Model Approach Issues

Jamaal Green

10/19/2017

I am currently sketching out ideal model specifications for my modeling chapter. I am debating two approaches, currently, and collecting the basic data needed for both.

Buffer Approach

The approach most recently discussed is a buffer-based approach. Two examples include recent work from Aaronson, Hartley and Mazumder (2017) on the effects of redlining on long term segregation and poverty and some older work by Neumark and Kolko (2010) estimating the employment effects of enterprise zones. In both studies the researchers find immediately neighboring neighborhoods or census geographies and use them as comparisons to the treatment areas.

This approach allows for more flexibility on geographies (I would prefer to stay with block groups) but there are some validity concerns:

1. How zones are designated matters

- In the case of the EZ buffer approach Neumark and Kolko used they could be fairly confident that the areas immediately outside of the zones were similar to those inside. Also, specific census tracts were chosen for specific zones meaning some similar, and contiguous, zones were not selected. In fact, Neumark and Kolko use this in the paper as a second model weighing relative differences in job growth in tracts that were initially rejected and then entered into zones. The industrial districts are slightly different as they are primarily based on already existing allowable *uses* of the zone. It is unlikely, at least it seems unlikely to me, that a city would designate a new zone or overaly on top of an existing zone and leave significant portions of the surrounding area with the exact same use immediately around it. In this case, the surrounding areas may be fundamentally different than the treated areas such as the transition from primarily industrial to primarily residential.
- This may be more of an empirical question. I have yet to check to see if this is the case but it is a concern. Some of these industrial zones are pretty massive

2. Variable needs

- This is not unique to this approach but is more general concern. Using the LEHD it is easy to get data for however many time points as you wish, but it is more difficult using the census data and one is limited to particular geographies and variables if you wish to use them.

Propensity Score Approach

Originally, we discussed a propensity score/matching approach. Bill has used this technique a lot and can offer particular guidance on its idiosyncrasies. Lester's used the technique examining the impact of living wages laws (as seen here) and examining TIF impacts in Chicago here. In short, the propensity score is a two-staged modeling approach. The first estimates the probability a particular unit would be treated compared to the population as a whole. From there you can then weight your observations by that score in order to estimate the effect of the treatment weighted by the probability of unit was going to be treated in the first place. After scoring, treatment and control units are matched by certain criteria and this new sample is what is estimated using the final model.

Propensity scores have a now long history of use with Bill I've got a great resource on its idiosyncracies but there are some hurdles to consider:

1. Comparing over time

- While the LEHD is a longitudinal dataset other census data are decidedly not. In order to estimate a propensity score I would need to model the probability of a particular unit being treated. Ideally, this would use earlier data, in this case year 2000 census data. While NHGIS now has some longitudinal matching of census data from 1990 all the way 2010, the variables are limited. In particular, income data such as median household income and poverty data are not yet available at the block group level. They are available at the census tract level and allow for easy matching of the geographies. There are census geography crosswalk files but I have no real experience with them, but I am looking into them as a potential approach for linking block group income data.

Model specifications

Buffer Approach

1. Dependent variable

- change in industrial employment (logged or no?)

2. Independent variables

- Infrastructure/Landscape
  - distance to CBD
  - more central areas, especially if industrial, are probably more likely to face conversion pressures and I suspect a positive relationship between distance from CBD and change in employment
  - intersection density (from Smart Location DB, needs to be at block group level)
  - again, central city districts may face greater conversion pressure and other push factors that make it harder for industrial users to stay and so I would expect a negative relationship between increased street density and employment
  - distance from population weighted centroid transit stop (Smart Location DB)
  - densely populated areas with greater transit density offer potential conflicts with industrial users so I would expect a negative relationship

3. Land Use/Population

- Total pop
  - likely negative relationship
- % of population that are renters
  - unsure, but in areas where industry and residents coincide renters may be more present (this may be more appropriate for propensity approach)
- housing units (note, may be too correlated, but will check)
  - industry may more easily cohabit with multifamily buildings than single family... but more units mean more population and potentially more conflict. I suspect a negative relationship
- ratio of population to jobs (jobs using latest LEHD and pop from census)

4. Economic Mix

- % of jobs that are retail (LEHD)
  - an indicator of conversion. because it is LEHD it should be doable to get the change in retail jobs also. I expect a negative relationship
- Employment entropy score
  - this is SLD indicator measuring the *mix* of employment in particular block groups. I would suspect such a measure to be negatively associated

Propensity Approach

Propensity Score

1. Dependent variable

- probability a block group would receive treatment (industrial protection zone or overlay)

2. Independent variables

- Total population
  - expect this would be negatively associated with treatment
- percent non-white
  - an ambiguous relationship as neighborhoods with large proportions of working class people of color may be in areas or peripheral to industrial areas and there may actually be political support in such neighrborhoods for maintaining industrial land like in Chicago in the 80s.
- median household income
  - expect a negative relationship
- Industrial employment in 2002
  - taken from the LEHD, clearly cities would not preserve areas that were devoid of industrial workers. I expect a positive relationship

Final Model

Same variables from the first model.