Kai McNamee
Gov 52: Models
May 2, 2022

Contents



**Project Summary**

This project attempts to identify the impact freeways have on local communities by examining California's Interstate 105 as a case study. Opened in 1993, I-105 is the latest addition to Los Angeles's freeway system — the roughly 20 miles of highly trafficked road traverses a wide range of LA communities from east to west, cutting through lower-income, historically minority neighborhoods (ex: Watts, Compton) as well as comparatively affluent, whiter neighborhoods (ex: Downey, Norwalk).

Informed by a wealth of literature that documents how freeway construction has disproportionately displaced LA's communities of color, this analysis aims to uncover relationships between freeway construction, race, and economic welfare, as measured by Census population counts and records of median home values. Previous research and reporting suggest the routing of LA's freeway system has catered toward those with political influence: While community opposition was able to stop freeway construction in wealthier, whiter neighborhoods like Beverly Hills and South Pasadena, Boyle Heights, a majority Latino and low-income neighborhood, was carved up by one of the world's busiest freeway interchanges.

Findings on the impact of freeway construction on local economic indicators are mixed. In a 2015 study of freeway development in the Netherlands, for example, Or Levkovich et al find that freeway accessibility can generally increase home values, but freeways can also act as a disamenity whose noise and air pollution lower home values.

My analysis treats the opening of California's I-105 as a natural experiment, and I use demographic and economic data from the 1990 and 2000 Census in an effort to find potentially causal relationships between the opening of the freeway and trends in the surrounding communities' welfare. My results are generally inconclusive: I found no sigificant relationship

between home values and distance to the I-105, and changes in demographics and the availble housing stock were more salient predictors in my modeling.

**Data and Methods**

To examine the impact I-105's opening had on local communities, I used demographic and home value data from the 1990 and 2000 census summarized at the block group level, and geographic data from the City of LA (I-105 shapefiles), Google Maps (I-105 ramp locations), and the National Oceanic and Atmospheric Administration (CA coastline shapefiles). The main independent variable of interest, distance to the I-105, was calculated as the minimum distance between a Census block group centroids and the nearest point on the I-105. I repeated these calculations to generate distances between block group centroids and I-105 on/off ramps, as well as the distance to the coastline.

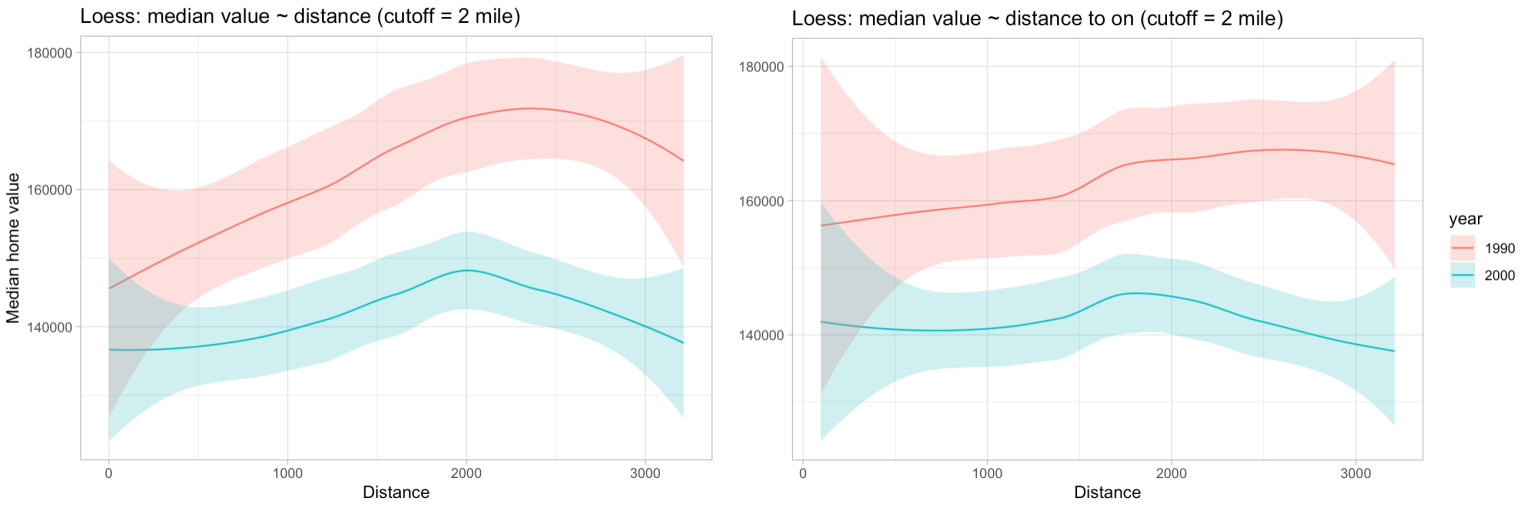
To account for shifting block group boundaries between the 1990 and 2000 censuses, I used the R package areal's areal interpolation functionality to conform 1990 data to 2000 boundaries. The resulting dataset contains a row for every 2000 block group, and columns for every demographic/economic variable in 1990 and 2000 for the specified geography.

| geoid.2000 | distance | distance_shoreline | off_east | off_west | on_east | on_west | population.1990 | population.2000 |
|---|---|---|---|---|---|---|---|---|
| 060371011101 | 36679.23 | 33545.17 | 36649.91 | 36645.02 | 36819.04 | 36704.14 | 1492 | 1773 |
| 060371012201 | 36122.20 | 33454.81 | 36135.39 | 36101.59 | 36226.93 | 36103.75 | 1261 | 1060 |
| 060371012202 | 35639.81 | 33092.52 | 35660.47 | 35619.56 | 35744.70 | 35621.75 | 1915 | 1968 |

Drawing from Linden and Rockoff's 2008 study of crime risk and property values, I used local polynomial regression to explore the relationships between median home value and distance to I-105 over both periods of observation to identify potential breakpoints in trends that could be identified as "treatment conditions" (ex: a block group would be considered treated if its centroid falls within 0.5 miles of I-105). I included several control variables like total demographic proportions, total housing units, demographic breakdowns by occupancy type (renter occupied housing vs. owner occupied housing), distance to the coast, and more (Linden and Rockoff; Diao et al; Hou; Keeler et al; Tita et al).

**Results**

Initial exploration of the data finds that median home value decrerases on average across LA county — block groups within 2 miles of I-105 also experience this trend. Plotting a local regression of median value on distance to I-105 yields no significant results: Binned by differing cutoff distances from I-105, trends in median home value vary by bin.




With a 2 mile cutoff, plotting median value and distance to I-105 (left) and distance to the nearest on-ramp (right) generated the most interesting results. For both, median home value increases with distance, but the trend tapers off between 1500 and 2000 meters. Filtering by different maximum distances dramatically changes the visualization, however, suggesting no clear cutoff effect of distance. Additionally, the parallel trends between 1990 and 2000 indicate there is likely no causal relationship between the opening of I-105 and changes in median home value between 1990 and 2000 — we would expect the 1990 data to show no trend between median value and distance, since distance would not be relevant prior to I-105 opening. Plotting changes in demographic breakdown by distance also generates no significant results, although at every bin tested, there is a positive correlation between distance and the percent White population.

Modeling the relationship between change in median value between 1990 and 2000 using linear regression is also inconclusive. After controlling for factors like changes in the number of occupied units, the percent population of racial groups, and the number of renter or owner occupied units, distance to I-105 (including distance to the nearest on/off ramps) had little to no significant effect on median home value (see Appendix: Table A). The 2 mile cutoff regression found the most significant results at the 95% and 99% confidence levels — the percent change in White population and White owner-occupied housing had the greatest significant positive impact on the change in median home value, and the percent change in Hispanic population had the greatest sigifnicant negative impact. The change in owner occupied units had a weak negative correlation with change in home value at the 90% confidence level. Further modeling finds little to no meaningful relationships when these variables are regressed as dependent variables (see Appendix: Table B).

Since the exploratory local regressions indicated a potential, but likely insignificant, treatment distance at which the increase in median home value by distance tapered off, I also tried

difference in difference models to further examine changes between the 1990 and 2000 data. I used the following model specification:

$$\text{Median home value} \sim \text{treatment} + \text{post} + \text{treatment} * \text{post}$$

where treatment is an indicator variable that returns TRUE if the block group falls within a specified treatament zone (1500m, 1750m, and 2000m for each model, respectively), and post is an indicator that returns TRUE if the block group is observed in 2000 and FALSE if observed in 1990. Like the exploratory plots, the data is filtered to block groups within a 2 mile buffer of I-105. The models fail to find a significant difference-in-difference coefficient, indicating there is no causal impact of I-105's opening on median home values within the treatment zones (see Appendix: Table C).

**Conclusions**

Moving forward, it might be interesting to use the I-105 as a natural experiment again, but keeping in mind the limitations of this analysis. First, the biggest takeaway in this analysis is that the trends between median home value and the distance independent variables in 1990 and 2000 data are mostly parallel — the opening of I-105 in 1993 did not change trajectories in home value or demographic proportions between the two periods of observation. This could be because the freeway had no impact on any of the dependent variables of interest, but it is also possible I'm looking at the wrong years: While the freeway opened in 1993, it's possible earlier planning or construction affected demographic and economic indicators before 1990. This analysis suggests changes in median home value are more correlated with demographic changes rather than the opening of I-105.

Furthermore, it might be interesting to analyze how the freeway impacted what types of buildings were constructed in the years following its opening. Looking at more recent ACS data might show how the volume of residential construction within x miles of the freeway has changed over time.

Overall, the part of this project I learned the most from was the data gathering process. I learned a lot about how to work with spatial data specifically in R, but I also learned about more general techniques for processing and interpolating spatial data.

## Appendix

Table A: Modeling change in median home value at different distance bins

| | 0.1mi | 0.25mi | 0.5mi | 1mi | 2mi |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| distance | 0.002 | -0.0001 | -0.0001 | -0.00002 | -0.00002 |
| | | (0.0002) | (0.0001) | (0.00003) | (0.00002) |
| distance_shoreline | 0.0001 | 0.00001 | 0.00001 | 0.00001*** | 0.00001*** |
| | | (0.00001) | (0.00000) | (0.00000) | (0.00000) |
| on_east | 0.0002 | -0.0001 | -0.00004 | -0.00003 | -0.00001 |
| | | (0.0001) | (0.00003) | (0.00003) | (0.00002) |
| on_west | -0.002 | -0.0003* | -0.00001 | 0.00004* | 0.00003* |
| | | (0.0001) | (0.00003) | (0.00002) | (0.00001) |
| off_east | 0.002 | 0.0003* | -0.00002 | -0.00004 | -0.00002 |
| | | (0.0001) | (0.0001) | (0.00004) | (0.00003) |
| off_west | -0.0005 | 0.0001 | 0.0001 | 0.00005 | 0.00003 |
| | | (0.0001) | (0.00005) | (0.00004) | (0.00003) |
| change_occupied | 1.086 | -0.020 | 0.008 | -0.047 | -0.042 |
| | | (0.066) | (0.062) | (0.048) | (0.035) |
| change_owner_occupied | 4.939 | -0.089 | 0.389*** | 0.166* | -0.096* |
| | | (0.209) | (0.121) | (0.086) | (0.052) |
| change_renter_occupied | 2.374 | -0.142 | 0.195 | 0.125 | -0.002 |
| | | (0.184) | (0.138) | (0.087) | (0.059) |
| change_white | 1.595 | 0.119 | 0.086 | 0.095** | 0.082*** |
| | | (0.107) | (0.070) | (0.040) | (0.025) |
| change_black | 0.239 | 0.047 | 0.022 | 0.030 | 0.001 |
| | | (0.058) | (0.034) | (0.029) | (0.017) |
| change_hispanic | 0.065 | 0.015 | -0.050 | -0.072** | -0.072*** |
| | | (0.080) | (0.054) | (0.036) | (0.023) |
| change_asian | 0.293 | 0.036 | 0.054 | 0.048 | -0.006 |
| | | (0.054) | (0.048) | (0.037) | (0.023) |
| change_white_owner | -0.292 | -0.022 | 0.026 | 0.055 | 0.044** |
| | | (0.115) | (0.054) | (0.035) | (0.020) |
| change_black_owner | 0.009 | -0.003 | -0.006 | -0.014 | -0.006 |
| | | (0.011) | (0.011) | (0.009) | (0.004) |
| change_hispanic_owner | -0.072 | -0.046 | -0.007 | -0.001 | 0.015 |
| | | (0.045) | (0.032) | (0.021) | (0.012) |
| change_asian_owner | -1.506 | -0.013 | 0.002 | -0.027 | -0.015 |
| | | (0.039) | (0.032) | (0.026) | (0.016) |
| change_white_renter | | 0.071 | 0.050 | 0.003 | 0.008 |
| | | (0.063) | (0.051) | (0.020) | (0.014) |
| change_black_renter | | -0.052 | -0.040 | -0.031 | -0.005 |
| | | (0.052) | (0.031) | (0.024) | (0.014) |
| Constant | -1.745 | -0.162 | -0.105 | -0.162*** | -0.204*** |
| | | (0.097) | (0.073) | (0.060) | (0.039) |
| Observations | 18 | 47 | 103 | 226 | 454 |
| R2 | 1.000 | 0.660 | 0.490 | 0.330 | 0.333 |
| Adjusted R2 | | 0.421 | 0.374 | 0.269 | 0.304 |
| Residual Std. Error | | 0.108 (df = 27) | 0.138 (df = 83) | 0.164 (df = 206) | 0.156 (df = 434) |
| F Statistic | | 2.760*** (df = 19; 27) | 4.201*** (df = 19; 83) | 5.349*** (df = 19; 206) | 11.406*** (df = 19; 434) |
| Note: | | | | | *p<0.1; **p<0.05; ***p<0.01 |

Table B: Additional exploratory modeling

| | Dependent variable: | | | | |
|---|---|---|---|---|---|
| | change_hispanic | change_white | change_owner_occupied | change_white_owner | change_owner_occupied |
| | (1) | (2) | (3) | (4) | (5) |
| distance | 0.00000 | 0.00001 | -0.00000 | 0.0001* | -0.00000 |
| | (0.00002) | (0.00003) | (0.00001) | (0.00003) | (0.00001) |
| change_occupied | -0.139* | | | -0.178 | |
| | (0.082) | | | (0.130) | |
| change_owner_occupied | -0.149 | | | 1.061*** | |
| | (0.108) | | | (0.172) | |
| change_renter_occupied | 0.096 | | | 0.429** | |
| | (0.118) | | | (0.187) | |
| change_value | -0.251*** | | | | |
| | (0.095) | | | | |
| Constant | 0.367*** | 0.137** | 0.021 | -0.030 | 0.021 |
| | (0.037) | (0.064) | (0.022) | (0.058) | (0.022) |
| Observations | 454 | 454 | 454 | 454 | 454 |
| R2 | 0.028 | 0.0003 | 0.0002 | 0.118 | 0.0002 |
| Adjusted R2 | 0.017 | -0.002 | -0.002 | 0.110 | -0.002 |
| Residual Std. Error | 0.376 (df = 448) | 0.671 (df = 452) | 0.232 (df = 452) | 0.598 (df = 449) | 0.232 (df = 452) |
| F Statistic | 2.551** (df = 5; 448) | 0.154 (df = 1; 452) | 0.068 (df = 1; 452) | 15.009*** (df = 4; 449) | 0.068 (df = 1; 452) |
| Note: | | | | | *p<0.1; **p<0.05; ***p<0.01 |

Table C: Modeling median home value with difference in difference design, defining different treatment zones with a cutoff of 2 miles.

| | Dependent variable: | | |
|---|---|---|---|
| | value | | |
| | 1500m | 1750m | 2000m |
| | (1) | (2) | (3) |
| treatment | -10,037.120*** | -7,304.676* | -9,774.522** |
| | (3,818.853) | (3,816.725) | (3,884.843) |
| post | -23,445.770*** | -23,256.880*** | -24,785.940*** |
| | (3,655.541) | (3,891.651) | (4,252.506) |
| treatmentTRUE:post | 6,084.075 | 4,998.868 | 6,889.434 |
| | (5,400.674) | (5,397.664) | (5,493.997) |
| Constant | 164,134.400*** | 163,333.000*** | 165,392.000*** |
| | (2,584.858) | (2,751.813) | (3,006.976) |
| Observations | 908 | 908 | 908 |
| R2 | 0.069 | 0.065 | 0.068 |
| Adjusted R2 | 0.066 | 0.062 | 0.065 |
| Residual Std. Error (df = 904) | 40,541.910 | 40,630.030 | 40,566.330 |
| F Statistic (df = 3; 904) | 22.306*** | 20.904*** | 21.917*** |
| Note: | | | *p<0.1; **p<0.05; ***p<0.01 |

**Documentation**

read-1.2.Rmd

- Used for reading and cleaning data
- Data inputs
    - Shape data
        - I-105 shapefiles from the LA Master Plan of Highways
        - 1990 census block group shapes from the Census Bureau website
        - 2000 census block group shapes via tidycensus::get_decennial
        - CA coastline shapefiles from NOAA
        - I-105 ramp coordinates from Google Maps
    - Census data
        - 1990 census block group data from the Census Bureau
        - 2000 census block group data via tidycensus::get_decennial
    - Note: The input data required to run read-1.2.Rmd was too big for GitHub, but the raw census and geospatial data can be found on Google Drive here. File paths should all work if the linked folder is in your working directory.
- Data outputs
    - i105.RDS: R data file containing shape data for only I-105
    - la_shoreline.RDS: R data file containing shape data for CA coastline clipped to LA county boundaries
    - i105_ramps.RDS: R data file containing shape data for all I-105 on/off ramps
    - c1990_bg_out.RDS: cleaned 1990 census data filtered to variables of interest
    - c2000_bg_out.RDS: cleaned 2000 census data filtered to variables of interest and combined with geospatial analysis (distances to I-105, ramps, shorelines, etc.)

align-1.3.Rmd

- Used for aligning 1990 and 2000 census data. Includes areal interpolation and summed variables (ex: 1990 census combines Asian and Pacific Islander counts while 2000 census splits them)
- Data inputs
    - Shape data
        - i105.RDS
        - i105_ramps.RDS
    - Census data
        - c1990_bg_out.RDS
        - c2000_bg_out.RDS
- Data outputs

    - c1990_interpolated.RDS: R data file containing results of areal interpolation using 1990 source data and 2000 target geometries cleaned to match corresponding data in 2000 census
    - c2000_cleaned.RDS: R data file cleaned to match corresponding data in 1990 census

models-1.1.Rmd

- Used for all modeling
- Data inputs
    - c1990_interpolated.RDS
    - c2000_clean.RDS
- Data outputs
    - model_data.RDS: 1990 and 2000 block group data combined with geospatial analysis. Every row includes data for a 2000 block group boundary, with columns storing variables recorded in 1990 and 2000 in that block group (ex: population.1990 and population.2000 in block group 060371011101)