

Causal Inference for the Effect of COVID-19 Mortality on Per-County Housing Prices

Sam Lee

Introduction

In this analysis I seek to estimate the causal effect that 2020 COVID-19 mortality rates had on per-county housing prices in the United States in 2021. According to elementary economic theory, all else equal, when consumers of a good leave a market, demand decreases. This implies that prices decrease at every quantity demanded (Law of Demand). This theory may be grimly applied when we analyze COVID-19 mortality deaths from 2020. According to the Law of Demand, all else equal, we would assume that on average, housing prices would decrease given the decrease in the number of consumers in the housing market per county. Not only can this be empirically tested using econometric methods, but additionally, this could provide compelling evidence for why providing early care for COVID-related illnesses may have real economic benefits and impacts in external markets.

Data

To estimate the effect that 2020 COVID-19 mortality rates had on per-county housing prices in 2021, I merged per-county COVID-19 mortality rates with the per-county housing prices. Per-county COVID-19 mortality rates was obtained from NYT's public database (<https://github.com/nytimes/covid-19-data>). The `us-counties-2020.csv` data set contains the **fips** code for each county and the corresponding **death count**. This was merged with data from the U.S. Census Bureau (<https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html>) to obtain the 2020 population totals for each county (see `co-est2022-pop.xlsx`). **Mortality rate** was then calculated as the death count for each county divided by the population for each county multiplied by 1000 (deaths per 1,000). This data set also contains the **state** for each county record, which will later be used to control for.

I then merged per-county housing prices to with the mortality data. To do this, I used zip-level housing price data (see `HPI_AT_BDL_ZIP3.xlsx`) maintained by the Federal Finance Housing Agency (FHFA) (<https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>) and merged it with data that mapped zip-codes in 2021 to the

corresponding counties. I used HUD’s ZIP-County Crosswalk [API](#) to obtain the zip codes that correspond to each county. The Python script, [zip-county.py](#), uses this API to compile [zip-county.csv](#). After combining these two data sets, the housing prices was condensed to create an average **HPI** (housing price index) for each county consisting of the **HPIs** from each zip code in each county. For the purposes of this analysis, to control for inflation, the **HPI** that’s been adjusted for inflation using 2000 year as a base year was used instead of raw **HPI**. From here on out in this analysis, **HPI** will be used to refer to as this **HPI** that was been adjusted for inflation using 2000 as the base year.

Additionally, 6 lag variables on the 2021 **HPI** were constructed using The housing prices data. (See [A.1](#) for the selection process for the number of lag variables).

Merging the housing prices data set with the mortality data set yields the final data set for the econometric analysis: The response variable HPI_{2021} will be measured on the independent metric of interest, Mortality Rate for each county, and this will be controlled for using a series of state dummy variables and the 6 lag variables on HPI_{2021} .

Exploratory Data Analysis

Econometric Model

For year t , state s , and county c , we wish to estimate for n # of states,

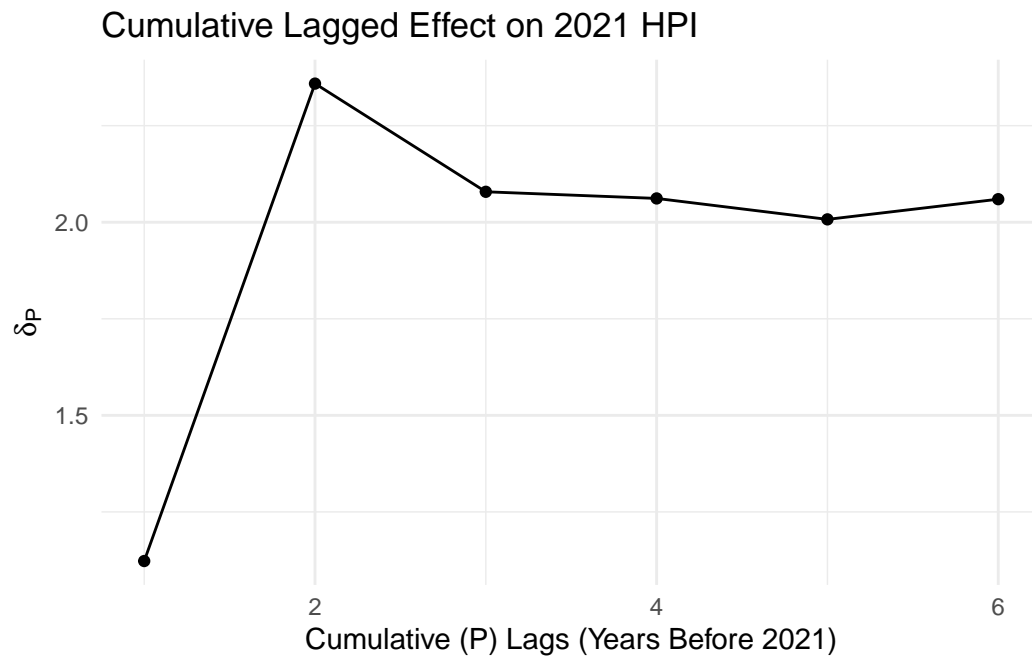
$$(1) HPI_{sct} = \beta_0 + \beta_1 Mortality_{sct} + \sum_{j=1}^{n-1} \beta_{j+1} I(State_s = j) + \sum_{p=1}^6 \delta_p HPI_{sct-p} + \eta_{sct}$$

Where β_1 is the parameter of interest. Hence, since we are only interested in the effect that 2020 COVID-19 mortality had on 2021 per-county housing prices, setting $t = 2021$ yields a regression on a cross-sectional data set.

Appendix

A.1

I determined a sufficient # of lags (P) for the lagged effects on $HPI_{sct=2021}$ by regressing $HPI_{sct=2021}$ on $HPI_{sct-1}, \dots, HPI_{sct-P}$ until the last the regression coefficient, δ_P , was no longer significant for a chosen level of α . Let $\alpha = 0.1$



Hence, let $P = 6$.