Senior Project

Department of Economics

Impact of Air Quality on Housing Prices

Emily Reese

Spring 2024

Advisor: *Dr. Ali Enami*

Abstract:

This paper examines the relationship between air quality and housing prices, investigating how the different categories of air quality based on the AQI index influence housing prices. Using a two-way fixed effects model looking at 940 counties across the United States from 2000-2022, we find a significant relationship between housing prices and 'good' air quality days and days classified as 'unhealthy for sensitive groups.' This analysis finds that, on average, a one percent increase in the number of 'good' days yields housing prices to increase by 0.0084 percentage points and that a one percent increase in the number of 'unhealthy for sensitive groups' days yields housing prices to decrease by 0.0444 percentage points. While good days are not economically significant, unhealthy for sensitive groups is economically significant. This may be due to people being more sensitive to worse air quality or being more responsive to deteriorating air quality at a certain level. This shows the importance of maintaining good air quality standards for public health and the stability of residential property values.

Table of Contents:

I. Introduction

The regulation of air quality in the U.S. has a significant history, dating back to the 1970s with the implementation of the Clean Air Act (CAA). This act regulates air emissions and authorizes the EPA to establish regulations to protect public health and welfare, and to regulate emissions of hazardous air pollutants. The CAA plays a crucial role in maintaining the cleanliness of the air we breathe, which affects the health and aesthetic appeal of our neighborhoods and outdoor spaces. Understanding how people value air quality and their willingness to pay for better air quality can help make regulations and to what extent.

My research question is what the impact of air quality on housing prices. Air quality is challenging to define, as its economic value can be inferred from its association with housing prices. Individuals implicitly assign value to air quality based on their willingness to pay for housing in areas with better air quality. Therefore, by examining how housing prices vary in relation to air quality, we can indirectly estimate the economic value people place on clean air.

Understanding the influence of air quality on housing prices is important for economists and policymakers due to health concerns about air quality and creating pollution regulations. Housing is a significant investment for individuals and families, making factors affecting these property values essential for economists and policymakers. If communities and policymakers understand the impact of air quality on housing prices, they might be more inclined to reduce pollution and improve their communities' air quality to improve their neighborhoods and increase their housing values. Many studies have established links between poor air quality and a range of health issues, including respiratory conditions, heart disease, and lung cancer.

Previous literature has examined whether air pollution or air quality impacts housing prices with varying results. Many of these papers examine one city or one specific area and how air quality or pollution levels impact housing prices over time. Some have found a significant relationship, and others have not found any significant relationship.

My contribution to this topic is an updated analysis using more recent data. This paper will examine the impact of air quality across the U.S., including 940 different counties.

The main databases used in this paper are house index data (HPI) from the Federal Housing Finance Agency for the dependent variable, which is calculated by annual percentage change by county and year. The air quality data is from the United States Environmental Agency (EPA). Air quality is determined by the type of day based on the air quality index (AQI). This separates air quality into six different groups based on the AQI index. The air quality variable is calculated by the proportion of days falling into each category by dividing the total number of days per category by the total number of days observed each year and looking at the annual percentage change by county and year.

This paper will analyze the impact of air quality on housing prices in the U.S. using a two-way fixed-effects model, looking at 940 counties across the U.S. A two-way fixed-effects model allows us to control unobserved heterogeneity across counties and over time. We will use the annual percentage change of the HPI by year, and the county will regress on the percentage change of the air quality variable, including control variables such as income, unemployment, and population.

The main results show that, on average, a one percent increase in the number of 'unhealthy for sensitive groups' days yields housing prices to decrease by 0.0444 percentage points. This result is significant at the 5% level. Although less economically significant, the model for good days shows that, on average, a one percent increase in several 'good' days yields housing prices to increase by 0.0084* at the 10% significance level. These results show that air quality impacts housing prices, although increasing the number of good days has lower economic results.

The remainder of this paper is organized as follows: Section II discusses previous literature on the impact of air quality on housing prices in the U.S. and other parts of the world; Section III is a theoretical discussion; Section IV describes the data; Section V discusses the empirical methodology of the analysis; Section VI discusses the results; and Section VII is the conclusion based on the findings.

II. Literature Review

The relationship between air quality and housing prices has been studied previously with varying results. Beginning with Ridker and Henning (1967) they are the first to link air pollution with property values. They used a multiple regression model on air pollution and other housing characteristics on housing prices in St. Louis and found that air pollution had a statistically significant and negative effect on median housing prices. Following this study, many economists and researchers continued this research in different cities and countries to better understand the impact.

Another study in the U.S. by Chay and Greenstone (2005) analyzed housing prices in counties heavily affected by the Clean Air Act due to pollution levels above the net ceiling declared by the government. They found evidence that stricter environmental regulations increased housing prices in areas with significant air quality improvements. Looking at a more recent study in the U.S. Choi et al. (2023) finds that air quality degradation has a direct and longer-term negative impact on house prices. Using a spatial difference-in-differences model they looked at the Aliso Canyon gas leak to examine the effect of air quality on housing prices of Los Angeles City and found that houses within the 5-mile radius of the gas well experienced an 8.6% discount in price during the leak, and an additional 13.1% discount after air quality was restored.

We also can see a significant impact of air quality's impact in other countries such as in Ecuador. Borja-Urbano et al. (2021) investigated the effects of air pollution on urban housing prices in one of the most polluted cities in Ecuador, the Metropolitan District of Quito (DMQ). This study finds the marginal willingness to pay for cleaner air in DMQ using the impact of air pollutants on price properties. Their results showed the economic impact on the housing market was statistically significant, with a decrease in property value between 1.1% and 2.8%.

A study done in China had found a significant but relatively small impact of air pollution on housing prices. Zou et al. (2022) employed a machine-learning approach to investigate the nonlinear impact of air pollution on housing prices in Shanghai. Their results show that their other variables contribute 97.21% of the influences on housing prices, and the contribution of air pollution variables is 2.79%. Another study in China, Cai et al. (2024), finds that that high-income dwellers tend to pay more for clean air using the hedonic framework. They found that households were willing to pay an extra 0.0852% per housing unit price for an average quarterly reduction in PM2.5 of 1 $\mu g/m^3$.

Lu et al. (2022) looks at the spatial relationship between the ambient air pollution level of an apartment and its property value in the housing market of South Korea using a two-stage spatial Durbin error model. They find that, holding other factors equal, a 1% increase in the air pollution level can, on average, cause a decrease in the value of a local real property by 0.32%.

While some papers have found a significant relationship, others have found no significant relationship between the two. Research by Wang et al. (2015), looking at city-level data from China found that air pollution does not necessarily negatively affect housing prices using a spatial Durbin model to estimate the spatial spillover effects of air pollution on housing prices among neighboring cities. They found a positive correlation between air pollution and housing prices in large cities, while there is a negative correlation in small and medium-sized cities.

Miłuch et al. (2024) investigated the impact of ambient air pollution on housing prices in Warsaw, Poland, and failed to find a consistent relationship between air pollution levels and housing values. They examined spatial dependencies, looking at the high concentration of particulate matter expecting to reduce real estate values, yet they did not find evidence of this impact.

A study by Tang et Niemeier (2021) looked at air pollution measurements and housing prices using spatial autocorrelation and endogeneity effects. Surprisingly, they found a positive relationship between pollution and housing prices in Oakland, California. However, they noted that this may be due to homeowners in high demand and low housing stocks in the area and that people may be insensitive to air pollution.

While some studies have reported a negative relationship between air quality and housing prices, others have failed to find a significant or positive relationship. These results also have differing economic significance. Results may have to do with the country or city the study focuses on, the specific model used, and the variables used within these models.

This paper will further the research done on the U.S. looking at air quality and housing prices with updated data from 2000-2022. Instead of focusing on one city, this study will look across the U.S. at 940 different counties to see on average in the U.S. how air quality impacts housing prices and see at what level it begins to affect these housing decisions.

III. Theoretical Discussion

The theory of environmental amenities suggests that people derive utility from clean air, green spaces, and other natural amenities, which influence their housing preferences. Higher air quality is associated with a range of health benefits, including reduced respiratory illnesses and improved overall well-being. Therefore, areas with higher air quality may be perceived as more desirable places to live, leading to increased demand for housing and an increase in prices.

In contrast, poor air quality can have adverse effects on human health and environmental quality, potentially making the area less desirable and decreasing property values. Pollution from industrial activities, vehicular emissions, and other sources can contribute to air quality degradation, leading to concerns about respiratory health, pollution-related health risks, and overall environmental quality.

In this theory we would expect a positive relationship between housing prices and air quality. People would be willing to pay more to live in an area with cleaner air, increasing the demand than the housing prices of these areas with better air quality.

Yet this is only the case if people value air quality enough for it to change their housing preferences or if it is drastic enough change from area to area to impact their health. In the United States, pollution and air quality is regulated by the EPA. According to the Environmental Performance Index (EPI)'s air quality rankings, which measures the direct impacts of air pollution on human health in each country, the United States ranks 16th out of 180 countries in 2022. The average AQI in the U.S hasn't been below the 'good' threshold since 1999. With this in mind, we can look at the question on whether home buyers in the United States consider air quality in housing preferences and if so, what is the impact.

IV. Data Description

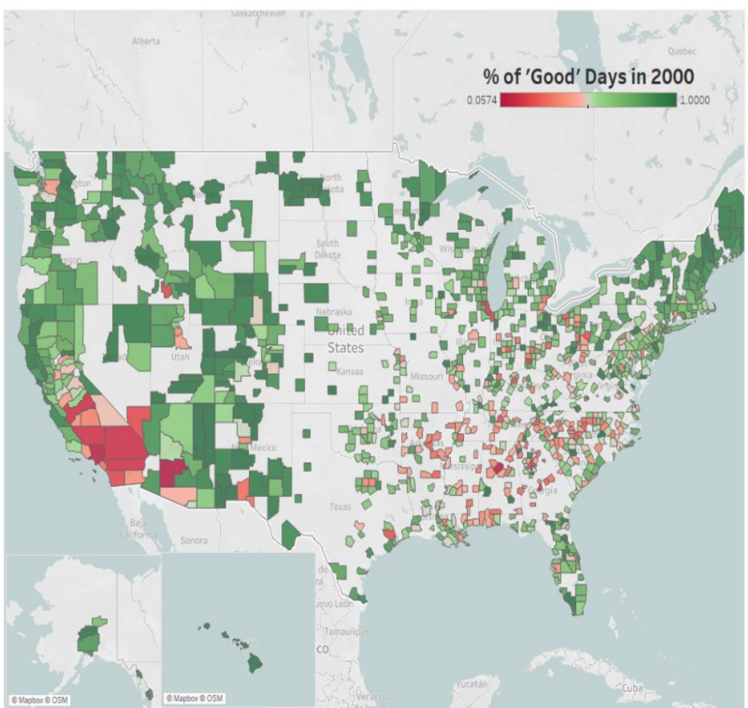
The data in this analysis spans from years 2000-2022 looking at 940 different counties over the United States. For my dependent variable this analysis uses house index data (HPI) from the Federal Housing Finance Agency to calculate the percentage change of house prices by county starting from 2000.

For the main regressor variable, this analysis uses air quality data from the United States Environmental Agency (EPA). The main regressor variable is the percentage change in the number of days of each category of type of days by year. Each model will look at a different category of day which is determined by where the AQI is in the range. These categories of days are defined by the EPA and include 'good', 'moderate', 'unhealthy for sensitive groups', 'unhealthy', 'very unhealthy', and 'hazardous'.

This analysis calculates the percentage change in type of air quality days for each county by dividing the count of type days by the total number of days recorded, and subsequently calculating the yearly change.

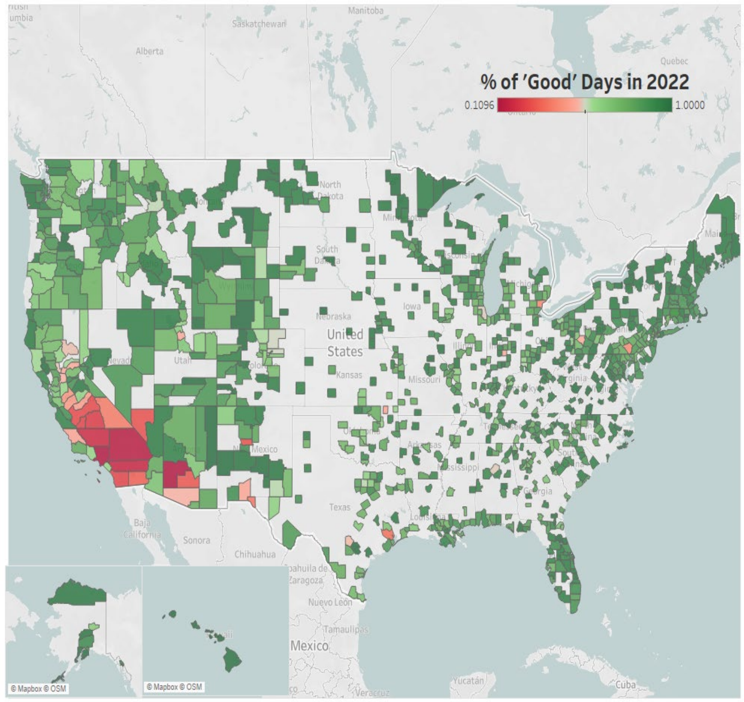
Figures 1 and 2 show the different counties used in this model showing percentage of good days in 200 and 2022. The green counties show the higher number of 'good' days and red shows lower number of good days. Counties in white are counties that didn't have air quality accounted for in the year shown. Comparing both graphs can see that in 2000 there were more red counties especially in east than in 2022. This shows the number of 'good' days has been increasing in the east while in California it has not.

Figure 1: Percentage of Good Days in 2000



Source: United States Environmental Agency 2000, and own calculations.

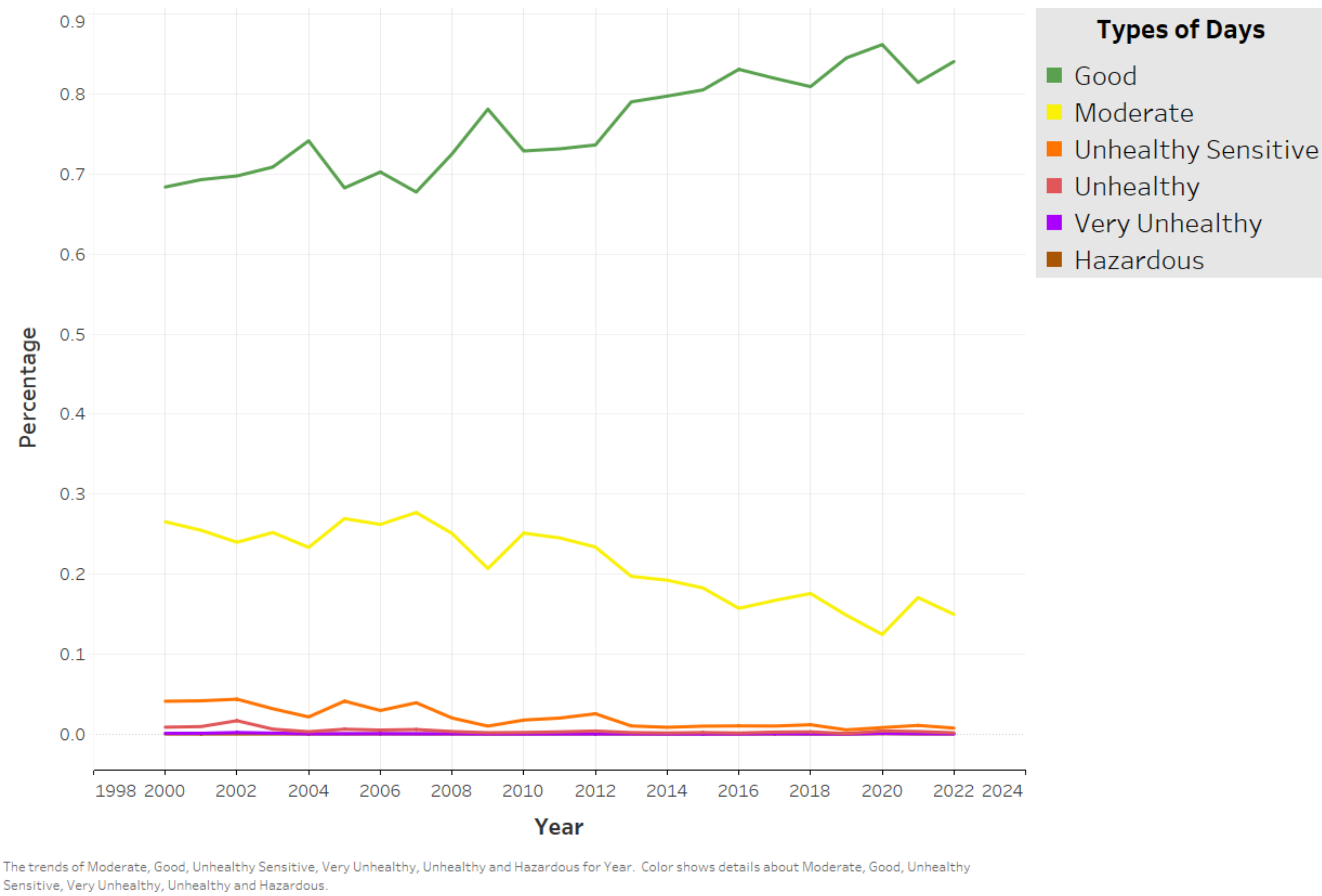
Figure 2: Percentage of Good Days in 2022



Source: United States Environmental Agency 2022, and own calculations.

Figure 3 shows the distribution and trends of the air quality categories overtime in the U.S. We can see a positive trend of 'good' days showing that air quality is improving over time. 'Moderate' and 'unhealthy for sensitive groups' days have a negative trend and are decreasing overtime. The last three groups 'unhealthy', 'very unhealthy', and 'hazardous' days are a small percentage overall in the U.S.

Figure 3: Percentage of Air Quality Index (AQI) Categories Overtime



Source: United States Environmental Agency 2000-2022, and own calculations.

For the other control variables, the analysis uses percent change in unemployment rate by county, and percent change in median income by county, and percent change in population by county.

Table 1: Summary Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum	Median
HPI	15355	2.647136	6.252642	-40.409	36.287062	2.547277
Good	15231	0.919729	9.107006	-90.7742	88.44086	0.877193
Moderate	15231	-0.65974	7.859407	-54.8306	66.153846	-0.59822
Unhealthy for Sensitive Groups	15231	-0.21445	2.629765	-41.9355	41.935484	0
Unhealthy	15231	-0.04317	1.076634	-17.005	24.657534	0
Very Unhealthy	15231	-0.00289	0.27127	-7.96167	6.25	0
Hazardous	15231	0.000516	0.14656	-4.936	10.136986	0
Unemployment	15188	0.025059	2.873078	-27.6839	24.222718	-0.12868
Income	15405	1.987863	4.703851	-37.9487	60	1.565558
Population	15406	0.744881	1.393812	-53.4342	16.760944	0.581553

Source Federal Housing Finance Agency, United States Environmental Agency. Own Calculations

V. Empirical Methodology

The model used in this study to estimate the impact of air quality on housing prices is as follows.

$$HousingPrice_{it} = \beta 0 + \beta 1 AirQuality_{it} + x_{it} + county_i + year_t + \epsilon_{it}$$

The dependent variable, $HousingPrice$, represents the percent change in the HPI for county i at time t. This variable is the main variable in our analysis and reflects the outcome of interest. The main independent variable of interest in the model is $AirQuality$ that represents the percent change in percentage of the category of day for county i at time t. x includes our other control variables $Unemployment$, and $Income$, and $Population$. These variables capture the percentage change in unemployment rate, income, and population for county i at time t. The model includes these variables as potential determinants of housing prices, based on economic theory suggesting their influence on the housing market.

This analysis employs a two-way fixed effects model to estimate the relationship between housing prices and our independent variables while accounting for both county-specific and time-specific effects. This allows us to control unobserved heterogeneity across counties and over time, providing more reliable estimates of the coefficients. By including county fixed effects and time fixed effects, this study mitigates the potential bias from omitted variables that are constant within counties or over time but vary across counties.

VI. Results

The results below show the estimated coefficients for the impact of each type of air quality's impact on housing prices. Each category has two models, one with and without control variables.

In table 1 model 2, we find that on average, a one percent increase of number of 'good' days yields housing prices to increase by 0.0084 percentage points at the 10% significance level. While these results are statistically significant and align with our a priori expectations that better air quality will increase air quality they are not economically significant.

In table 2 model 6, we find that on average, a one percent increase of number of 'unhealthy for sensitive groups' days yields housing prices to decrease by 0.0444 percentage points. This result is significant at the 5% level. These results are statistically significant and align with our prior expectations that worse levels of air quality would decrease housing prices. These results are economically significant.

This analysis does not see significant results for moderate, unhealthy, very unhealthy, and hazardous air quality days but the coefficients still align with expectations and are negative except for hazardous which may be due to the small number observations for this category.

These results show that air quality does have an impact on housing prices although increasing number of good days has lower economical results, limiting days where the AQI is below the 'moderate' level can impact housing prices.

Table 2: Good and Moderate Days on % Change HPI

	Good Days		Moderate Days	
Regressors	**Model1**	**Model2**	**Model3**	**Model4**
Good Days	0.0035	0.0084*		
	(0.0050)	(0.0049)		
Moderate Days			-0.0016	-0.0055
			(0.0055)	(0.0054)
Unemployment		-0.5861***		-0.5854***
		(0.0499)		(0.0498)
Income		0.0075		0.0073
		(0.0083)		(0.0083)
Population		1.2511***		1.2500***
		(0.0735)		(0.0734)
Intercept	5.6057**	4.0281*	5.5937**	4.0080*
	(2.3994)	(2.3485)	(2.3983)	(2.3467)
County and Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	15184	15175	15184	15175
Adjusted R-Square	0.4578	0.5041	0.4577	0.5041
Overall Significance	60384081966.20**	10305182691.05**	15638790398.41**	5066726822

Source: Federal Housing Finance Agency, United States Environmental Agency, and own calculations.
.Notes: Robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance levels, respectively.

Table 3: Unhealthy for Sensitive and Unhealthy Days on % Change HPI

	Unhealthy for Sensitive		Unhealthy	
Regressors	**Model5**	**Model6**	**Model7**	**Model8**
Unhealthy for Sensitive Days	-0.0275	-0.0444**		
	(0.0187)	(0.0180)		
Unhealthy Days			0.0135	-0.0084
			(0.0554)	(0.0534)
Unemployment		-0.5856***		-0.5842***
		(0.0498)		(0.0496)
Income		0.0078		0.0074
		(0.0083)		(0.0084)
Population		1.2526***		1.2499***
		(0.0731)		(0.0733)
Intercept	5.6100**	4.0189*	5.5830**	3.9889*
	(2.3965)	(2.3433)	(2.3980)	(2.3434)
County and Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	15184	15175	15184	15175
Adjusted R-Square	0.4579	0.5043	0.4577	0.504
Overall Significance	17,259,772.51***	15791402488.32**	373,967,855.45**	236.41***

Source: Federal Housing Finance Agency, United States Environmental Agency, and own calculations.
Notes: Robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance levels, respectively.

Table 4: Very Unhealthy and Hazardous Days on % Change HPI

	Very Unhealthy		Hazardous	
Regressors	**Model9**	**Model10**	**Model11**	**Model12**
Very Unhealthy Days	-0.3592	-0.4225		
	(0.3208)	(0.3014)		
Hazardous Days			0.4879	0.2821
			(0.3963)	(0.4217)
Unemployment		-0.5829***		-0.5842***
		(0.0498)		(0.0496)
Income		0.0078		0.0072
		(0.0083)		(0.0083)
Population		1.2522***		1.2486***
		(0.0733)		(0.0733)
Intercept	5.6119**	4.0128*	5.5835**	3.9858*
	(2.3979)	(2.3466)	(2.3970)	(2.3452)
County and Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	15184	15175	15184	15175
Adjusted R-Square	0.458	0.5044	0.4579	0.5041
Overall Significance	37838244722.16**	27918013314.47**	10942519213.81**	13350126031.62**

Source: Federal Housing Finance Agency, United States Environmental Agency, and own calculations.
Notes: Robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance levels, respectively.

VII. Conclusion

In conclusion, this analysis generally aligns with the hypothesis that better air quality increases housing prices while poorer air quality decreases them. Our main results show that an increase of one percent in the number of 'unhealthy for sensitive groups' days leads to a decrease in housing prices by 0.0444 percentage points on average significant at the 5% level. Conversely, a one percent increase in the number of 'good' days correlates with an average increase in housing prices by 0.0084, significant at the 10% level. These results show the impact of air quality on housing prices, with a notable impact on days classified as ‘good’ or ‘unhealthy for sensitive groups.’ This suggests that efforts to reduce the number of days with air quality below the 'moderate' level can positively impact housing prices.

We can see with these results that initiatives and policies aimed at improving air quality have the potential to positively impact housing markets. Relatively minor when increasing number of ‘good’ days but at a larger level when decreasing number of ‘unhealthy for sensitive groups’ days. This shows that people are paying less for housing when the pollution levels fall below the moderate level. Overall, the trend of number of ‘good’ days is increasing while other lower air quality day categories are decreasing so keeping current regulations and perhaps state and local governments can focus on pollution levels in counties where the levels fall below ‘moderate’ levels to improve housing prices and quality of life. Governments and local authorities may consider implementing stricter regulations and investing in cleaner technologies to reduce pollution levels especially when days start to fall below the moderate level. These results show that policies that improve air quality not only benefit people’s health but also contribute economically to these communities.

References

Borja-Urbano, S., Rodríguez-Espinosa, F., Luna-Ludeña, M., & Toulkeridis, T. (2021). Valuing the Impact of Air Pollution in Urban Residence Using Hedonic Pricing and Geospatial Analysis, Evidence From Quito, Ecuador. Air, Soil and Water Research, 14.

Cai, Y., Smit, M., & Helbich, M. (2024). Economic effects of air quality on housing prices: evidence from Beijing, China. Journal of Housing and the Built Environment, 1-24. https://doi.org/10.1007/s10901-023-10108-z.

Chay, K. Y., & Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. Journal of Political Economy, 113(2), 376–424.

Choi, A., Kim, P., & Park, A. (2023). The effects of air quality on housing prices: Evidence from the Aliso Canyon gas leak. Journal of Housing Economics, 62(C).

Lu, J., & Lee, Y. (2022). Spatial Effects of Air Pollution on the Housing Market: Evidence from South Korea. Journal of Real Estate Research, 44(1), 131–150. https://doi.org/10.1080/08965803.2021.2008109.

Miłuch, O., & Kopczewska, K. (2024). Fresh air in the city: The impact of air pollution on the pricing of real estate. Environmental Science and Pollution Research, 31, 7604–7627.

Ridker, R. G., & Henning, J. A. (1967). The determinants of residential property values with special reference to air pollution. The Review of Economics and Statistics, 49(3), 246–257.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 82(1), 34–55.

Smith, V. K., & Huang, J.-C. (1995). Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. Journal of Political Economy, 103(1), 209–227.

Tang, M., & Niemeier, D. (2021). How Does Air Pollution Influence Housing Prices in the Bay Area? International Journal of Environmental Research and Public Health, 18(22), 12195.

Wang, J., Wu, K., & Du, Y. (2022). Does air pollution affect urban housing prices? Evidence from 285 Chinese prefecture-level cities. Journal of Cleaner Production, 370, 133480.

Zou, G., Lai, Z., Li, Y., Liu, X., & Li, W. (2022). Exploring the nonlinear impact of air pollution on housing prices: A machine learning approach. Economics of Transportation, 31, 100272. https://doi.org/10.1016/j.ecotra.2022.100272.

Appendix:

```
/*air quality variable- good days*/
PROC IMPORT OUT=airquality /* Specify the name for your first dataset */
        DATAFILE="/home/u60662301/econ/class/airquality.xlsx" /* Path to your Excel file */
        DBMS=XLSX /* Specify the Excel file format */
        REPLACE; /* Replace dataset if it already exists */
   SHEET="airquality"; /* Specify the name of the first sheet */
RUN;

proc sort data=airquality;
   by state county year;
run;




/* % change good days*/
data changeairquality;
   set airquality;
   by state county;

   /* Calculate percentage change for subsequent observations */
   lag_percgooddays = lag(percgooddays);
   if first.County then percent_change_air = .; /* Missing value for the first observation */
   else percent_change_air = (percgooddays - lag_percgooddays) * 100;
```

```
   /* Replace missing values with 0 or any appropriate value */
  *  if missing(percent_change_air) then percent_change_air = 0;
run;




/*HPI- dependent variable*/
PROC IMPORT OUT=county /* Specify the name for your first dataset */
            DATAFILE="/home/u60662301/econ/class/county.xlsx" /* Path to your Excel file */
            DBMS=XLSX /* Specify the Excel file format */
            REPLACE; /* Replace dataset if it already exists */
   SHEET="county"; /* Specify the name of the first sheet */
RUN;
proc sort data=county;
   by state county year;
run;
/*change in HPI variable*/
data hpi;
   set county;
   by State County;

   /* Calculate percentage change for subsequent observations */
   lag_hpi = lag(HPI);
   if first.County then percent_change_hpi = .; /* Missing value for the first observation */
   else percent_change_hpi = ((HPI - lag_hpi) / lag_hpi) * 100;

   /* Replace missing values with 0 or any appropriate value */
  *  if missing(percent_change_hpi) then percent_change_hpi = 0;
```

```
run;



/*merge air quality and HPI*/
data fixed;
   set changeairquality (rename=(Year=Year2));
   length Year $ 5.;
   Year= put(Year2,f5. -L);
   drop Year2;
run;
PROC SORT data=Fixed;
    BY State County Year;
RUN;

data fixed2;
   set hpi (rename=(Year=Year2));
   length Year $ 5.;
   Year= put(Year2,f5. -L);
   drop Year2;
run;
/* Sort dataset2 by State, County, and Year */
proc sort data=fixed2;
  by State County Year;
run;

DATA merged_dataset;
    MERGE Fixed(IN=a) fixed2(IN=b);
```

```
   BY State County Year;
   IF a AND b;
RUN;




/*unemployment variable*/
PROC IMPORT OUT=unem /* Specify the name for your first dataset */
        DATAFILE="/home/u60662301/econ/class/Unemployment.xlsx" /* Path to your Excel
file */
        DBMS=XLSX /* Specify the Excel file format */
        REPLACE; /* Replace dataset if it already exists */
   SHEET="unemploy"; /* Specify the name of the first sheet */
RUN;

proc sort data=unem;
  by State County;
run;

proc transpose data=unem
   out=data (rename=(col1=Unemployment _name_=Year));
   var Y2000 - Y2022;
   by State County ;
run;

data dropvar(drop=_Label_);
   set data;
run;
```

```
data dropy;
   set dropvar;
   kyear = substr(year, 2);
run;

data newest(drop=year);
   set dropy;
run;



data unemploymentpremerge;
   set newest(rename=(kyear=year));
run;


/*merge unemployment*/
DATA merged_dataset2;
   MERGE unemploymentpremerge(IN=a)  merged_dataset(IN=b);
   BY State County Year;
   IF a AND b;
RUN;


/*% change unemployment*/
data newmerge;
   set merged_dataset2;
   by State County;
```

```
   /* Calculate percentage change for subsequent observations */
   lag_unemployment = lag(unemployment);
   if first.County then percent_change_unemployment = .; /* Missing value for the first observation */
   else percent_change_unemployment = ((unemployment - lag_unemployment) * 100);

   /* Replace missing values with 0 or any appropriate value */
  *  if missing(percent_change_unemployment) then percent_change_unemployment = 0;
run;



/*income variable*/
PROC IMPORT OUT=income /* Specify the name for your first dataset */
        DATAFILE="/home/u60662301/econ/class/incomes.xlsx" /* Path to your Excel file */
        DBMS=XLSX /* Specify the Excel file format */
        REPLACE; /* Replace dataset if it already exists */
   SHEET="income"; /* Specify the name of the first sheet */
RUN;

proc sort data=income;
   by state county year;
run;


/*% change income*/
data changeincome;
   set income;
```

```
    by state county;

    /* Calculate percentage change for subsequent observations */
    lag_median= lag(median);
    if first.County then percent_income = .; /* Missing value for the first observation */
    else percent_income = ((median - lag_median)/lag_median) * 100;

    /* Replace missing values with 0 or any appropriate value */
    *if missing(percent_income) then percent_income = 0;
run;

data num;
   set changeincome (rename=(Year=Year2));
   length Year $ 5.;
   Year= put(Year2,f5. -L);
   drop Year2;
run;
PROC SORT data=num;
    BY State County Year;
RUN;



/*merge income*/
DATA variablesmerge;
    MERGE newmerge(IN=a)  num(IN=b);
    BY State County Year;
    IF a AND b;
```

```
RUN;



/*other day variables*/
PROC IMPORT OUT=fullairquality /* Specify the name for your first dataset */
        DATAFILE="/home/u60662301/econ/class/fullairquality.xlsx" /* Path to your Excel file
*/
        DBMS=XLSX /* Specify the Excel file format */
        REPLACE; /* Replace dataset if it already exists */
   SHEET="fullairquality"; /* Specify the name of the first sheet */
RUN;

proc sort data=fullairquality;
   by state2 county year;
run;



/*% change moderate days*/
data changefullairquality;
   set fullairquality;
   by state2 county;

   /* Calculate percentage change for subsequent observations */
   lag_percmoderate= lag(percmoderate);
   if first.County then percent_moderate = .; /* Missing value for the first observation */
   else percent_moderate = ((percmoderate - lag_percmoderate)) * 100;
```

```
   /* Replace missing values with 0 or any appropriate value */
  *  if missing(percent_moderate) then percent_moderate = 0;
run;
/*% change unhealthy sensitive days*/
data changefullairquality1;
   set changefullairquality;
   by state2 county;

   /* Calculate percentage change for subsequent observations */
   lag_percunhealthysens= lag(percunhealthysens);
   if first.County then percent_unhealthysens = .; /* Missing value for the first observation */
   else percent_unhealthysens = ((percunhealthysens - lag_percunhealthysens)) * 100;

   /* Replace missing values with 0 or any appropriate value */
  * if missing(percent_unhealthysens) then percent_unhealthysens = 0;
run;

/*% change unhealthy days*/
data fullairquality2;
   set changefullairquality1;
   by state2 county;

   /* Calculate percentage change for subsequent observations */
   lag_percunhealthy= lag(percunhealthy);
   if first.County then percent_unhealthy = .; /* Missing value for the first observation */
   else percent_unhealthy = ((percunhealthy - lag_percunhealthy)) * 100;

   /* Replace missing values with 0 or any appropriate value */
```

```
  *  if missing(percent_unhealthy) then percent_unhealthy = 0;
run;
/*% change very unhealthy days*/
data fullairquality3;
    set fullairquality2;
    by state2 county;

    /* Calculate percentage change for subsequent observations */
    lag_percveryunhealthy= lag(percveryunhealthy);
    if first.County then percent_veryunhealthy = .; /* Missing value for the first observation */
    else percent_veryunhealthy = ((percveryunhealthy - lag_percveryunhealthy)) * 100;

    /* Replace missing values with 0 or any appropriate value */
  *  if missing(percent_veryunhealthy) then percent_veryunhealthy = 0;
run;
/*% change moderate days*/
data changefullairquality4;
    set fullairquality3;
    by state2 county;

    /* Calculate percentage change for subsequent observations */
    lag_perchazardous= lag(perchazardous);
    if first.County then percent_hazardous = .; /* Missing value for the first observation */
    else percent_hazardous = ((perchazardous - lag_perchazardous)) * 100;

    /* Replace missing values with 0 or any appropriate value */
   * if missing(percent_hazardous) then percent_hazardous = 0;
run;
```

```
/*merge all days*/
data fix;
  set changefullairquality4 (rename=(Year=Year2));
  length Year $ 5.;
  Year= put(Year2,f5. -L);
  drop Year2;
run;
PROC SORT data=fix;
   BY State2 County Year;
RUN;

PROC SORT data=variablesmerge;
   BY State2 County Year;
RUN;
DATA variablesmerge2;
   MERGE variablesmerge(IN=a)  fix(IN=b);
   BY State2 County Year;
   IF a AND b;

RUN;

PROC SORT data=variablesmerge2;
   BY State County Year;
RUN;

/*population variable*/
PROC IMPORT OUT=population
```

```
        DATAFILE="/home/u60662301/econ/class/pop.xlsx" /* Path to your Excel file */
        DBMS=XLSX /* Specify the Excel file format */
        REPLACE; /* Replace dataset if it already exists */
    SHEET="Population by Age and Sex - US,"; /* Specify the name of the first sheet */
RUN;
proc sort data=population;
    by state2 county year;
run;

data changepop;
    set population;
  by state2 county;

    /* Calculate percentage change for subsequent observations */
    lag_population=lag(population);
    if first.County then percent_population = . ;
    else percent_population = ((population - lag_population)/lag_population) * 100;

    /* Replace missing values with 0 or any appropriate value */
    *if missing(percent_income) then percent_income = 0;
run;






proc sort data=population;
    by state2 county year;
```

```
run;

data fixpop;
   set changepop (rename=(Year=Year2));
   length Year $ 5.;
   Year= put(Year2,f5. -L);
   drop Year2;
run;



data changepop;
    set population;
  by state2 county;

    /* Calculate percentage change for subsequent observations */
    lag_population=lag(population);
    if first.County then percent_population = . ;
    else percent_population = ((population - lag_population)/lag_population) * 100;

    /* Replace missing values with 0 or any appropriate value */
    *if missing(percent_income) then percent_income = 0;
run;



PROC SORT data=variablesmerge2;
    by state2 county year;
RUN;
```

```
data changepop;
   set changepop (rename=(Year=Year2));
   length Year $ 5.;
   Year= put(Year2,f5. -L);
   drop Year2;
run;



DATA merged_datasetpop;
    MERGE variablesmerge2(IN=a)  changepop(IN=b);
     by state2 county year;
    IF a AND b;
RUN;

/*model 1* good days*/
ods output ParameterEstimates=PEforModel1 DataSummary=ObsModel1
FitStatistics=AdjRsqModel1 Effects=OverallSigModel1;
Proc SurveyReg data= merged_datasetpop;
                                                    Class State County Year;
                                                    Model1: Model percent_change_hpi = percent_change_air
    County State Year/Solution AdjRsq;
Run;

/*model2* good days with variables*/
ods output ParameterEstimates=PEforModel2 DataSummary=ObsModel2
FitStatistics=AdjRsqModel2 Effects=OverallSigModel2;
Proc SurveyReg data= merged_datasetpop;
```

```
Class State County Year;

Model2: Model percent_change_hpi = percent_change_air

percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;

Run;

/*model3* moderate days*/

ods output ParameterEstimates=PEforModel3 DataSummary=ObsModel3
FitStatistics=AdjRsqModel3 Effects=OverallSigModel3;

Proc SurveyReg data= merged_datasetpop;

Class State County Year;

Model3: Model percent_change_hpi = percent_moderate

County State Year/Solution AdjRsq;

Run;


/*model4* moderate days with variables*/

ods output ParameterEstimates=PEforModel4 DataSummary=ObsModel4
FitStatistics=AdjRsqModel4 Effects=OverallSigModel4;

Proc SurveyReg data= merged_datasetpop;

Class State County Year;

Model4: Model percent_change_hpi = percent_moderate

percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;

Run;


/*model5* unhealthy for sensitive days */

ods output ParameterEstimates=PEforModel5 DataSummary=ObsModel5
FitStatistics=AdjRsqModel5 Effects=OverallSigModel5;

Proc SurveyReg data= merged_datasetpop;

Class State County Year;
```

```
                                        Model5: Model percent_change_hpi =
percent_unhealthysens
   County State Year/Solution AdjRsq;
Run;


/*model6* unhealthy for sensitive days with variables */
ods output ParameterEstimates=PEforModel6 DataSummary=ObsModel6
FitStatistics=AdjRsqModel6 Effects=OverallSigModel6;
Proc SurveyReg data= merged_datasetpop;
                                        Class State County Year;
                                        Model6: Model percent_change_hpi =
percent_unhealthysens
  percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;
Run;


/*model7* unhealthy days */
ods output ParameterEstimates=PEforModel7 DataSummary=ObsModel7
FitStatistics=AdjRsqModel7 Effects=OverallSigModel7;
Proc SurveyReg data= merged_datasetpop;
                                        Class State County Year;
                                        Model7: Model percent_change_hpi = percent_unhealthy
   County State Year/Solution AdjRsq;
Run;


/*model8* unhealthy days with variables */
ods output ParameterEstimates=PEforModel8 DataSummary=ObsModel8
FitStatistics=AdjRsqModel8 Effects=OverallSigModel8;
Proc SurveyReg data= merged_datasetpop;
                                        Class State County Year;
```

```
                                  Model8: Model percent_change_hpi = percent_unhealthy
  percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;
Run;

/*model9* very unhealthy days */
ods output ParameterEstimates=PEforModel9 DataSummary=ObsModel9
FitStatistics=AdjRsqModel9 Effects=OverallSigModel9;
Proc SurveyReg data= merged_datasetpop;
                                  Class State County Year;
                                  Model9: Model percent_change_hpi =
percent_veryunhealthy
   County State Year/Solution AdjRsq;
Run;
/*model10* very unhealthy days with variables */
ods output ParameterEstimates=PEforModel10 DataSummary=ObsModel10
FitStatistics=AdjRsqModel10 Effects=OverallSigModel10;
Proc SurveyReg data= merged_datasetpop;
                                  Class State County Year;
                                  Model10: Model percent_change_hpi =
percent_veryunhealthy
  percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;
Run;
/*model11* hazardous days */
ods output ParameterEstimates=PEforModel11 DataSummary=ObsModel11
FitStatistics=AdjRsqModel11 Effects=OverallSigModel11;
Proc SurveyReg data= merged_datasetpop;
                                  Class State County Year;
                                  Model11: Model percent_change_hpi =
percent_hazardous
```

```
   County State Year/Solution AdjRsq;
Run;
/*model12* hazardous days with variables */
ods output ParameterEstimates=PEforModel12 DataSummary=ObsModel12
FitStatistics=AdjRsqModel12 Effects=OverallSigModel12;
Proc SurveyReg data= merged_datasetpop;
                    Class State County Year;
                    Model12: Model percent_change_hpi =
percent_hazardous
   percent_change_unemployment percent_income percent_population County State
Year/Solution AdjRsq;
Run;


ods output ParameterEstimates=PEforModel10 DataSummary=ObsModel10
FitStatistics=AdjRsqModel10 Effects=OverallSigModel10;


proc surveymeans data=merged_datasetpop;
   class State County Year;
   var percent_change_hpi percent_veryunhealthy percent_change_unemployment
percent_income percent_population;
run;


******************************;
/* Build the results table */
******************************;
/*Table 1: Good and Mod days*/
/* Step 1: clean-up the output of the regression analysis you have saved */
Data Table_Long;
                    length Model $10; /* Makes sure the variable Model has
the right length and its values are not truncated */
```

run;

```
length Parameter $30; /* Makes sure the variable Parameter has the right length and its values are not truncated */
set PEforModel1 PEforModel2 PEforModel3 PEforModel4   indsname=M; /*"indsname" creates an indicator variable (here I call it "M") that tracks the name of databases use in the "set" statement */
keep Model Parameter EditedResults;
if      M="WORK.PEFORMODEL1" then Model="Model1";
        else if M="WORK.PEFORMODEL2" then Model="Model2";
else if M="WORK.PEFORMODEL3" then Model="Model3";
else if M="WORK.PEFORMODEL4" then Model="Model4";


if Probt le 0.01 then Star="***";
        else if Probt le 0.05 then Star="**";
        else if Probt le 0.1 then Star="*";


Results=Estimate;
EditedResults=Cats(put(Results,comma16.4),Star);
output;

Results=stderr;
EditedResults=Cats("(",put(Results,comma16.4),")");
output;

run;
```

```
/* We sometimes need this sorting step when we have multiple regression models */

proc sort data=Table_Long out=Table_Long_Sorted;

                              by Model Parameter;

run;



/* Step 2: Create separate results columns (in the form of separate databases) corresponding to
each model */

data Model1Results(rename=(EditedREsults=Model1))

                                   Model2Results(rename=(EditedREsults=Model2))

                                   Model3Results(rename=(EditedREsults=Model3))

                                   Model4Results(rename=(EditedREsults=Model4));



                              set Table_Long_Sorted;

                              if Model="Model1" then output Model1Results;

                                   else if Model="Model2" then output
Model2Results;

                                   else if Model="Model3" then output
Model3Results;

                                   else if Model="Model4" then output
Model4Results;



                              drop Model;

run;


data Table_Wide;

                              merge Model1Results Model2Results Model3Results
Model4Results Model5Results Model6Results;
```

```
                                    by Parameter;
                                    if mod(_n_,2)=1 then Regressors=Parameter;


                                    length Order 3;
run;
/* Order the variables in the results table */
proc sort data=Table_Wide out=Table_Wide_Sorted(drop=Order Parameter);
                                    by Order;
run;



Data NumofObs;
                                    merge ObsModel1(rename=(Nvalue1=NVModel1)
drop=CValue1) ObsModel2(rename=(Nvalue1=NVModel2) drop=CValue1)
ObsModel3(rename=(Nvalue1=NVModel3) drop=CValue1)
ObsModel4(rename=(Nvalue1=NVModel4) drop=CValue1);
                                    where Label1="Number of Observations";
                                    Model1=put(NVModel1,comma16.);
                                    Model2=Put(NVModel2,comma16.);
                                    Model3=put(NVModel3,comma16.);
                                    Model4=put(NVModel4,comma16.);



                                    keep Label1 Model1 Model2 Model3 Model4;
run;


/* The row for the adjusted R-Squared */
Data AdjRsq;
                                    merge AdjRsqModel1(rename=(cvalue1=Model1)
drop=nvalue1)
```

```
                                        AdjRsqModel2(rename=(cvalue1=Model2)
drop=nvalue1)
                                        AdjRsqModel3(rename=(cvalue1=Model3)
drop=nvalue1)
                                        AdjRsqModel4(rename=(cvalue1=Model4)
drop=nvalue1)
;
                                  Where Label1="Adjusted R-Square";
run;


/* The row for the F-test related to the Overall Significance of the model */
Data OSM1(rename=(EditedValue=Model1)) OSM2(rename=(EditedValue=Model2))
OSM3(rename=(EditedValue=Model3)) OSM4(rename=(EditedValue=Model4));
                                  set OverallSigModel1 OverallSigModel2
OverallSigModel3 OverallSigModel4  indsname=M;
                                  where Effect="Model";
                                  if ProbF le 0.01 then Star="***";
                                       else if ProbF le 0.05 then Star="**";
                                       else if ProbF le 0.1 then Star="*";
                                  THisISM=M;


                                  Label1="Overall Significance";
                                  EditedValue=cats(put(FValue,comma16.2),Star);


                                       if M="WORK.OVERALLSIGMODEL1" then
output OSM1;
                                       else if M="WORK.OVERALLSIGMODEL2" then
output OSM2;
                                       else if M="WORK.OVERALLSIGMODEL3" then
output OSM3;
```

```
        else if M="WORK.OVERALLSIGMODEL4" then
output OSM4;


    keep Label1 EditedValue;
run;

Data OverallSig;
    merge OSM1 OSM2 OSM3 OSM4;
    by Label1;
run;


/* Combine all rows for other statistics */
data OtherStat;
    set NumofObs AdjRsq OverallSig;
    rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStat;
    set Table_Wide_Sorted OtherStat;
run;

/* create new name for variables in the regression results table through defining a new format*/
proc format;
    value $VariableName(default=50) ;
```

```
/* Print the clean results table */

ods excel file="/home/u60662301/MySAS/capstonemanymods.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */

Title "Table 1: Air Quality on Hosuing Prices";

footnote "Source: Federal Housing Finance Agency, United States Environmental Agency, and
own calculations.

Notes: robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1%
significance levels, respectively";

proc print data=Table_Wide_Sorted_withStat noobs;

                    var regressors;

                    var Model1-Model4 / style(header)={Just=Center}
style(data)={Just=Center};

                    format Regressors $VariableName.;

run;

ods excel close;
```

```
/* Table 4 Unhealhy Sens and Unhealthy Days*/

/* Step 1: clean-up the output of the regression analysis you have saved */

Data Table_Long;

                    length Model $10; /* Makes sure the variable Model has
the right length and its values are not truncated */

                    length Parameter $30; /* Makes sure the variable
Parameter has the right length and its values are not truncated */
```

```
set PEforModel5 PEforModel6 PEforModel7
PEforModel8   indsname=M; /*"indsname" creates an indicator variable (here I call it "M") that
tracks the name of databases use in the "set" statement */

keep Model Parameter EditedResults;

if      M="WORK.PEFORMODEL5" then
Model="Model5";

        else if M="WORK.PEFORMODEL6" then
Model="Model6";

else if M="WORK.PEFORMODEL7" then
Model="Model7";

else if M="WORK.PEFORMODEL8" then
Model="Model8";



if Probt le 0.01 then Star="***";
        else if Probt le 0.05 then Star="**";
        else if Probt le 0.1 then Star="*";



Results=Estimate;
EditedResults=Cats(put(Results,comma16.4),Star);
output;


Results=stderr;
EditedResults=Cats("(",put(Results,comma16.4),")");
output;


run;


/* We sometimes need this sorting step when we have multiple regression models */
```

```
proc sort data=Table_Long out=Table_Long_Sorted;
        by Model Parameter;
run;


/* Step 2: Create separate results columns (in the form of separate databases) corresponding to
each model */
data Model5Results(rename=(EditedREsults=Model5))
          Model6Results(rename=(EditedREsults=Model6))
          Model7Results(rename=(EditedREsults=Model7))
          Model8Results(rename=(EditedREsults=Model8));


        set Table_Long_Sorted;
        if Model="Model5" then output Model5Results;
          else if Model="Model6" then output
Model6Results;
          else if Model="Model7" then output
Model7Results;
          else if Model="Model8" then output
Model8Results;


        drop Model;
run;

data Table_Wide;
        merge Model5Results Model6Results Model7Results
Model8Results ;
        by Parameter;
```

```
                    if mod(_n_,2)=1 then Regressors=Parameter;

                    length Order 3;
run;
/* Order the variables in the results table */
proc sort data=Table_Wide out=Table_Wide_Sorted(drop=Order Parameter);
                    by Order;
run;



Data NumofObs;
                    merge ObsModel5(rename=(Nvalue1=NVModel5)
drop=CValue1) ObsModel6(rename=(Nvalue1=NVModel6) drop=CValue1)
ObsModel7(rename=(Nvalue1=NVModel7) drop=CValue1)
ObsModel8(rename=(Nvalue1=NVModel8) drop=CValue1) ;
                    where Label1="Number of Observations";
                    Model5=put(NVModel5,comma16.);
                    Model6=Put(NVModel6,comma16.);
                    Model7=put(NVModel7,comma16.);
                    Model8=put(NVModel8,comma16.);



                    keep Label1 Model5 Model6 Model7 Model8;
run;

/* The row for the adjusted R-Squared */
Data AdjRsq;
                    merge AdjRsqModel5(rename=(cvalue1=Model5)
drop=nvalue1)
```

```
AdjRsqModel6(rename=(cvalue1=Model6)
drop=nvalue1)
AdjRsqModel7(rename=(cvalue1=Model7)
drop=nvalue1)
AdjRsqModel8(rename=(cvalue1=Model8)
drop=nvalue1);

Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM5(rename=(EditedValue=Model5)) OSM6(rename=(EditedValue=Model6))
OSM7(rename=(EditedValue=Model7)) OSM8(rename=(EditedValue=Model8));
set OverallSigModel5 OverallSigModel6
OverallSigModel7 OverallSigModel8 indsname=M;
where Effect="Model";
if ProbF le 0.01 then Star="***";
    else if ProbF le 0.05 then Star="**";
    else if ProbF le 0.1 then Star="*";
THisISM=M;

Label1="Overall Significance";
EditedValue=cats(put(FValue,comma16.2),Star);

    if M="WORK.OVERALLSIGMODEL5" then
output OSM5;
    else if M="WORK.OVERALLSIGMODEL6" then
output OSM6;
    else if M="WORK.OVERALLSIGMODEL7" then
output OSM7;
```

```
else if M="WORK.OVERALLSIGMODEL8" then
output OSM8;

keep Label1 EditedValue;
run;

Data OverallSig;
merge OSM5 OSM6 OSM7 OSM8 ;
by Label1;
run;

/* Combine all rows for other statistics */
data OtherStat;
set NumofObs AdjRsq OverallSig;
rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStat;
set Table_Wide_Sorted OtherStat;
run;

/* create new name for variables in the regression results table through defining a new format*/
proc format;
value $VariableName(default=50) ;
```

```
/* Print the clean results table */

ods excel file="/home/u60662301/MySAS/capstonemanymods2.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */

Title "Table 1: Air Quality on Housing Prices";

footnote "Source: Federal Housing Finance Agency, United States Environmental Agency, and
own calculations.

Notes: robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1%
significance levels, respectively";

proc print data=Table_Wide_Sorted_withStat noobs;

        var regressors;

        var Model5-Model8 / style(header)={Just=Center}
style(data)={Just=Center};

        format Regressors $VariableName.;

run;

ods excel close;




/* Table 4 Very Unhealhy Sens and hazardous Days*/

/* Step 1: clean-up the output of the regression analysis you have saved */

Data Table_Long;

        length Model $10; /* Makes sure the variable Model has
the right length and its values are not truncated */

        length Parameter $30; /* Makes sure the variable
Parameter has the right length and its values are not truncated */

        set PEforModel9 PEforModel10 PEforModel11
PEforModel12   indsname=M; /*"indsname" creates an indicator variable (here I call it "M") that
tracks the name of databases use in the "set" statement */

        keep Model Parameter EditedResults;

        if      M="WORK.PEFORMODEL9" then
Model="Model9";
```

```
else if M="WORK.PEFORMODEL10" then
Model="Model10";
else if M="WORK.PEFORMODEL11" then
Model="Model11";
else if M="WORK.PEFORMODEL12" then
Model="Model12";


if Probt le 0.01 then Star="***";
    else if Probt le 0.05 then Star="**";
    else if Probt le 0.1 then Star="*";


Results=Estimate;
EditedResults=Cats(put(Results,comma16.4),Star);
output;

Results=stderr;
EditedResults=Cats("(",put(Results,comma16.4),")");
output;

run;

/* We sometimes need this sorting step when we have multiple regression models */
proc sort data=Table_Long out=Table_Long_Sorted;
    by Model Parameter;
run;
```

```
/* Step 2: Create separate results columns (in the form of separate databases) corresponding to
each model */

data Model9Results(rename=(EditedREsults=Model9))

                                        Model10Results(rename=(EditedREsults=Model10))

                                        Model11Results(rename=(EditedREsults=Model11))

                                        Model12Results(rename=(EditedREsults=Model12));



                                        set Table_Long_Sorted;
                                        if Model="Model9" then output Model9Results;
                                             else if Model="Model10" then output
Model10Results;
                                             else if Model="Model11" then output
Model11Results;
                                             else if Model="Model12" then output
Model12Results;



                                        drop Model;
run;


data Table_Wide;
                                        merge Model9Results Model10Results Model11Results
Model12Results ;

                                        by Parameter;
                                        if mod(_n_,2)=1 then Regressors=Parameter;


                                        length Order 3;
```

```
run;

/* Order the variables in the results table */

proc sort data=Table_Wide out=Table_Wide_Sorted(drop=Order Parameter);

                                        by Order;

run;



Data NumofObs;

                                        merge ObsModel9(rename=(Nvalue1=NVModel9)
drop=CValue1) ObsModel10(rename=(Nvalue1=NVModel10) drop=CValue1)
ObsModel11(rename=(Nvalue1=NVModel11) drop=CValue1)
ObsModel12(rename=(Nvalue1=NVModel12) drop=CValue1) ;

                                        where Label1="Number of Observations";

                                        Model9=put(NVModel9,comma16.);

                                        Model10=Put(NVModel10,comma16.);

                                        Model11=put(NVModel11,comma16.);

                                        Model12=put(NVModel12,comma16.);



                                        keep Label1 Model9 Model10 Model11 Model12;

run;


/* The row for the adjusted R-Squared */

Data AdjRsq;

                                        merge AdjRsqModel9(rename=(cvalue1=Model9)
drop=nvalue1)

                                                AdjRsqModel10(rename=(cvalue1=Model10)
drop=nvalue1)

                                                AdjRsqModel11(rename=(cvalue1=Model11)
drop=nvalue1)
```

```
                                        AdjRsqModel12(rename=(cvalue1=Model12)
drop=nvalue1);

                                        Where Label1="Adjusted R-Square";
run;

/* The row for the F-test related to the Overall Significance of the model */
Data OSM9(rename=(EditedValue=Model9)) OSM10(rename=(EditedValue=Model10))
OSM11(rename=(EditedValue=Model11)) OSM12(rename=(EditedValue=Model12));
                                        set OverallSigModel9 OverallSigModel10
OverallSigModel11 OverallSigModel12  indsname=M;
                                        where Effect="Model";
                                        if ProbF le 0.01 then Star="***";
                                              else if ProbF le 0.05 then Star="**";
                                              else if ProbF le 0.1 then Star="*";
                                        THisISM=M;

                                        Label1="Overall Significance";
                                        EditedValue=cats(put(FValue,comma16.2),Star);

                                              if M="WORK.OVERALLSIGMODEL9" then
output OSM9;
                                              else if M="WORK.OVERALLSIGMODEL10"
then output OSM10;
                                              else if M="WORK.OVERALLSIGMODEL11"
then output OSM11;
                                              else if M="WORK.OVERALLSIGMODEL12"
then output OSM12;

                                        keep Label1 EditedValue;
```

```
run;

Data OverallSig;
        merge OSM9 OSM10 OSM11 OSM12 ;
        by Label1;
run;


/* Combine all rows for other statistics */
data OtherStat;
        set NumofObs AdjRsq OverallSig;
        rename Label1=Regressors;
run;

/* Add rows for other statistics to the table */
Data Table_Wide_Sorted_withStat;
        set Table_Wide_Sorted OtherStat;
run;

/* create new name for variables in the regression results table through defining a new format*/
proc format;
        value $VariableName(default=50) ;


/* Print the clean results table */
ods excel file="/home/u60662301/MySAS/capstonemanymods3.xlsx"
options(Embedded_Titles="ON" Embedded_Footnotes="ON"); /*Use the path to your MySAS
folder */
Title "Table 1: Air Quality on Housing Prices";
```

```
footnote "Source: Federal Housing Finance Agency, United States Environmental Agency, and own calculations.

Notes: robust standard errors are in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance levels, respectively";

proc print data=Table_Wide_Sorted_withStat noobs;

                                var regressors;

                                var Model9-Model12 / style(header)={Just=Center} style(data)={Just=Center};

                                format Regressors $VariableName.;

run;

ods excel close;



/* Example using Proc Means */

proc means data=merged_datasetpop n mean std min max  median;

  var percent_change_hpi percent_veryunhealthy percent_change_unemployment percent_income percent_population  percent_change_air percent_moderate percent_unhealthysens percent_unhealthy percent_hazardous ;

run;
```