Research Paper

Real Estate Project The Impact of Macroeconomic Factors on House Price

I. Introduction and Literature Review

In the years following the 2008 housing market crash, homes in the United States have been increasingly difficult for prospective home buyers to acquire. Similarly, the prices of houses have gone up tremendously since then, with some saying the housing market is far exceeding sustainable levels of price growth. This study aims to answer the question of what the significant determinants of U.S. residential house prices are by using monthly data from 1987-2021 and looking at macroeconomic factors such as inflation, employment, and income. This will allow us to understand the state of the housing market and determine the significance of the different independent variables studied. The results of this study will be significant to prospective home buyers. Purchasing a home is likely the biggest purchase one would make, so understanding the underlying effects of different economic variables on the housing market can be extremely beneficial. The results of this study can help educate a prospective buyer on the underlying factors of the housing market in 2023 and allow them to make intelligent decisions on the fair market value of a property.

"The Macroeconomic Determinants of House Prices and Rents", by Shida, Jakob, undertook a comprehensive analysis of the macroeconomic drivers influencing house prices and rent across 21 countries. In the context of housing prices significantly positive effects of income and bank lending, along with noteworthy negative impacts of housing stock and interest rates. Remarkably, contrary to the DiPasquale and Wheaton (1992) model, stating that rent drives housing prices, this study challenged that notion, attributing it to the discrepancies of other macroeconomic variables in attempts to estimate future rent expectations. The study emphasizes the need for continued research on the macroeconomic determinants of housing prices and rent, particularly the growing availability of data. Algieri, Bernardina. "House Price Determinants: Fundamentals and Underlying Factors", focuses on the five main Euro areas, while also examining the same macroeconomic factors Shida, Jakob did with a few additional variables between 1970 and 2010. Both papers emphasize the correlation between income, bank lending, interest rates and the increases in housing prices globally, although each paper describes independent variables that affect the price of homes that can't be explained through traditional means, requiring more data to be explained. Regardless of the independent variables that affect the results of the data from the two research papers, they both come to the same conclusion of the effects of interest rates, income and other macroeconomic factors on housing prices. The consistent emphasis on these determinants underscore the significant impact they have on the real estate market across diverse countries.

II. The econometric model: Macroeconomic Factors - Determinants of House Prices

In this study, we delve into the intricate relationships between house prices and key macroeconomic variables, seeking to uncover the underlying determinants that shape housing market trends. This research employs a robust linear regression model to analyze the relationships between the house price index, our dependent variable, and a set of carefully selected independent variables. These independent variables include stock, CPI, the unemployment rate, population size, disposable income, mortgage rate, and real interest rate. The time-series data under consideration spans an extensive period from January 1, 1975, to January 1, 2021, offering a comprehensive view of the long-term trends and fluctuations in the housing market. By incorporating data from diverse sources,

including Kaggle, Data World Bank, and FRED, we aim to capture the nuanced interactions between macroeconomic forces and house prices, providing valuable insights for both academic discourse and practical policy considerations.

The expectations regarding the signs of these coefficients (positive or negative) will be based on economic theory and prior research. According to Bernardina Algieri's conclusions about the study (2013), changes in real income, long-run interest rates, and stock prices have an impact on real house prices (Algieri, 2013). An increase in real income is expected to improve house affordability, leading to higher demand for housing and driving up prices. A decline in real long-term interest rates reduces mortgage rates, boosting liquidity, demand for houses, and house prices. Stock prices also have a positive relationship with real house prices, indicating that the wealth effect dominates the substitution effect. Higher stock prices increase investors' total wealth and their capability to invest in other assets, including houses. Jakob Shida conducted a study focusing on the macroeconomic determinants of house prices and rents. The research paper suggests that the unemployment rate has a negative effect on both house prices and rents (Shida, 2022). The estimated coefficient of the unemployment rate is significant in all specifications, indicating that higher unemployment leads to lower house prices and rents. In the research article titled "Factors Affecting House Prices in Cyprus: 1988-2008", the authors investigate the impact of various macroeconomic variables on house prices in Cyprus during the period 1988-2008. In this article, the factors that contributed most to the increase in house prices were the increase in the cost of materials and labor, the population, and the per capita GDP (Pasharde and Savva, 1988-2008). Therefore, we expect the positive impacts of the population size and GDP on the house price index. Ahmed, in his study "Impact of Macroeconomic Variables on Housing Prices in Saudi Arabia", his variance decomposition analysis reveals that CPI is the variable with the highest explanatory power over the variation in housing prices, followed by GDP and UNEMP, indicating that CPI is the most influential determinant for housing prices (Ahmed, 2020). The study found that CPI has a significant and negative relationship with housing prices. In the short run, a shock to CPI leads to a decrease in housing prices for the first four quarters, and this decrease continues for the next six quarters.

Hence, based on the results and conclusions of those empirical studies, the formulation of our econometric model will be built with the expectation that changes in the independent variables (stock, CPI, unemployment rate, population size, disposable income, mortgage rate, real interest rate) will be associated with changes in the dependent variable, the house price index (HPI). The relationship can be expressed mathematically in a single linear regression equation:

House Price Index

```
= \beta_0 + \beta_1 * Macroeconomic Factor 1 + \beta_2 * Macroeconomic Factor 2 + \beta_3 * Macroeconomic Factor 3 + \cdots + \beta_7 * Macroeconomic Factor 7 + \varepsilon
```

Here, β 0 represents the intercept, and β 1 through β 7 represent the coefficients for each independent variable. The ϵ term represents the

error term. The coefficients indicate direction and strength of the relationship between each independent variable and the house price index.

III. Analysis of Results

1. The descriptive statistics of variables

The statistics table (see table I) provides summary information on the eight economic indicators over the 420 month period. The average house price index over the time frame was 134, while the stock price index had an average of 65. Consumer price index had a higher mean of 191. Population and real disposable income are very large numbers indicating their scale over the national economy. Unemployment rate averaged close to 6 percent. The maximum mortgage rate observed was over 11 percent, much higher than the mean of 6.3 percent. There is considerable variation across the variables as seen in the large standard deviation values. The maximum values tend to be between 2-3 times higher than the minimum values showing the degree of fluctuation over the time.

2. Correlation table

The correlation table (see Table II) shows the relationship between 7 different macroeconomic variables. There are very strong correlations between most variables, especially consumer price index, population, and real disposable income, indicating these factors tend to move together over time. The exception is unemployment rate, which has little correlation between mortgage rate and factors like housing price index and real disposable income.

3. Assumption 3: Multicollinearity

Looking at table(I) - our correlation matrix table, some pairs of variables have very high correlation coefficients. For example, the Stock Price Index has strong positive correlations with Consumer_Price_Index (0.942818) and Real_Disposable_Income (0.9581239). Mortgage_Rate has strong negative correlations with Stock_Price_Index (-0.8715697) and Real_Disposable_Income (-0.93448901) and strong positive correlation with Real_Interest_Rate (0.84228). These correlations provide insights into the relationships between economic indicators. These high correlations suggest a potential issue with multicollinearity - when two or more independent variables in a regression model are highly correlated, making it difficult to isolate the individual effects of each variable on the dependent variable. If we try to put all of those highly correlated independent variables in one linear regression model, our regression model will violate Assumption 3: no perfect collinearity. To address multicollinearity, we consider variable selection, in which we choose a subset of variables that are most relevant to our research question or have lower intercorrelation. And this process, often referred to as univariate regression or simple linear regression, involves running separate regression models for each independent variable to assess their individual impact on the dependent variable. During the process, we try to analyze the significance of each variable based on p-values, coefficients, and goodness-of-fit measures (e.g., R-squared). The second approach of dealing with multicollinearity is data transformation, in which we transform variables to reduce correlation (e.g., using ratio, percentage change instead of raw values). In this dataset, we try to transform population and real disposable income into one variable by dividing Real_Disposable_Income by population. With this calculation, we achieve a new variable called Income_Per_Capita.

4. Regression models

Before running regression models, we take the logarithm of the dependent variable (House_Price_Index) and logarithm of Income_Per_Capita. As running a series of simple linear regression models, each with one independent variable and the dependent variable, as shown in table (III), we notice that all those variables show significant relationships with the dependent variable and contribute meaningfully to the model. Looking at all the models, we can see all those variables have significant impacts on House_Price_Index at 1%. In terms of magnitude, in Model 1, on average, a one-unit increase in the Stock_Price_Index is associated with a 1% increase in the dependent variable (loghpi). In Model 2, on average, a one-unit increase in Consumer_Price_Index is associated with a 0.8% increase in the dependent variable (loghpi). In Model 3, on average, a one-unit increase in Mortgage_Rate is associated with a 15.7% decrease in the dependent variable (loghpi). In Model 4, on average, a one-unit increase in Real_Interest_Rate is associated with a 11.5% decrease in the dependent variable (loghpi). In Model 5, on average, 1% increase in Income_Per_Capita is associated with a 2.091% increase in the dependent variable (loghpi).

Although, overall, it appears that all the models have good performance of how each factor impacts on house price and they are also significant with very small p-values, we want to identify the most preferable model out of those 5 models and use this model for further analysis. To determine which model fits the data most, we typically look at goodness-of-fit measures such as the R-squared values. In our results, we compare the R-squared values across the models since a higher R-squared indicates that a larger proportion of the variance in the dependent variable is explained by the independent variables. In this case, Model 5 has the highest R-squared value so it would be considered the best-fitting model among the ones we've presented.

Since Unemployment_Rate has a very weak correlation with other independent variables, we consider adding this variable in Model 5, which is Model 6. Table IV demonstrates a comparison table of Model 5 and Model 6. We can notice that Model 6 has a better performance since all the independent variables in this model are significant at 1% and it also has a slightly higher R-squared. Since model 5 is violating Assumption 4 (Zero Conditional Mean), model 5 is preferred. With Model 6, in terms of magnitude, on average, 1% increase in Income_Per_Capita is associated with a 2.091% increase in the dependent variable (loghpi), holding all else as constant. Here, when we add another independent variable, the coefficient estimate of in Income_Per_Capita does not change. On average, a one-unit increase in Unemployment_Rate is associated with a 1.2% decrease in the dependent variable (loghpi), holding all else as constant. Therefore, we achieve a complete linear regression model to help predict house price based on those significantly determinant factors as following:

loghpi = -17.932 + 2.091*logipc - 0.012*Unemployment_Rate

The results of our model are consistent with Bernardina Algieri's conclusions about the study (2013) and Jakob Shida's conclusion about his study (2022). The impact of income per capita and the unemployment rate on house prices can be attributed to various economic, demographic, and housing market dynamics. Income per capita is a key determinant of the purchasing power of individuals and households. Higher income levels generally mean more disposable income that can be allocated to housing expenses. As income increases, individuals may be able to afford more expensive homes, if the supply of housing does not keep pace with this

demand, prices tend to rise. The unemployment rate is a crucial indicator of the health of the job market. When the unemployment rate is low, there is typically greater job stability and confidence among workers. This can lead to increased demand for housing as individuals feel more secure in their employment, contributing to higher house prices. Overall, Income stability and low unemployment can enhance consumer confidence in the economy. Confident consumers are more likely to make significant financial commitments, such as purchasing a home.

5. Assumption 6: No serial correlation

Time series data often exhibits autocorrelation due to the inherent temporal structure of the observations. Autocorrelation, or serial correlation, refers to the correlation of a time series with its own past and future values. Serial correlation can contribute to spurious regression results, where the model appears to fit well, but the relationships are misleading due to the autocorrelation. This can lead to overconfidence in the model's predictive power. The ordinary least squares (OLS) estimates of regression coefficients are no longer efficient. T-tests and F-tests that rely on standard errors become invalid in the presence of autocorrelation. Therefore, we try to conduct testing for serial correlation by using T-test for AR1 serial correlation.

$$\mu t = \rho \mu t - 1 + 0 \mathbf{e}_t$$

Hypothesis Test: $H_0=0$, H_a : # 0

Using the T-test, we have a very small p-value (<2e-16e), thus, we reject the null hypothesis (H₀). Since we reject the null hypothesis, there is autocorrelation of residuals. Our model violates Assumption: no serial correlation.

6. Assumption 5: Homoscedasticity

Ensuring homoscedasticity is crucial for the validity of statistical inferences and hypothesis testing in regression analysis. If homoscedasticity is not met, it is essential to explore and apply appropriate corrective measures to enhance the reliability of the regression results. Since we previously detected autocorrelation in Model 6, we, then, test for homoscedasticity by using the Breusch-Pagan test:

Hypothesis Test: H₀: Homoscedasticity,

H_a: Heteroscedasticity

bptest(model6)

The result of the Breusch-Pagan test indicates that we also reject the null hypothesis which means we detect heteroscedasticity since we achieve a very small p-value (4.141e-07).

8. Solutions for serial correlation and heteroscedasticity

We attempt to correct for serial correlation by undertaking a differencing method: Differencing a time series with AR(1) model. After that, we use the Durbin-Watson test to test serial correlation, and it appears that DW value is 0.1517 (<2), thus, we reject the null hypothesis, there is still a positive correlation of residuals across the data. Therefore, it does not eliminate positive serial correlation.

Since our regression model is not strictly exogenous, it's essential to calculate serial correlation-robust standard errors in order to fix serial correlation and heteroscedasticity. We try to use the Newey-West method.

Model8 <- coeftest(mod6,vcov=NeweyWest)

T-test of coefficients	Estimate	Std.Error	T-value	Pr(> T)
intercept	-17.932337	0.575812	-31.1427	<2e-16e
logic	2.09132	0.050696	41.2519	<2e-16e
Unemployment Rate	-0.011659	0.017005	-0.6856	0.4934

The results show that the standard errors of Model 8 are likely to be bigger. And, in the Model 8, we can notice that the magnitude of Unemployment_Rate also changes. Unemployment_Rate no longer has a significant impact on House_Price_Index at any significant level. However, the coefficients of both independent variables are likely to be the same. More importantly, shown in figure (iii) the difference between two models is that one is under the assumptions of no serial correlation and homoscedasticity, holding the assumptions were true, and another one is when we realized that it's not true and we tried to fix autocorrelation and heteroscedasticity by using Newey-West method.

Both homoscedasticity and the absence of serial correlation are important assumptions for unbiased and efficient estimation in linear regression. The Newey-West estimator calculates robust standard errors that take into account potential autocorrelation and heteroskedasticity in the residuals of a regression model. These robust standard errors provide more accurate estimates of the standard errors of the coefficient estimates, especially when the assumptions of homoscedasticity and no autocorrelation are violated. When we use the "coeffest" function with corrected standard errors obtained from the Newey-West estimator, the robust standard errors adjust for potential correlations between residuals at different time points, leading to more accurate t-statistics and p-values. This approach also improves the accuracy of statistical inference, making it more robust to violations of assumptions related to autocorrelation and

heteroskedasticity. Technically, heteroskedasticity is automatically corrected for when we use the serial correlation-robust formulas for standard errors and test statistics.

IV. Conclusion

Our research paper examines the impact of macroeconomic factors on the house price index over the period 1987-2021. This study focuses on macroeconomic factors like the stock price index, consumer price index, unemployment rate, income per capita, mortgage rate, and interest rate. We employ a robust linear regression model to analyze the relationship between the house price index and various macroeconomic variables. Our findings indicate that all those macroeconomic factors significantly impact the house price index. This indicates a complex interplay of economic variables in influencing the housing market. While regression models for each independent variable indicate that the stock price index, consumer price index, income per capita, mortgage rate, and interest rate strongly impact the house price index, the regression model with the independent variable income per capita is the best fitting model out of those models. It also highlights income per capita is a stronger predictor among them. Compared to this simple linear regression model, the multiple regression model is preferred. The conclusion is that 92% of the house price index is explained by income per capita and the unemployment rate, emphasizing the dominant role these two factors play in shaping housing market trends over the studied period. Before the study, we have not known the extent to which income per capita and the unemployment rate dominate in explaining house price variations. This insight into the dominance of these variables provides a clearer understanding of the primary drivers of housing market dynamics.

Before the study, we did not know the extent to which income per capita and the unemployment rate dominate in explaining house price variations. This insight into the dominance of these variables provides a clearer understanding of the primary drivers of housing market dynamics. The intercorrelation of various macroeconomic factors underscores the complexity of the housing market.

Policymakers can use our findings to prioritize interventions that target income growth and employment stability, recognizing their substantial impact on housing market stability. The findings suggest policy implications such as the effectiveness of monetary controls and direct methods like employment incentives or tax adjustments for affordable housing. The identified influential factors, including the stock price index, consumer price index, mortgage rate, and interest rate, can also serve as indicators for policymakers to monitor and respond to changes in economic conditions that may affect the housing market. The lack of data on tax policy or on any government subsidy could affect the module that could be out. In the future, looking at macroeconomic policies like tax policy and government spending could give insets into relations that affect housing affordable also holding the values of the house for current investors and billions of people, many investors own homes that keep the value of the house, we won't return to the housing crash like the 2008 housing crash.

While our model explains a significant portion of the house price variability, it may not capture all relevant factors. Consideration of additional variables or external events might provide a more complete picture. The quality and availability of data for certain variables may pose limitations. Sensitivity analyses or robustness checks can help assess the impact of data uncertainties on your results.

V. Data Appendix

The data set we use is from 01/1987 - 12/2021

Data	Description	Source
House Price Index	House price change according to the index base period set	Kaggle
Stock Price Index	Stock price change according to the index base period set	Kaggle
Consumer Price Index	The Consumer Price Index measures the overall change in consumer prices based on a representative basket of goods and services over time.	Kaggle
Population	Population of the U.S	Kaggle
Unemployment Rate	Unemployment rate of the U.S	Kaggle
Mortgage Rate	Interest charged on mortgages - 30 years fixed (unit: percentage).	Kaggle
Real Disposable Income	Money left from salary after all the taxes are paid (unit: chain 2012 dollars).	Kaggle
Real Interest Rate	The lending interest rate adjusted for inflation as measured by the GDP deflator.	FRED

VI. References

Shida, Jakob. "The Macroeconomic Determinants of House Prices and Rents." Jahrbucher Fur Nationalokonomie Und Statistik, vol. 242, no. 1, 2022, pp. 39–86. https://www-degruyter-com.offcampus.lib.washington.edu/document/doi/10.1515/jbnst-2020-0043/html

Algieri, Bernardina. "House Price Determinants: Fundamentals and Underlying Factors." Comparative Economic Studies, vol. 55, no. 2, June 2013, pp. 315–41.

 $https://web-s-ebscohost-com.off campus.lib.washington.edu/ehost/detail/vid=6\&sid=ff71fe0b-bd41-4e07-8fa3-724995d27015\% \\ 40redis\&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ\%3d\%3d\#AN=1387633\&db=ecn$

Pashardes, Panos and Savva, S. Christos. "Factors Affecting House Prices in Cyprus: 1988-2008". January, 2009. https://www.researchgate.net/publication/228632026 Factors Affecting House Prices in Cyprus 1988-2008

Ahmed, D.E.E.A. Impact of Macroeconomic Variables on Housing Prices in Saudi Arabia: A VAR Approach. Preprints 2020, 2020120715. https://www.preprints.org/manuscript/202012.0715/v1

Data set: Kaggle https://www.kaggle.com/datasets/faryarmemon/usa-housing-market-

factors?select=Monthly Macroeconomic Factors.csv

FRED

https://fred.stlouisfed.org/series/REAINTRATREARAT10Y

VII. Tables

Table I: Descriptive Statistics

Variables	N	Mean	St. Dev.	Min	Max
Date	420	2004.065	10.112	1987.01	2021.12
House_Price_Index	420	133.991	51.096	63.735	278.681
Stock_Price_Index	420	64.743	35.396	13.358	159.943

Consumer_Price_Index	420	191.199	44.707	111.4	280.887
Population	420	185029124	17678878	153968000	206780000
Unemployment_Rate	420	5.875	1.648	3.5	14.7
Mortgage_Rate	420	6.341	2.207	2.684	11.26
Real_Disposable_Income	420	10,494,868,333,333,000	2,807,838,670,022.000	5,967,200,000,000	19,119,500,000
Real_Interest_Rate	420	3.216	2.176	0.25	7

Table II: Correlations

Variables	Date	House_Pric e_Index	Stock_Price_ Index	Consumer_P rice_Index	Population	Unemploym ent_Rate	Mortgage_R ate	Real_Disposa ble_Income	Real_Inter est_Rate
Date	1	0.9325364	0.9503324	0.99776888	0.98225788	0.04031626	-0.95593204	0.99080224	-0.75778
House_Price_Ind ex	0.932536	1	0.9462049	0.9342313	0.9119139	-0.0767721	-0.8515976	0.9524269	-0.60115
Stock_Price_Inde x	0.950332	0.9462049	1	0.942818	0.9091475	-0.1648879	-0.8715697	0.9581239	-0.6056
Consumer_Price _Index	0.997769	0.9342313	0.942818	1	0.9821235	0.05472914	-0.94785992	0.9872744	-0.75355
Population	0.982258	0.9119139	0.9091475	0.9821235	1	0.07401456	-0.94785992	0.96572914	-0.75355

Unemployment_ Rate	0.040316	-0.0767721	-0.1648879	0.05472914	0.07401456	1	-0.16302033	0.03049829	-0.45125
Mortgage_Rate	-0.95593	-0.8515976	-0.8715697	-0.94785992	-0.94785992	-0.16302033	1	-0.93448901	0.84228
Real_Disposable _Income		0.9524269	0.9581239	0.9872744	0.96572914	0.03049829	-0.93448901	1	-0.72983
Real_Interest_Ra te		-0.6011545	-0.6056035	-0.7535463	-0.7535463	-0.45125091	0.84228	-0.7298341	1

Table III:Regression results

Variables	1	2	3	4	5
	0.010 ***				
	(0.0002)				
Stock Price					
		0.008 ***			
		0.0001			
Consumers Price index					
			-0.157 ***		
			(0.004)		
Mortgage Rate					
				-0.115 ***	
				(0.007)	
Real interest rate				,	

					2.091 ***
					(0.029)
Logic					
	4.153 ***	3.229 ***	5.818 ***	5.193 ***	-17.999 ***
	(0.014)	(0.026)	(0.027)	(0.026)	0.318
Constant					
Observations	420	420	420	420	420
R2	0.874	0.906	0.782	0.409	0.925
Adjusted R2	0.874	0.906	0.781	0.408	0.925
Residual Std. Error (Cdf = 418)	0.139	0.120	0.183	0.301	0.107
F Statistic (Cdf = 1; 418)	2,912.196	4,040.834	1,497.806	289.183	5,147.734

Table IV

Variable	C1	Std_Error_C1	C2	Std_Error_C2
logipc	2.091	(0.029)	2.091	(0.029)
Unemployment_Rate			-0.012	(0.003)
Constant	-17.999	(0.318)	-17.932	(0.314)
Observations	420		420	
R2	0.925		0.927	
Adjusted R2	0.925		0.927	

Residual Std. Error	0.107	(df = 418)	0.106	(df = 417)
F Statistic	5,147.734	(df = 1; 418)	2,659.672	(df = 2; 417)

VIII. Figures

Figure (i)

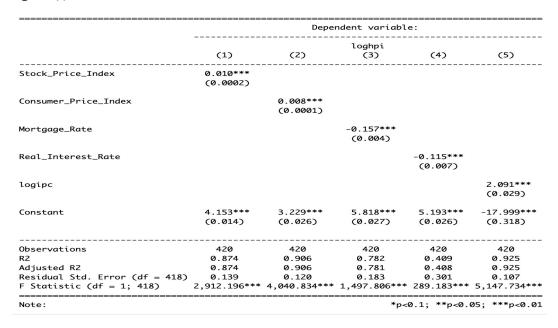


Figure (ii)

Dependent variable:						
	loghpi					
	(1)	(2)				
logipc	2.091***	2.091***				
	(0.029)	(0.029)				
Unemployment_Rate		-0.012***				
		(0.003)				
Constant	-17.999***	-17.932***				
	(0.318)	(0.314)				
Observations	420	420				
R2	0.925	0.927				
Adjusted R2	0.925	0.927				
Residual Std. Error	0.107 (df = 418)	0.106 (df = 417)				
F Statistic	5,147.734*** (df = 1;	418) 2,659.672*** (df = 2; 417)				
Note:		*p<0.1; **p<0.05; ***p<0.01				

Figure (iii)

	Dependent variable	e:
	loghpi OLS (1)	coefficient test (2)
logipc	2.091*** (0.029)	2.091*** (0.051)
Unemployment_Rate	-0.012*** (0.003)	-0.012 (0.017)
Constant	-17.932*** (0.314)	-17.932*** (0.576)
Observations R2 Adjusted R2 Residual Std. Error F Statistic	420 0.927 0.927 0.106 (df = 417) 2,659.672*** (df = 2; 417)	