Does the introduction of the Climate Innovation Hub affect Housing affordability In Sunset Park?

In March 2024, the New York City Economic Development Corporation (NYCEDC) issued a Request for Proposal (RFP) for the consortium that will operate the new Climate Innovation Hub (CIH) at the Brooklyn Army Terminal, part of the Sunset Park Climate District. Following the release of the RFP, residents and local activists raised concerns about whether new investments in the neighborhood would affect its affordability. Thus, the project's guiding research question could be formulated as follows: Did the introduction of Climate Innovation Hubs in Sunset Park cause statistically significant changes in rental prices in Sunset Park compared to similar neighborhoods without such interventions (H1)?

To analyze it, I use a Difference in Difference (DiD) design for regression analysis based on the Zillow Observed Rent Index (ZORI) and American Community Survey data. First, I upload and clean the ZORI data, checking for missing values, creating the necessary variables, and filtering for the zip codes used in the study (Sunset Park, Greenpoint, and Bushwick as a control group). I then create the treatment, post, and treatment post variables needed for the DiD analysis. Then, I collect the ACS data using the API and the tidy census library. Similar to the ZORI data, I filter it and compute variables such as the percentages of racial groups, the unemployment rate, and the vacancy rate, in addition to other variables already collected. Next, since the ZORI data is not adjusted for inflation, I collected the Consumer Price Index (CPI) using the "blscrapeR" library, adjusted the ZORI variable, and merged it with the ASC variables to create a final dataset (Table 1).

I then perform explanatory data analysis using the "ggplot" library. To do this, I create graphs of histograms of numerical variables (Fig. 1), rent trend by zip code (Fig. 2), and, most importantly, plot adjusted ZORI distribution (Fig. 3) and ZORI distribution by zip code (Fig. 4), which shows that the distribution of the dependent variable is skewed. Therefore, I apply log transformation to the data, after which the distribution appears more normal (Fig. 5). In addition, this is supported by the Q-Q plot (Fig. 6), which shows that the distribution is close to normal, although the heavier tails of the distribution that deviate from normal are also present. I then establish the need for a regression model using scatter plots comparing the dependent variable to the independent variable (Fig. 7), which shows that there is a need to move to a regression model. Finally, I check the parallel trends in the adjusted ZORI data (Fig. 8), which shows that there are changes after treatment, thus justifying the DiD design of the study.

Before running the regression model, I checked the variance and correlation between the independent variables to account for potential multicollinearity (Table 2). The table shows that some of the variables, such as the percentage of the population that is white and the unemployment rate, have a high correlation, so we need to drop them to avoid multicollinearity. Finally, I scale the data and run the regression model. The summary (Table 3) shows a high value for R squared (0.917), a good fit of the model based on the residual standard error (0.052), as well as the histogram of residuals (Fig.) and the overall statistical significance of the model (F-statistic = 557, $p < 0.001$). Based on the coefficient, we see that some of them affect the price, but the treatmentpost variable is not statistically significant. In addition, I ran the fixed effects model to account for zip code fixed effects, which shows that all covariates are fixed effects of zip codes. The summary (Table 4) shows a positive and statistically significant coefficient on post in all neighborhoods, meaning that rents increased by about 6.1% after the RFP was announced. However, since the treatment coefficient is not statistically significant, it likely reflects citywide or macroeconomic trends rather than anything specific to the introduction of the CIH. Therefore, the CIH does not appear to have a significant short-term impact on rental prices, at least within the time frame and covariates of the model presented.

Data & Code

All of the code and data for the project are available in the GitHub Repository:
https://github.com/temapankin/sunset-park-regression

Appendix

Figure 1. Histograms of numeric variables

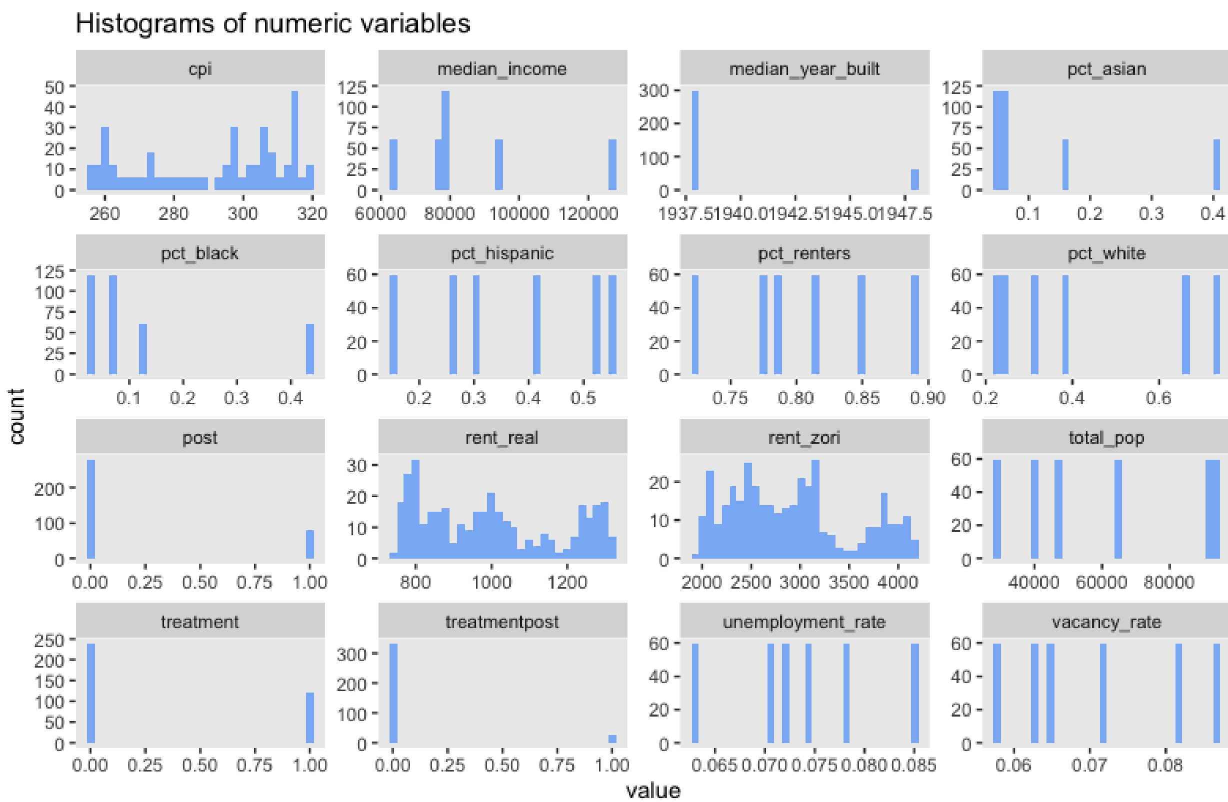

Figure 2. Rent Trends by ZIP code

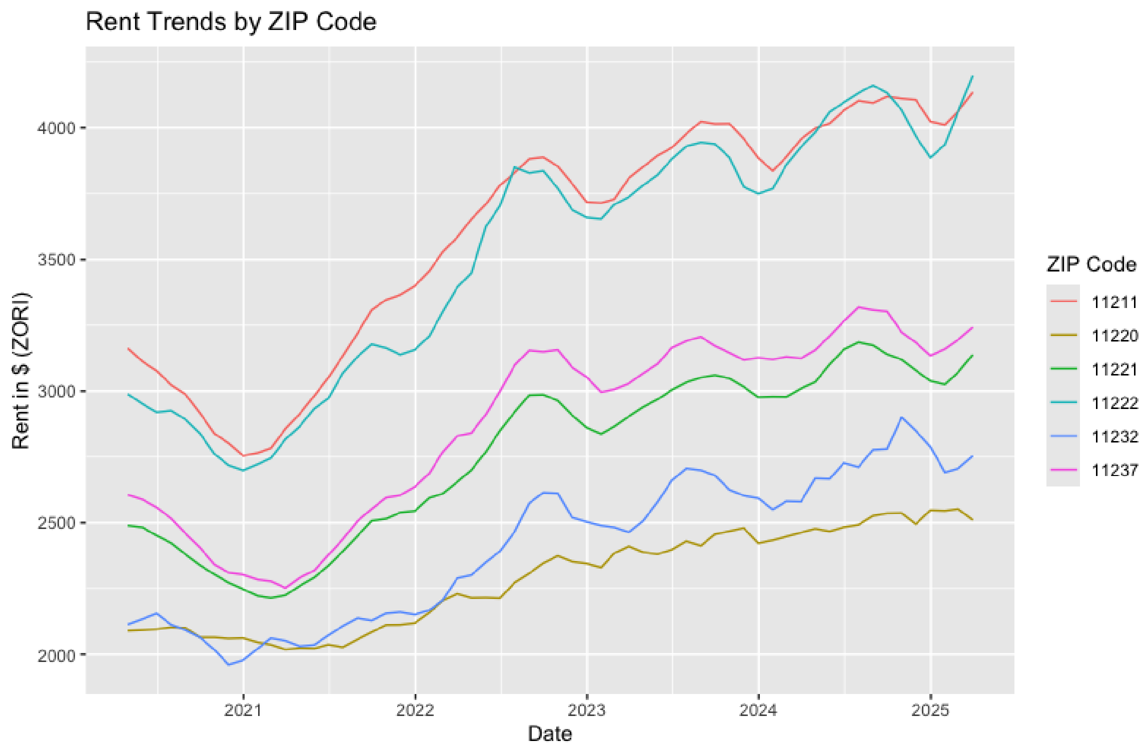

Figure 3. Distribution of adjusted ZORI variable

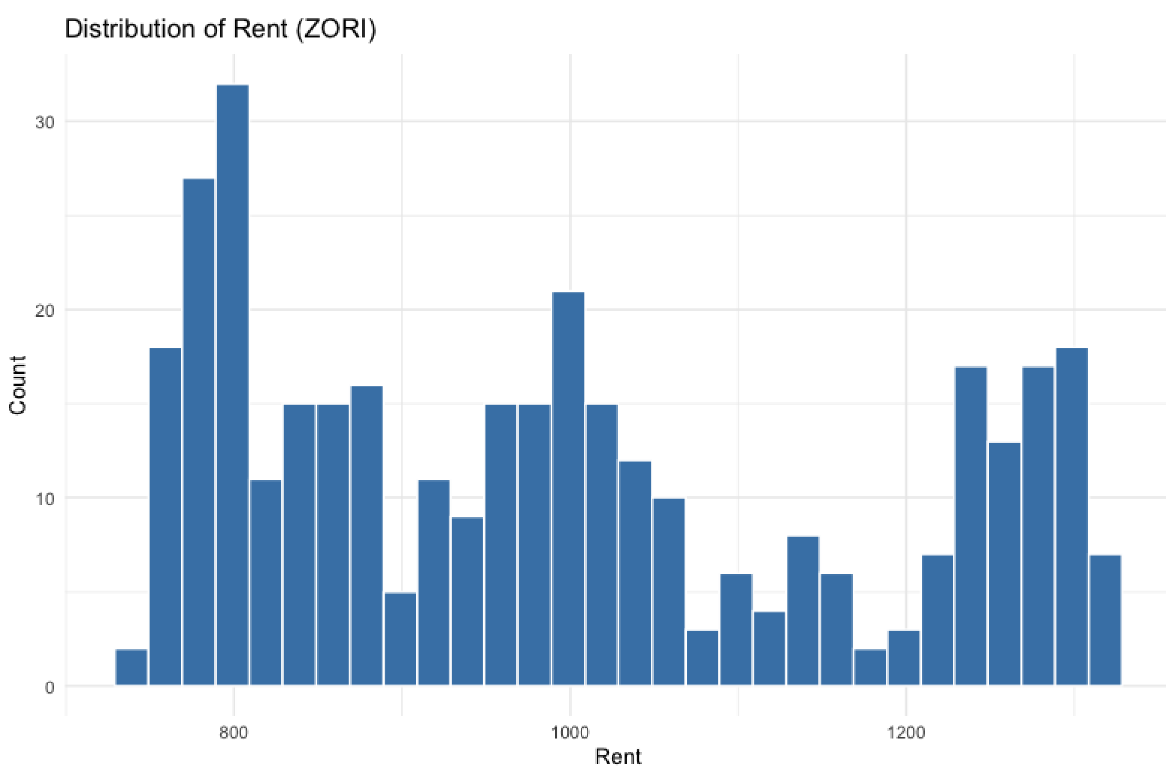

Figure 4. Distribution of adjusted ZORI variable by ZIP code

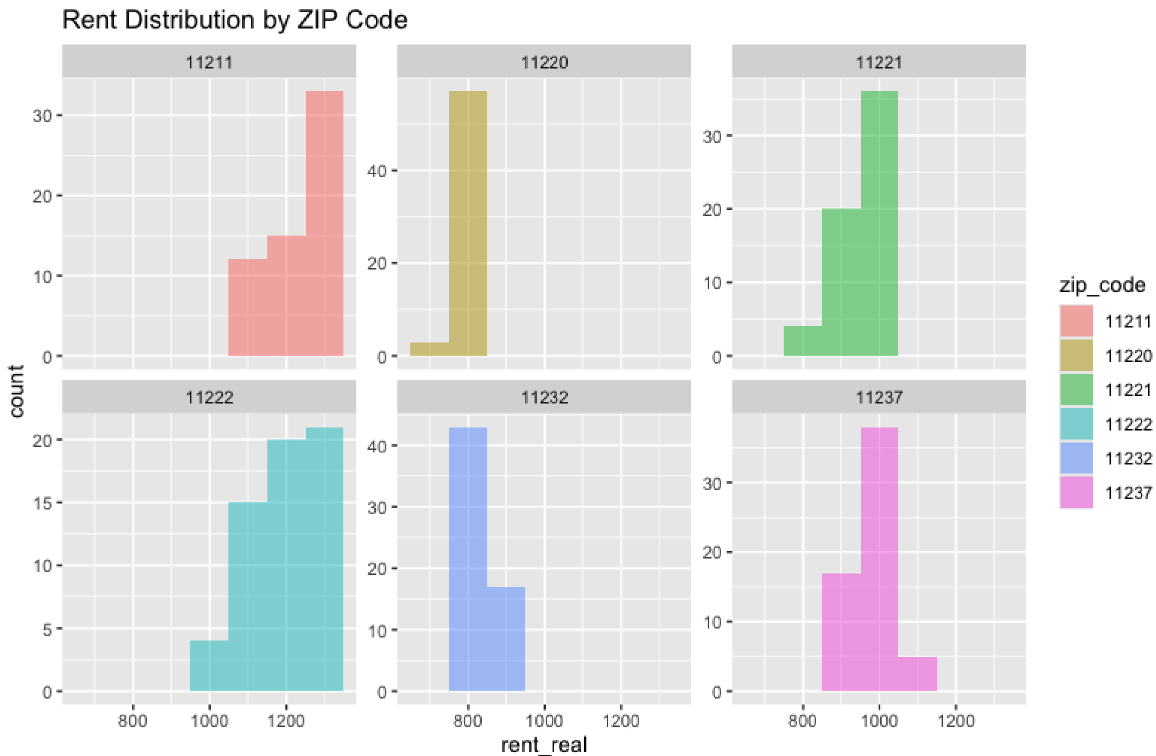

Figure 5. Distribution of adjusted ZORI variable after log transformation

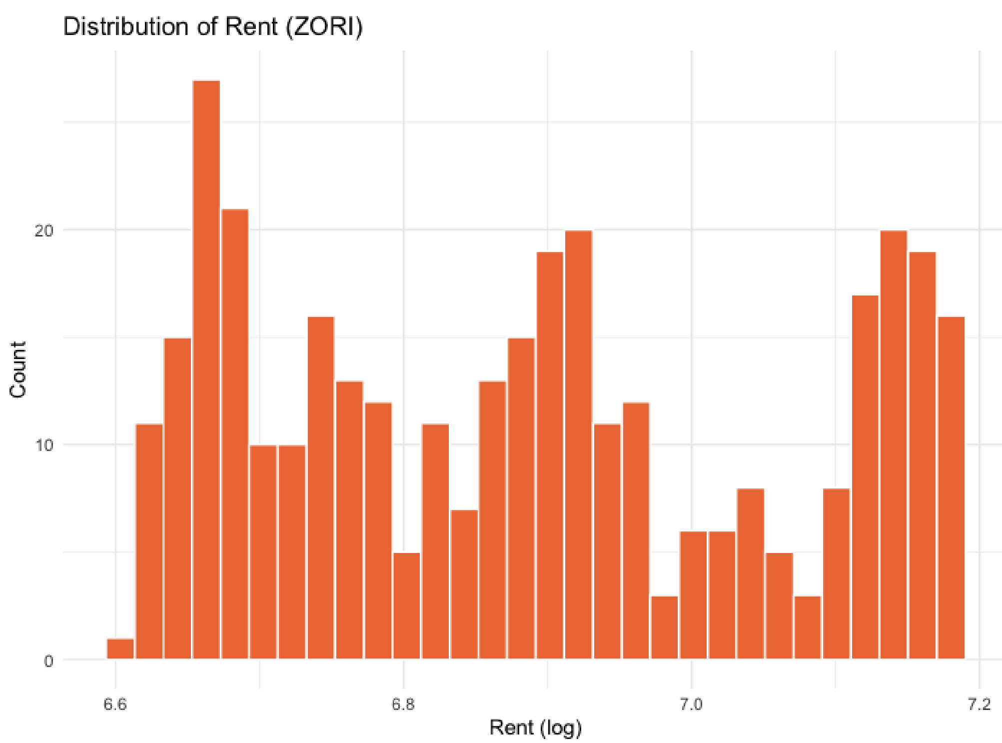

Figure 6. Q-Q plot of dependent variable distribution.

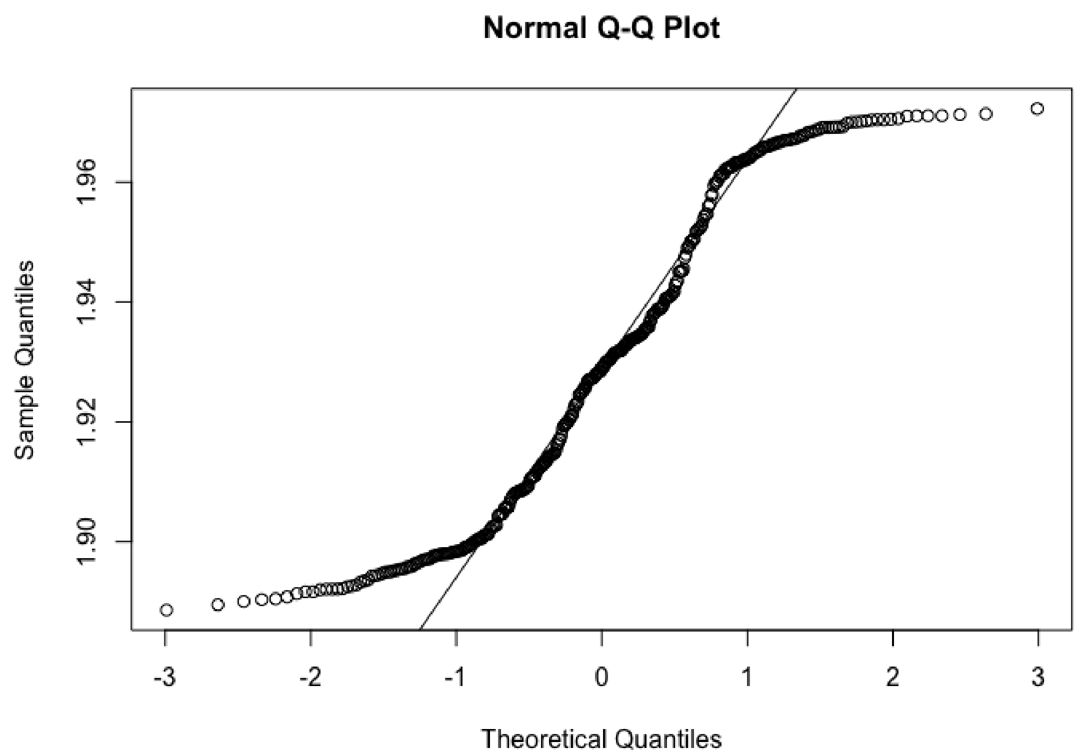

Figure 7. Scatter Plots comparing the dependent variable to the independent variable

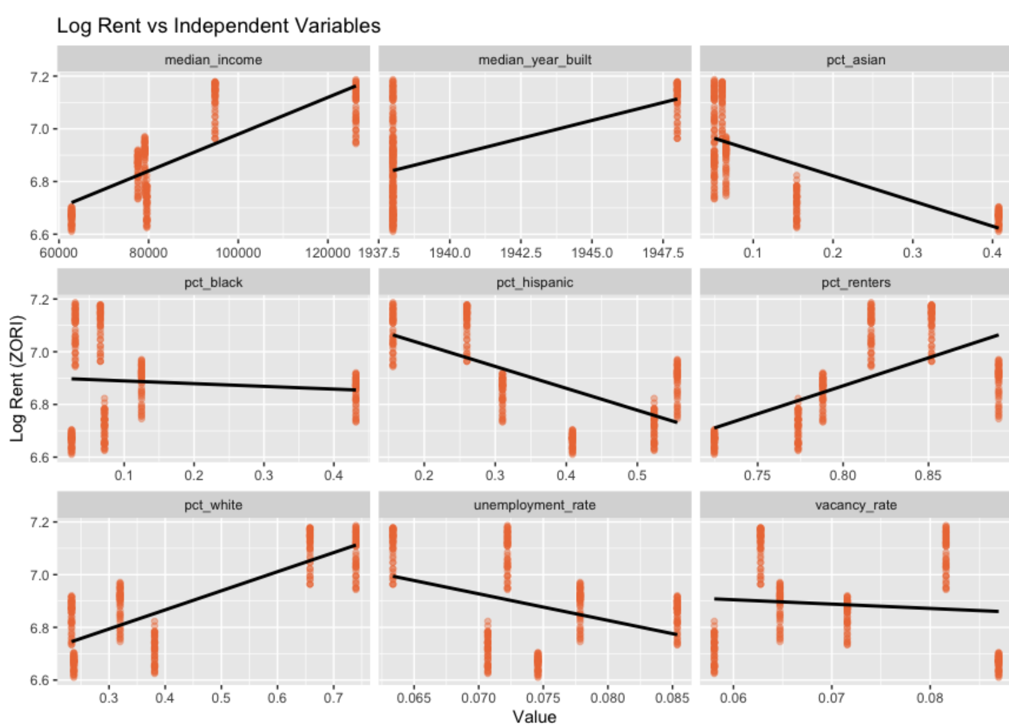

Figure 8. Parallel Trends Chart

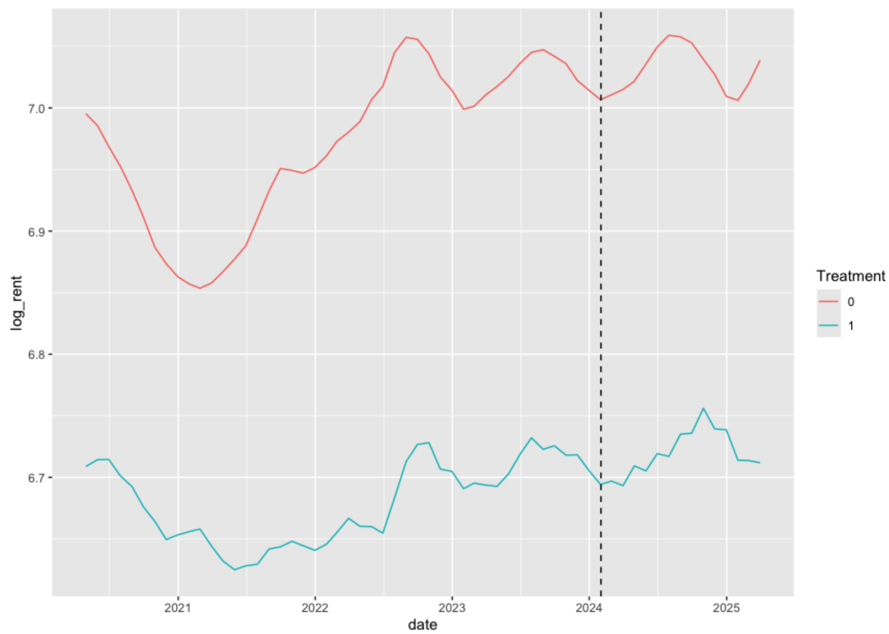

Figure 9. Histogram of residuals

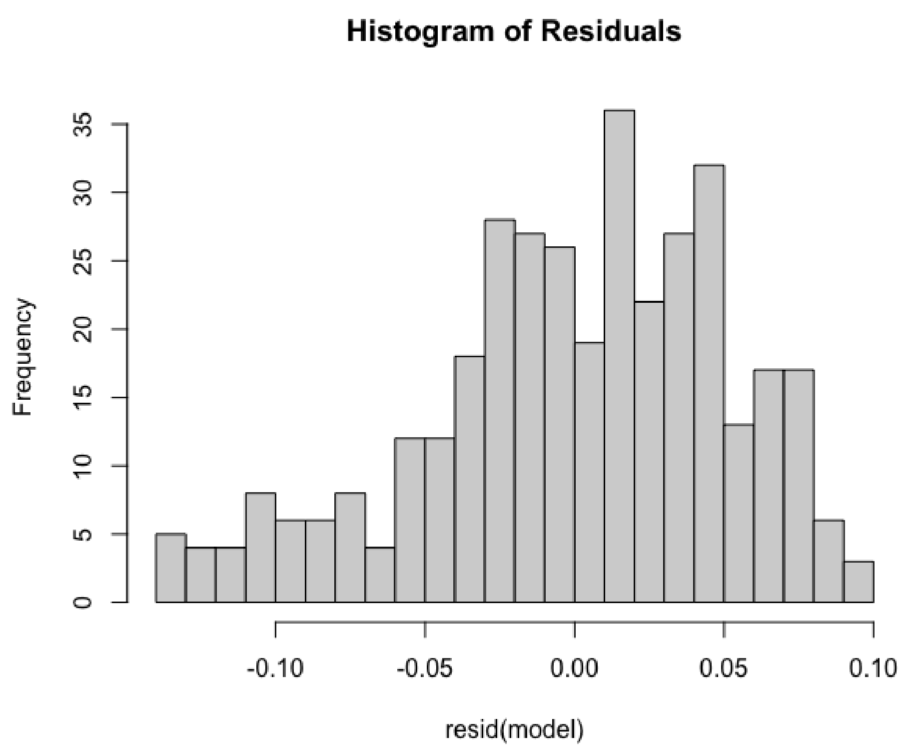

Table 1. Sample of final dataset

zip_code	date	rent_zori	treatment	post	treatmentpost	median_income	pct_renters	total_pop	pct_white	pct_black	pct_asian	pct_hispanic	vacancy_rate	median_year_built	unemployment_rate	rent_real
11220	4/30/20	2089.68338	1	0	0	62804	0.72462106	93008	0.23725916	0.02470755	0.40762085	0.40869603	0.08696221	1938	0.07458611	815.044087
11220	5/31/20	2092.34157	1	0	0	62804	0.72462106	93008	0.23725916	0.02470755	0.40762085	0.40869603	0.08696221	1938	0.07458611	816.06495
11221	10/31/21	2515.01606	0	0	0	77600	0.78814016	91236	0.23330703	0.43151826	0.05116401	0.31017362	0.07157157	1938	0.08537176	909.297209
11221	11/30/21	2538.67455	0	0	0	77600	0.78814016	91236	0.23330703	0.43151826	0.05116401	0.31017362	0.07157157	1938	0.08537176	913.363128
11237	9/30/24	3302.68188	0	1	0	79136	0.89105236	47183	0.31967022	0.1248331	0.06584999	0.55628086	0.0647263	1938	0.07785188	1047.46952
11237	10/31/24	3223.51094	0	1	0	79136	0.89105236	47183	0.31967022	0.1248331	0.06584999	0.55628086	0.0647263	1938	0.07785188	1021.18421
11232	2/29/24	2581.48618	1	0	0	79599	0.77369903	28137	0.3811707	0.07182713	0.154423	0.52372321	0.05805501	1938	0.07070383	831.862681
11232	3/31/24	2579.97552	1	1	1	79599	0.77369903	28137	0.3811707	0.07182713	0.154423	0.52372321	0.05805501	1938	0.07070383	826.036242

Table 2. Correlation matrix

	pct_white	pct_black	pct_asian	pct_hispanic	pct_renters	unemployment_rate	vacancy_rate	median_year_built
pct_white	1	-0.48	-0.47	-0.68	0.39	-0.69	-0.04	0.52
pct_black	-0.48	1	-0.38	-0.03	0.02	0.77	-0.14	-0.19
pct_asian	-0.47	-0.38	1	0.29	-0.76	-0.04	0.53	-0.25
pct_hispanic	-0.68	-0.03	0.29	1	0.03	0.19	-0.45	-0.34
pct_renters	0.39	0.02	-0.76	0.03	1	-0.15	-0.51	0.37
unemployment_rate	-0.69	0.77	-0.04	0.19	-0.15	1	0.25	-0.71
vacancy_rate	-0.04	-0.14	0.53	-0.45	-0.51	0.25	1	-0.35
median_year_built	0.52	-0.19	-0.25	-0.34	0.37	-0.71	-0.35	1

Table 3. Regression Results

term	*estimate*	*std.error*	*statistic*	*p.value*
(Intercept)	-2.7771	6.9822	-0.3977	0.6911
median_income	0.0000	0.0000	-7.7630	0.0000
pct_white	5.1329	0.5163	9.9412	0.0000
pct_hispanic	0.4941	0.0417	11.8360	0.0000
unemployment_rate	63.8894	5.2715	12.1197	0.0000
median_year_built	0.0026	0.0037	0.7078	0.4795
post	0.0610	0.0082	7.4813	0.0000
treatment	NA	NA	NA	NA
treatmentpost	-0.0192	0.0141	-1.3575	0.1755

Num.Obs.	360
R2	0.917
R2 Adj.	0.916
AIC	-1096.8
BIC	-1061.9
Log.Lik.	557.416
RMSE	0.05

Table 4. Fixed Effect Model Results

Observations: 360
Fixed-effects: zip_code: 6
Standard-errors: Clustered (zip_code)

Estimate	*Std.*	*Error*	*t- value*	*Pr(>|t|)*	
post	0.060980	0.008198	7.438178	0.00069238	***
treatmentpost	-0.019165	0.020335	-0.942470	0.38922515	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.051441
Adj. R2: 0.915548
Within R2: 0.164149