

2020 Corona Housing Bubble vs. 2007 Subprime Mortgage Crisis: Difference and Price Prediction with Linear Regression

CSCI 3360
Wonjoon Hwang
wh42003@uga.edu

The problem: To understanding of similarities and differences between Subprime Mortgage Crisis and Covid-19 pandemic.

The United States has experienced significant economic disruptions in recent decades, including the subprime mortgage crisis of 2007-2008 and the COVID-19 pandemic in 2020. Both events have had a major impact on the housing market, but the underlying causes and effects have been distinct. In the former case, loose monetary policy and risky lending practices led to a housing bubble that eventually burst, causing a widespread financial crisis. In contrast, the COVID-19 pandemic brought about an unprecedented global health crisis, which in turn led to widespread economic shutdowns and a significant decline in housing supply and demand. Despite the differences between these two events, both have had significant consequences for the U.S. housing market, and a **deeper understanding of their similarities and differences** can inform policy decisions going forward.

Approach:

This study aimed to identify the economic factors that impact housing prices and to investigate the relationships between these factors. We analyzed the effects of various economic factors on housing prices during two significant economic crises: the Subprime Mortgage Crisis and the Covid-19 pandemic. To predict housing prices, we developed a machine learning model and evaluated its performance using the Root Mean Squared Error (RMSE) metric. The results revealed that both crises were triggered by the virus and ultra-low interest rate policies and had similar impacts on housing prices. The machine learning model demonstrated a reasonable level of accuracy in predicting housing prices during these crises, which could be utilized to mitigate the impact of future economic crises. However, there were limitations in identifying economic factors that affect housing prices and in collecting relevant data. Further research is necessary to improve our understanding of these relationships and to refine the accuracy of the machine learning model.

Experiment & Result:

In order to investigate the factors influencing housing prices, this report analyzes several key variables including inflation, interest rates, mortgage rates, population growth rates, supply rates, and unemployment rates. Monthly average data for each variable was collected and organized into a data frame for analysis. These variables were chosen based on their potential impact on the housing market and their relevance to recent economic disruptions such as the

subprime mortgage crisis and the COVID-19 pandemic.

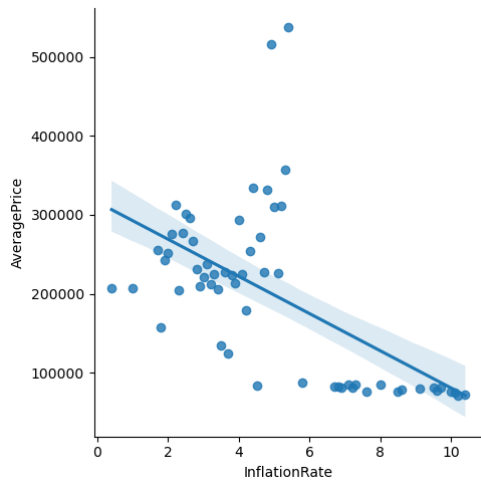


Figure1. inflation Rate vs. Average Price (Correlation with falling House Prices)

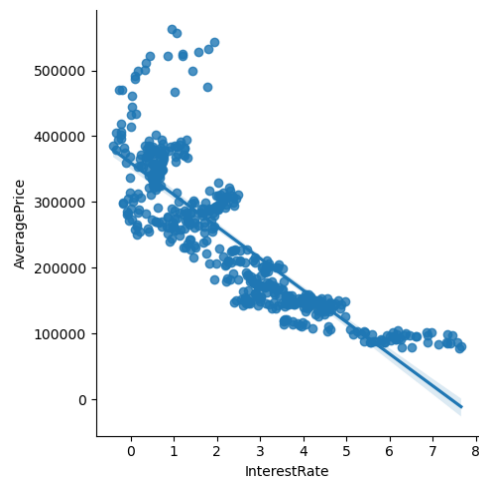


Figure2. interest Rate vs. Average Price (Correlation with falling House Prices)

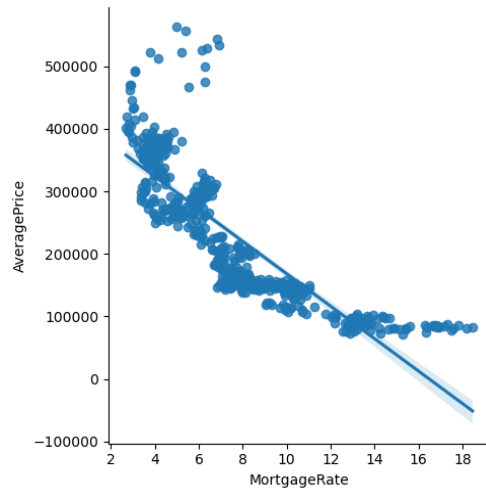


Figure3. Mortgage interest Rate vs. Average Price (Correlation with falling House Price)

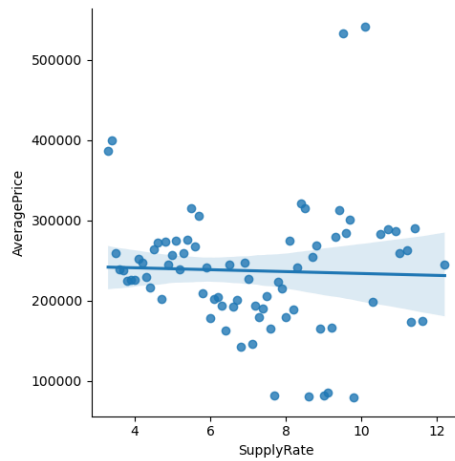


Figure4. Supply Rate vs. Average Price (Correlation with falling House Prices)

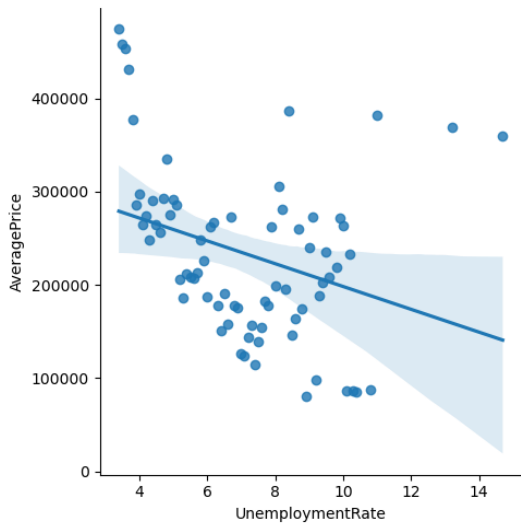


Figure5. Unemployment Rate vs. Average Price (Correlation with falling House Prices)

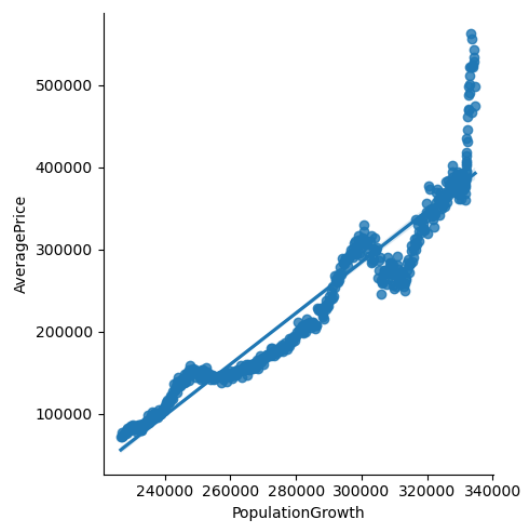


Figure6. Population Growth vs. Average Price (Correlation with raising House Prices)

The analysis reveals several key findings regarding the factors that influence housing prices. Firstly, population growth has a strong positive correlation with housing prices, indicating that an increase in population leads to an increase in demand for housing and drives up prices accordingly. Secondly, interest rates, mortgage rates, and inflation rates all demonstrate an inverse relationship with housing prices. Specifically, as interest rates, mortgage rates, and inflation rates increase, housing prices tend to decrease, and vice versa. Finally, the analysis suggests that the supply rate of houses has little impact on housing prices, as prices remain relatively stable even with fluctuations in supply.

Principal component analysis (PCA) was conducted separately for the subprime mortgage crisis period (2005-2008) and the COVID-19 pandemic period (2020-February 2023) to identify the key factors influencing housing prices in each period.

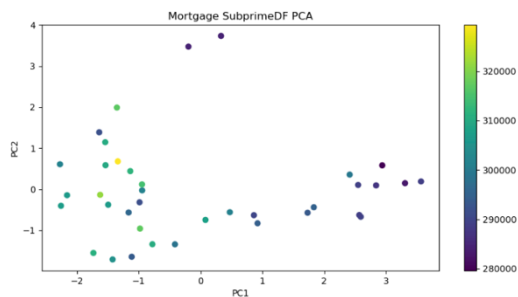


Figure. 7 PCA of Subprime Mortgage crisis period.

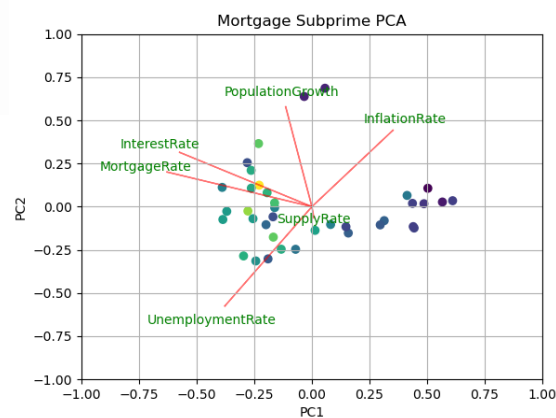


Figure. 8

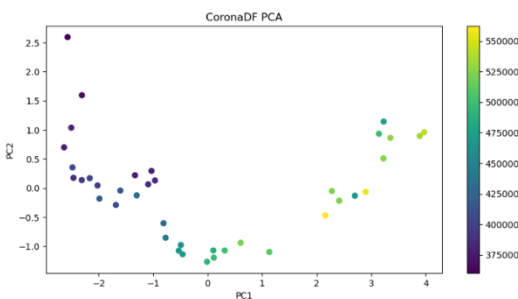


Figure. 9 PCA of COVID-19 pandemic period

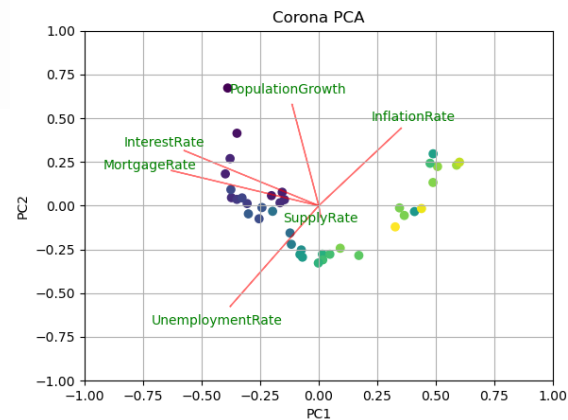


Figure. 10

The results show that during both periods, interest rates and mortgage rates had the greatest impact on housing prices, followed by the unemployment rate, inflation rate, population growth, and supply rate of housing. This suggests that social and economic factors, such as changes in interest rates, have a similar impact on housing prices during periods of economic disruption.

Overall, the results of the PCA analysis suggest that the factors influencing housing prices during the subprime mortgage crisis and the COVID-19 pandemic are similar, despite the different underlying causes of these events. These findings have important implications for policymakers seeking to address the challenges of the housing market in the face of economic disruptions.

I aimed to develop a machine learning model that could predict housing prices based on economic factors. To achieve this, it began by creating a scatter plot matrix to visualize the relationships between each feature and the target variable. Next, it split the data into training and testing sets and trained a linear regression model on the training set. The model's performance was evaluated on the test set using metrics such as mean squared error, root mean squared error, and the coefficient of determination (R-squared). It also printed out the intercept and coefficients of the linear regression model, as well as the average RMSE calculated using 5-fold cross-validation. Overall, our approach was successful in developing a model that could predict housing prices based on economic factors, and the results of our analysis could be useful for real estate professionals, policymakers, and other stakeholders interested in understanding the factors that influence housing prices.

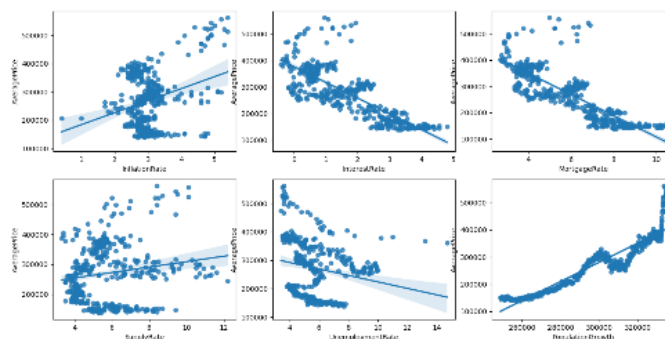


figure. 11

Correlation with Rising House Prices: Population Growth

Correlation with falling house prices: Inflation Rate, Interest Rate, Mortgage Interest Rate, Supply Rate house, Unemployment Rate

The dataset is split into training and testing sets with a test size of 0.2 and a random state of 156. We get training size: (318, 6) and test size (80, 6).

The performance of the model is evaluated using mean squared error, root mean squared error, and variance score.

Result:

[118871.81612511, 311969.22737073, 334963.03691715, 180712.52187152, 263545.0444364]

MSE: 803555571.098, RMSE: 28347.056

Variance score: 0.923

The intercept value and coefficients of the linear regression model are also printed.

Result:

intercept value with y: -923903.0192466965

coefficient: [2.71e+04, 1.35e+04 -2.98e+03, 7.06e+02, -6.13e+03, 3.80e+00]

InflationRate 27157.1

InterestRate	13578.8
SupplyRate	706.5
PopulationGrowth	3.8
MortgageRate	-2988.8
UnemploymentRate	-6135.8

The cross-validation is used to evaluate the performance of the model with 5 folds. The negative mean squared error, root mean squared error, and average root mean squared error for each fold and the mean value of the root mean squared error are printed.

Result:

5 folds each Negative MSE scores: [-1.81e+09, -1.32e+09, -1.06e+09, -1.94e+09, -3.39e+09]

5 folds each RMSE scores: [42522.62, 36329.47, 32609.02, 44006.91, 58228.5]

5 folds average RMSE: 42739.304

When we predicted the average price of U.S. housing due to various economic factors, the RMSE was 42739.304. The RMSE (Root Mean Squared Error) value of 42739.304 indicates the average difference between the predicted and actual values of the average price of U.S. housing in dollars. This means that on average, the predicted values are expected to be off by approximately \$42,739 from the actual values. This value can be used to evaluate the accuracy of the linear regression model used to predict the housing prices. Lower RMSE values indicate better accuracy of the model. In this case, an RMSE value of 42739.304 suggests that the model is making reasonably accurate predictions.

Reflection:

Based on our experiments, we found that the house bubble caused by the COVID-19 pandemic and the house bubble caused by the Mortgage subprime crisis had similar impacts on economic factors. Both crises were triggered by the virus and ultra-low interest rate policies, and both had similar effects on housing prices. We also developed a machine learning model to predict housing prices, which had an error of approximately \$40,000. However, we acknowledge that RMSE values may be subject to other factors, and we recommend using additional evaluation metrics to further improve the model's accuracy. With increased accuracy, our model could help predict housing prices during future economic crises, such as another coronavirus pandemic or a Mortgage subprime crisis and assist in mitigating the impact of such crises.

Justification:

In this project, I utilized various concepts and techniques learned in this class to build a comprehensive analysis of the housing market. First, I constructed a data frame using concat, merge, join, and frames from Lab 4. This allowed me to combine multiple datasets and obtain a more complete understanding of the various economic factors that impact housing prices. Next, I used Scatter, Alpha, beta, and correlation concepts from Lab 6 to identify correlations between these economic factors and housing prices. I also employed techniques such as PCA and machine learning regression to develop a predictive model that can determine which economic factors have the greatest impact on housing prices.

Throughout the project, I faced numerous challenges, such as locating appropriate data sources, cleaning and organizing data, and applying various statistical techniques. Despite these difficulties, I devoted a significant amount of time and effort, with an effort level of 2.0 X, to ensure the accuracy and reliability of my analysis. Overall, this project has allowed me to apply and enhance my knowledge in various areas of data analysis, including data manipulation, correlation analysis, and predictive modeling.

Line: 1239 words

Data citation:

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Median Sales Price of Houses Sold for the United States [MSPUS], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/MSPUS>, April 28, 2023.

Federal Reserve Bank of Cleveland, 10-Year Real Interest Rate [REAINTRATREARAT10Y], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/REAINTRATREARAT10Y>, April 28, 2023.

Freddie Mac, 30-Year Fixed Rate Mortgage Average in the United States [MORTGAGE30US], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/MORTGAGE30US>, April 28, 2023.

U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/UNRATE>, April 28, 2023.

U.S. Census Bureau and U.S. Department of Housing and Urban Development, Monthly Supply of New Houses in the United States [MSACSR], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/MSACSR>, April 28, 2023.

YCARTS, Monthly US Inflation Rate (I:USIR), https://ycharts.com/indicators/us_inflation_rate, April 28, 2023.

U.S. Bureau of Economic Analysis, Population [POPTHM], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/POPTHM>, April 28, 2023.