Gamze Bilsen

December 20, 2021

Data Analysis Final Project: the Impact of Natural Disasters on Housing Price Changes

1. Introduction

As climate change becomes more prevalent, available resources to sustain the continually growing population across the world are going to become scarcer. Urbanization has been seen as a critical solution as compactness of cities are able to more efficiently meet human needs with fewer resources than rural areas. However, highly in demand cities like Austin, Texas are already starting to feel the impacts of climate change through increased instances of extreme heat. So, it's important to increase the number of cities or promote migration to existing ones that are less vulnerable to natural hazards.

As the occurrences of climate related disasters increase with climate change, more cities are going to see higher costs to recover from the more frequent and significant disasters. Thus, it's important to understand if there has so far been an impact of frequency of climate disasters in housing prices, which would be indicative of whether people already are influenced by natural hazards in their choice of city to settle down in. This would help understand if there should be governmental or market-based nudges to get people to purchase homes hence settle down in more climate resilient areas, and also inform the financial changes in the housing market. So, this study uses the average housing sale price change in 806 cities in the US for the last decade to uncover whether natural disasters are an indicator for internal migration patterns within the United States. It's expected that if a city has seen higher number of natural disasters such as flooding or drought, their housing market is going to be in less demand. This is an important topic since climate change is predicted to increase natural disasters, and this relationship could uncover which cities organizations such as insurance companies should hedge against possible decreases in housing prices. Lastly, this study is inspired by the study "Disaster on the Horizon: The Price Effect of Sea Level Rise" which discovered the relationship between forecasted sea level rise and its impact on current housing prices (Bernstein).

2. Hypothesis

Controlling for the population density, average temperature, air pollution levels, smoking ban, federal expenditure and renewable power capacity, the higher the number of historical climate related disasters in a city's periphery, the lower the increase in housing prices over the last decade.

3. Data Sources, variables and measurements

The dataset used in this study is a combination of multiple datasets. The original dataset which other variables are merged onto is from Zillow's dataset of housing prices for selected 806 cities from 2000 to 2021 that displays the average annual sale price of homes within each city's periphery. Using QGIS, this dataset was merged with the number of disasters data from Columbia University's Geocoded Disasters dataset, with each city having the frequency of natural disasters from 1960-2018 within 200 miles from the city center whose coordinates were identified through geocoding. These coordinates were also used to merge each city center with air pollution and the climate indicator average temperature data using Open Weather Map's API. Next, the amount of renewable generation capacity per each county each city is located in is pulled from World Resources Institutes energy generation database. Next, each sample is the county each city belongs to, with the boundaries imported from the census tract's county dataset. Lastly, the merged county level information is from Open Intro Organization.

The dependent variable in the study is the percent change in average sale prices per city from 2011 to 2021 as a continuous variable, representing the change in desirability of each city across the last decade. The main independent variable is the frequency of natural disasters within the 200 miles radius per county of cities that have a population density mean of 200K in 2019 to analyze the impacts of climate related disasters on housing prices.

- **percent_dif:** continuous variable, percentage change in average sale price of a home per city in from January October 31 2011 to October 31 2021
- **number_disasters:** continuous variable, amount of natural disasters from 1960 to 2018 within 200 miles periphery of each city center
- **percapitaincome2019:** continuous variable, average per capita income within each county the cities reside in, in 2019
- **popdensity:** continuous variable, average per population density within each county the cities reside in, in 2019
- **avgnovtemp:** continuous variable, average temperature in Celsius for two days late November 2021 at 1am.
- **PM10:** a continuous variable, average ten particulates levels, indicator for air pollution levels
- **loggeneration2019:** continuous variable, logarithmic version of the amount of GW that is produced within the county of each city
- **fed_spending2009:** continuous variable, total federal spending per county in 2009

- **smoking_ban2010:** binary variable, 1 when there was a complete or a partial ban, 0 when there was no ban in the county in 2010

4. Data Analysis

4.1 Descriptive Statistics

Table 1 provides the descriptive information of the variables in this study. Mean value, standard deviation, minimum and maximum values are listed in the table.

Table 1

	percent_dif	number_disasters	avgnovtemp	loggeneration2019	PM10	percapitaincome2019	popdensity	smoking_ban_2010	fed_spending_2009
count	806.000000	806.000000	806.000000	806.000000	806.000000	806.000000	8.060000e+02	806.000000	8.060000e+02
mean	72.537618	80.818859	5.655849	3.992686	5.292246	28463.873449	2.087014e+05	0.349876	8.083920e+05
std	38.192590	67.794772	6.439380	2.005386	8.195011	6608.778370	4.410465e+05	0.477227	3.479059e+06
min	5.727935	0.000000	-24.397500	-0.954512	0.510000	0.000000	3.058000e+03	0.000000	2.505000e+03
25%	46.428874	19.000000	1.238542	2.846063	1.550000	24842.500000	4.123125e+04	0.000000	9.845275e+04
50%	62.922567	77.000000	5.082083	2.846063	3.255000	27809.000000	7.549300e+04	0.000000	1.953620e+05
75%	85.949995	127.750000	9.991875	5.271636	5.330000	31533.250000	1.799878e+05	1.000000	4.265245e+05
max	266.090185	322.000000	23.955625	10.454848	98.230000	68883.000000	5.223719e+06	1.000000	7.950814e+07

On average cities save a 73-percentage point increase in housing prices in the last 10 years. Most of the cities increased less than the average with the median being 63%. The average is mostly skewed by highly in demand cities such as the max city which increased by 266percentage points. Looking through the independent variables, every county on average saw around 81 natural disasters between 1960 and 2018 with the median also being close to the average value at 77. Third, average November temperature of the selected cities across the US is 6 degrees Celsius, with the median also being close to the mean at 5 degrees Celsius. Forth, the logarithmic version of the renewable generation in MW per county, logarithmic version due to high right skewness of the original dataset, is at 18MW, with the median being below the mean at 12MW. Fifth, the average PM10 concentration around the city center is 5.2 micrometers. This value is significantly different across the chart with the 75th percentile being lower than the mean at 5.3, and the max particulate being 98 micrometers, meaning some cities see an extremely high air pollution compared to the average expected value, most likely heavily polluted cities like New York City or historically industrial cities like Detroit. Sixth, the per capita income in 2019 on average in $28K per county, with the median being close to the median. The max value is significantly larger than the 75th percentile at $69K per capita. Seventh, population density on average is 200K people per county, with the 75th percentile being lower than the average. This displays existence of heavily dense cities like New York City, whose county most likely has the max value at 5.2M population density in 2019. Eighth, smoking ban in 2010 on average is 0.34 per county, showing that most counties didn't have complete or partial smoking bans in 2010. Lastly, federal expenditure per county in 2009 which is right after the 2008 recession is very right skewed with the mean being lower than the 75th percentile. This displays a possible inequality in federal expenditure among different cities.

4.2 Initial Model

The initial analysis is based on the Ordinary Least Squares regression model. Its goal is to find the relationship between percentage difference in average housing prices across a county for cities that have on average 200 thousand population density in 2019 from October 2011 to October 2021 to the total number of disasters within the 200 miles periphery of each city center from 1960 to 2018. The OLS model tells us the relationship between our independent variable of interest and the dependent variable on average, controlling for alternative variables that can explain the variation in the percentage change in housing prices such as the average temperature of a county this November and federal expenditure per county in 2009.

My assumptions prior to running the model is that first and foremost I expect higher number of historical natural disasters will increase immigration out of the counties, thus having a strong historical negative impact on the percentage difference in the housing price. Next, I expect higher temperatures in November means higher migration to cities, increasing their housing prices percentage difference due to people preferring warm November temperatures to colder ones in the places they settle down in, hence buy a house in. Third, I expect higher generation of renewable energy to be a positive indicator of increasing prices due to it being an indicator of high long-term investment in the county as renewables are relatively new technology, and higher renewable generation means higher expectation of private and public investors to meet the increasing energy demand from the county and also indicates higher level of natural resources that are applicable for renewable generation. Forth, I expect higher air pollutant PM10 to have a negative effect on the housing prices due to people looking out for their health in the cities they settle down in. Fifth, I expect higher per capita income to have a higher impact in increases in housing prices due to higher income individuals more likely buying houses that are more expensive and are deemed to be in valuable areas, thus driving up the housing prices. Next, I expect population density to increase prices since population density is indicative of urbanization, and urban areas tend to have higher housing demand due to many benefits such as access to public transportation. Also, I expect smoking ban to increase housing prices since it's indicative of the government being involved in the wellbeing of its residents, thus most likely having housing laws that have maintained the health of the buildings. Lastly, I expect federal spending in 2009, after the 2008 crisis, to have a positive impact on housing prices since it displays that there is high investment from the government to maintain the wellbeing of the population, possibly impacting migration to cities like these for people who are looking for jobs hence positively impacting the increase in housing prices in the next decade.

Table 2

OLS Regression Results

Dep. Variable:	percent_dif	R-squared:	0.275
Model:	OLS	Adj. R-squared:	0.268
Method:	Least Squares	F-statistic:	37.84
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	4.33e-51
Time:	14:31:33	Log-Likelihood:	-3949.4
No. Observations:	806	AIC:	7917.
Df Residuals:	797	BIC:	7959.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	41.5648	6.340	6.556	0.000	29.121	54.009
number_disasters	-0.1435	0.018	-8.085	0.000	-0.178	-0.109
loggeneration2019	2.2465	0.585	3.843	0.000	1.099	3.394
avgnovtemp	0.7193	0.187	3.850	0.000	0.353	1.086
percapitaincome2019	0.0008	0.000	4.487	0.000	0.000	0.001
PM10	1.1190	0.157	7.132	0.000	0.811	1.427
popdensity	6.579e-06	3.02e-06	2.181	0.029	6.58e-07	1.25e-05
fed_spending_2009	-6.065e-07	3.32e-07	-1.824	0.069	-1.26e-06	4.61e-08
smoking_ban_2010	-2.6894	2.432	-1.106	0.269	-7.463	2.084

Omnibus:	110.190	Durbin-Watson:	1.870
Prob(Omnibus):	0.000	Jarque-Bera (JB):	229.424
Skew:	0.791	Prob(JB):	1.52e-50
Kurtosis:	5.081	Cond. No.	1.97e+07

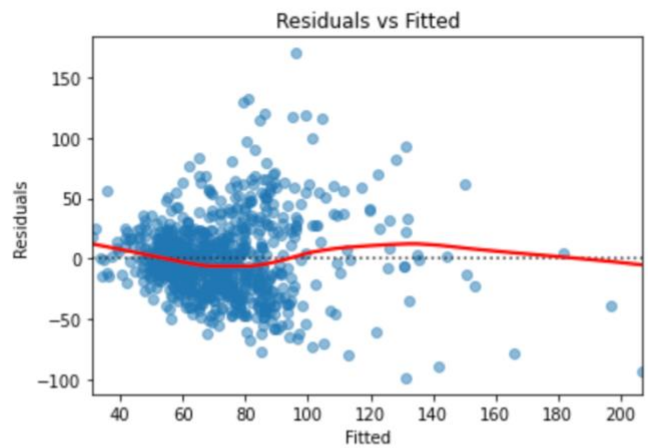
Table 2 displays the results of the initial OLS regression. For number of disasters, the coefficient is what I expected: on average, an increase in the historical amount of natural disasters in a city centers 200-mile periphery decreases the percentage increase in housing prices by 0.14 percentage points at a 99.9% confidence level, controlling for 2019 renewable generation, mean November temperature, mean per capita income in 2019, PM10 levels, population density, total federal spending in a county in 2009 and the existence of a smoking ban in 2010. For the control variables, firstly, renewable generation's average impact has the highest magnitude and is in the direction I expected, controlling for the other variables. Contrarily, the direction of air pollutant PM10 is also opposite to my expectations. This is more likely due to the variable also being an endogenous variable, as air pollution mostly increases when the amount of people increases in a given area due to increases in pollutants from more fossil fuel generation or traffic exhaust. In addition, smoking ban has a negative effect on the percentage increase in housing prices. This could be due to my initial assumption being flawed as more urbanized areas which have higher housing demand thus increase in prices more likely see higher number of smokers thus higher pushback into laws that ban smoking. However, this value is statistically insignificant thus the coefficient is most likely unindicative of the direction. Also, population density seemingly has a coefficient close to 0 at a 95% confidence level, however this coefficient is expected since the average population density is 200K people, so the changes in population density tend to be in high values. Similarly, although federal expenditure on average seems to have a little impact, considering the high average and median value of this variable, the coefficient has a strong negative impact on the housing prices which is against my initial assumption of a positive relationship. However, due to its confidence level being only at 90%, the coefficient is most likely unreliable. It could also be due to federal expenditure being an endogenous variable, thus most likely is being influenced by other variables such as renewable generation as the driver behind renewable integration into the grid is in large parts due to subsidies thus higher investment by the federal government.

Additionally, the R-squared shows that the model is able to explain 27.5% of the variation in y. With the Adjusted R-squared being close in value, we can state that despite some variables such as federal expenditure and smoking bans having low to no statistical significance and oppositional direction to my initial expectation, all of the variables add valuable information to the model.

Next, due to possible endogeneity in the model and smoking ban being statistically insignificant, I wanted to test the five Gauss-Markov assumptions to understand how the model can be improved.

4.3 Testing OLS Assumptions

Below are the five OLS Assumption tests on the initial model (Subramanian)   ***Table 3***



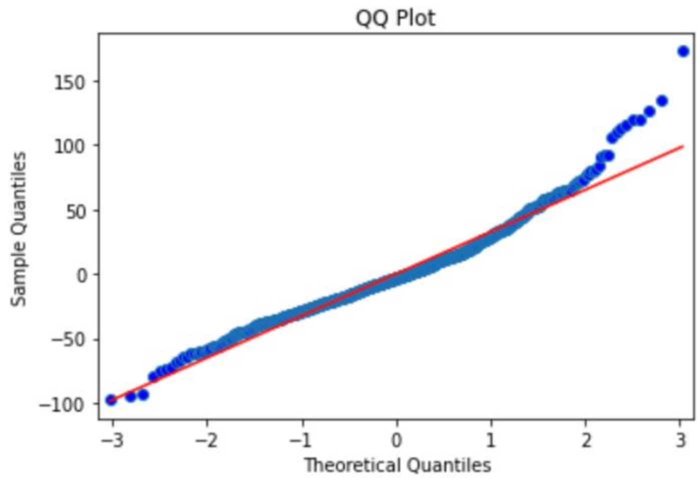
1. Linearity: plotting the residuals to the fitted values displays an although not perfectly linear fitted line, a slightly convex fitted line around the y-axis. Consequent to this, looking over bivariate graph of each independent variable to the dependent variable, I've transformed PM10, population density, federal spending variables to be their logarithmic versions. Also, I've conducted reset_ramsey to test the quadratic version of the number of disasters, logarithmic version of the 2019 renewable generation, average November temperature and per capita income variables. The results indicated a statistically insignificant improvements in the log generation and per capita income degree increases. However, number of disasters and average November temperature saw the lowest p-value for the second-degree increase, so, the next model will include continue with these changes.
2. Multicollinearity: Testing the VIF factor across all of the independent variables lead to low VIF factors ranging from 1.06 to 6.3, indicating limited collinearity, thus this not being a threat to our assumptions (Table 4).

Table 4

VIF_Factor	2.372889	1.811948	4.366445	1.729562	6.290947	1.513427	1.552925	1.058864

3. Normal Distribution of the Errors: As can be seen from the QQ Plot, the errors are mostly normally distributed until after the second Theoretical Quantile. Due to it being mostly normally distributed and the sample size being 806, not meeting this assumption doesn't lead to much hesitance in using this model due to Asymptotic Normality.

Table 5



4. Heteroskedasticity: Using the Breusch Pagan test, the low p-value indicates high heteroskedasticity in the model. Due to this, I will use cov_type='HC3' for all of the next models to increase the robustness of standard errors.

Table 6

Langrange multiplier statistic	9.467007e+01
p-value	5.222229e-17
f-value	1.325897e+01
f p-value	4.363383e-18

5. Autocorrelation: To test this assumption, we can look at the Durbin-Watson statistics in Table 2, which shows 1.87 value, indicating possible slight positive autocorrelation. However, due to the value being close to 2, removing non-statistically significant values after using cov_type='HC3' can be a solution.

4.4 Model Improvements

Table 7

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.3558	16.684	-0.021	0.983	-33.056	32.344
number_disasters	-0.2241	0.049	-4.584	0.000	-0.320	-0.128
number_disasterssecond	0.0004	0.000	2.032	0.042	1.4e-05	0.001
loggeneration2019	2.2927	0.604	3.795	0.000	1.109	3.477
avgnovtemp	0.7863	0.191	4.117	0.000	0.412	1.161
avgnovtempsecond	-0.0216	0.016	-1.318	0.188	-0.054	0.011
percapitaincome2019	0.0008	0.000	3.669	0.000	0.000	0.001
logPM10	8.9192	1.521	5.863	0.000	5.938	11.901
logpopdensity	4.7864	1.355	3.532	0.000	2.130	7.443
logfedspending_2019	-0.8958	0.871	-1.029	0.304	-2.602	0.811
smoking_ban_2010	-2.6496	2.415	-1.097	0.273	-7.382	2.083

With the additions of the second-degree order of number of disasters and average November temperature in addition to interchanging PM10, population density and federal expenditure with their logarithmic versions, the model improved only slightly with a 0.5-percentage point increase to 28% in the R-squared and 0.4 percentage point increase in the adjusted R-squared to 27.1% from the initial OLS model. Additionally, using robust standard errors has improved the reliance on the t-statistic in this model. This model has Durbin-Watson value of 1.897, close to the initial model.

In this model, number of disasters' coefficient's impact increases from -.14 to -0.22, indicating an on average strong negative impact of disaster numbers to housing prices, controlling for other factors at a 99.9% confidence level. Looking at the coefficient of the second order number of disasters, we can see that there is a slight positive relationship with the number of disasters to the housing prices at a 95% confidence level. This could be due to a third omitted confounding variable that increases people's migration into cities, disrupting the local ecosystem, thus consequently increasing the number of disasters after the demand to the city has increased. Next, logarithmic version of renewable generation shows that an average 1% increase in renewable generation in 2019 we expect to see on average a 2.3 percentage point increase the housing prices from 2011 to 2021. Next, average November 2021 temperature's magnitude slightly increased from the original OLS regression; now the displays that on average a 1 Celsius increase in average November temperature increased housing prices increase in percentage points by 0.78 at a 99.9% significance level. However, the second order average November temperature is statistically insignificant, although it displays on average a negative impact of higher temperature at the maximum point of its quadratic line. Next, on average $1000 increase in average household income leads to an 8-percentage point increase in housing prices in a given city controlling for other variables at a 99.9% confidence level, and remaining at the same coefficient level as the initial OLS model. In addition, the logarithmic population density now has a higher statistical significance at 99.9% significance, and a 1% increase in population density now leads to a 4.7 percentage point increase in housing prices in a decade, controlling for other factors. Lastly, smoking ban continues to be statistically insignificant and the in this model the logarithmic version of federal expenditure's coefficient increased in magnitude however is now statistically insignificant.

With these factors in mind, my next model will remove the statistically insignificant variables from my model.

5. Final Model Results and Discussion

To conclude, my final model is as follows:

Percent_dif = B0 + B1*number_disasters + B2* number_disasterssecond + B3*loggeneration2019 + B4*avgnovtemp + B5*percapitaincome2019 + B6*logPM10 + B7*logpopdensity + μ

Table 8

OLS Regression Results

Dep. Variable:	percent_dif	R-squared:	0.276
Model:	OLS	Adj. R-squared:	0.270
Method:	Least Squares	F-statistic:	34.55
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	3.45e-42
Time:	16:54:40	Log-Likelihood:	-3948.7
No. Observations:	806	AIC:	7913.
Df Residuals:	798	BIC:	7951.
Df Model:	7		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-9.7243	13.580	-0.716	0.474	-36.341	16.892
number_disasters	-0.2331	0.049	-4.717	0.000	-0.330	-0.136
number_disasterssecond	0.0004	0.000	2.211	0.027	4.98e-05	0.001
loggeneration2019	2.2540	0.605	3.727	0.000	1.069	3.439
avgnovtemp	0.5622	0.176	3.189	0.001	0.217	0.908
percapitaincome2019	0.0008	0.000	4.066	0.000	0.000	0.001
logPM10	8.6757	1.484	5.847	0.000	5.768	11.584
logpopdensity	4.4715	1.329	3.364	0.001	1.866	7.077

Omnibus:	95.392	Durbin-Watson:	1.901
Prob(Omnibus):	0.000	Jarque-Bera (JB):	173.265
Skew:	0.744	Prob(JB):	2.38e-38
Kurtosis:	4.717	Cond. No.	3.44e+05

By dropping the statistically insignificant variables smoking ban, federal spending's logarithmic version and the second order of average November temperature, the final model ends up having only statistically significant variables. To improve the Table 7 model, as the odd directions of some coefficients and slightly autocorrelation could be due to omitted variables bias, I've also tried adding the area in square feet per county and 2011 unemployment rate to the regression, in addition to testing various interaction terms across variables such as the interaction between the area of the county to population density as my assumption was that smaller counties are more likely to have lower population density, however all of these trials led to statistically insignificant results. Thus, the model is open for improvement through testing different variables such as variables that measure transportation access, walkability or education level of each county, which all should make a city more attractive thus hike the house prices.

All in all, the hypothesis is supported under this model with all of the coefficients except the second-degree number of disasters variable being statistically significant at a 99.9% confidence level, and with the former being statistically significant at a 95% confidence level.

To interpret the model, the amount of disasters from 1960-2018 have a negative relationship with the average percentage point increase in the housing prices in each county until the maximum point. On average, one unit increase in number of disasters across all counties leads to a 0.23 percentage point decrease in housing prices from 2011 to 2021 controlling for other factors. Although the second-degree number of disasters has a positive coefficient, its magnitude is small. This coefficient is most likely due to migration to a small number of desirable yet vulnerable cities like Austin, Texas who see high number of natural disasters however due to other factors such as having a high number of young population and education institutions make the city desirable despite the high instances of natural disasters. So, considering the average percentage increase in prices across the last decade is 72, number of disasters has a significant impact on housing prices changes across time, displaying a possible trajectory of housing price decrease even in vulnerable cities like Austin or New York as the second order coefficient is significantly lower in magnitude. In addition, on average, 1% increase in renewable generation in MW's per county leads to a 2.25 percentage points increase in the dependent variable controlling for other factors. This again confirms my first assumption since high investment in a new technology like renewables indicates investors see a high current and future value in these counties. Third, on average, a 1 Celsius increase in November temperature leads to on average a 0.56 percentage point increase in the dependent variable controlling for other factors, again confirming my initial assumption of warmer weather being more desirable for people in settling down in a place through purchasing a home. Fourth, a $1K increase in per capita income in 2019 leads to on average an 8-percentage point increase in the dependent variable, controlling for other factors, again confirming my initial assumption of large positive association between these two variables. Fifth, on average, 1% increase in the average PM10 concentration of a county leads to 8.67 percentage point increase in the dependent variable, controlling for other factors. This is the variable that is different than my initial assumption, however, upon thinking is understandable. There is likely a confounding factor that causes people to increase their demand in the housing market in a given county (dependent variable) and increases their migration patterns, increasing air pollutants as a result. Lastly, on average, 1% increase in the population density leads to 4.47 percentage point increase in the dependent variable, controlling for other factors. This is again in accordance to my expectations as higher density signals more compact county, which indicates walkability and transportation networks which are desirable factors in cities.

The R-squared for this model is almost the same as the initial model at 27.6%, thus all six variables are able to explain 27.6% of the variation in the dependent variable. Because the adjusted R-squared is close to the R-squared at 27%, we can conclude that all of the variables more or less add valuable information with each being statistically significant compared to previous models who despite having a higher R-squared has variables that were unable to reject the null hypothesis.

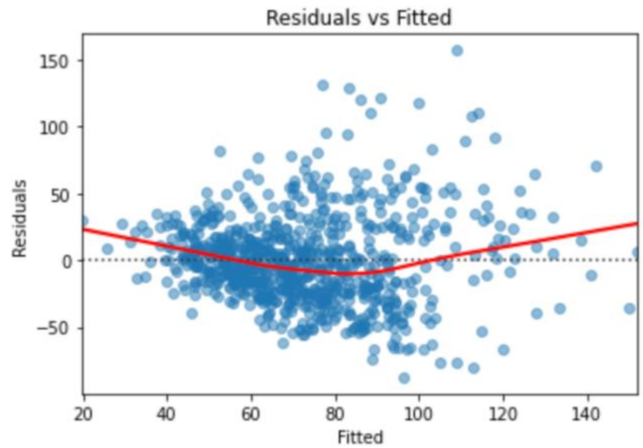
<u>6. Final Diagnostics, Limitations and Conclusion</u>

6.1 Final Model and OLS Assumptions

Below are the five OLS Assumption tests on the initial model

1. Linearity: plotting the residuals to the fitted values on the right displays a non-linear distribution, however we can see that the residuals are spread across the x axis more than the initial model. The disparity still exists, thus future research is needed in applying variables that can best explain the variation in percentage difference in housing prices in cities across the US while maintaining the linearity assumption, or testing different regression models.

Table 9



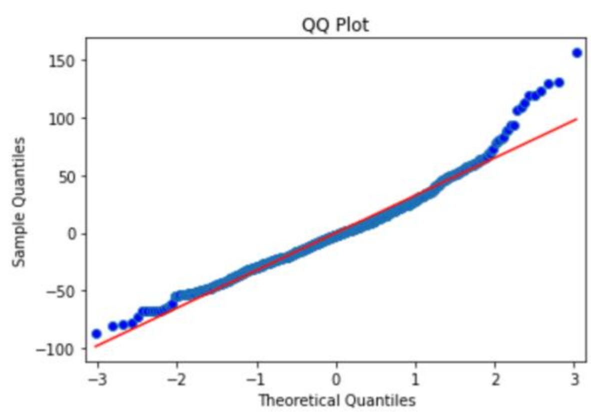
2. Multicollinearity: Testing the VIF factor across all of the independent variables lead to low VIF factors ranging from 2.56 to 39.3, indicating some multicollinearity, more than before, however because none are above 100, no variables are a cause of concern.

Table 10

VIF_Factor	19.93442	1.976246	5.181278	24.082985	39.264019	2.568261	12.852671

3. Normal Distribution of the Errors: As can be seen from the QQ Plot, the errors are again mostly normally distributed until after the second Theoretical Quantile. Due to it being mostly normally distributed and the sample size being 806, not meeting this assumption doesn't lead to much hesitance in using this model due to Asymptotic Normality.

Table 11



4. Heteroskedasticity: The model uses robust standard errors, thus there is limited impact of Heteroskedasticity on the t-statistic values.
5. Autocorrelation: The Durbin-Watson statistics which shows 1.90 value, a slight increase from the previous models, indicating possible slight positive autocorrelation. However, due to the value being close to 2, this is not a cause of concern.

6.2 Final Model Limitations

Firstly, the model explains only 27% of the variation in the dependent variable, thus there is likely omitted variables from the model. In addition, PM10 and increases in housing prices in the last decade are most likely driven by the same phenomenon not uncovered in this model, again implying omitted variable bias and confounding factors in the model. There are many, if not unending list of factors which increase the demand in a particular city that can be controlled for to uncover the relationship of interest, the impact of number of disasters on housing price changes, such as a variable which measures the public education quality on a given county which the dataset used did not have.

Secondly, the variables used such as population density and renewable energy production capacity are proxies to other phenomenon such as city walkability and long-term investment levels, indicating a possible measurement error thus decreasing the confidence level of the findings.

Finally, as previously mentioned, the model does not meet the linearity assumption of OLS. Due to this, different non-linear predictive models can be used to predict changes in housing prices such as Support Vector Machines, or variables can be alternated to be linear variables to meet this criterion.

6.3 Conclusion

The findings of all of the models display a negative relationship between the number of historical disasters that's within 200-mile proximity to a city's center to the percentage change in housing prices over the last decade. This displays that people tend to choose cities that are more naturally resilient, as disasters have a high emotional and economic impact on the wellbeing of its residents. With natural disasters projected to increase as climate change gets more pronounced, this project displays a possible housing price increase in cities that see a small number of disasters and visa versa for cities who see a higher number of disasters. Additionally, it's also important to look at the impact of control variables which are the alternative explanations to the number of disasters variable: all of the other variables also had a high average impact on percentage change in the last decade, displaying a possible future tendency that people might follow in the upcoming decades as the personality of each city changes as time passes. Lastly, there is much more to be done in this arena, with adding more variables that could explain a desirability of a city, to possibly identify aspects in which governments or market agents can enact into increase the desirability of cities that fit the characteristics people are looking for in purchasing a new home, making sure these areas are projected to have low frequency natural disasters thus high climate resilience for the long-term wellbeing of people.

Works Cited

Bernstein, Asaf, et al. “Disaster on the Horizon: The Price Effect of Sea Level Rise.” *Journal of Financial Economics*, North-Holland, 23 Mar. 2019, https://www.sciencedirect.com/science/article/abs/pii/S0304405X19300807.

“Geocoded Disasters (GDIS) Dataset, v1 (1960−2018).” *SEDAC*, https://sedac.ciesin.columbia.edu/data/set/pend-gdis-1960-2018/data-download.

“Global Power Plant Database - DATA: World Resources Institute.” Data, https://datasets.wri.org/dataset/globalpowerplantdatabase.

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

OpenWeatherMap.org. “Weather API.” OpenWeatherMap, https://openweathermap.org/api.

Subramanian, Niranjan B. “5 Assumptions Made by OLS: Python.” *AI ASPIRANT*, 27 June 2021, https://aiaspirant.com/ols-assumptions/.

“United States Counties.” Data Sets, https://www.openintro.org/data/?data=county_complete.