Difference-in-Difference Study for GGRF

Andrew Bartnof

April 25, 2025

1 Introduction

This paper documents my difference-in-difference (DiD) analyses for the GGRF insurance project, in which I attempt to quantify the impact of a green energy project on a neighborhood's home appreciation rates. I present a mixed-effects model in the main part of this paper. In the appendix, I present another model, which uses a more conventional two-way fixed-effect DiD schema.

2 Brief Background

If green energy projects (eg solar farms and wind turbines) are built near residential areas, these projects can impact local home values, lowering the rate at which they appreciate vis-à-vis homes in other areas. This gives home-owners an incentive to resist green projects in their neighborhoods, even if they support green projects in general.

We want to quantify this potential relative change in the rate in which homes appreciate near green energy projects, so that we can help to develop a new kind of financial project; a parametric risk insurance payout for home-owners near green energy projects. The idea is that, for a limited amount of time, while the green energy projects artificially depress local homes' values, any home-owner that sells their house can be compensated for this drop in home-value.

3 Data

Naturally, there are a few dimensions we'd like to better understand– do all green energy projects lower the rates at which homes appreciate? how far need a house be from a project before its value is impacted? and how long does this change in home value endure? This is a rather 'cheap-and-cheerful' analysis using readily-available datasets, so I'll attempt to answer these, while making sensible assumptions where necessary.

In this paper, I'll refer to homes that are near green energy projects as **manipulation** homes, and homes that are not near green energy projects as **control** homes.

I use three key datasets:

3.1 Zillow Home Prices

This dataset lists the home prices for 26,344 zip-codes, between the years 2000 and 2024. This was collected before I jumped on the GGRF project so I can't give much more insight into it.

3.2 Large-Scale Solar Photovoltaic Sites

The solar dataset[1] lists the locations and years of operation of thousands of large-scale solar photovoltaic sites. Note that Ben Hoen is an author of both this dataset and the turbine dataset.

3.3 Wind Turbines

This dataset[2] lists the locations and years of operation of thousands of wind turbines.

3.4 Data Notes

Each observation in the solar and turbine datasets is represented as a polygon in a shapefile. The Zillow file commits us, however, to using zip-codes as our finest geographic granularity. Consequently, I assign each solar and turbine project to a single zip-code. Any homes that share a zip-code with a green project are considered manipulations. (We'd prefer not to say that a home either *is*, or *isn't*, near a solar project as if it's a binary metric, but it'll do.)

I also round all dates to the year. If there are multiple home values noted in a given zip-code within this year, I take the median home price.

4 Mixed-Effects Model

The formula for our mixed-effects model looks like this:

```
rate of home appreciation ~

year + solar_manipulation + turbine_manipulation + (1|zip_code)
```

Dependent variable

- The rate of home appreciation is calculated as: *(current home price - last year's home price) / (last year's home price).*

Fixed effects

---

[1] Fujita, K.S., Ancona, Z.H., Kramer, L.A. et al. Georectified polygon database of ground-mounted large-scale solar photovoltaic sites in the United States. Sci Data 10, 760 (2023). https://doi.org/10.1038/s41597-023-02644-8

[2] Rand, J.T., Kramer, L.A., Garrity, C.P. et al. A continuously updated, geospatially rectified database of utility-scale wind turbines in the United States. Sci Data 7, 15 (2020). https://doi.org/10.1038/s41597-020-0353-6

- Year: the calendar year, represented as an unordered factor. This allows us to represent general market fluctuations (eg the Great Recession) easily.

- Solar manipulation: if a zip code ever had a solar plant built there, we encode the zip-code each year as one of the following:

  - Construction: if only one solar plant is built in this zip-code, then 'construction' is synonymous with 'year of operation'. If multiple solar plants are built in this zip-code, this represents the duration between the first plant's year of operation and the last plant's year of operation. 'Construction' indicates that we have faith that during this time period, workers and trucks were around.

  - -5:-1: Years prior to year of first plant's operation

  - 1:5: Likewise, years after last plant's operation

  - Censored: if we are more than five years before construction, or less than five years after, construction in a zip-code

  - Control: a zip-code that never gets a solar project

- Turbine manipulation: encoded exactly like solar manipulation

Random effects

- Zip-code: here, I model all zip-codes as random intercepts. This is common practice in models that have repeated measures per subject; this compensates for the mean difference in home appreciation in each zip-code

4.1 Goodness-of-Fit Metrics

In order to verify whether it's even sensible to try to calculate the impact of green energy projects, we also fit a null model, which omits turbine manipulation and solar manipulation. This means that the null model's formula is:

```
change in home price v. last year ~ year + (1|zip_code)
```

.

An ANOVA comparing these two models indicates that the full model is statistically significant (see figure 1). This indicates that it the turbine and solar manipulations are statistically significant.

The unadjusted intra-class correlation (ICC) is 0.011, which is quite low. A high ICC would that a lot of the variance that is explained in the full model is done by the random-effects. Rather, our fixed effects are explaining a lot of the variance– and the effects we're trying to model are all fixed-effects.

Variable	Estimate	2.5%	97.5%
(Intercept)	0.053	0.047	0.058
year2002	0.015	0.013	0.016
year2003	0.014	0.013	0.015
year2004	0.021	0.02	0.023
year2005	0.047	0.046	0.049
year2006	0.05	0.049	0.052
year2007	-0.013	-0.014	-0.012
year2008	-0.071	-0.072	-0.069
year2009	-0.131	-0.132	-0.13
year2010	-0.116	-0.117	-0.115
year2011	-0.09	-0.091	-0.089
year2012	-0.097	-0.098	-0.096
year2013	-0.029	-0.031	-0.028
year2014	0.005	0.004	0.006
year2015	-0.012	-0.013	-0.01
year2016	0.001	-0	0.002
year2017	-0.014	-0.016	-0.013
year2018	-0.003	-0.004	-0.002
year2019	-0.003	-0.004	-0.002
year2020	-0.006	-0.008	-0.005
year2021	0.055	0.054	0.057
year2022	0.082	0.081	0.083
year2023	0.015	0.014	0.016
year2024	-0.03	-0.031	-0.029
solar_manipulation-2	-0.001	-0.004	0.003
solar_manipulation-3	-0.001	-0.005	0.002
solar_manipulation-4	-0.002	-0.005	0.002
solar_manipulation-5	-0.003	-0.007	0.001
solar_manipulation1	0.004	0	0.007
solar_manipulation2	0.004	0	0.007
solar_manipulation3	0.006	0.003	0.01
solar_manipulation4	0.01	0.006	0.014
solar_manipulation5	0.009	0.005	0.013
solar_manipulationCensored	0.004	0.001	0.006
solar_manipulationConstruction	0.003	-0	0.006
solar_manipulationControl	0.006	0.001	0.011
turbine_manipulation-2	0	-0.006	0.007
turbine_manipulation-3	0	-0.006	0.007
turbine_manipulation-4	0.001	-0.006	0.007
turbine_manipulation-5	0.001	-0.006	0.008
turbine_manipulation1	-0.001	-0.007	0.004
turbine_manipulation2	-0.004	-0.01	0.002
turbine_manipulation3	-0	-0.006	0.006
turbine_manipulation4	0	-0.006	0.006
turbine_manipulation5	-0.003	-0.009	0.004
turbine_manipulationCensored	0.003	-0.002	0.008
turbine_manipulationConstruction	0.002	-0.003	0.007

```
Data: JoinedData
Models:
mod0: home_price_change ~ year + (1 | zip_code)
mod1: home_price_change ~ year + solar_manipulation + turbine_manipulation + (1 | zip_code)
      npar      AIC      BIC logLik deviance  Chisq Df Pr(>Chisq)
mod0    26 -1361360 -1361072 680706 -1361412
mod1    49 -1361440 -1360898 680769 -1361538 126.26 23  3.069e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 1: The full regression model is statistically significant, compared to the null model which omits green energy projects.

4.2   Estimates

The fixed effects can be seen in figure 2 and figure 3, but they are listed below in their entirity. Note that any variable whose 95% CI includes 0.0 can plausibly be said to be negligible.

Do note that while this model gives quite sensible estimates, we suffer from a relatively low sample size for a model of this complexity. For this reason, this model was slightly rank-defficient. (One notable missing value is turbine manipulation for the year just before operation.)

Figure 2 shows the annual intercepts for overall home appreciation in the national market. Figure 3 shows what happens when solar and turbine projects are built in a zip-code. I interpret these in the following way:

- test

5   Appendix

How did the rate of change of home appreciation change for zip-codes that had solar or turbine projects in them?

I think this study is really easy to understand with an example. In 1979, Jimmy Carter put solar panels on his house, in zip-code 20500. If we wanted to see if this solar panel project impacted the rate of home appreciation in zip-code 20500, we'd do the following: Look at the year-over-year rate of home appreciation in 20500 in 1979, and for reference, include the previous few years (say, 1974-1979), and a few years after (say, 1979-1984). Let's define the rate of home appreciation here as:
(current home price - last year's home price) / (last year's home price).

Then we'd compare this to the median year-over-year rate of home appreciation of homes across the USA in zip-codes that didn't have any solar panels installed. We'd see two trend lines– that of our manipulation (zip-code 20500), and that of our control (all the other zip-codes without solar panels).

The problem with the above study is that our manipulation group has an n of 1, so it would be a good idea to expand our manipulation group to all other zip-codes in the USA that had solar panels installed in 1979.

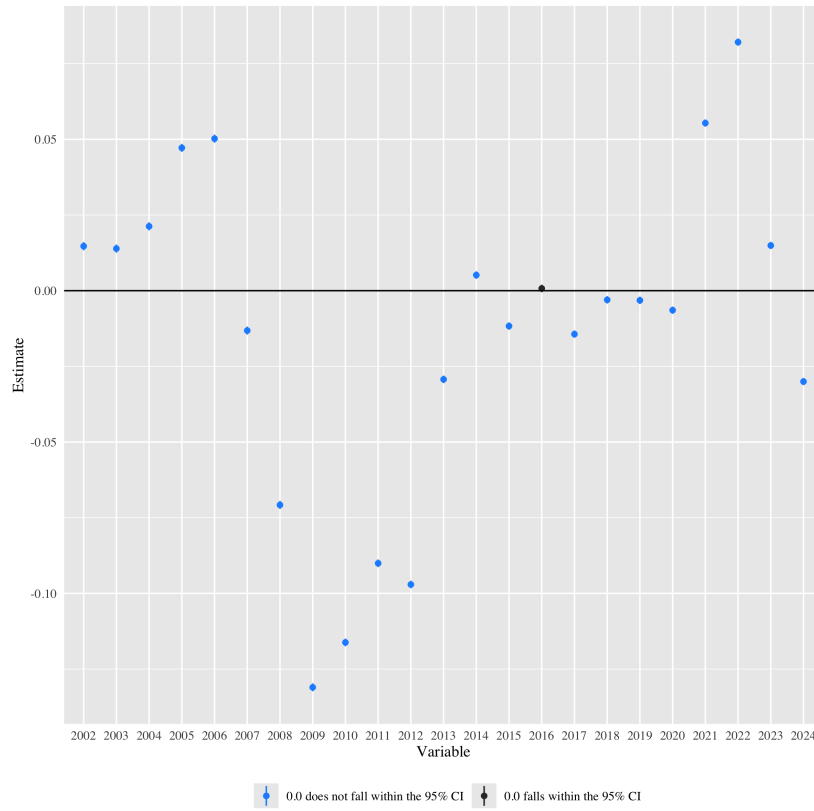This is the nature of study 1. The results can be seen in figure 4.

Figure 2: Some of the fixed effects from the second study.

Figure 4 provides no useful insight, because it doesn't account for general changes in the real estate market. For example, we can see the market tanking in 2008, but this completely confounds our ability to compare these lines to those a decade later.

Figure5 shows us the DiD for turbine facilities, but it suffers from the same shortcomings as figure4.

The other issue with study 1 is small sample size (see figure 6). The sample sizes for the manipulation groups pictured here are very small, only a fraction of the size of the control group. In addition to this problem, while the observations in the manipulation group change in each of the diagrams, the observations in the manipulation group changes completely. This means that individual outliers in the control group might be skewing our results.
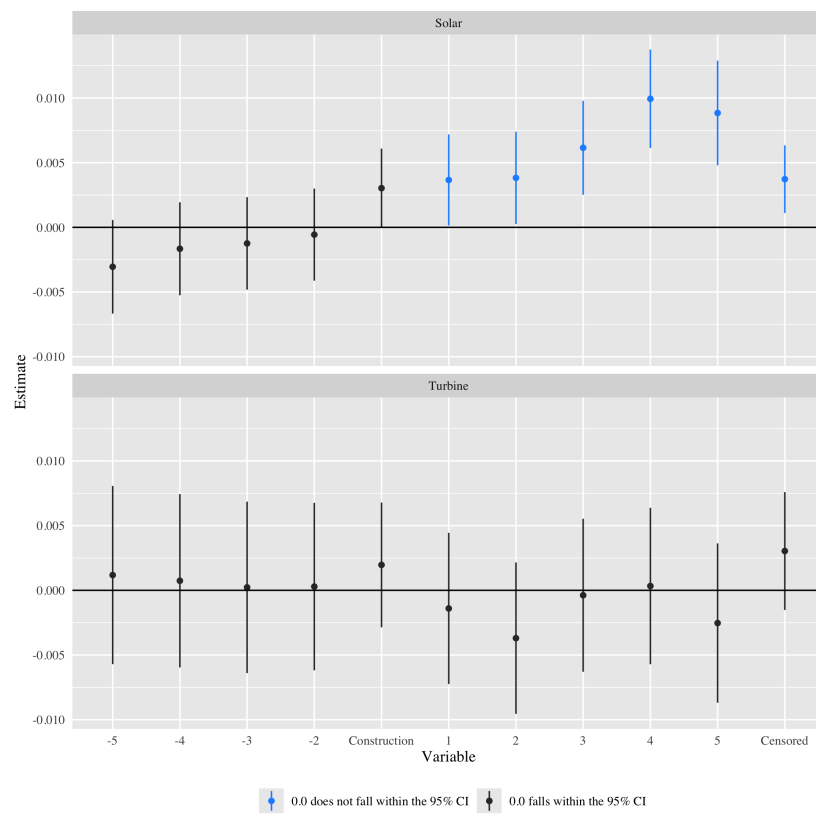
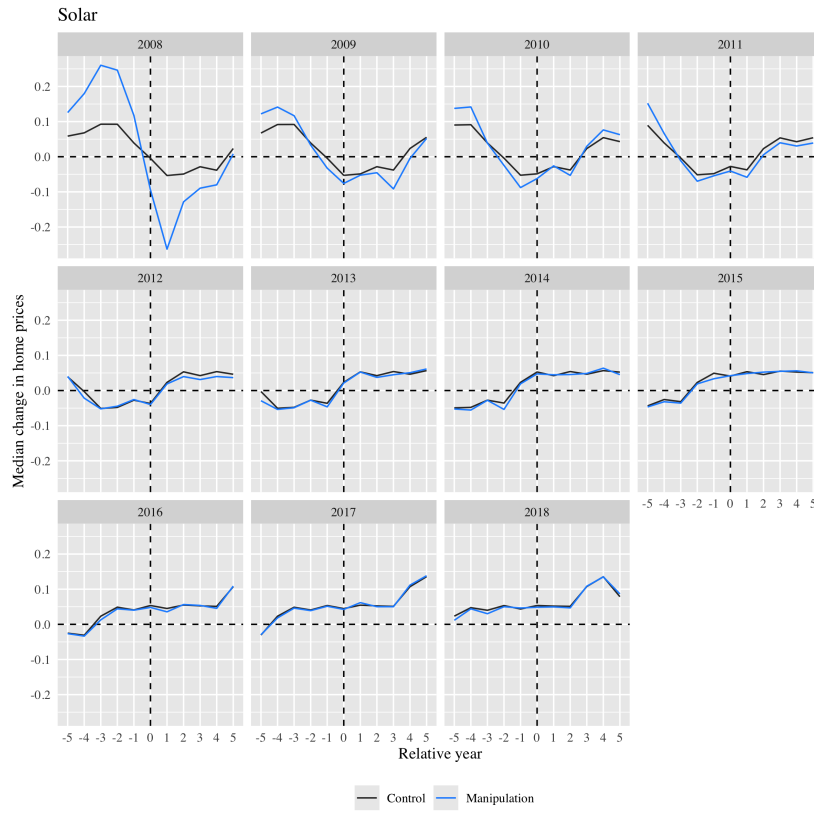Figure 3: The rest of the fixed effects from the second study.

Figure 4: Median change in home prices in zip-codes that had solar panels installed
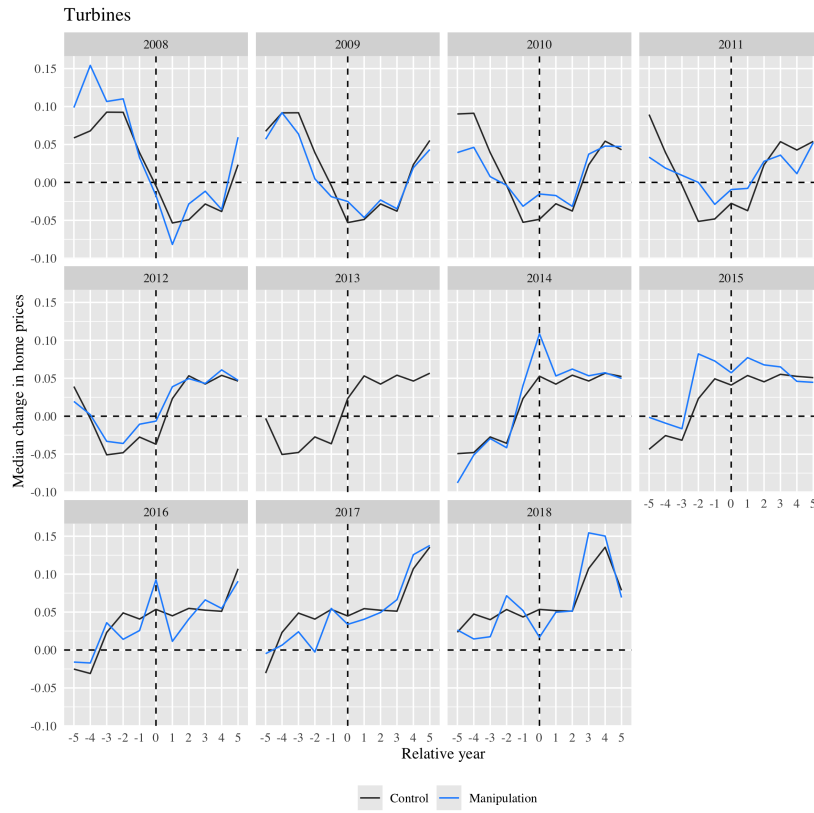
8

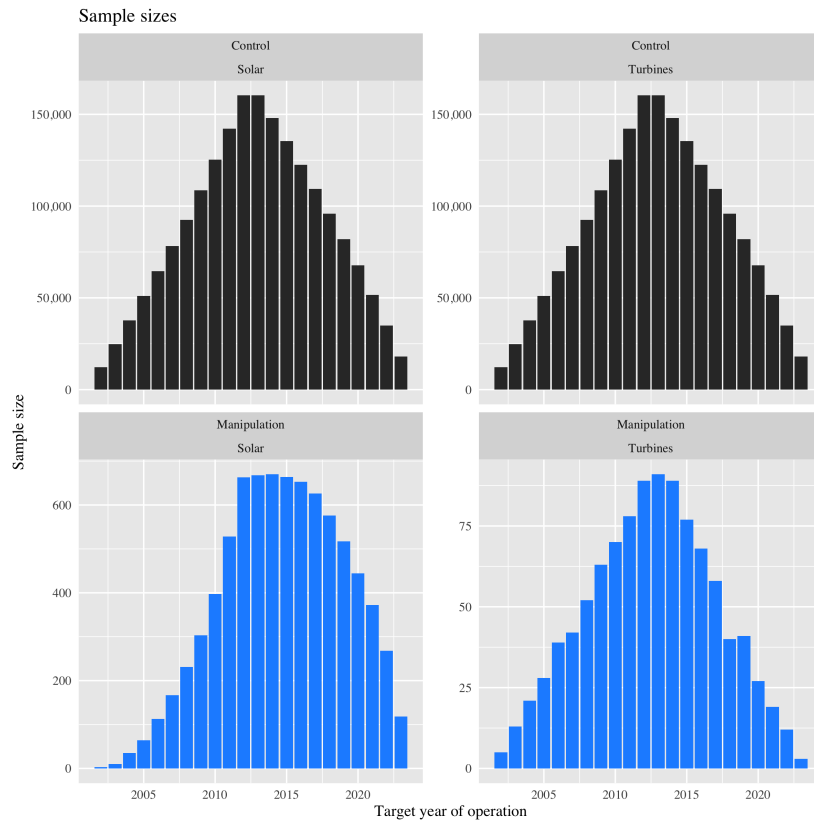Figure 5: Median change in home prices in zip-codes that had wind turbines installed

Figure 6: The manipulation groups in study 1 were extremely small