Spencer Andersen

Write Up

Econ 388

Data Assignment 3

**Introduction**

This project analyzes the effect that Covid-19 had on house prices. Specifically, the death rate per county during 2020 (the height of the pandemic) was measured to see what effect, if any, that it had on the house prices of the following year.

The fear of death and contagion caused an uproar nationally and internationally, greatly affecting the economy. Housing prices are one effective measure of economic stability, so this question has high relevance for today’s politics and economics. Did the number of deaths in 2020 affect the housing prices for 2021?

**Data**

The data used in this project came from four sources. The New York Times provided data on mortality rates in 2020 due to Covid-19. Population data by county was found at Census.gov. Fhfa.gov gave us data for Zip-level housing price data, and data showing zip-codes with their corresponding county was at huduser.gov. Many of the data files were cleaned and prepared in Excel for use in Stata.

Each source provided a file of data needed to examine the question of interest. First we merged the first two datasets based on the fips code so that we had one dataset with combined mortality rates and population rates per county. Then we merged the last 2 datasets into one so that we could have a dataset showing the housing price indexes based on zip code with its corresponding county. Then we merged the two integrated datasets into one final dataset which was organized and collapsed to have one observation per county (of which there were 3130 nationally), with all the variables of interest present.

The following table shows key statistics for each variable of interest. Population, deaths, and cases data is all from 2020, whereas there is house price data from both 2020 and 2021 to view the difference.

**Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. Dev. | Min | Max |
| population | 3130 | 103069.64 | 328354.41 | 64 | 10014009 |
| deaths | 3130 | 101.625 | 360.585 | 0 | 10345 |
| cases | 3130 | 6192.079 | 22309.75 | 1 | 770915 |
| deathrate | 3130 | .001 | .001 | 0 | .009 |
| caserate | 3130 | .067 | .028 | .002 | .278 |
| IndNSA21 | 3130 | 255.342 | 46.109 | 161.206 | 534.797 |
| IndNSA20 | 3130 | 231.174 | 39.083 | 147.496 | 500.1 |
| IndNSAChange | 3130 | 24.168 | 10.19 | -3.562 | 81.334 |
|  | | | | | |

**Empirical Analysis**

The empirical strategy used was to regress the housing price change from 2020 to 2021 (IndNSAChange) on the death rate (deathrate) as well as on other variables. The outcome of interest is the IndNSAChange, whereas the main regressor is the deathrate. A simple scatterplot of the primary two variables of interest shows there is some negative correlation, as shown below:

****

There was little data to provide controls, but one control used was the case rate (caserate) which was found by dividing the number of 2020 Covid cases recorded in each county by the 2020 population of that county. (The death rate was similarly found by dividing the total deaths per county by the population of the county).

IndNSAChange was calculated by taking the house price index of each county in 2021 and subtracting the house price index of the county in 2020 from that, giving the change in house price index from one year to the next. These values are heavily positive, presumably because of inflation and increased demand for houses due to the quarantine and commonplace of working from home. We do not have data on those variables, but it seems reasonable to infer those were instrumental in changing the price. The housing price index of each year is highly correlated with the next year, which makes sense because houses are a good store of value, unlike cryptocurrency for example.

When the regression of IndNSAChange on deathrate and caserate was performed, it was found that both deathrate and caserate have negative partial effects on IndNSAChange. This is interesting because even though both of these variables had negative effects on house prices, house prices still skyrocketed during the period, leading us to believe we have a serious omitted variable bias problem.

In order to perform the regressions, we assumed that the deathrate was uncorrelated with the error term, and that there were no omitted variables (said assumption was unlikely to hold). There is no strong argument that the deathrate is correlated with the error term, so that assumption is likely valid. Because of these results, it appears that deaths do decrease house prices, possibly because of the slight decrease in people to live in them.

**Conclusion**

Like every paper, this one has limitations. For example, when collapsing on "collapse (mean) IndNSA21 IndNSA20, by(fips county state deaths cases population y21 y20)", it weights each zip code equally even though there might be more of one zip code than others per county. This could affect the population numbers and bias the results. Another limitation would be that we don’t have population data for 2021 to see the change in population for each county.

It could be called into question whether the deathrate is a good indicator for the price change. The quarantine and economic shutdowns involved may have been more instrumental. Further analysis could be done to see what effect that specifically the quarantine or people having a higher propensity of working at home had on housing prices of the following year. In conclusion, the pandemic had a large effect on housing prices, one of the reasons for which was the high deathrate caused by Covid-19.