Do Homeowners Value Air Quality: Evidence From Wildfire Smoke

Tristan Misko

1 December 2021

Abstract

Main Idea and Setting

Chronic air pollution poses a significant threat to public health.

Housing prices should reflect significant determinants of quality of life like air quality Argal, housing prices should respond to air quality changes

For many households in the United States, the decision to purchase a home is the most important financial transaction

Purchasing a home is a fundamental asset that affects all facets of life.
The decisions about where to live depend on many factors, among them employment opportunities, neighborhood characterisitics, cultural amenities, local school quality, proximity to friends and family, ideosyncratic locational preferences, and environmental amenities. Since families base their decisions on where to live on the

Economists have long sought to represent housing prices as a hedonic function of the particular amenities or goods that a home provides to understand the relative value homeowners place on those goods.

determine whether housing prices respond to

changes to in air quality. The naive regression of housing prices on air quality may suffer from significant endogeneity. If homeowners respond to air quality, then it is likely that taste-based sorting occurs, producing a selection bias.
To overcome endogeneity, we use wildfire smoke as a source of exogenous variation in air quality and exploit the heterogeneous increase in wildfire smoke since 2015 to obtain casual estimates. If homeowners value air

quality, then we would expect increases in wildfire smoke to cause decreases in home values.

The data take the form of panel data, with monthly observations at the United States county level of the number of days in each month in which the county is covered by wildfire smoke plumes, the mean air quality index (AQI) over the month, and the level of the Zillow housing price index in that month. I also have associated to each observation a set of controls for unemployment level and housing characteristics. The data range from June 2010 to July 2019 and contain counties outside of the geographic west of the United States (such counties may suffer potentially significant confounding because wildfire events are heavily correlated with smoke events and may also affect housing prices).

Equations to be estimated

We first estimate the naive OLS equation both with and without unemployment controls:

$$\mathrm{ZHVI}_{c,t} = \beta_{\mathrm{OLS}} \cdot \mathrm{smoke}_{c,t} + \gamma \cdot \mathrm{unemp}_{c,t} + F_c + T_t + \epsilon_{c,t}.$$

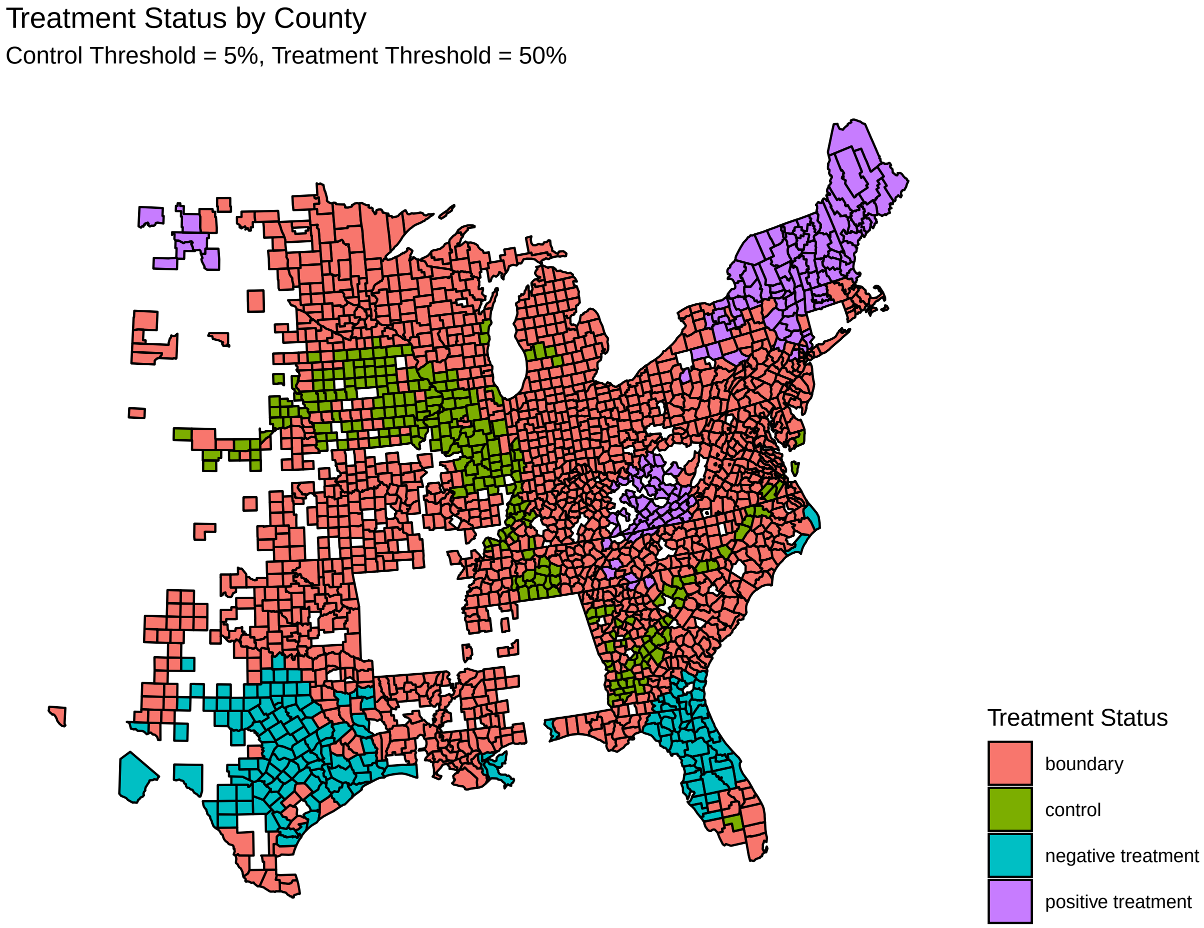
This regression certainly suffers from significant endogeneity and hence has a biased coefficient, but it gives a starting point for understanding the more sophisticated designs.

Next, we run our main regression of interest, given by

$$\mathrm{ZHVI}_{c,t} = \beta \cdot \mathrm{smoke\_treatment}_{c,t}^{(t^*=55)} + \gamma \cdot \mathrm{unemp}_{c,t} + F_c + T_t + \epsilon_{c,t}.$$

The dummy $\mathrm{smoke\_treatment}_{c,t}^{(t^*=55)}$ turns on if $t > 55$, corresponding to the months following January 2015 and later, and if county c is in the treatment group. To characterize treatment status, we choose threshold values for the percentage change in smoke score across the pre- and post-treatment periods. The threshold values we select are arbitrary, therefore all results must be carefully analyzed for sensitivity and robustness (see below). For the main regression, we characterize a county as "treated positive" if the increase in the mean smoke score from the pre- to post-treatment period is greater than 50%, as "control" if the magnitude of the change in means is less than 5%, as "treated negative" if the decrease is greater than 50% in magnitude, and as "boundary" otherwise.

The final set of models which we estimate are those which have buckets for smoke exposure instead of omitting the "boundary" counties. I am still in the process of setting up this model. The bucket selections will also be subject to the robustness checks discussed below.



Methodology

Results

We summarize the results of the two regressions above in Table 1.

The coefficient from equation (1), the standard OLS equation, is not significantly different from zero. Its interpretation is that a one point increase in the smoke score in a given month, roughly corresponding to a 1 μg increase in wildfire smoke exposure over the course of the month, is associated with a -0.6412 drop in ZHVI.

The coefficient on equation (2) is negative and highly significant. Its interpretation is that the ZHVI (roughly interpreted as the smoothed median home price in the county in dollars) is expected to fall by 1484.0340 given a 50% long term increase in wildfire smoke exposure. Here, long term refers to the difference between the 2010-2014 and the 2015-2019 means.

Table 1: Effect of Wildfire Smoke on Housing Prices

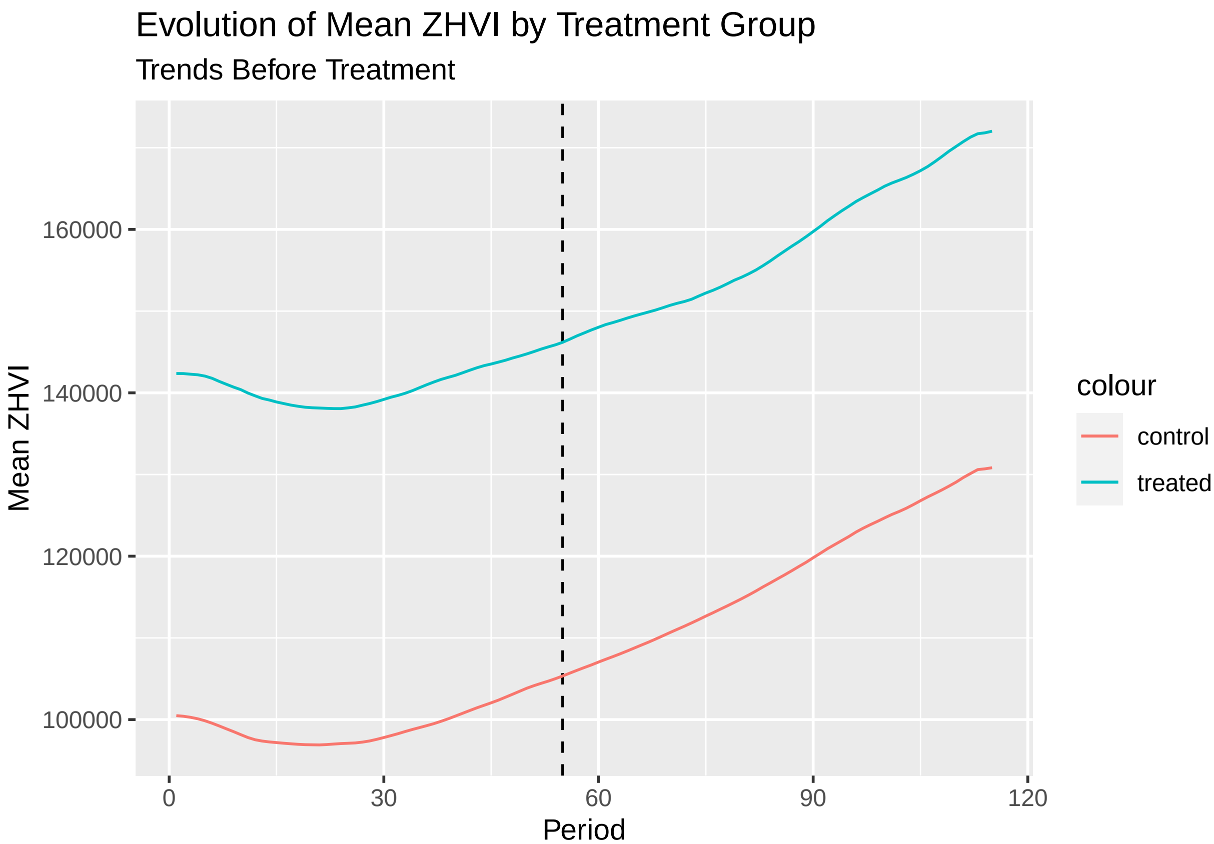
	(1)	(2)
(Intercept)	183687.00***	73976.901
	(1100.8998)	(1113.0914)
n.score	−0.6631	
	(0.6412)	
treat.k5.t50		−1484.0340***
		(168.9167)
unemp	1114.7302***	1563.9206***
	(27.7190)	(48.1914)
County Fixed Effects	Yes	Yes
Time Fixed Effects	Yes	Yes
Observations	234600	45655
R^2	0.9804	0.98367

Robustness Checks

Robustness checks will be a major feature of my paper because of the number of threshold choices I have had to make in the course of my analysis.

Parallel Trends: We plot the mean ZHVI across the treatment and control groups in each period and examine the evolution in time. From the graph, it is visually clear that the trends in the means are close to parallel, and it seems that there is a small dip in the ZHVI for the treatment group after treatment.

```
## `summarise()` has grouped output by 'status.k5.t50'. You can override using the `.groups` argument.
```



Effect Size Threshold for Treatment Status: One major choice which will likely have significant bearing on the resultant regression coefficient is the treatment status. Changing this threshold affects which counties are included in the treatment and control groups. To address robustness, I plan to produce a plot which varies the threshold continuously on the x axis and plot the resultant regression coefficient on the y axis. Tbis plot will give visual insight into the sensitivity of the effect size estimate to the threshold choice.

Treatment Start Time: I will also run the models with different periods in which the treatment begins.

Table 2: Summary Statistics of Main Datafile Variables

Statistic	N	Mean	St. Dev.	Min	Max
year	234,600	2,014.696	2.773	2,010	2,019
month	234,600	6.652	3.434	1	12
zhvi.score	234,600	135,041.500	79,504.160	22,372	1,646,548
SizeRank	234,600	1,328.630	787.017	1	3,071
n.light	234,600	2.748	4.662	0.000	30.714
n.medium	234,600	0.584	1.625	0.000	25.995
n.heavy	234,600	0.130	0.550	0	12
n.score	234,600	25.749	55.318	0.000	820.000
unemp	234,600	5.908	2.505	1.100	21.648
post43	234,600	0.626	0.484	0	1
post48	234,600	0.583	0.493	0	1
post55	234,600	0.522	0.500	0	1
post60	234,600	0.478	0.500	0	1
post67	234,600	0.417	0.493	0	1
m.s.pre43	234,600	28.083	15.841	5.602	61.858
m.s.post43	234,600	24.355	15.515	2.782	75.967
m.s.delta43	234,600	−3.729	10.237	−36.251	30.297
m.s.pch43	234,600	−0.059	0.403	−0.820	1.634
m.s.pre48	234,600	25.528	14.411	5.129	56.054
m.s.post48	234,600	25.907	16.633	2.911	81.352
m.s.delta48	234,600	0.378	10.704	−31.451	39.691
m.s.pch48	234,600	0.108	0.493	−0.796	2.159
m.s.pre55	234,600	26.099	14.801	4.814	60.081
m.s.post55	234,600	25.428	15.898	2.941	79.808
m.s.delta55	234,600	−0.671	8.164	−25.744	29.185
m.s.pch55	234,600	0.031	0.338	−0.754	1.295
m.s.pre60	234,600	24.754	14.237	4.637	57.425
m.s.post60	234,600	26.834	16.644	2.963	83.996
m.s.delta60	234,600	2.081	8.551	−22.197	34.476
m.s.pch60	234,600	0.156	0.387	−0.728	1.536
m.s.pre67	234,600	26.489	15.616	4.153	66.977
m.s.post67	234,600	24.716	14.460	3.395	74.525
m.s.delta67	234,600	−1.773	6.198	−19.957	17.780
m.s.pch67	234,600	−0.012	0.265	−0.683	0.992