Overview of Greenstone, Hornbeck, and Moretti (2010)

GHM (2010)

Hussain Hadah (he/him)
26 January 2024

Outline for Today

1. **Introduce you to the GHM paper**
2. **Basics of "Difference-in-Differences"**
3. **Summary of the results of GHM**
4. **Tech clusters**
5. **Activity on COVID-19 and agglomeration**
6. **Practice quiz on Canvas in pairs**

Tulane University

Next Week

1. Economic development incentives
2. Economic development incentives briefing note

Readings

1. O'Flaherty (2005) (pages after 525 and after are optional reading)
2. Bartik (2017) just pages 1 to 5 (the "overview")
3. Briefing note readings:
    1. Last Names A to D -> Neumark, Kolko - 2010
    2. Last Names E to G -> Button - 2019
    3. Last Names H to J -> Holmes - 1998
    4. Last Names K to L -> Coates and Humphreys - The Stadium Gambit
    5. Last Names M to O -> Moretti, Wilson - 2014
    6. Last Names P to S -> Strauss-Kahn, Vives - 2009

Agglomeration Spillovers in Action

Paper's abstract

> We quantify agglomeration spillovers by comparing changes in total factor productivity (TFP) among incumbent plants in "winning" counties that attracted a large manufacturing plant and "losing" counties that were the new plant's runner-up choice. Winning and losing counties have similar trends in TFP prior to the new plant opening. Five years after the opening, incumbent plants' TFP is 12 percent higher in winning counties. This productivity spillover is larger for plants sharing similar labor and technology pools with the new plant. Consistent with spatial equilibrium models, labor costs increase in winning counties, indicating that profits ultimately increase less than productivity.

Spillovers from agglomeration

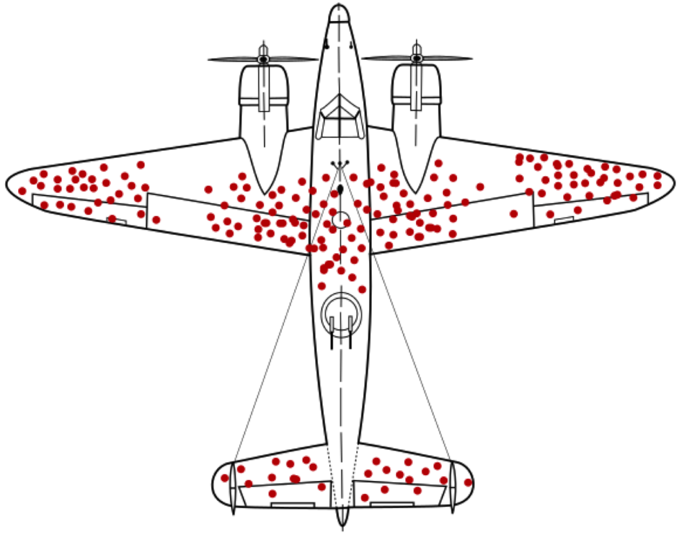
- Greenstone et al. (2010) try to measure the spillovers of agglomeration.
- We talked about these earlier. Locating near other firms, especially of a similar type, can lead to benefits.
- These benefits include:
    1. Sharing similar intermediate input markets (labor, capital) that are “thick” markets, so it’s easier to find what you need. Costs of capital with high fixed costs (e.g., production studio) can be better shared among firms. Thick markets also allow a better match between employees and employers.
    2. Spread of ideas, often through workers that move more frequently between firms or otherwise communicate often (“happy hour”, industry groups).
- These spillovers will increase productivity.
- Greenstone et al. (2010) seek to figure out what happens to existing manufacturing firms when a large manufacturing firm moves to the area.
- How does the productivity of the existing manufacturing firms change?

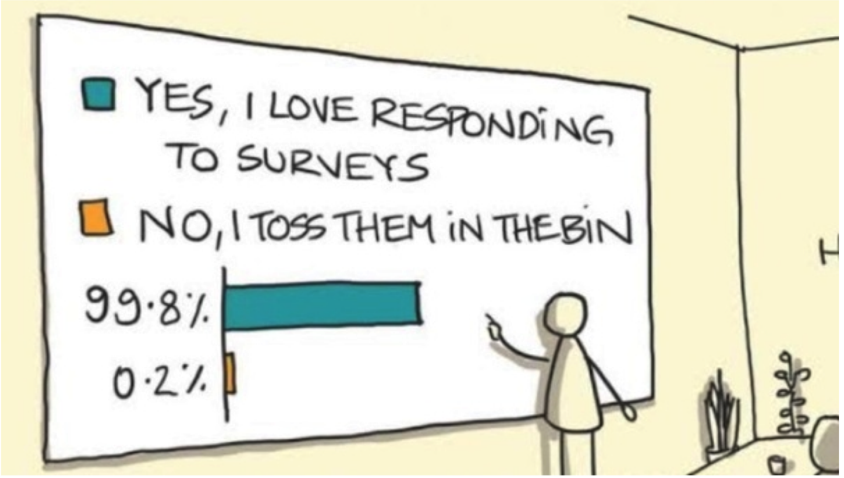
Selection Bias

What is selection bias?

- Suppose I want to know the effect of a college degree on earnings.
- I compare the earnings of people with a college degree to those without a college degree.
- I find that people with a college degree earn more than those without a college degree.
- Does this mean that a college degree causes higher earnings?
- No, it could be that people who go to college are more motivated, smarter, etc. than those who don't go to college.
- This is called selection bias.
- We need to control for these other factors that affect earnings.

Example of selection bias





Measuring the effect of agglomeration is hard

- Suppose I looked at firms in Silicon Valley.
- I would see that they're all (mostly!) very productive.
- Is this because of the benefits of agglomeration?
- Or is it selection bias, where the productive firms self-selected into Silicon Valley? Or these firms are just more productive anyways, and it doesn't have to do with the benefits of agglomeration?
- To capture the spillovers from agglomeration we need to separate the selection bias from the actual spillover effect. I will show you soon how selection bias could operate in the context of GHM and how they get around it.

Measuring the effect of agglomeration is hard

- Suppose that selection bias exists.
- If this is the case, then we will get a biased estimate of the causal effect.
- My estimate of the causal effect = actual spillover effect + selection bias
- Depending on the context, selection bias could be positive, so I incorrectly overstate the causal effect, or it could be negative, so I incorrectly understate the causal effect.
- To capture the spillovers from agglomeration I need to separate the selection bias from the actual spillover effect. GHM does this in a unique way using an approach called “Difference-in-Differences” that I will introduce you to this week.

What are "causal effects"?

- About half of economics research nowadays tries to quantify what are called causal effects. They try to find the causal effect for some policy or event.
- Causal effects means: what is the actual effect of this policy or event on some outcome(s)?
- Examples:
- What is the effect of tax incentives for the film industry on filming location choice?
- What effect do rent control policies have on rent prices and the supply of housing?
- What effect do “ban the box” policies have on discrimination in hiring on the basis of criminal record, race, or ethnicity?
- What effect do Section 8 housing vouchers have on families?
- We call the these “causal effects” since we want to know how some exist or policy causes some sort of outcome.
- We want an estimate that is “causal”, showing the effect of some X on some Y.

What are "causal effects"?

- We want to avoid an estimate that gives us simply a correlation. Many things can be correlated but not related through causation.
- E.g., greenhouse gas emissions have risen over time at the same time that the number of people who are pirates decreases. This does not mean that the decrease in pirates is causing climate change! We also want to estimate the effect of a policy or program that controls for other things that could be going on.
- E.g., Seattle could increase it's minimum wage. After that Seattle could experience some change in employment (positive or negative). That could be due to the minimum wage or it could be due to something else that we haven't controlled for.

What are "causal effects"?

- The gold standard (i.e. the best approach) to estimate causal effects is through a randomized control trial (called an RCT).
- This is done in the social sciences sometimes, but it’s more common in medicine.
- For example, a common RCT is studying the causal effect of a drug on some sort of health outcome(s).
- Researchers will randomize subjects (people) into a treatment or control group. The treatment group gets the pill (called the “treatment”) and the control group gets a placebo.
- Since the treatment and control groups are on-average identical due to the randomization, any differences in outcomes between the two groups is due to the treatment (the pill).

What are "causal effects"?

- While social scientists can sometimes randomize “treatments” to study their effects (we will see a few examples), often time it’s not possible to use an RCT to estimate causal effects.
- For example, it is unethnical or not feasible to randomize things like state laws. It would be lovely, from an estimation standpoint, to randomize, say, which states have tax credits to see what effect tax credits have, but it’s just not possible.
- Social scientists often have to use other methods to try to get an estimate of something close to the true causal effect.
- One such method we will discuss is the “Difference-in-Differences” (aka Diff-in-Diff, DiD, or DD), which is incredibly common in economics and also in the quantitative social sciences.
- The DiD approach gives us a causal estimate of some policy or event X on some outcome Y, but under certain assumption only.

What do Greenstone et al. (2010) do?

- GHM look at large manufacturing firms ("million dollar plants" or MDPs) that chose to set up in certain counties ("winner" counties).
- They then use information to determine their runner-up (loser) counties. They determine what the loser counties are by using published information in the corporate real estate journal Site Selection.
- In Site Selection they published an article in each issue called "Million Dollar Plants".
- Each article mentions the winning county and the loser counties.
- This gives then a "treatment group" (the winning counties) and a set of "control groups") (counties that just barely did not win the MDP).
- GHM compare other, existing manufacturing firms the winning county to existing manufacturing firms in the counties that just barely lost out on getting the MDP.
- Firms in the winning county = "treated" group -> these firms gets the agglomeration spillovers from the large firm that moves in.
- Firms in the losing counties = "control" group -> these firms do NOT get the spillovers. Similar to a randomized trial (e.g., a study of the effects of a drug).

Are the "losers" a good control group?

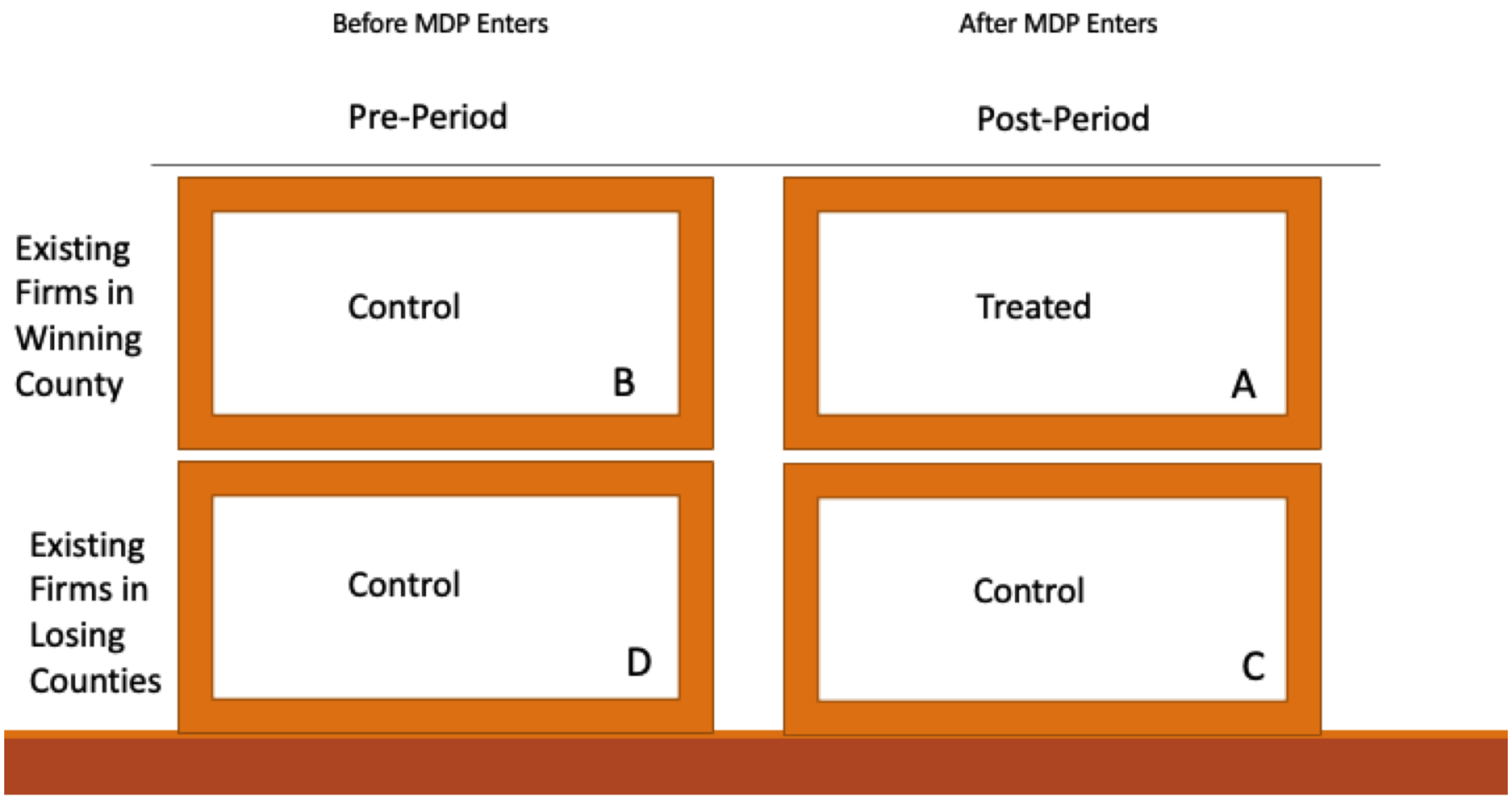
- A fundamental assumption is required to get a proper estimate of the actual (causal) effect of spillovers.
- The treatment group (firms in winning county) must be as close as identical as possible to the control group (firms in losing counties).
- Why might this not hold?
- Winning counties might be better. Firms in winning counties might be more productive, independent of the spillover effect from the new firm.
- For this reason they use a “difference-in-differences” methodology.

Intro to Difference-in-Differences

A Brief Intro to Difference-in-Differences

- Also called "Diff-in-Diff" or just DD or DID or DiD.
- This is a particular model used in regression analysis (more on what that is later).
- Instead of just comparing the "treated" firms (firms in the winning county) to the "control" firms (firms in the losing counties), they make this comparison over time.
- Compare the pre-period (the large firm hasn't moved in yet, no firm is "treated") to the post-period (the large firm has moved in, only firms in the winning county are "treated")

How does DiD work?



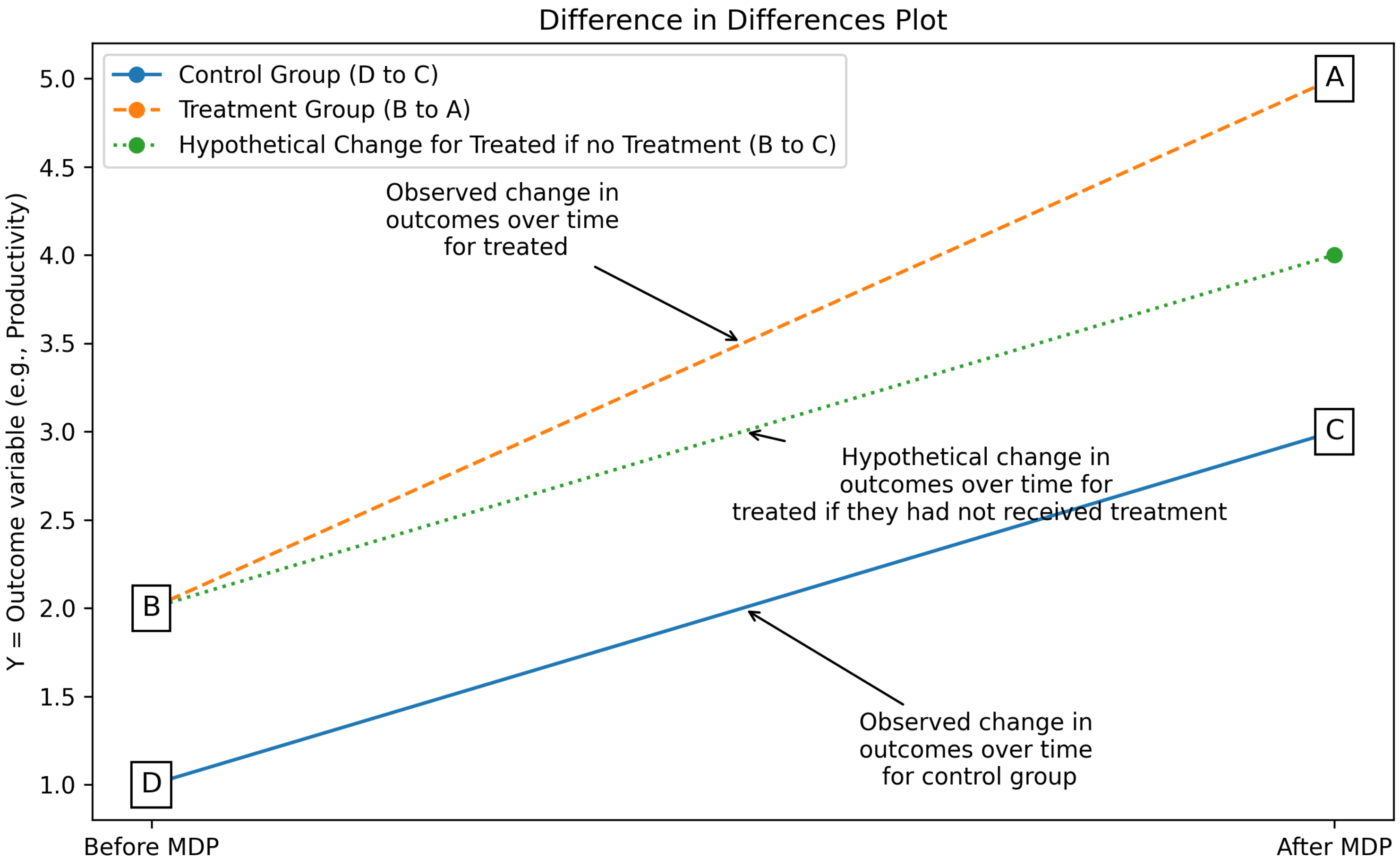
The Difference-Difference Estimate

- Step 1: Take the before and after difference for the firms in the winning county: A – B
- Where A = productivity of existing firms in winning counties AFTER the MDP plant moves in.
- B = productivity of existing firms in winning counties BEFORE the MDP plant moves in.
- Step 2: Take the before and after difference for the firms in the losing counties: C – D
- Where C = productivity of existing firms in losing counties AFTER the MDP plant moves into the winning county (but not the losing counties).
- D = productivity of existing firms in losing counties BEFORE the MDP plant moves in.
- Step 3: Take the difference-in-difference (hence the name).
- The difference-in-differences estimate of the causal effect is: (A – B) – (C – D)
- That is, the before vs after in winning counties (A – B) compared to the before vs after in losing counties (C – D).

The Difference-Difference Estimate

- The difference-in-differences estimate of the causal effect is: (A – B) – (C – D)
- The A – B tells us the change over time in the winning county. How did productivity change for existing firms?
- The C – D tells give us an estimate of the counterfactual. If winning counties are similar to losing counties in their economic trends (we will get into that) then C – D gives us an estimate of what would have happened without a MDP moving in.

The Difference-Difference Estimate: Plot



DiD vs "Naïve Comparisons"

To better illustrate how DiD can estimate causal effects better than other comparisons, consider two "naïve" comparisons:

1. Naïve comparison 1: No control group (A – B)
2. Naïve comparison 2: No pre-period (A – C)

Naïve Comparison 1: No Control Group

- Suppose I didn't have a control group, and I decided just to compare the treated group before and after.
- That is, I compare the productivity of the existing firms in the county that wins the MDP before the MDP arrives ("B") to the productivity of the existing firms in the county that wins the MDP after the MDP arrives ("A").
- That is, I calculate A – B and use that as my estimate of the effect of spillovers on productivity.
- The problem with using A – B as the estimate is that it could include bias from **uncontrolled time trends**.
- Time trends = an existing trend where the outcome variable would have decreased or increased anyways, independent of the effect of the "treatment" (MDP moving in).
- The estimate A - B does not control for these existing time trends, and it could overstate or understate the true causal effect.