Analysis of Factors Influencing US Home Prices

Swayam Patil
swayam.work25@gmail.com

***Abstract*—This study investigates key factors influencing US home prices over the past 20 years using the SP Case-Schiller Home Price Index. Through exploratory data analysis and machine learning techniques, we identify significant economic indicators affecting home prices and develop predictive models. The findings reveal that lagged home prices, rolling mean prices, and residential investment are among the most crucial factors in predicting home prices. The study provides insights for understanding the dynamics of the US housing market.**

***Index Terms*—Home prices, economic indicators, machine learning, time series analysis**

I. Introduction

The US housing market plays a critical role in the overall economy, influencing personal wealth, consumer spending, and broader economic indicators. Understanding the factors that drive home prices is essential for various stakeholders, including policymakers, investors, and homeowners. This study aims to:

- Identify and analyze key economic indicators affecting US home prices
- Examine relationships between these factors and home prices over 20 years
- Develop and compare predictive models
- Provide insights into the most influential factors for future predictions

II. Data and Methodology

A. Data Collection and Preprocessing

We collected data on various economic indicators from sources such as the Federal Reserve Economic Data (FRED) and the US Census Bureau. Key variables include:

- YearMonth: Month and year of observation
- Federal_Funds_Rate: Interest rate for federal funds trading
- Disposable_Income: Households' available money after taxes
- Housing_Starts: New residential construction projects
- Mortgage_Rate: Interest rate on mortgages
- Housing_Supply: Total housing inventory for sale
- Residential_Investment: Investment in residential construction
- Unemployment_Rate: Percentage of unemployed labor force
- Personal_Savings_Rate: Percentage of saved disposable income
- Home_Price: Price index of homes (target variable)

Preprocessing steps included:

- Converting YearMonth to datetime format
- Creating additional time-based features (Year, Month, TimeIndex, Quarter)
- Generating lag features and rolling statistics
- Handling missing values using backward fill method
- Scaling numerical features using StandardScaler

Preprocessing steps included:

- Converting YearMonth to datetime format and extracting Year and Month:

```
df['Year'] = df['YearMonth'].apply(lamb
df['Month'] = df['YearMonth'].apply(lam
df['YearMonth'] = pd.to_datetime(df['Ye
df['TimeIndex'] = (df['YearMonth'] - df
df['Quarter'] = df['YearMonth'].dt.quar
```

- Creating lag features and rolling statistics:

```
df['Lag1'] = df['Home_Price'].shift(1)
df['Lag3'] = df['Home_Price'].shift(3)
df['Lag12'] = df['Home_Price'].shift(12
df['RollingMean_3'] = df['Home_Price'].
df['RollingStd_3'] = df['Home_Price'].r
```

- Handling missing values using backward fill method:

```
df['Residential_Investment'].fillna(met
df['Federal_Funds_Rate'].fillna(method=
df = df.dropna()
```

- Scaling numerical features using StandardScaler:

```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X
X_test_scaled = scaler.transform(X_test
```

B. Exploratory Data Analysis

1) Time Series Analysis of Key Factors: We conducted a comprehensive time series analysis of the key factors influencing home prices.

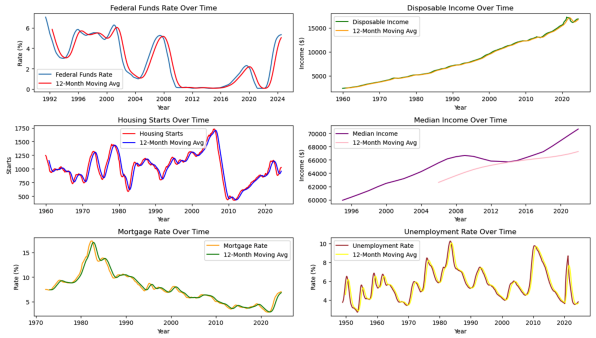

Fig. 1: Multiple Factors Over Time (1990 - 2020)

Fig. 1 shows the trends of various economic indicators from 1990 to 2020. Notable observations include:

- Home prices show a general upward trend with a significant dip around 2008-2010, likely due to the housing market crash.
- The federal funds rate exhibits significant fluctuations, with a sharp decline post-2008 and remaining low through 2020.
- Disposable income shows a steady increase over time, reflecting overall economic growth.
- The federal funds rate shows dramatic changes, particularly the near-zero rates from 2008 to 2016 in response to the financial crisis.
- Housing starts exhibit cyclical patterns with a sharp decline around 2008-2009, clearly showing the impact of the subprime mortgage crisis.
- Mortgage rates show a general downward trend from the 1980s, with stability at low levels in the 2010s.

2) Correlation Analysis: We analyzed the correlations between various factors using a heatmap.

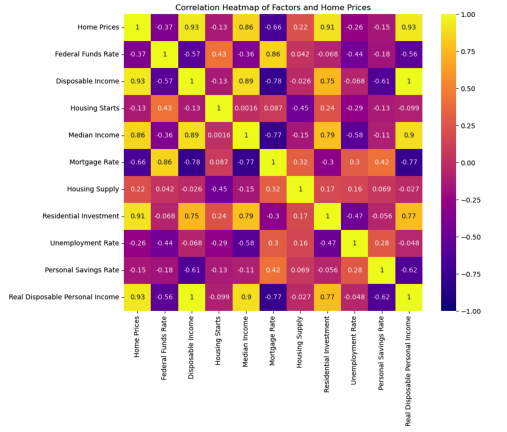

Fig. 2: Correlation Heatmap of Factors Influencing Home Prices

Fig. 2 reveals several important relationships:

- Home prices show strong positive correlations with disposable income (0.93), median income (0.86), and residential investment (0.91).
- There's a notable negative correlation between home prices and mortgage rates (-0.66), suggesting that lower borrowing costs are associated with higher home prices.
- The federal funds rate shows a strong positive correlation with mortgage rates (0.86), reflecting the influence of monetary policy on borrowing costs.

3) Pairwise Relationships: To further explore relationships between variables, we created a pair plot.

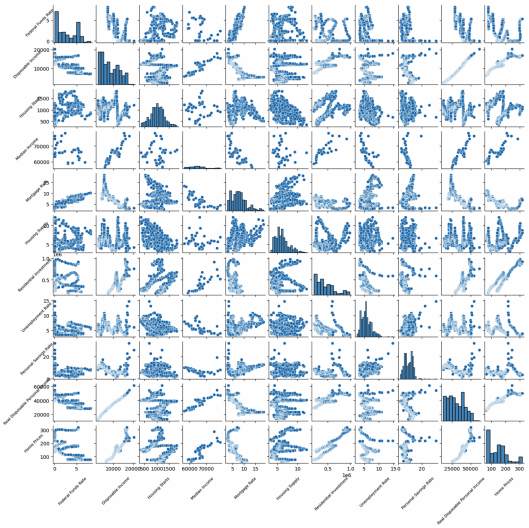

Fig. 3: Pair Plot of Key Variables

Fig. 3 highlights several important relationships:

- Home prices vs. federal funds rate shows a negative relationship with a slight curve, suggesting non-linear effects at higher rates.
- Home prices vs. disposable income exhibits a strong positive relationship with tight clustering of points.
- Home prices vs. mortgage rate displays a negative relationship with a pronounced curve, indicating stronger effects at higher rates.
- Residential investment vs. home prices shows a strong positive relationship, particularly at higher values.

C. Insights from Exploratory Data Analysis

The comprehensive time series analysis revealed complex dynamics in the housing market and related economic indicators. The general upward trend in home prices, interrupted by the 2008 financial crisis, underscores the long-term appreciation of housing as an asset. The federal funds rate's dramatic fluctuations, particularly the near-zero rates post-2008, highlight the significant role of monetary policy in shaping the economic environment for housing.

The correlation analysis provided quantitative support for several economic theories. The strong positive correlations between home prices and factors like disposable income and residential investment suggest that economic growth and investment in housing go hand in hand with rising home prices. The negative correlation with mortgage rates confirms the

intuitive notion that lower borrowing costs tend to drive up home prices by making purchases more affordable.

The pair plot analysis revealed nuanced relationships between variables. Of particular interest is the curved relationship between home prices and the federal funds rate, suggesting that the impact of interest rate changes on home prices may not be linear. This non-linearity could have important implications for predicting the effects of monetary policy on the housing market.

The cyclical nature of housing starts, as seen in the time series plot, emphasizes the sensitivity of new construction to economic conditions. The sharp decline in housing starts during the 2008 crisis illustrates how economic downturns can rapidly impact the supply side of the housing market.

These insights from the exploratory data analysis provide a rich context for interpreting our model results and understanding the complex interplay of factors influencing home prices.

III. Modeling Approach

We selected three models for our analysis:

- Linear Regression: To capture linear relationships
- Random Forest Regressor: To capture non-linear relationships
- Lasso Linear Regressor: For feature selection and regularization

Models were trained on a temporal split of the data and evaluated using Mean Squared Error (MSE) and R-squared (R2) metrics.

IV. Results and Discussion

A. Model Performance

Model	MSE	R2
Linear Regression	0.0622	0.9999
Random Forest	1.0799	0.9998
Lasso Regression	0.3047	0.9999

TABLE I: Model Performance Comparison

All models showed excellent performance, with R2 values very close to 1, indicating they explain almost all variance in home prices. The Linear Regression model showed the lowest MSE.

B. Interpretation of Model Performance

- **Mean Squared Error (MSE)**: This metric measures the average of the squares of the errors. The Linear Regression model has a significantly lower MSE than the Random Forest Regressor, suggesting that its predictions are closer to the actual values.
- **R-squared (R2):** This metric indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. R2 values range from 0 to 1, with higher values indicating better model performance. Both models have extremely high R2 values, but the Linear Regression model has a slightly higher R2, indicating a better fit to the data.

C. Model Comparison

- **Linear Regression**: The exceptionally high R2 value (0.99999) and very low MSE (0.062) suggest that the Linear Regression model explains nearly 100% of the variance in the home prices and has an extremely low prediction error. This is an outstanding performance for a simple model.
- **Random Forest Regressor:** Although the Random Forest model also performs very well, with an R2 value of 0.99980 and an MSE of 1.072, it does not perform as well as the Linear Regression model. The higher complexity of Random Forest might not be necessary for this dataset, especially considering that it results in higher prediction errors compared to Linear Regression.

D. Feature Importance

The Lasso model provided insights into the most important features for predicting home prices.

Feature	Importance
Lag1	51.72
RollingMean_3	6.33
Residential_Investment	1.02
TimeIndex	0.56
RollingStd_3	0.49
Personal_Savings_Rate	0.05
Housing_Supply	-0.29

TABLE II: Feature Importance from Lasso Regression

Interpretation of Feature Importance Results

The Lasso Regression model identified the following features as the most important for predicting the housing price index:

- Lag1: The value from the previous time period is the most significant predictor of the housing price index.
- RollingMean_3: The 3-period rolling mean of the housing price index also plays a significant role.
- Residential_Investment: Investment in residential properties is another critical factor.
- TimeIndex: The time index, which could be an ordinal representation of time, shows moderate importance.
- RollingStd_3: The 3-period rolling standard deviation indicates some variability in the housing prices.
- Personal_Savings_Rate: The rate at which households save their disposable income shows minor importance.
- Housing_Supply: The total housing inventory available for sale negatively impacts the housing price index slightly.

The Lasso Regression model has successfully identified the key features influencing the housing price index, providing insights into the most significant economic indicators.

V. Conclusion

This study examined key factors influencing US home prices using the SP Case-Schiller Home Price Index and various economic indicators. Our analysis revealed:

- Strong positive correlations between home prices and disposable income, and residential investment
- Significant negative correlation between home prices and mortgage rates
- Non-linear relationships between home prices and key factors like the federal funds rate and mortgage rates

Predictive modeling using Linear Regression, Random Forest, and Lasso Regression highlighted the importance of lagged home prices, rolling mean prices, and residential investment in predicting home prices. The Linear Regression model provided the best performance, suggesting that simpler models can effectively capture the dynamics of the US housing market.

Future research could explore more advanced machine learning techniques and incorporate additional economic indicators to enhance predictive accuracy and provide deeper insights into the factors driving home prices.

REFERENCES

[1] Federal Reserve Economic Data (FRED), https://fred.stlouisfed.org/.
[2] US Census Bureau, https://www.census.gov/.