2020 Corona Housing Bubble vs. 2007 Subprime Mortgage Crisis:

Difference and Price Prediction with Linear Regression

Wonjoon Hwang
wh42003@uga.edu
CSCI 3360

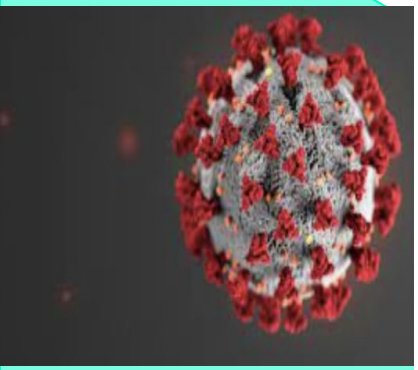
Problem: Movie:" The big Short"

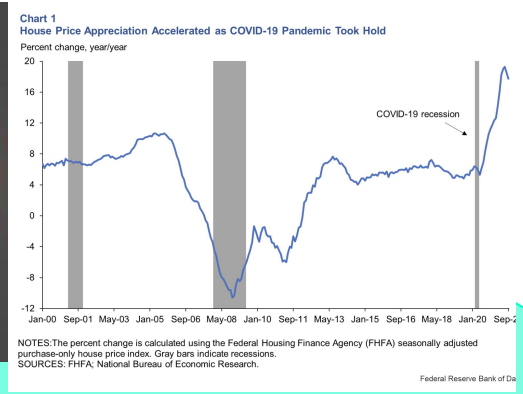
"The Big Short" is a movie about the 2008 financial crisis, following a group of investors who bet against the housing market and try to profit from the collapse of the subprime mortgage industry.

Subprime Mortgage Crisis

a sharp increase in defaults on subprime mortgages (loans given to people with poor credit histories) and a subsequent collapse of the housing market in the U.S.

Problem: COVID-19 pandemic





The COVID-19 pandemic has had a significant impact on the world economy, leading to widespread job losses, business closures, supply chain disruptions, and reduced economic activity. Many countries have experienced recessions or significant slowdowns in economic growth, with industries such as travel, hospitality, and retail particularly hard hit.

Covid-19 Pandemic

The COVID-19 recession led to a steep but short-term decline in the housing market, followed by a rapid increase in house prices, reaching a peak of 19.3% in July 2021, which is much higher than the pre-pandemic rate of about 5%. However, it is uncertain if this increase will lead to a subsequent collapse of the housing market in the United States.

Introduction

1. Subprime Mortgage Crisis (2007-2008)
   - Loose monetary policy and risky lending practices
   - Housing bubble burst, causing a financial crisis

2. COVID-19 Pandemic (2020)
   - Global health crisis led to economic shutdowns
   - Significant decline in housing supply and demand

3. Consequences for U.S. Housing Market
   - Both events had significant impacts on housing market

- But how they are different in Housing bubble?

- Creating machine learning to predict average house prices as an economic factor in preparation for another looming economic collapse

Approach.

- Investigated economic factors impacting housing prices during the Subprime Mortgage Crisis and COVID-19 pandemic
- Developed machine learning model to predict housing prices with evaluation using RMSE
- Findings can inform preparation for future economic crises
- Limitations include challenges in finding relevant economic factors and collecting data

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.

DATA CLEANING AND PROCESS

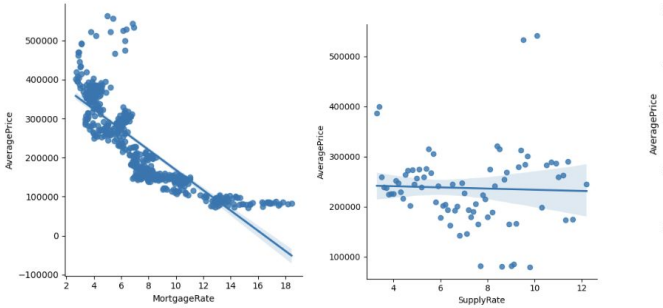
the most influential factors on housing prices

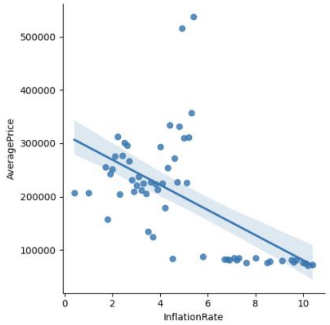
- Population Growth
- Interest Rate
- Mortgage Interest Rate
- Unemployment Rate
- Inflation Rate
- Supply Rate

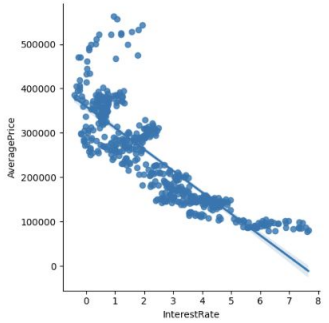
House

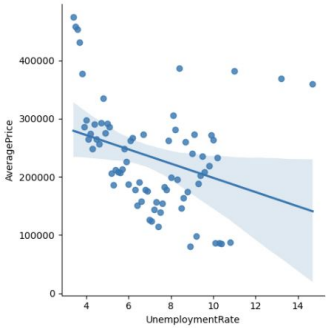
date	AveragePrice	InflationRate	InterestRate	MortgageRate	SupplyRate	UnemploymentRate	PopulationGrowth
1990-01-01	151700.0	4.1	4.1276234	9.895	7.0	5.4	248743.0
1990-02-01	150900.0	4.1	4.36692739	10.1975	7.6	5.3	248920.0
1990-03-01	144600.0	3.7	4.55196762	10.268	7.8	5.2	249146.0
1990-04-01	153400.0	3.6	4.55216196	10.37	8.3	5.4	249436.0
1990-05-01	150600.0	3.4	4.80506176	10.4775	8.2	5.4	249707.0
1990-06-01	150400.0	3.8	4.38261115	10.164	7.9	5.2	249990.0
1990-07-01	149800.0	3.4	4.3418185	10.035	7.8	5.5	250285.0
1990-08-01	144700.0	4.6	4.17053207	10.1	8.2	5.7	250595.0
1990-09-01	142100.0	4.7	4.3987447	10.175	8.4	5.9	250904.0
1990-10-01	153000.0	4.8	4.20053902	10.1775	8.7	5.9	251201.0
1990-11-01	143300.0	4.7	4.05885498	10.014	8.2	6.2	251486.0
1990-12-01	153400.0	4.7	3.8939207	9.6725	8.5	6.3	251758.0
1991-01-01	148600.0	3.9	3.92293115	9.6375	9.4	6.4	252012.0
1991-02-01	147800.0	3.4	3.90705382	9.365	7.9	6.6	252253.0
1991-03-01	156400.0	3.3	4.17583618	9.5	7.3	6.8	252507.0
1991-04-01	150800.0	3.2	4.1098331	9.4925	7.3	6.7	252778.0
1991-05-01	145400.0	3.1	4.04991631	9.472	7.0	6.9	253060.0
1991-06-01	145900.0	3.3	4.18764426	9.615	7.0	6.9	253350.0
1991-07-01	148200.0	3.1	4.22225375	9.575	7.1	6.8	253650.0
1991-08-01	141800.0	3.2	4.13991739	9.244	6.8	6.9	253966.0
1991-09-01	147300.0	3.0	3.87849555	9.0075	7.4	6.9	254280.0
1991-10-01	147400.0	3.2	3.63786206	8.855	6.7	7.0	254576.0
1991-11-01	141700.0	2.9	3.56505772	8.712	6.2	7.0	254841.0
1991-12-01	143000.0	2.7	3.41033346	8.497500000000000	6.2	7.3	255089.0
1992-01-01	144200.0	2.7	3.19854245	8.432000000000000	5.2	7.3	255331.0

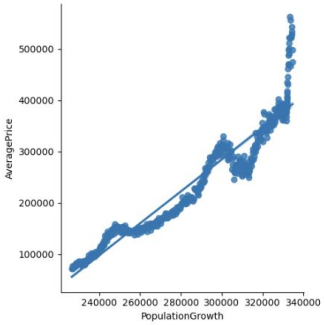
Correlation between House Prices and other feather











Correlation with Rising House Prices: Population Growth

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

PCA Analyze 1. How I did

1. Data preprocessing

```
start_date = '2005-01-01'
end_date = '2008-01-01'

mortgageSubprimeDF = df[(df['date'] >= start_date)&(df['date'] <= end_date)]

start_date1 = '2019-12-01'
end_date1 = '2023-02-01'

CoronaDF = df[(df['date'] >= start_date1)&(df['date'] <= end_date1)]

# print("--------------------------------")
# print(mortgageSubprimeDF.head(10))
# print("--------------------------------")
# print(CoronaDF.head(10))

data_cols = ['InflationRate', 'InterestRate', 'MortgageRate', 'SupplyRate', 'UnemploymentRate', 'PopulationGrowth']
label = 'AveragePrice'
features = ['InflationRate', 'InterestRate', 'MortgageRate', 'SupplyRate', 'UnemploymentRate', 'PopulationGrowth']
# Standardize the data
scaler = StandardScaler()
mortgageSubprimeScaled = scaler.fit_transform(mortgageSubprimeDF[data_cols])
coronaScaled = scaler.fit_transform(CoronaDF[data_cols])
```

- Filter data based on specific date ranges.
- Select columns to be used for PCA.
- Standardize the data using StandardScaler from sklearn.preprocessing.

PCA Analyze 1. How I did

2. PCA:

```
pca = PCA()
mortgageSubprimePCA = pca.fit_transform(mortgageSubprimeScaled)
coronaPCA = pca.fit_transform(coronaScaled)
```

- Perform PCA on the standardized data using PCA from sklearn.decomposition.

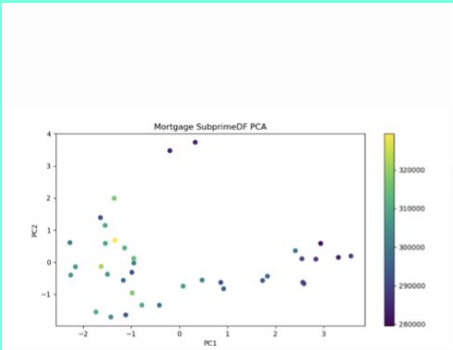
PCA Analyze 1. How I did

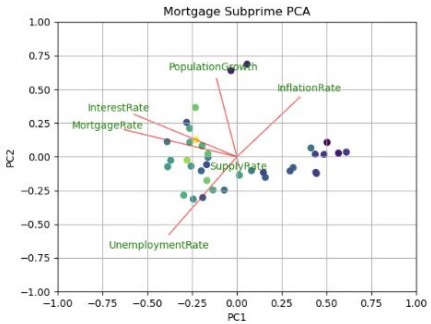
3. Biplot:

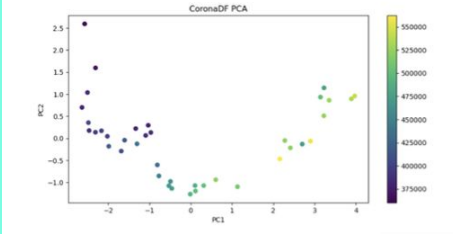
```
def biplot(score,coeff,pcax,pcay,labels=None):
  pca1=pcax-1
  pca2=pcay-1
  xs = score[:,pca1]
  ys = score[:,pca2]
  n=score.shape[1]
  scalex = 1.0/(xs.max()- xs.min())
  scaley = 1.0/(ys.max()- ys.min())
  plt.scatter(xs*scalex,ys*scaley, c= mortgageSubprimeDF['AveragePrice'])
  for i in range(n):
    plt.arrow(0, 0, coeff[i,pca1], coeff[i,pca2],color='r',alpha=0.5)
    if labels is None:
      plt.text(coeff[i,pca1]* 1.15, coeff[i,pca2] * 1.15, "Var"+str(i+1), color='g', ha='center', va='center')
    else:
      plt.text(coeff[i,pca1]* 1.15, coeff[i,pca2] * 1.15, labels[i], color='g', ha='center', va='center')
  plt.xlim(-1,1)
  plt.ylim(-1,1)
  plt.xlabel("PC{}".format(pcax))
  plt.ylabel("PC{}".format(pcay))
  plt.grid()
```

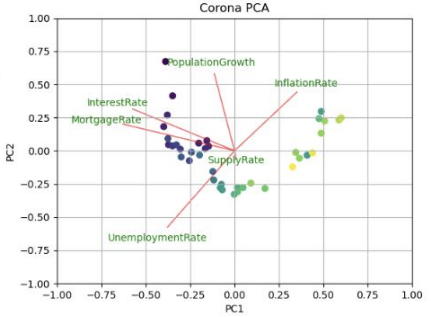
- Create a biplot to visualize the principal components and their relationships with the original variables.
- Show arrows representing the variables and their relationships with the principal components.
- These functions take the PCA scores, PCA components, and the principal components to be plotted as arguments.
- Calculate the scaling factor for the x and y axis and plot the data points and arrows on the biplot.

PCA Analyze 2. Results









PCA of Subprime Mortgage crisis period (2005-01 ~ 2008-01)

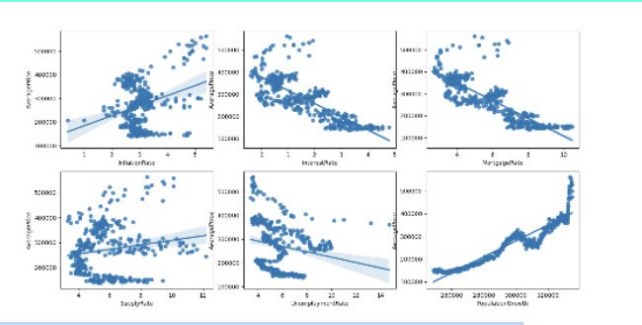
PCA of COVID-19 pandemic period (2020-01~2023-02)

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.

Developing a Machine learning model for predicting house price

1. Visualizing Feature-Target Relationships:
    - Created a scatter plot matrix to visualize the relationships between each feature and the target variable.
2. Model Training and Testing:
    - Split the data into training and testing sets.
    - Trained a linear regression model on the training set.
    - Evaluated model performance on the test set using metrics such as mean squared error, root mean squared error, and R-squared.
3. Model Analysis and Results:
    - Printed the intercept and coefficients of the linear regression model.
    - Calculated the average RMSE using 5-fold cross-validation.
    - Developed a successful model for predicting housing prices based on economic factors.
    - Results of the analysis could be useful for real estate professionals, policymakers, and other stakeholders interested in understanding the factors that influence housing prices.

Linear Regression Model for Predicting housing prices 1.



Correlation with Rising House Prices:
Population Growth

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

Result:

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

```
[118871.81612511, 311969.22737073, 334963.03691715,
180712.52187152, 263545.0444364]
```

```
MSE: 803555571.098,
```

```
RMSE: 28347.05
```

```
Variance score: 0.923
```

Linear Regression Model for Predicting housing prices 2.

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

```
intercept value with y: -923903.0192466965
coefficient: [ 2.71571e+04, 1.35788e+04
-2.98880e+03, 7.06500e+02, -6.13580e+03,
3.80000e+00]
InflationRate       27157.1
InterestRate        13578.8
SupplyRate          706.5
PopulationGrowth      3.8
MortgageRate        -2988.8
UnemploymentRate    -6135.8
dtype: float64
```

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.

```
5 folds each Negative MSE scores:
[-1.80817282e+09, -1.31983058e+09,
-1.06334844e+09, -1.93660796e+09,
-3.39055833e+09]
 5 folds each RMSE scores: [42522.62
36329.47 32609.02 44006.91 58228.5]
 5 folds average RMSE: 42739.304
```

Linear Regression Model for Predicting housing prices 3.

Conclusion:

- The RMSE (Root Mean Squared Error) value of our linear regression model for predicting US housing prices is 42739.304.
- This value indicates the average difference between the predicted and actual values of the average price of US housing in dollars.
- On average, the predicted values are expected to be off by approximately $42,739 from the actual values.
- Lower RMSE values indicate better accuracy of the model, and an RMSE value of 42739.304 suggests that our model is making reasonably accurate predictions.

Conclusion

- The COVID-19 pandemic and Mortgage subprime crisis had similar impacts on economic factors, triggering house bubbles and causing similar effects on housing prices.
- Our machine learning model was able to predict housing prices with an error of approximately $40,000, but additional evaluation metrics should be used to further improve accuracy.
- Improved accuracy in predicting housing prices during economic crises, such as another pandemic or Mortgage subprime crisis, could assist in mitigating their impact.
- Future research could explore other factors that may impact housing prices and incorporate them into the model to enhance accuracy.

Citation.

Slide 3, chart 1, and cite contents.

- Duca, John V., and Anthony Murphy. "Why House Prices Surged as the COVID-19 Pandemic Took Hold." *Www.dallasfed.org*, 28 Dec. 2021, www.dallasfed.org/research/economics/2021/1228.

Thank You
For watching my presentation.