JOURNAL OF
EMERGING INVESTIGATORS

Article

The effect of COVID-19 on the USA house market

William G. Xiao[1], Liying Gong[1]
[1] Godwin High School, Richmond, Virginia
[2] Finance Service Industry, Richmond, Virginia

SUMMARY

The purpose of this study was to understand the varying effects of COVID-19 on the median house price in metropolitan areas in the USA. Recently, the demand for houses has increased due to the COVID-19 pandemic. The pandemic increased the desire to stay indoors and introduced working from home. Four different factors were tested for their impact on the housing market: number of COVID-19 deaths, number of available houses for sale, unemployment rate, and mortgage interest rates. The control in this study was 2018-2019 house price data. We hypothesized that the house supply would cause the greatest change in house prices. Ckmean clustering method was applied to separate 93 metropolitan areas of the USA into three groups with different median house price levels. We then conducted linear regression modeling for each group to determine the varying level of effect each of the variables had. The results suggested that unemployment rate had the largest correlation with house prices, especially for metropolitan areas with extremely high house prices. This result did not support the research hypothesis. Further research could clarify the findings of our study by using fine-turned modeling techniques and including more economic factors like immigration or tourism. Additionally, we can research the effect of COVID-19 on house prices globally.

INTRODUCTION

Since the beginning of 2020, COVID-19 caused a world-wide pandemic that resulted in millions of deaths. Stores and restaurants across the world shut down and unemployment rates hit an all-time high, even surpassing those of the great economic crisis of 2008-2009 (1). This pandemic also caused lifestyle changes such as working from home, heavily relying on social media for networking and entertainment, and increased online shopping. The COVID-19 pandemic affected many economic sectors like tourism, international trade, and the housing market (2). With the inflation of house prices and a lack of steady income, many American citizens found it much more difficult to afford houses during this time. House prices rose by 13% between March of 2020 and March of 2021 in the United States alone (3). We observed this phenomenon throughout the entire globe, with Australia seeing an annual house price increase of 6% (3).

The purpose of this research was to understand how the selected economic factors affected the housing market in the United States during the COVID-19 pandemic and to determine which one had the greatest impact. We used the median house prices between March of 2020 to July of 2021 and March of 2018 to July of 2019 over 93 USA metropolitan areas to compare house prices during the COVID-19 pandemic and before the pandemic. With the results from our study, proper adjustments could be made to the most impactful factor to help maintain a stable housing market and enhance house affordability in the USA.

Unemployment rate, house mortgage rate, supply of houses, and COVID-19 deaths were the economic factors tested in the study. We selected these four factors based on other studies, personal understanding, and data availability. These factors were significantly influenced due to COVID-19 (4). We selected 'number of deaths' assuming it reflects the joint effects of virus transmission, virus transformation, society quarantine policy, and vaccine's creation and distribution. The unemployment rate was defined as the number of unemployed people divided by the total number in the civilian labor force. It is typically used to measure the effect of economic events, such as a recession. The house supply affects the price of houses based on the law of supply and demand; when there is more supply and low demand, the price will go down and vice versa. Lastly, the mortgage rates for 30-year fixed rate mortgages were used in this analysis as the most popular mortgage product. Given the same initial house prices, the lower mortgage rate allows people to purchase house with lower total costs. Other economic factors were considered but not included due to data limitation, such as immigration rates, tax rates, and the amount of stimulus checks passed out.

Our hypothesis was that house supply would have the greatest impact on house prices. This is because house supply is the most direct factor to the price compared to the other factors. Due to COVID-19, group activities were inhibited, including construction, which may have reduced the supply of available houses. Along with the introduction of working from home, people desire more space at home, thus causes an increase in the demand for houses. With the low supply and high demand of houses, the price could increase significantly. However, the results of our paper suggested that unemployment rate had the greatest effect on house prices, thus initial research hypothesis was rejected. House supply ended up being the second most impactful factor.

RESULTS

The pandemic re-ordered the real estate markets across the board on an unprecedented scale (**Figure 1**). The median house price was very steady from 2008 to 2019 but started rising during the COVID-19 pandemic period. This

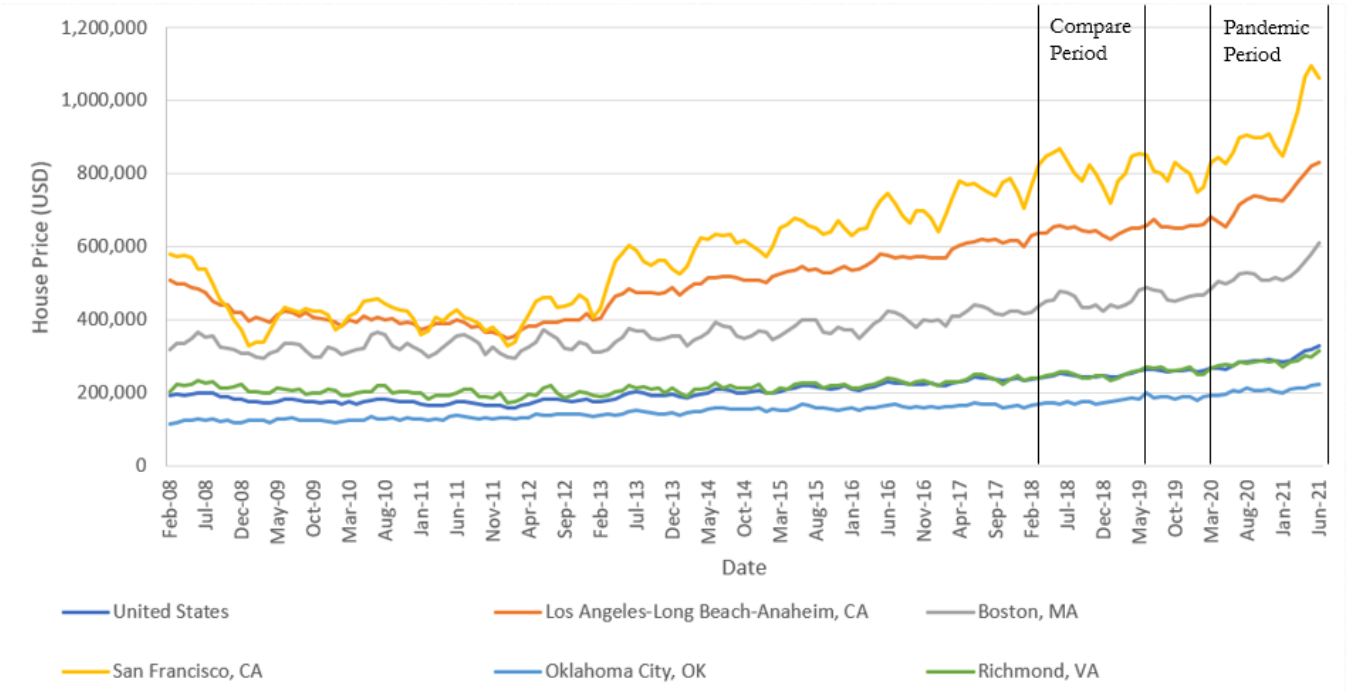

Figure 1: Median house price over time. Median house prices in different cities in the USA over the past 13 years. Data was obtained through the Zillow database (6). The compare period is from March 2018 to July 2019. The pandemic period is from March 2020 to July 2021

trend is more obvious for higher house price areas such as San Francisco and Los Angeles. To prove the statistically significant increase in house prices, we conducted a one-tailed t-test to compare 2018-2019 house prices to 2020-2021 house prices. The test revealed that house prices from 2020-2021 were significantly higher ($p < 0.000001$).

Starting from March of 2020, the overall house supply was lower on average than in the pre-COVID period. We inferred this was probably caused by the working-from-home lifestyle, which reduced the chance of relocation brought by job changes and the number of second-hand house sales. House construction companies also reduced the supply of new houses due to the economic uncertainty brought by COVID-19. Starting from September 2018, the mortgage started to decline until middle of 2019. It stabilized for a half year but then started to decline again during the COVID-19 period until March 2021.

We observed that the median house price change over time had similar trends over different metropolitan areas, but the magnitude of these changes was different (**Figure 1**). To understand whether COVID-19 deaths, number of houses available for sale, mortgage rate, and unemployment rate had different impacts on house prices in different metropolitan areas, we used Ckmean clustering to class 93 metropolitan areas into three groups based on their house price in March 2018. We selected House prices from March of 2018 because this month is when the COVID-19 pandemic started in 2020 so the data incorporates the same seasonal changes. Ckmean is a function to cluster a one-dimensional array of floats into k optimal clusters (5). Using Ckmean, we chose three different clusters to compare normal, expensive, and super expensive house price regions. These group names were chosen by comparing each group to the median house price in the USA. Most metropolitan areas resided in the normal group, with less in the expensive group and only 2 in super expensive group. Some example cities in the normal group were Oklahoma City OK, Richmond VA, Houston TX, and Miami-Fort Lauderdale FL. The metropolitan areas in the expensive group included cities such as Los-Angeles, Boston, San Diego, and New York. Last group consisted of two places: San Francisco and San Jose.

We used a linear regression model to understand the different effects each of the factors had on the house price. In total, four simple linear regression models were built. One was built on all house price changes across the 93 metropolitan areas and the other three were separately built on each of the three metropolitan groups identified by Ckmean clustering. Each simple linear regression model returned four coefficients representing how much each factor affected the house price for different metropolitan groups (**Table 1**). Since all factors have been standardized, the coefficients can be directly used to compare their impact on the house price change. For every percent the independent variable (factor) increases, the percent change in house price due to COVID-19 increases or decreases by the value of the coefficient.

We found that unemployment rate with an absolute coefficient of 0.495 had the greatest impact on house price changes during the COVID-19 pandemic out of the four factors tested. The absolute coefficients of other three factors ranged from 0.0044 to 0.1327. The negative coefficient for unemployment rate suggested that the higher unemployment rate was, the lower house price was. In addition, the unemployment rate was more strongly correlated with house prices in the super expensive group than others. The house supply was ranked as the second most impactful factor. More houses available in the market may have made house prices

	House Supply Change	Unemployment Rate Change	Standardized COVID-19 Deaths Change	Mortgage Rate Change
Overall	-0.1327	-0.4950	-0.0148	0.0044
Normal	-0.1107	-0.4229	-0.0185	0.0015
Expensive	-0.2668	-0.4734	-0.0304	0.0173
Super Expensive	-0.2497	-1.8385	0.0699	0.0048

Table 1: Linear regression coefficients. The percentage of change each economic factor has on the different groups. Overall represents all 93 metropolitan areas in the USA. Values obtained by using Python correlation table feature.

drop. The house price changes were more sensitive to house supply changes for both the expensive and super expensive groups compared to the normal house price zones. For the full model, the coefficients for COVID-19 deaths and mortgage rate changes were close to zero. The *p*-values of those coefficients were 0.200 and 0.298 which are both significantly larger than the set significance level of 0.05 when comparing their effects on pre and post COVID-19 house prices. We observed similar trends in the other models. COVID-19 deaths and mortgage rate changes were not significant factors to explain the house price change during COVID-19.

DISCUSSION

The purpose of our study was to determine which of the chosen economic factors that may have been affected by COVID-19 had the greatest impact on metropolitan area house prices in the USA. The linear regression model showed that the unemployment rate had the greatest correlation with the house price. Thus, the initial research hypothesis was rejected because the house supply did not have as much of an impact. This might be explained by COVID-19's more severe effect on unemployment rate compared to the other factors (1). The unemployment rate during COVID-19 even surpassed that of the 2008-2009 economic crisis, hitting a high of 20.6 million citizens (1). With this information, many things could be done to alter the unemployment rate which may in turn affect the house prices. Such things include supporting small businesses, increasing public infrastructure construction, allowing service business operation under health guidelines, etc. House supply was found to be the second most important factor. To enhance the house supply while maintaining the profitability of the construction companies, building more smaller-size houses or providing financial support to house construction companies are some options the government could consider.

Deaths caused by COVID-19 did not have as much of an impact on house prices as expected. What impacts house prices was more related to the behavior changes brought by the pandemic instead of direct COVID-19 factors. Mortgage rate is also insignificant in our study. This could be caused by a data limitation. Due to a lack of mortgage rates at the metropolitan level, the average rate across the United States is applied for all metropolitan areas. There were only 16 unique observations for mortgage rate changes, while the other three factors have 1488 records (93 areas for 16 months). Another source of error could be the use of only metropolitan cities, which means the data cannot be generalized to all of the USA. The use of Ckmean in our study caused a difference in group size which could possibly affect the coefficients found via linear regression due to bias. The standardization of unemployment and mortgage rate is different from the standardization of COVID-19 deaths and house supply, which is also a potential source of error. While the normalization was different, the significance of variables remains the same. Lastly, Ckmean clustering minimizes the sum of squares of within-cluster distances from each element to its corresponding cluster center. The method is sensitive to outlier observations, which resulted in only 2 metropolitan areas included in the super expensive group.

For future studies, we can fine-tune the linear model through transforming independent variables, introducing the interaction terms, and applying a model selection algorithm. A linear regression model assumes that the input variables are independent with each other and have a linear relationship with the target variable. Transforming inputs and introducing interaction terms can address the limitations of observed data where linear regression model assumptions may not work. Different types of modeling techniques can also be tried, like Gradient Boosting or time series model. We can use these modeling techniques to verify the conclusion of this study and to predict future house prices in these metropolitan areas. Future research can also include more factors like immigration rate, travel, stock market index, and international trades. Those factors were also influenced dramatically by COVID-19 (2). More COVID-19 health factors could also be considered for future research like hospitalizations, vaccinations, and new cases.

MATERIALS AND METHODS

This research contains five steps described in **Figure 2.** The first two steps introduce how the dependent variable and independent variables are selected, collected, and prepared. Excel was used as the main analysis tool for the first two steps. All data was collected from trusted and authorized websites. Median house prices and house supplies in the metropolitan areas were extracted from the Zillow Research Center (6).

Figure 2: Research procedure for experiment.

Mean unemployment rates were extracted from the U.S. Bureau of Labor Statistics website (7). The monthly 30-year fixed mortgage rate was provided by the Fred Economic Data website (8) and the number of COVID-19 death is downloaded from the National Center for Health Statistics (9).

After the data was collected, it was observed that 1-2% of the data was missing. Any missing points were imputed using one of two methods: one for a singular missing point and one for several continuous missing data points. A singular missing point was imputed with an average calculated from the pre- and post-month observations. For continuous missing points, a linear trend was found using pre- and post- data and then applied to impute the missing values.

To better capture the change of house prices, ratio changes of house prices pre- (March 2018 – July 2019) and post-COVID (March 2020 – July 2021) were calculated as the dependent variable of this analysis, as shown by equation (1). Choosing median price instead of average house price is to reduce the effect of outliers and eliminate the regional differences within metropolitan areas. Additionally, the change in house prices is calculated by months to take seasonality changes into account. There are 16 months and 93 areas, resulting in a total of 1488 observations.

$$Y_{t,l} = \frac{P_{t,l} - P_{t-24,l}}{P_{t-24,l}} \quad (1)$$

Where:

$Y_{t,l} = House\ price\ change\ ratio\ at\ time\ t\ and\ metropolitan\ area\ l$

$t = Months\ between\ March\ 2020\ and\ June\ 2021$

$P_{t,l} = Median\ house\ price\ at\ time\ t\ and\ metropolitan\ area\ l$

$P_{t-24,l} = Median\ house\ price\ 24\ months\ before\ time\ t\ at\ metropolitan\ area\ l$

The four factors in this analysis have different scales: mortgage rate and unemployment rate are percentage levels; house supply and the number of COVID-19 deaths have a scale from a few thousand to a few-hundred thousand. To ensure that the impacts of these variables on the house price change are comparable, a standardization has been applied to ensure all independent variables have a value from 0 to 1. Equations (2), (3), (4), and (5) describe how the final independent variables are formed and standardized.

$$X1_{t,l} = \frac{HS_{t,l} - HS_{t-24,l}}{HS_{t-24,l}} \quad (2)$$

Where:

$X1_{t,l} = House\ supply\ change\ ratio\ at\ time\ t\ and\ metropolitan\ area\ l$

$HS_{t,l} = Number\ of\ houses\ available\ at\ time\ t\ and\ metropolitan\ area\ l$

$HS_{t-24,l} = Number\ of\ houses\ available\ 24\ months\ before\ time\ t\ at\ metropolitan\ area\ l$

$$X2_{t,l} = UE_{t,l} - UE_{t-24,l} \quad (3)$$

Where:

$X2_{t,l} = Difference\ in\ unemployment\ rate\ pre\ and\ post\ COVID-19\ at\ metropolitan\ area\ l$

$UE_{t,l} = Unemployment\ rate\ at\ time\ t\ and\ metropolitan\ area\ l$

$UE_{t-24,l} = Unemployment\ rate\ 24\ months\ before\ time\ t\ at\ metropolitan\ area\ l$

$$X3_{t,l} = \frac{DE_{t,l} - Min(DE_{i,j})}{Max(DE_{i,j}) - Min(DE_{i,j})} \quad (4)$$

Where:

$X3_{t,l} = COVID-19\ death\ number\ change\ ratio\ at\ time\ t\ and\ metropolitan\ area\ l$

$DE_{t,l} = Number\ of\ COVID-19\ deaths\ at\ time\ t\ and\ metropolitan\ area\ l$

$DE_{i,j} = Number\ of\ COVID-19\ deaths\ at\ time\ i\ and\ metropolitan\ area\ j,$

$for\ i\ from\ March\ 2020\ to\ June\ 2021,\ for\ j\ the\ 93\ metropolitan\ areas$

$$X4_t = MR_t - MR_{t-24} \quad (5)$$

Where:

$X4_t = Mortgage\ rate\ change\ from\ time\ t$

$MR_t = Morgtage\ rate\ at\ time\ t$

$MR_{t-24} = Mortgage\ rate\ 24\ months\ before\ time\ t$

For step 3, a one-tailed t-test was conducted to compare pre- and post- COVID-19 house prices. This type of t-test was chosen to determine the significance of directional effect. The SciPy package in Python conducted the t-test using standard methods and obtained the t-value and *p*-value of the sample.

Step 4 helped us understand whether the factors have different impacts on house prices in different metropolitan areas, Ckmean clustering method was used to group the 93 metropolitan areas into three buckets (**Figure 3**).

In step 5, four linear regression models were built to identify the different effects each of the economic factors had on the house price. A linear regression model is a common method to discover whether a relationship between variables exists. Equation (6) describes the simple linear regression model. Ordinary least squares (OLS) method was used for estimating the intercept and coefficients of independent variables. OLS estimates the coefficients by the principle of least squares:

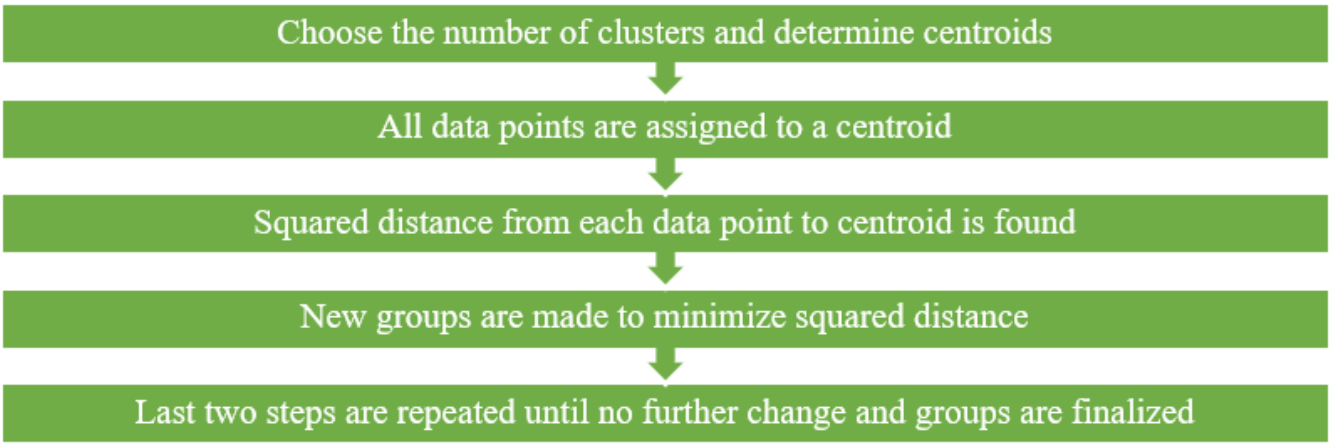

Figure 3: Ckmean procedure.

minimizing the sum of the squares of the differences between the observed dependent variable values and those predicted by the linear function of the independent variables. Python packages including Pandas, Numpy, Statsmodel, and CKmean are used to conduct analysis in the last three steps.

$$Y_{t,l} = \beta_0 + \beta_1 * X1_{t,l} + \beta_2 * X2_{t,l} + \beta_3 * X3_{t,l} + \beta_4 * X4_t \quad (6)$$

Where:

$Y_{t,l} = Dependent\ Variables\ introduced\ in\ equation\ (1)$

$\beta_0 = Intercept\ of\ the\ regression\ line\ (Offset)$

$X1_{t,l}, X2_{t,l}, X3_{t,l}, X4_t = Independent\ Variables\ introduced\ in\ equations\ (2) - (5)$

$\beta_1, \beta_2, \beta_3, \beta_4 = Coefficients\ of\ independent\ variables$

Received: March 6, 2022
Accepted: August 29, 2022
Published: November 19, 2022

REFERENCES

1. Smith, Sean M., and Roxanna Edwards. "Unemployment Rises in 2020, as the Country Battles the Covid-19 Pandemic." Monthly Labor Review, 2021, doi:10.21916/mlr.2021.12.
2. Siddiquei, Mohd Imran, and Waseem Khan. "Economic Implications of Coronavirus." *Journal of Public Affairs*, 2020, doi:10.1002/pa.2169.
3. Heywood, Andrew. "Paying for Covid? House Prices and the Pandemic." Housing Finance International, Summer 2021, pp. 4–5.
4. Morris, Alan. "The Impact of the Covid-19 Pandemic on the Housing Market and Policy in Australia." Housing Finance International, Summer 2021, pp. 33–38.
5. Wang, Haizhou, and Mingzhou Song. "Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming." The R journal vol. 3,2 (2011): 29-33.
6. "Housing Data." Zillow, 8 Feb. 2006, Accessed 4 Sept. 2021.
7. "Local Area Unemployment Statistics." U.S. Bureau of Labor Statistics, Sept. 1995, Accessed 4 Sept. 2021.
8. "30-Year Fixed Rate Mortgage Average in the United States." FRED Economic Data, ST. Louis Fed, 1991, Accessed 24 Sept. 2021.
9. "Provisional COVID-19 Death Counts by Week Ending Date and State." Center for Disease Control and Prevention, 1 May 2020, Accessed 5 Sept. 2021.