -1.797
## as.factor(country)Argentina    0.247010   0.049443   4.996
## as.factor(country)Armenia     0.102673   0.049372   2.080
## as.factor(country)Australia   2.033781   0.049356  41.206
## as.factor(country)Azerbaijan  -0.365574   0.063219  -5.783
## as.factor(country)Belarus     -0.458276   0.049357  -9.285
## as.factor(country)Belgium     1.775695   0.049372  35.966
## as.factor(country)Benin       0.242399   0.049370   4.910
```

```
## as.factor(country)Bolivia     0.222254   0.049345   4.505
```

Unit fixed effects models in R

```
# If you don't want to see the country fixed effects  
# in your output  
screenreg(country_fe_model,  
           custom.coef.map = list("(Intercept)"="Intercept",  
                                  "oil"="Oil"))
```

```
##  
## =====  
##              Model 1  
## -----  
## Intercept    -0.44 ***  
##              (0.03)  
## Oil          -0.00  
##              (0.00)  
## -----  
## R2           0.98  
## Adj. R2      0.98  
## Num. obs.    682  
## =====
```

Unit fixed effects models in R

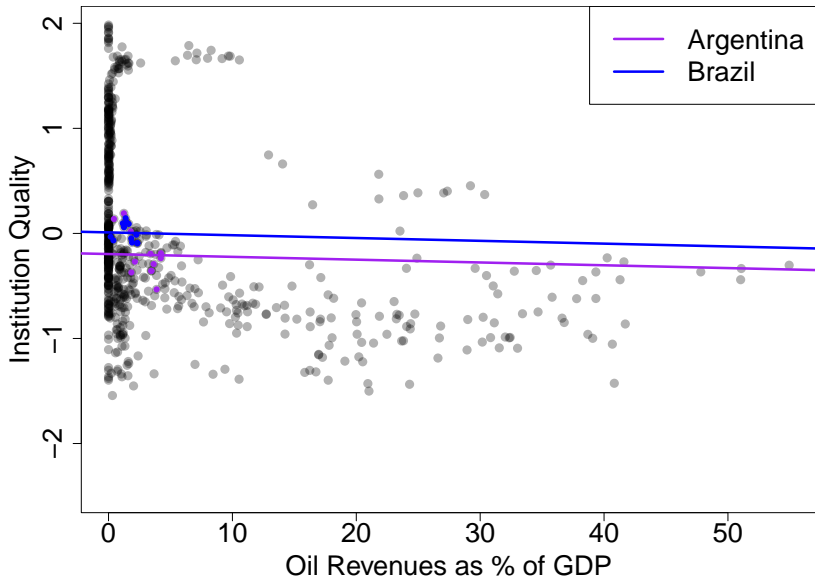
- We can interpret the coefficient estimates using the same logic as before, remembering that we are now also *holding country constant*.
- Note that the intercept in the Country FE model is the expected value of Y for the baseline country with oil rent being 0% of GDP.

```
##
## =====
##              Pooled      Pooled      Country FE
## -----
## Intercept      0.106 **   -3.356 ***  -0.443 ***
##                (0.034)   (0.099)    (0.035)
## Oil            -0.028 ***  -0.015 ***  -0.003
##                (0.003)   (0.002)    (0.001)
## Log(GDP pc)                0.403 ***
##                          (0.013)
## Polity IV                0.031 ***
##                          (0.003)
## Aid                0.012 ***
##                          (0.003)
## -----
## R^2              0.114      0.744      0.980
## Adj. R^2         0.113      0.742      0.979
## Num. obs.        682        672        682
## =====
```

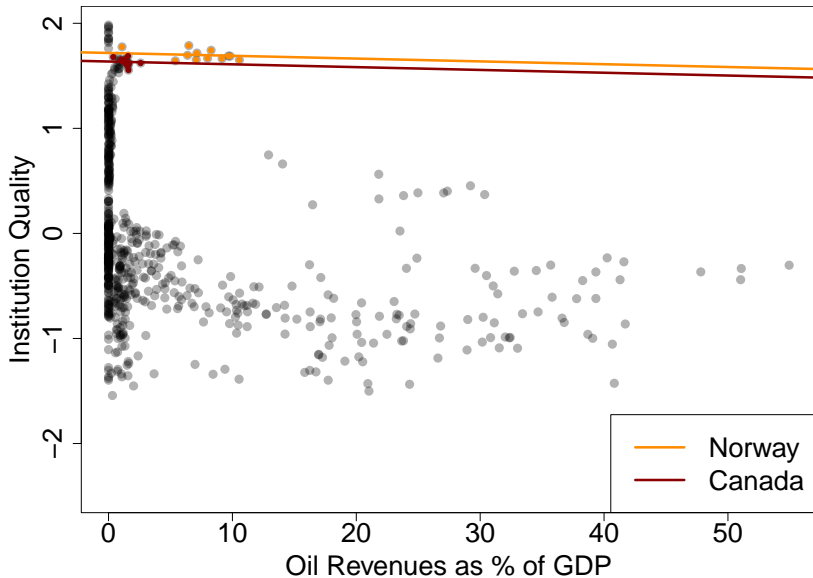
What does it mean to hold country constant?

- The country fixed effects represent the constant α_i for each country. This means that each country's regression line has the same slope and is parallel to all others.
- For Argentina, the constant α_{ARG} of -0.2 is the level of institutional quality when all X variables are 0, compared to an α_{BRA} of 0.01 for Brazil.
- In other words, we are using variation in oil and institutions over time *within* countries to estimate the effect of oil revenue (as % of GDP) on institutional quality.
- See [this GIF](#) for a visualisation of what unit fixed effects do
- The absolute level of the fixed effects becomes difficult and less meaningful as we add more variables to the model (analogous to the intercept), but the differences across units can provide useful information.

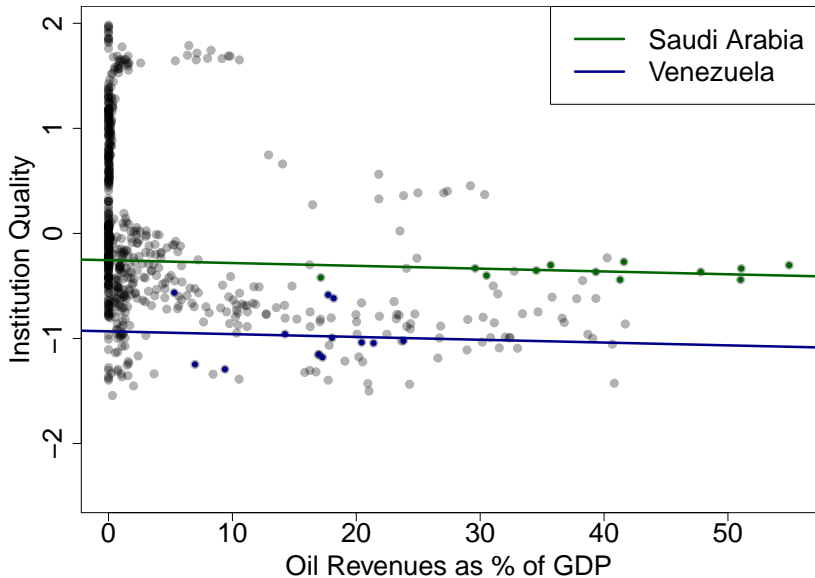
Fitted values from a fixed effect model



Fitted values from a fixed effect model



Fitted values from a fixed effect model



What do country fixed effects control for?

- Fixed effects control for any *constant* features of countries.
 - Colonial heritage (mostly)
 - Weather / climate (mostly)
 - Culture (mostly)
 - Wealth (not exactly)
- Over longer time periods, it is less likely that potential confounders remain truly constant.
 - If an unobserved confounder only remains *mostly* constant, fixed effects control for much of its potential for confounding but not all

Unit fixed effects model

Advantages

- Allows us to control for all “time invariant” (ie constant over time) confounders
 - This is particularly important in the case of variables which are difficult or impossible to observe.

Disadvantages

- In exchange for this, we have changed what it is that we are estimating
 - No longer comparing different countries.
 - We are looking for “within country” relationships between X and Y over time.

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of times when oil rents are higher/lower.

- Global oil price shocks
- Conflict
- Etc

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of times when oil rents are higher/lower.

- Global oil price shocks
- Conflict
- Etc

→ If only there was a way to control for *time-variant* omitted variables in observational data without observing all of the confounders...

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of times when oil rents are higher/lower.

- Global oil price shocks
- Conflict
- Etc

→ If only there was a way to control for *time-variant* omitted variables in observational data without observing all of the confounders...

→ With panel data there is! We can just control for the time.

Time fixed effects

- The country fixed effects model in the resource curse example prevents omitted variable bias from variables that, within countries, do not change over time
- However, panel data also allows us to control for omitted variable bias from variables that change over time across all countries
 - For example, the global oil price affects all countries, and could confound the relationship between oil revenues and institution quality.
- Time fixed effects control for any common shocks that affect all countries at the same time²

²Remember the intuition behind diff-in-diff design.

The time fixed effect model

Rather than write out all the dummy variables, we define the time fixed effects model as follows:

$$Y_{i,t} = \delta_t + \beta_1 X_{1i,t} + \epsilon_{i,t}$$

where

- $Y_{i,t}$ is institutional quality for country i in year t
- $X_{1i,t}$ is oil revenue for country i in year t
- δ_t **is the intercept for time t**
- β_1 is the coefficient estimate for oil revenue
- $\epsilon_{i,t}$ is the error term or residual for country i in year t

Time fixed effects model in R

Estimating time fixed effects in R is just as straightforward as estimating unit fixed effects.

```
time_fe_model <- lm( institutions ~ oil +  
  as.factor(year), # the time FE  
  data = resource_curse  
)
```

The only difference to the unit fixed effects model is that we now control for year since we are interested in time fixed effects.

Time fixed effects models in R

```
summary(time_fe_model)
```

```
##
```

```
## Call:
```

```
## lm(formula = institutions ~ oil + as.factor(year), data = resource_curse)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -1.6598 -0.5339 -0.2172  0.5678  1.8682
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value      Pr(>|t|)  
## (Intercept)    0.1252002  0.1048933   1.194    0.233  
## oil            -0.0282864  0.0030299  -9.336 <0.0000000000000002 ***  
## as.factor(year)1998 -0.0755021  0.1471882  -0.513    0.608  
## as.factor(year)2000  0.0555346  0.1470753   0.378    0.706  
## as.factor(year)2002 -0.0280016  0.1469693  -0.191    0.849  
## as.factor(year)2003 -0.0221837  0.1469729  -0.151    0.880  
## as.factor(year)2004 -0.0100149  0.1470042  -0.068    0.946  
## as.factor(year)2005  0.0004953  0.1471453   0.003    0.997  
## as.factor(year)2006  0.0080277  0.1471435   0.055    0.957  
## as.factor(year)2007  0.0160554  0.1471418   0.109    0.914
```

Time fixed effects models in R

```
# If you don't want to see the time fixed effects  
# in your output  
screenreg(time_fe_model,  
           custom.coef.map = list("(Intercept)"="Intercept",  
                                  "oil"="Oil"))
```

```
##  
## =====  
##           Model 1  
## -----  
## Intercept    0.13  
##              (0.10)  
## Oil          -0.03 ***  
##              (0.00)  
## -----  
## R2          0.12  
## Adj. R2     0.10  
## Num. obs.    682  
## =====
```

Time fixed effects models in R

- We can interpret the coefficient estimates using the same logic as before, remembering that we are now *holding year constant*.
- The time fixed effects model yields an estimate for the `oil` variable that is very similar to the pooled model

```
##
## =====
##           Pooled      Pooled      Country FE   Time FE
## -----
## Intercept    0.106 **   -3.356 ***   -0.443 ***    0.125
##              (0.034)   (0.099)     (0.035)       (0.105)
## Oil          -0.028 *** -0.015 ***   -0.003        -0.028 ***
##              (0.003)   (0.002)     (0.001)       (0.003)
## Log(GDP pc)                0.403 ***
##                          (0.013)
## Polity IV                0.031 ***
##                          (0.003)
## Aid                  0.012 ***
##                          (0.003)
## -----
## R^2           0.114      0.744      0.980      0.116
## Adj. R^2      0.113      0.742      0.979      0.100
## Num. obs.    682        672        682        682
## =====
```

What does it mean to hold time constant?

- The time fixed effects mean that the regression lines for each year have the same slope and are parallel to each other.
- In other words, we are using variation in oil and institutions across countries, within the same year to estimate the effect of oil revenue (as % of GDP) on institutional quality.
- The absolute level of the fixed effect becomes difficult and less meaningful as we add more variables to the model, but the differences across years can provide useful information.

What do time fixed effects control for?

- Fixed effects control for any features of years that are *constant* across countries:
 - Oil price shocks (mostly)
 - Weather / climate (mostly)
 - Culture (mostly)
 - Wealth (not exactly)
- Across a more diverse set of countries, it is less likely that potential confounders are truly constant.
 - If an unobserved confounder is only *mostly* constant across units, fixed effects control for much of its potential for confounding but not all

Time fixed effects model

Advantages

- Allows us to control for all “time variant” confounders that affect *all units* (also called “unit invariant”)
 - This is particularly important in the case of time variant variables which are difficult or impossible to observe.

Disadvantages

- In exchange for this, we have changed what it is that we are estimating
 - No longer comparing different years
 - We are looking for “within year” relationships between X and Y across countries.

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil.

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: Probably not! There might be many other possible confounding features of countries with more/less oil.

→ If only we could employ unit and time fixed effects at the same time!

Two-way fixed effects

- In some applications we can use unit and time fixed effects at the same time.
 - This model is sometimes referred to as the *two-way fixed effects* model.
- When we include both sets of fixed effects, we rule out omitted variable bias that comes from:
 - factors that do not vary within country over time and
 - factors that vary over time across countries

The two-way fixed effect model

We define the two-way fixed effects model as follows:

$$Y_{i,t} = \alpha_i + \delta_t + \beta_1 X_{1i,t} + \epsilon_{i,t}$$

where

- $Y_{i,t}$ is institutional quality for country i in year t
- $X_{1i,t}$ is oil revenue for country i in year t
- α_i is the intercept for country i
- δ_t is the intercept for time t
- β_1 is the coefficient estimate for oil revenue
- $\epsilon_{i,t}$ is the error term or residual for country i in year t

Two-way fixed effects models in R

```
twoway_fe_model <- lm( institutions ~ oil +  
                        as.factor(country) + # the unit FE  
                        as.factor(year), # the time FE  
                        data = resource_curse  
                      )
```

Two-way fixed effects models in R

```
summary(twoway_fe_model)
```

```
##  
## Call:  
## lm(formula = institutions ~ oil + as.factor(country) + as.factor(year),  
##     data = resource_curse)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.42456 -0.06482  0.00200  0.06699  0.45720   
##  
## Coefficients:  
##  
##              Estimate Std. Error t value  
## (Intercept)   -0.408568   0.037613 -10.862  
## oil           -0.001730   0.001648  -1.049  
## as.factor(country)Argentina    0.245153   0.048805   5.023  
## as.factor(country)Armenia      0.103465   0.048717   2.124  
## as.factor(country)Australia    2.033903   0.048698  41.766  
## as.factor(country)Azerbaijan  -0.390617   0.065489  -5.965  
## as.factor(country)Belarus     -0.458492   0.048699  -9.415  
## as.factor(country)Belgium      1.776487   0.048717  36.466  
## as.factor(country)Bolivia      0.242157   0.048715   4.971
```

Two-way fixed effects models in R

```
# If you don't want to see the fixed effects  
# in your output  
screenreg(twoway_fe_model,  
           custom.coef.map = list("(Intercept)"="Intercept",  
                                  "oil"="Oil"), digits=3)
```

```
##  
## =====  
##           Model 1  
## -----  
## Intercept   -0.409 ***  
##              (0.038)  
## Oil         -0.002  
##              (0.002)  
## -----  
## R2          0.981  
## Adj. R2     0.979  
## Num. obs.   682  
## -----
```

Two-way fixed effects models in R

We can interpret the coefficient estimates using the same logic as before, remembering that we are now *holding year and country constant*.

```
##
## =====
##           Pooled           Country FE       Time FE           Twoway FE
## -----
## Intercept    0.106 **      -0.443 ***      0.125             -0.409 ***
##              (0.034)       (0.035)         (0.105)           (0.038)
## Oil          -0.028 ***     -0.003          -0.028 ***       -0.002
##              (0.003)       (0.001)         (0.003)           (0.002)
## -----
## R^2           0.114         0.980           0.116             0.981
## Adj. R^2      0.113         0.979           0.100             0.979
## Num. obs.    682           682             682               682
## =====
```


Two-way fixed effects in R

Question: What is the interpretation of the intercept in the different models?

What does it mean to hold time *and* unit constant?

- The time fixed effects mean that the regression lines for each year have the same slope and are parallel to each other.
- The unit fixed effects mean that the regression lines for each country have the same slope and are parallel to each other.
- In other words, we are simultaneously adjusting for time-specific and unit-specific unobserved confounders.
- The absolute level of the fixed effects becomes difficult and less meaningful as we add more variables to the model, but the differences across years and countries can still provide useful information.

What does it mean to hold time *and* unit constant?

- Two-way FE models control for absolute trends: country and year averages
- In other words, fixed effects de-mean (subtract the average of) the X and Y variables
- The remaining variation, which we try to explain using an explanatory variable (e.g. oil), are **relative deviations**: how countries differ from their own baseline and from the global trend
- The coefficient on the explanatory variable measures the average relationship between X and Y across all such relative (**within-country, over-time**) deviations in the data

Two-way fixed effects model

Advantages

- Allows us to control for all “time variant” confounders that affect all countries and “time invariant” country-specific confounders at the same time
 - This helps us to account for many potential confounders

Disadvantages

- In exchange for this, we have changed once again what it is that we are estimating
 - We are looking for “within country” relationships between X and Y , while accounting for time-specific global shocks
- The ability to adjust for unobserved confounding relies upon the assumption of linear additive effects (you can read more here [Imai and In Song, 2020](#))

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

What about unconfoundedness assumption?

Question: Are we comparing observations that are identical on average in terms of all confounders?

Answer: It depends, we might!

Summary of what we learned by using fixed effects

- Institution quality is quite “sticky”
 - Very little within-country variation in Y
- The apparent effect of oil revenue on institution quality mostly disappeared when controlling for country. This tells us that the relationship in the pooled model...
 - ...mostly resulted from a tendency for countries with consistently higher oil revenues to also have lower institutional quality.
 - ...mostly did not reflect a tendency for countries with changing oil revenues to have much change in their institutional quality.
- Not much evidence that time shocks (eg global oil prices) are important in this particular application.

Conclusion

Observational research designs

1. **Cross-sectional designs**

- Compare outcomes for treated and control units, possibly adjusting for pre-treatment characteristics

2. **Before and after designs**

- Compare outcomes before and after treatment for treated units using over-time data

3. **Difference-in-differences designs**

- Compare outcomes before and after treatment for treated units, and compare to the same change over time for control units

4. **Panel data designs**

- Compare outcomes over time and across units

Regression and causality

When can we interpret a regression coefficient causally?

1. **Randomized experiments**: Coefficient on a binary treatment is estimate of the average treatment effect
2. **Observational studies**: We can only interpret coefficients causally when we have controlled for **all confounders** as additional X variables
3. **Before and after design**: Assuming no time-varying confounding, you can estimate the average causal effect by taking the difference of the mean values of the outcome before and after the treatment.
4. **Difference in differences designs**: Assuming parallel trends between control and treated groups for the latter in the absence of treatment, you can estimate the average causal effect by taking the difference of the mean change in the treated and control groups.
5. **Panel data designs**: Assuming no time *and* unit varying confounders, you can focus on within-unit variation in treatment and outcome with unit and time fixed effects

Seminar

In seminars this week, you will ...

1. ...calculate difference-in-differences.
2. ...learn how to make nice plots.