Highway Through Property Values: The impact of freeway removal on local real estate

Brendan J Semien

Professor Ida Johnsson

University of Southern California - ECON 570

6 May 2022

Introduction

Large, urban freeways are attributed with assisting the economic development of cities across the United States; they are also associated with incredible human and environmental violations. Easy access in urban areas undoubtedly leads to economic and city growth, allowing for easy access for commuters on the way to work in downtown areas or for quick distribution of resources across densely populated terrains. This same reconstitution of land leads to environmental destructions with the paving of mass areas previously with more greenspace and more porous for rainfall. Further, the increased concentration of air pollutants from modern transportation methods exacerbates environmental destruction. This said, the most notable detriment of urban freeways is the displacement of people, who disproportionately tend to be impoverished and or minority populations. However, the simple displacement of people is not the only impact for these low income communities of color. The impact of these freeways only compounds over time; being a neighborhood associated with a freeway decreases property values given the pure association as well as the environmental factors like the aforementioned air pollutants and noise pollution.

Endless examples exist of the exact phenomena described above, especially in 'concrete jungles', cities like Los Angeles or New York City. A Los Angeles Times article in 2018 describes the impact of the most recent Los Angeles freeway, I 105. Running east to west from Lynwood to the PCH, the freeway directly displaced 25,000 people and ran directly through some of Los Angeles' most impoverished areas (Sahagun). The article also describes a proposal for another large freeway from Victorville to Palmdale, speaking to the urgency of information against this expansion.

The creation of major freeways has stalled in recent years, this has created a historical data limitation. Therefore, this study will not explore the creation of a freeway to gauge the economic impact of freeway construction but will instead focus on the destruction of major urban freeways. This study aims to provide an economic basis and incentive for the destruction of irrelevant and pollutant creating freeways as well as providing a rationale against the further development of urban freeways. The study will use a difference in difference model to examine the effect of the deconstruction of the Sheridan Expressway in downtown New York City on local housing prices. The closest neighborhood in which we will measure the most direct effect is Tremont. We will be answering the question: What is the effect of the deconstruction of a freeway on local neighborhood housing prices?

Data

My data originates from Zillow.com. It came in a csv format with over 15 thousand observations. I moved it to excel and eliminated all observations outside of New York City. I started with variables like neighborhood name, county, state and temporal data on average house price from 2000 to March 2022; data was monthly. I identified the Tremont neighborhood as the closest to the Sheridan Expressway removal project and thus made it my tremont group. I then attempted to find the most similar other New York neighborhood that could be used for comparison in the dif in dif. It was important to identify a neighborhood that had similar demographics and trending housing prices without having the possibility of a spillover effect. I identified the Wakefield neighborhood as this similar neighborhood. I used the 2000 to 2010 period to calculate a simple slope and compare Tremont to all other NY neighborhoods. I later used the correlation command in my jupyter notebook to confirm this correlation. I narrowed the time frame to from 2014 to 2022 allowing for proper time before and after the event. Through research I found that the project concluded on December 2nd 2019 and I would later find that importantly, the deconstruction begin date was October of 2018. Using these dates I created the required dummy variables for a difference in difference model: time, treatment and time*treatment (or DID coefficient). I also used a control variable to attempt to account for the general housing trend. This was the New York Housing Price Index which provides a general assessment of the entirety of the New York housing market. In our regression this will help determine whether our results were actually due to the housing market as a whole or the freeway removal. I attempted to use other demographic control variables like income, but was unable to find specific data for the Tremont neighborhood and thus could not include this nor other variables.

Modeling

My model was a difference in difference model to assess the treatment effect of the Sheridan Expressway deconstruction on local housing prices. The model is as follows

$$\text{Housing Prices} = a + x(\text{Time}) + \gamma\,(\text{Treatment}) + \lambda(\text{Time*Treatment (DID)}) + z(\text{NYHPI}) + e$$

$$\ln(\text{Housing Prices}) = a + x(\text{Time}) + \gamma\,(\text{Treatment}) + \lambda(\text{Time*Treatment (DID)}) + z(\text{NYHPI}) + e$$

The model specification that I performed was that after plotting the housing prices over time I realized that the impactful date was in fact not the end date and rather was the start date of the destruction. As such, I created a second DID model with time dummy variables that marked October 2018 as the treatment date. I also used a log transformation for a variety of results.

Findings

I had a variety of findings, the most important of which is the difference in difference coefficient as displayed in the following regression results. However, I will first reveal some of the more exploratory, summary statistics and visual results.

```
In [13]: TremontData.describe()
Out[13]:
```

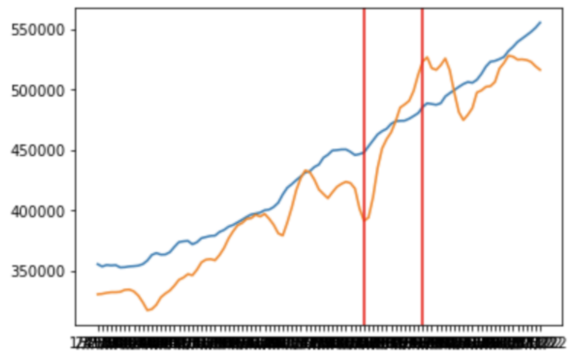
	Y	time	treated	DID	NYHPI
count	99.000000	99.000000	99.0	99.000000	99.000000
mean	421262.656566	0.282828	1.0	0.282828	671.732424
std	69981.527087	0.452666	0.0	0.452666	84.141121
min	317080.000000	0.000000	1.0	0.000000	560.890000
25%	358705.500000	0.000000	1.0	0.000000	597.550000
50%	413390.000000	0.000000	1.0	0.000000	658.620000
75%	493920.500000	1.000000	1.0	1.000000	719.080000
max	527860.000000	1.000000	1.0	1.000000	857.970000

```
In [15]: WakefieldData.describe()
Out[15]:
```

	Y	time	treated	DID	NYHPI
count	99.000000	99.000000	99.0	99.0	99.000000
mean	436543.414141	0.282828	0.0	0.0	671.732424
std	61530.062450	0.452666	0.0	0.0	84.141121
min	352504.000000	0.000000	0.0	0.0	560.890000
25%	378179.000000	0.000000	0.0	0.0	597.550000
50%	437693.000000	0.000000	0.0	0.0	658.620000
75%	487525.500000	1.000000	0.0	0.0	719.080000
max	555298.000000	1.000000	0.0	0.0	857.970000

The above two tables depict the summary statistics of Tremont and Wakefield neighborhoods in New York. We see that the mean, min and max average housing price of Treamont of the entire study period is marginally lower than that of Wakefield. We also know that all of Tremont is in the treatment group and that 28% of our observation period is after the treatment event of December 2019.

```
In [45]: plt.plot(Date, HousP_W)
         plt.plot(Date, HousP_T)
         plt.axvline(x= 72, color = 'r')
         plt.axvline(x= 59, color = 'r')
         plt.show()
```



This plot is both the house prices of Wakefield in blue and Tremont in orange across time as well as the beginning of reconstruction and ending of construction. While my inclination was that the end of the reconstruction period would be the event that sparked an increase in the rate of increase of housing prices, this graph would suggest that is not the case. When I saw my plot of the end date, I was drawn to implement the start date in the graph and a regression as well. Thus, I ran two regressions as follows.

```
t/anaconda3/lib/python3.9/site-packages (0.0.5)
```

In [53]:
```python
DID01 = smf.ols(formula="Y ~ time + treated + DID + NYHPI", data=dat
print(DID01.summary())
```

```
                            OLS Regression Results
==================================================================
===========
Dep. Variable:                      Y   R-squared:
0.908
Model:                            OLS   Adj. R-squared:
0.906
Method:                 Least Squares   F-statistic:
473.8
Date:                Thu, 05 May 2022   Prob (F-statistic):
1.40e-98
Time:                        17:03:38   Log-Likelihood:
-2242.5
No. Observations:                 198   AIC:
4495.
Df Residuals:                     193   BIC:
4511.
Df Model:                           4
Covariance Type:            nonrobust
==================================================================
===========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------
-----------
Intercept  -3.437e+04   1.93e+04     -1.784      0.076   -7.24e+04
3638.214
time        1732.3669   6500.849      0.266      0.790   -1.11e+04
1.46e+04
treated    -1.984e+04   3411.008     -5.815      0.000   -2.66e+04
-1.31e+04
DID          1.61e+04   6413.887      2.511      0.013    3454.029
2.88e+04
NYHPI        700.3105     30.422     23.020      0.000     640.309
760.312
==================================================================
===========
Omnibus:                       11.595   Durbin-Watson:
0.127
Prob(Omnibus):                  0.003   Jarque-Bera (JB):
12.459
Skew:                           0.506   Prob(JB):
0.00197
Kurtosis:                       3.698   Cond. No.
9.27e+03
==================================================================
===========
```

```
In [73]: DIDLn01 = smf.ols(formula="lnY ~ time + treated + DID + NYHPI", data
         print(DIDLn01.summary())

                            OLS Regression Results
==================================================================
===========
Dep. Variable:                    lnY   R-squared:
0.888
Model:                            OLS   Adj. R-squared:
0.885
Method:                 Least Squares   F-statistic:
381.2
Date:                Thu, 05 May 2022   Prob (F-statistic):
2.07e-90
Time:                        23:59:14   Log-Likelihood:
304.97
No. Observations:                 198   AIC:
-599.9
Df Residuals:                     193   BIC:
-583.5
Df Model:                           4
Covariance Type:            nonrobust
==================================================================
===========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------
-----------
Intercept     11.8384      0.050    237.692      0.000      11.740
11.937
time          -0.0187      0.017     -1.114      0.267      -0.052
0.014
treated       -0.0525      0.009     -5.950      0.000      -0.070
-0.035
DID            0.0456      0.017      2.752      0.006       0.013
0.078
NYHPI          0.0017   7.86e-05     21.654      0.000       0.002
0.002
==================================================================
===========
Omnibus:                        1.414   Durbin-Watson:
0.108
Prob(Omnibus):                  0.493   Jarque-Bera (JB):
1.191
Skew:                           0.186   Prob(JB):
0.551
Kurtosis:                       3.072   Cond. No.
9.27e+03
==================================================================
===========
```

This first regression is our difference in difference model as described above this time with the date of treatment as December of 2019, the end of the reconstruction. As we can see we have a great regression. Our R-squared is .906 which reveals our model accounts for 90% of the variation in the data points. Further we see that our Dif in Dif coefficient (DID) is significant at a 95% confidence level and almost at the 99%. Our DID coefficient is 16,100. This means that the treatment effect of the Sheridan Expressway is $16,100 or that the reconstruction of the expressway led to a 16,100 dollar increase in housing prices locally.

The second regression was of the log transformation of our housing prices. This gives us our results in a percentage form. Thus, we find that the treatment effect of the reconstruction led to a 4.56% increase on average and is highly significant.

In [57]:
```
DID_START = smf.ols(formula="Y ~ time + treated + DID + NYHPI", data
print(DID_START.summary())
```

```
                            OLS Regression Results
==============================================================================
===========
Dep. Variable:                      Y   R-squared:
0.918
Model:                            OLS   Adj. R-squared:
0.917
Method:                 Least Squares   F-statistic:
542.1
Date:                Thu, 05 May 2022   Prob (F-statistic):
9.95e-104
Time:                        17:24:11   Log-Likelihood:
-2230.3
No. Observations:                 198   AIC:
4471.
Df Residuals:                     193   BIC:
4487.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
===========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept   1.613e+04   1.78e+04      0.908      0.365   -1.89e+04
5.11e+04
time        2.058e+04   5583.678      3.686      0.000    9566.830
3.16e+04
treated    -2.061e+04   3579.965     -5.757      0.000   -2.77e+04
-1.35e+04
DID         1.256e+04   5496.317      2.285      0.023    1720.854
2.34e+04
NYHPI        612.8703     28.695     21.358      0.000     556.274
669.467
==============================================================================
===========
Omnibus:                        6.927   Durbin-Watson:
0.149
Prob(Omnibus):                  0.031   Jarque-Bera (JB):
8.101
Skew:                           0.280   Prob(JB):
0.0174
Kurtosis:                       3.818   Cond. No.
9.04e+03
==============================================================================
===========
```

In [76]:
```
DID_STARTln = smf.ols(formula="lnY ~ time + treated + DID + NYHPI",
print(DID_STARTln.summary())
```

```
                            OLS Regression Results
==============================================================================
===========
Dep. Variable:                    lnY   R-squared:
0.899
Model:                            OLS   Adj. R-squared:
0.897
Method:                 Least Squares   F-statistic:
431.3
Date:                Fri, 06 May 2022   Prob (F-statistic):
5.00e-95
Time:                        00:00:56   Log-Likelihood:
315.89
No. Observations:                 198   AIC:
-621.8
Df Residuals:                     193   BIC:
-605.3
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
===========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept     12.0055      0.046    259.950      0.000      11.914
12.097
time           0.0432      0.015      2.974      0.003       0.015
0.072
treated       -0.0556      0.009     -5.966      0.000      -0.074
-0.037
DID            0.0377      0.014      2.639      0.009       0.010
0.066
NYHPI          0.0014   7.46e-05     19.008      0.000       0.001
0.002
==============================================================================
===========
Omnibus:                        0.511   Durbin-Watson:
0.118
Prob(Omnibus):                  0.774   Jarque-Bera (JB):
0.650
Skew:                           0.063   Prob(JB):
0.723
Kurtosis:                       2.749   Cond. No.
9.04e+03
==============================================================================
===========
```

Our secondary regression provides similar results but this time at a more significant level. Our R squared increased and our AIC also decreased which means our regression explains more of the variation and is a better model. We see, however, that our coefficients decrease in both instances thus revealing that our first model is likely over estimating the coefficient.

In order to count our regressions as the best linear unbiased estimator, we must prove the standard errors are robust. I used the .bse command and received the following results giving me the standard errors of the following parameters. We see that these values are rather small and thus we lack great outliers and thus we can call the model robust.

```
In [77]: DID_START.bse

Out[77]: Intercept    17756.075212
         time          5583.678178
         treated       3579.965207
         DID           5496.316814
         NYHPI           28.695239
         dtype: float64
```

Conclusion

The above regression results are statistically significant and positive leading to an answer to our research question. Yes, the reconstruction of a freeway does increase the housing prices of the local area. However, our coefficient in perspective, is rather small. However, I believe that if our study was of a more grand freeway and in the creation of the freeway, rather than its destruction, our results would be far bigger in scale. We learned a small amount regarding our question, but we also learned that there is a need for further research, modeling and data exploration. I believe with the proper data to test impactful freeways like that of the aforementioned 105 freeway, our results would be far more eye-catching and revealing. I had a number of additional limitations. I was unable to have the constants that I desired including numerous demographic statistics, most notably income. These would have ensured that our results are meaningful and not misattributions from the error term.

Work Cited

Sahagun, Louis. "L.A. County Set to Build Its First New Freeway in 25 Years, despite Many Misgivings." *Los Angeles Times*, Los Angeles Times, 10 Feb. 2018, https://www.latimes.com/local/california/la-me-high-desert-freeway-20180210-htmlstory.html.

Zillow. "Housing Data." *Zillow Research*, 25 Mar. 2021, https://www.zillow.com/research/data/.