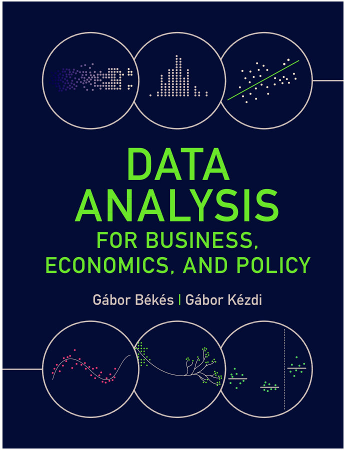
# 24. Appropriate Control Groups for Panel Data

**Gábor Békés**

Data Analysis 4: Causality

2020

# Slideshow for the Békés-Kézdi Data Analysis textbook




- Cambridge University Press, 2021
- **gabors-data-analysis.com**
  - Download all data and code: gabors-data-analysis.com/data-and-code/
- This slideshow is for **Chapter 24**

# When and why to select a control group in xt panel data?

- Consider a binary intervention
  - Treated observations (units, time periods)
  - Untreated observations (units, time periods
- With xt panel data we can use diff-in-diffs or FD or FE panel regressions to estimate the effect
- So far we used all observations in the data
- That meant using all untreated observations to estimate the counterfactual

# When and why to select a control group in xt panel data?

- But sometimes it makes sense to select a subset of the observations to estimate the counterfactual
  - Example: using a diff-in-diffs strategy the parallel trends assumption is more likely to hold for some units than for others
- In such cases we can select a more appropriate control group
- There are many ways to do that that are applicable in may situations
- We consider two specific methods to be used in specific situations
  - The synthetic control method in comparative case studies
  - Constructing a control group with pseudo-interventions in event studies

# Comparative case studies

- An event happened
  - Earthquake in Haiti in 2010
  - Germany reunified in 1990
  - California increased taxes on and restricted use of tobacco in 1988
- What was it's effect?
  - The effect of the earthquake on GDP in Haiti after 2010
  - The effect of the German reunification on the GDP of the Western part after 1990
  - The effect of the California tobacco taxes and restrictions on tobacco use in California after 1988

# Comparative case studies

- With appropriate data, it's straightforward to see what happened to the variable of interest after the intervention
  - What the Haiti GDP was after 2010
  - What the West German GDP was after 1990
  - What tobacco use in California was after 1988
- The big question: How to estimate the counterfactual?
  - What the Haiti GDP would have been after 2010 without the earthquake
  - What the West German GDP would have been after 1990 without the reunification
  - What tobacco use would have been in California after 1988 without the taxes and restrictions

# The synthetic control method

- A method to estimate the effect in a comparative case study
  - One subject one intervention (treated subject)
- By constructing a counterfactual
- Data is xt panel
  - Outcome variable $y$
  - Time series of $y$ and other variables for treated subject before and after the intervention
  - Time series of same variables for several untreated subjects
  - These untreated subjects are called the donors
  - The set of untreated subjects is called the donor pool

# The synthetic control method

- The method creates a single control subject
- From the donor pool of untreated subjects
- It's a synthetic control subject because it is not one of the actual untreated subjects
- Instead, it's a subject with a weighted average of the variables of several untreated subjects

- In essence, the method creates the synthetic control
  - As a weighed average of the subjects in the donor pool
  - By assigning weights to each subject in the donor pool
  - Making sure that the pre-intervention $y$ and selected other variables are similar

# More on the synthetic control algorithm

- The goal of the algorithm is to assign weights to each subject in the donor pool
  - The weights add up to one
  - Zero weights for some donors are OK, it just means that that subject won't add to the synthetic control subject
  - In fact most subjects in the donor pool tend to end up with zero weight
- The result is a set of weights assigned to each subject in the donor pool
  - For example, 0 for donor one, 0.1 for donor two, 0 for donor three, 0.5 for donor four, 0.4 for donor five, 0 for donor six
- The synthetic control subject is that weighted average

# More on the synthetic control algorithm

- Select variables that should be similar for treated subject and synthetic control subject
  - Pre-intervention values of $y$. Typically select values of $y$ in specific time periods
  - Potential confounders that don't change with time
  - Pre-intervention average values of potential confounders that change with time
- Intuitively, a search algorithm
  - Try out all possible weighted averages of the donors
  - Select the one for which the selected variables are the closest to their values for the treated subject
- A lot simpler minimum-distance procedure in practice

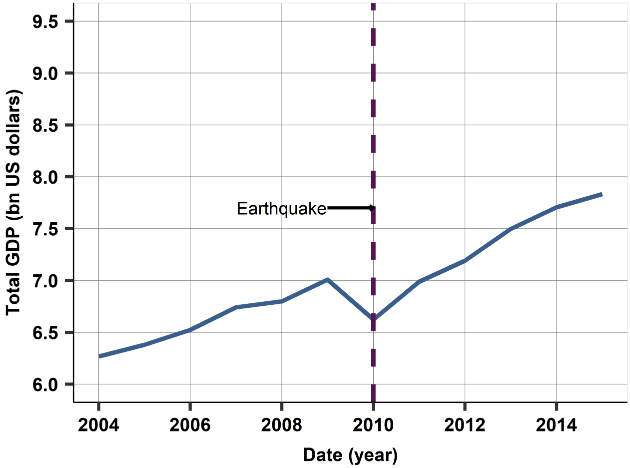
# The role of the analyst in the synthetic control method

- The event is given
- Need to choose the outcome(s) of interest
- Need to choose the donor pool
  - Data availability may be a constraint
- Need to choose the variables that should be similar between the treated subject and the synthetic control subject
  - That includes when those variables should be measured
  - For example, $y$ in what pre-intervention time periods
- The rest is done by the algorithm

# Estimating the effect of the 2010 Haiti earthquake on GDP

- A severe earthquake hit Haiti in January 2010
- What was the effect of this earthquake on Haitian GDP in 2010 and subsequent years?
  - Total GDP
  - Constant 2010 USD prices
- This is a comparative case study
  - An event happened in one country
  - What was the effect of this event on that country
  - Need counterfactual
- Case study based on Best and Burke (2019)
  - Use same data sources
  - `haiti-earthquake` dataset
  - Use their approach in selecting the variables

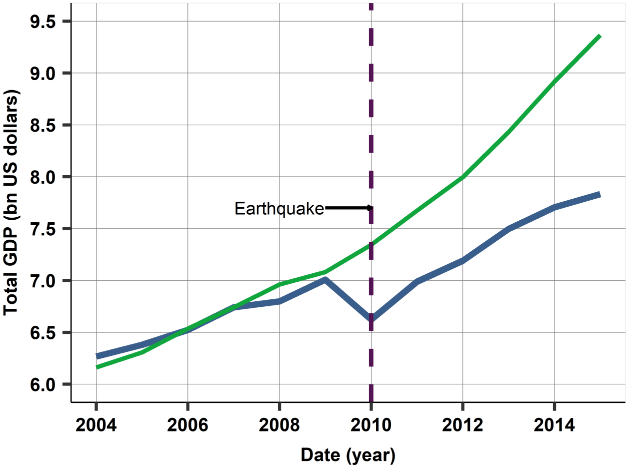
# Total GDP in Haiti



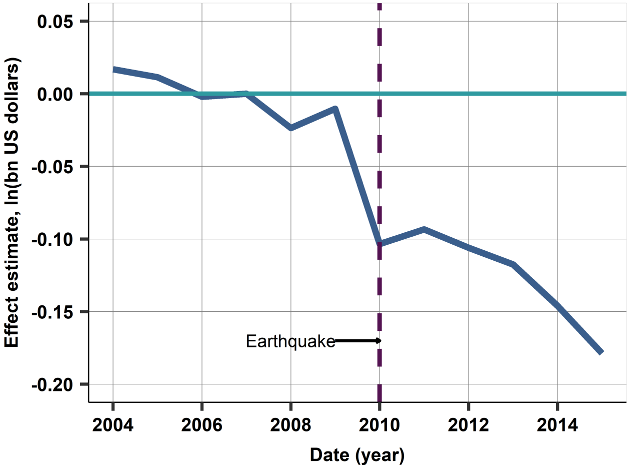
# The synthetic control for Haiti

- Donor pool
  - Countries with less than USD 4000 GDP per capita (2009 PPP USD)
  - Appropriate data available in 2004 through 2015
  - 21 countries altogether (plus Haiti)
- Variables
  - Land size and pre-intervention (2004-9) average values of population, GDP per capita, imports, exports, consumption, gross capital formation, inflation
  - Total GDP in 2005, 2007, 2009
- The synthetic control subject
  - 5 countries with nonzero weight
  - Burundi 23%, Cameroon 21%, Moldova 9%, Togo 47%, Liberia 0.2%

# Total GDP in Haiti and in the synthetic control country



# Log difference of total GDP in Haiti and the synthetic control country



# Answer question of case study

- Severe and permanent negative effect of the earthquake on Haitian total GDP
  - Total GDP dropped by 10% in 2010
  - Remained at least 10% below what it would have been in subsequent years
  - May have dropped even more after three years

# Answer question of case study

- How much should we believe these results?
- Found reasonably good synthetic control country
  - Likely imperfect synthetic control
  - Total GDP trended downwards in Haiti compared to it before 2010
  - Although that was nothing compared to the drop in 2010
  - And donor pool is limited to 21 countries
    - There are just so many countries...

Maybe magnitude of effect is smaller

But the result of a permanent negative effect is likely real

# Event studies setup

- Many subjects observed multiple times (xt panel data)
- Binary intervention (treatment)
- Some subjects remain untreated throughout the time we observe them
- Other subjects become treated, at different time periods
- Question is the average effect of the intervention
  - Focus on ATET: average effect on subjects that become treated
- This setup is included in the more general setup of FE and FD xt panel regressions
  - Redefine the time structure – more intuitive and transparent
  - More explicit about treated and control units

# Event studies

- Event studies re-define time around the time of the intervention
  - This re-defined time is called event time
- So event time is defined only for subjects that become treated at one point
  - Pre-intervention event time periods are negative
    - -1 for the time period before the intervention
    - -2 for two time periods before, etc.
  - Post-intervention event time periods are positive
    - 1 for the time period after the intervention
    - 2 for two time periods after, etc.
  - The intervention may take place within one time period, which has event time zero

# Event studies

- Look at this setup in two steps
- First, consider treated subjects in a first difference model, comparing treated and untreated differences
- Second, combine with idea seen before: find a better control group.

# Event study regression, treated subjects only

- xt panel data, observations indexed by $i$ and $t$
- For now restrict attention to subjects that became treated at one point in time
- Outcome is $y_{it}$
- Binary indicators $D_{is}$ one if the event time period is $s$, zero otherwise
  - $s$ is event time: it may be negative or positive or zero

- We will look at a simple example, see what event time implies, and how it helps get a more realistic ATET.

# Event study regression, treated subjects only

- This is an FD panel regression in event time

$$\Delta y_{it}^{E} = \alpha + \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_3 D_{i3} \tag{1}$$

Interpreting the coefficients

- $\alpha$ is the average change in $y$ outside the event time window $[1,3]$
  - before the intervention AND 4 or more time periods after the intervention
- $\beta_{i1}$ is how much more $y$ tends to change 1 time period after the intervention (how much more compared to $\alpha$); $\beta_{i2}$ and $\beta_{i3}$ analogously
- Sum = cumulative coefficient: $\beta_{cumul} = \beta_1 + \beta_2 + \beta_3$ – shows how $y$ changes, on average, within three time periods after the intervention
  - Compared to how $y$ tends to change outside the $[1,3]$ window of event time

# Event study regression, treated subjects only

$$\Delta y_{it}^{E} = \alpha + \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_3 D_{i3} \tag{2}$$

- Do we estimate an ATET?
- Is $\beta_{cumul}$ a good estimate of the cumulative effect of the intervention?
- Only if $\alpha$ is a good estimate of the counterfactual
  - Without the intervention, $y$ would have changed the way it did outside the $[1, 3]$ event time window among treated subjects
- There is no control group here
- The counterfactual is how $y$ changed before the intervention (and after the last post-intervention period included in the regression)

# Event study regression, treated subjects only

- With an event study regression that includes treated subjects only
- We don't have a control group: The effect is estimated assuming that what happened before the intervention is a good counterfactual
- Can we do more?
- We can add pre-intervention binary indicators
  - Such as $D_{i(-1)}$ or $D_{i0}$ for the time period of the intervention only, to take care of reverse causality or anticipation effects
  - But that would still not give us a control group
- To compare to a control group we need to define event time for subjects that were never treated
  - That's tricky but doable = second step.

# Estimating the impact of replacing football team managers

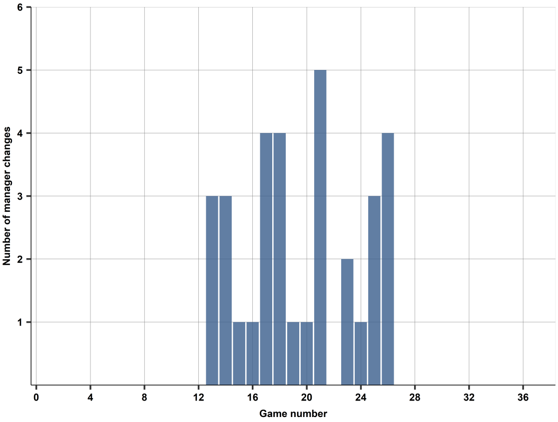
- What is the effect of replacing the manager of a football (soccer) team on team performance?
  - Professional football
  - Managers are coaches with broad responsibility
  - The situation we look at is replacing the manager within the season
  - Typically, that happens after poor performance
- Outcome is points per game
  - 0 for loss, 1 for draw, 3 for win
- Intervention is replacing the manager
  - Happens between games, there is no game with event time zero

# Football Data

- `football` dataset
- English Premier League, 11 seasons, 20 teams
- Every team plays with every other team twice in a season: 38 games
- We denote time within season by game number, 1 through 38.
- 8360 observations in total ($20 \times 38 \times 11$)
- Outcome is points per game
  - 0 for loss, 1 for draw, 3 for win
  - A little less than a quarter of the games results in draw (1pt)
  - The rest result in one team winning (average thus 1.5pt)
  - Overall average is 1.38

# Manager replacements

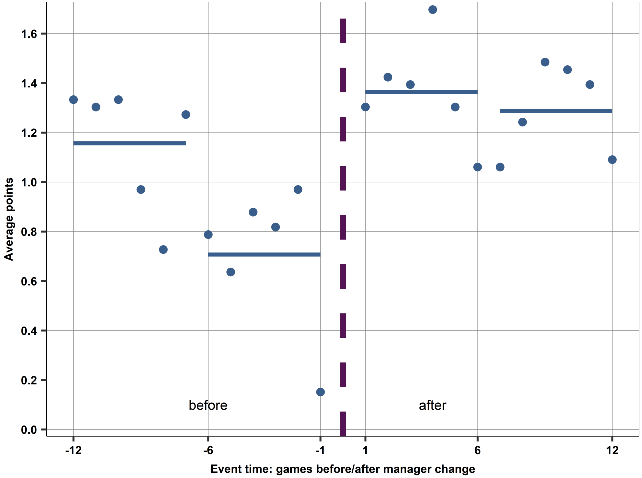
- 94 within-season manager changes during these 11 seasons
- We require 12 games played before and after the manager change
- restrict data to changes happened after 12 and before 27 game number.
- 33 manager changes analyzed



# Average outcomes before and after the intervention

- 33 interventions analyzed here
  - 33 teams replaced their managers *within* one of the 11 seasons in the data
  - AND replacement happened between game numbers 12 and 27
- Event time positive after replacement, negative before replacement
- Outcome (points) 12 weeks before intervention and 12 weeks after intervention
- Calculated average points for each event time across those 33 teams (24 such averages)
- Also calculated averages across six event time periods (four such averages)

# Average points before and after manager replacement



# Interpreting the results

- These results are for the 33 teams that replaced their manager within-season
- They performed below average already 7-12 before the intervention (1.16 here compared to 1.38 overall average)
- Then their performance worsened significantly, to 0.71 points on average for 1-6 games before intervention
- The game result before intervention was especially poor, average very close to zero
- After management replacement outcome increases substantially, to 1.38 1-6 games after
  - By coincidence, this is the overall average outcome in the League
- This increase looks permanent, persisting to 7-12 games after the intervention

# Is this the effect?

- We see a large and permanent increase of team performance after the manager is replaced
  - After below-average and then worsening performance
  - An increase to the overall league average
  - Persisting to at least 12 games after replacement
- Is this the effect of the manager replacement?
- It depends on the counterfactual
  - Is the very low before-intervention outcome the counterfactual?
  - In other words, would the very bad performance have continued had the manager not been replaced?

# Mean reversion

- When we examine a sequence of random values,
- a series of unusually large values are followed by a smaller value
  - closer to the mean
- or, a series of unusually small values are followed by a larger value
  - closer to the mean
- Importantly, we can have this even without any intervention.

- Here: after a run of bad luck, things may go back to normal
- without any intervention, such as replacing the manager

# What's the right counterfactual?

- How can we tell whether, or how much, of the improved team performance after the intervention is
  - Due to replacing the manager
  - Or due to other factors that would lead to mean reversion?

- We can try to find a control group that we can use to estimate the counterfactual
  - This is the second step of the event study approach

# Selecting a control group for event studies

- Selecting a control group of untreated subjects is necessary to estimate the counterfactual
  - In other words, to uncover what would have happened to the treated subjects without the treatment
- In event studies, we define the control group by defining pseudo-interventions
  - Pseudo-interventions are event time periods, or instances between event time periods, for untreated subjects
  - That are preceded by changes in outcomes that are similar, on average, to pre-intervention changes in outcomes among treated subjects

# Estimating the impact of replacing football team managers

- We found that team performance increases substantially after the manager is replaced during the season
  - But that may not show the effect we are after
  - Mean reversion may play a role here
- Need a control group to get a better estimate of the counterfactual
- Let's create such a control group
  - Identify subjects (teams and time periods) that can constitute a control group
  - Define pseudo-interventions for them
  - Compare post-intervention outcomes between treated and control

# Defining a good control group

- A control group should consist of teams playing 24 games within the season (24-game spells)
  - That's what the treatment group consists of
  - With 12 games played before the intervention, 12 games played after
- A good control group is similar to the treatment group in pre-treatment performance
  - Somewhat below-average performance (1.16) in games the first 6 games (event time window $[-12, -7]$)
  - Substantial drop in performance (to 0.71) in the net 6 games (event time window $[-6, -1]$)
  - An especially low performance (0.15) on game 12 (event time -1)
- Pseudo intervention is between games 12 and 13

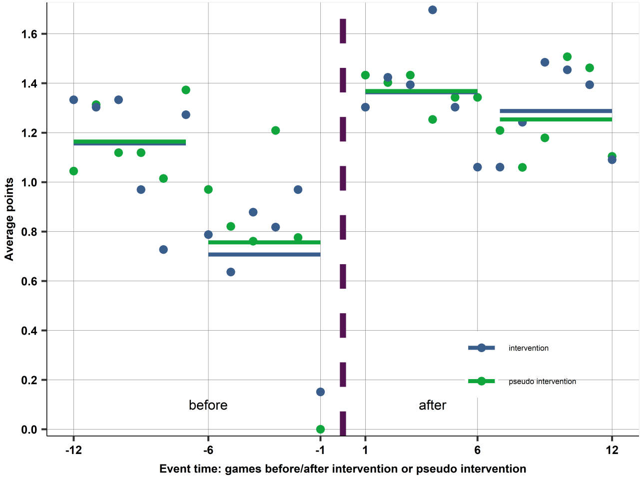
# Finding a control group

- But how to find such subjects (24-game spells)?
- Look at all 24-game spells in the data (exclude when manager was replaced)
- Apply criteria that makes the first 12 game outcomes similar to the pre-treatment outcomes in the treatment group
- The following criteria make sure the patterns are similar
  1. Average points in first 6 games was 0.83 to 1.33
  2. Average points in next 6 games was 0.17 to 1.33
  3. A loss at game 12 (0 point)
- There are 132 such 24-game spells in the data
- Overlap - randomly select. Result is 67 sets of 24-game spells

# Finding a control group

- We have 33 treated and 67 control teams
- Each with outcomes for 24 games
- The first 12 game (event time window $[-12, -1]$) outcomes are similar across the treated and control teams, on average
- After the 12th game manager is replaced in treatment group
- Pseudo-intervention is defined to take place after game 12 in the control group
- And we look at what happens on games 13-14
  - This is event time window $[1, 12]$

# Average points before and after manager change and pseudo-intervention



# Interpreting the results

- Blue dots and lines show the 33 treated teams that replaced their manager within-season
- Green dots and lines show the 67 control teams
- Their pre-intervention performance is very similar
  - By design; that's how the control group was selected
- After the intervention their performance is very similar
  - This is the key result of the exercise

# Interpreting the results

- Teams that experienced bad performance similar to the treatment group
  - but didn't replace their managers
- Experienced a similar increase in performance
  - than teams that did replace their manager

# Regression estimation

- The same results estimated in a regression
  - Aggregated the observations to 6-week event time windows
    - Average points in event time window $[-12, -7]$, $[-6, -1]$, $[1, 6$, $[7, 12]$
    - Then take first differences
    - Number of observations is 300: $33 \times 3$ in the treatment group, $67 \times 3$ in the control group
  - The regression replicates what we see on graph (those 6-game average lines)
  - Except it's in first differences to focus on changes between those lines
  - So coeff estimates are direct estimates of the effect (+ we get SEs)

$$\Delta y^E = \beta_0 + \beta_1 post_{1-6} + \beta_2 post_{7-12} + \beta_3 treat + \beta_4 treat \times post_{1-6} + \beta_5 treat \times post_{7-12} \tag{3}$$

# Regression estimation

$$\Delta y^E = \beta_0 + \beta_1 post_{1-6} + \beta_2 post_{7-12} + \beta_3 treat + \beta_4 treat \times post_{1-6} + \beta_5 treat \times post_{7-12} \quad (4)$$

- $post_{1-6}$ is a binary indicator for event time window $[1, 6]$
- $post_{7-12}$ is a binary indicator for event time window $[7 - 12]$
- $treat$ is a binary indicator for the treatment group

# Interpreting the regression coefficients

$$\Delta y^E = \beta_0 + \beta_1 post_{1-6} + \beta_2 post_{7-12} + \beta_3 treat + \beta_4 treat \times post_{1-6} + \beta_5 treat \times post_{7-12} \quad (5)$$

- $\beta_0$ shows the average change in points in the reference time period: from event time window $[-12, -7]$ to event time window $[-6, -1]$, for the control group
- $\beta_1$ shows the average change in points from event time window $[-6, -1]$ to event time window $[1, 6]$, compared to the change in the reference time period (captured by $\beta_0$), for the control group
- $\beta_2$ shows the average change in points from event time $[1, 6]$ to $[7, 12]$, again compared to the change in the reference time period, for the control group

# Interpreting the regression coefficients

$$\Delta y^E = \beta_0 + \beta_1 post_{1-6} + \beta_2 post_{7-12} + \beta_3 treat + \beta_4 treat \times post_{1-6} + \beta_5 treat \times post_{7-12} \tag{6}$$

- $\beta_3$ shows the treatment-control difference in the change in the reference time period (from $[-12, -7]$ 7-12 to $[-6, -1]$)
- If we selected the control group well, $\beta_3$ should be close to zero.
  - Because we want the control group to have the same pre-treatment changes in the outcome

# Interpreting the regression coefficients

$$\Delta y^E = \beta_0 + \beta_1 post_{1-6} + \beta_2 post_{7-12} + \beta_3 treat + \beta_4 treat \times post_{1-6} + \beta_5 treat \times post_{7-12} \tag{7}$$

- $\beta_4$ and $\beta_5$ are the effect estimates
- $\beta_4$ shows treatment-control difference in the change right after the intervention or pseudo-intervention, from the average of event time window $[-6,-1]$ to $[1,6]$
- $\beta_5$ shows treatment-control difference in the subsequent change, from the average of event time window $[1,6]$ to $[7,12]$

# Regression results

- Column (3) in the following table shows the coefficient estimates for the regression
- Columns (1) and (2) are analogous regressions separately for the treatment group and the control group
  - Interpreting the coefficient estimates of regressions (1) and (2) is a good exercise
  - So is showing the equivalence of some of those estimates to the estimates in column (3)

# Regression results

| Variables | (1) treatment | (2) control | (3) treatment+control |
|---|---|---|---|
| $post_{1-6}$ | 1.11** | 1.06** | 1.06** |
| | (0.19) | (0.09) | (0.09) |
| $post_{7-12}$ | 0.37* | 0.34** | 0.34** |
| | (0.16) | (0.09) | (0.09) |
| $treated$ | | | -0.00 |
| | | | (0.10) |
| $treated \times post_{1-6}$ | | | 0.04 |
| | | | (0.20) |
| $treated \times post_{7-12}$ | | | 0.04 |
| | | | (0.18) |
| Constant | -0.45** | -0.45** | -0.45** |
| | (0.10) | (0.03) | (0.03) |

# Interpreting the results

- The regression coefficients show the same as the figure
- Coefficient estimate on *treated* is zero as we wanted

- The effect estimates are very close to zero
  - The improvement from before to after the intervention (or pseudo-intervention) is the same in the two groups, on average
  - So is the subsequent change

# Conclusion of case study

- Is the effect zero then?
- Yes if the control group's performance after the pseudo intervention is a good counterfactual
  - Teams that replaced their manager
  - Would have performed as well as the control teams
  - Even if they hadn't replaced their manager
- Recall: we need exogenous variation in the causal variable to have a good effect estimate
  - Potential outcomes should be independent of treatment
  - Here: which team replaces their managers after a similar trajectory of bad outcomes through 12 games should be independent of how they would perform later

# Conclusion of case study

- Some of the variation may be exogenous indeed
  - To replace a manager there has to be available managers to take their places
  - Decisions to replace a manager may be in part arbitrary
- But some of the variation may be endogenous
  - Some teams may have performed badly for reasons other than the manager
  - and they may have remedied those causes

- So, perhaps, the effect is not zero after all
- But it's likely smaller than what we would identify by looking at the treated teams only!

# Main takeaway

- When estimating the effect of an intervention using xt panel data
  - It is sometimes better to select a subset of all non-treated observations to serve as a control group
- To estimate the effect of an intervention on a single subject
  - We can estimate the counterfactual using the synthetic control method
- With an intervention on affecting many subjects at different times
  - We can carry out an event study
  - With the help of a control group of comparable pseudo-interventions