STUDYING THE IMPACT OF MORTGAGE RATES, GDP PER CAPITA, AND POPULATION GROWTH ON HOUSING PRICES IN USA FROM 2008 TO 2022.

PROJECT BY:

YASH GUPTA

ASHUTOSH KUMAR

MOKSH AGGARWAL

ANUBHAV DHARMANI

DATASET

YEAR	Housing Price Index (HPI)	Mortgage Rates (MR)	GDP per capita (GDP)	Population Growth (PG)
2008	358.72	6.03	55105	0.945865
2009	338.81	5.04	53213	0.876651
2010	324.99	4.69	54188	0.829617
2011	312.44	4.45	54603	0.726787
2012	311.61	3.66	55422	0.7336
2013	323.3	3.98	56172	0.69286
2014	339.43	4.17	57138	0.733362
2015	356.86	3.85	58363	0.736217
2016	375.52	3.65	58968	0.724676
2017	395.88	3.99	60001	0.632644
2018	417.81	4.54	61418	0.526435
2019	437.16	3.94	62606	0.455381
2020	459.71	3.11	60983	0.964348
2021	522.89	2.96	64409	0.156747
2022	611.13	5.34	65420	0.377565

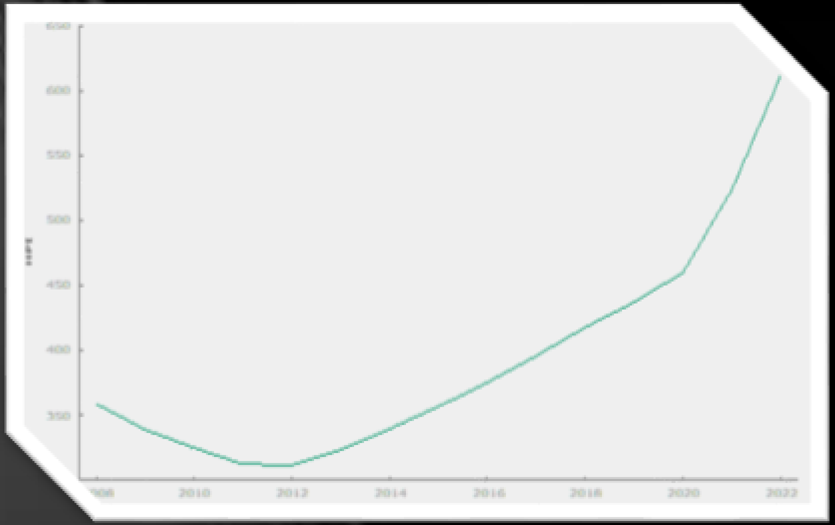
Sources Of Data

(extracted from fred.stlouis.fed.org)

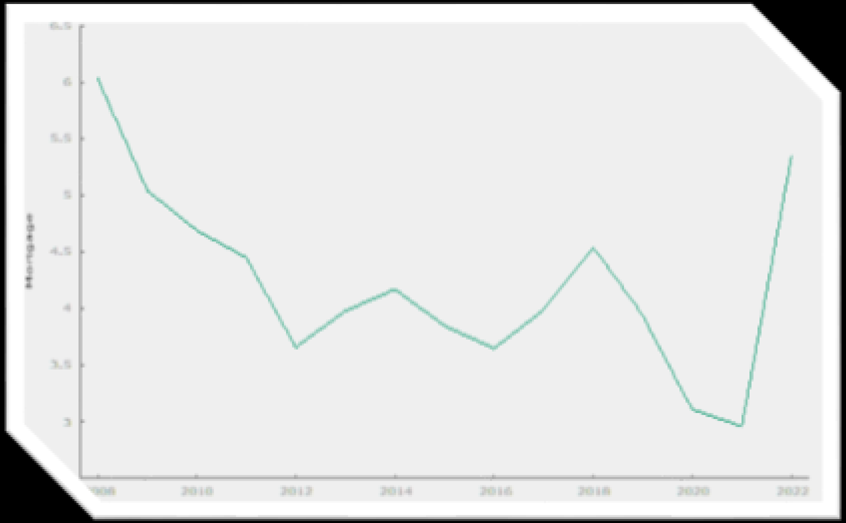
1. US Federal Housing Finance Agency.
2. US Bureau of Economic Analysis.
3. Freddie Mac.
4. World bank

PLOT OF DATA (2008-2022)

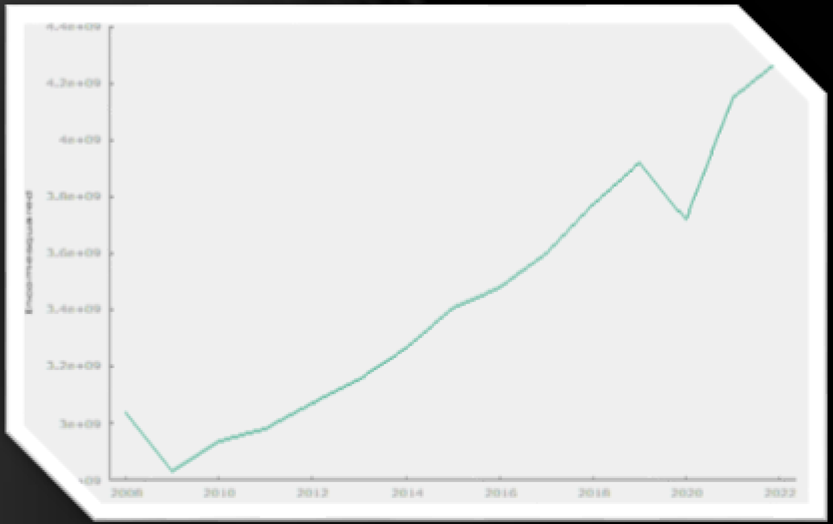
House price index



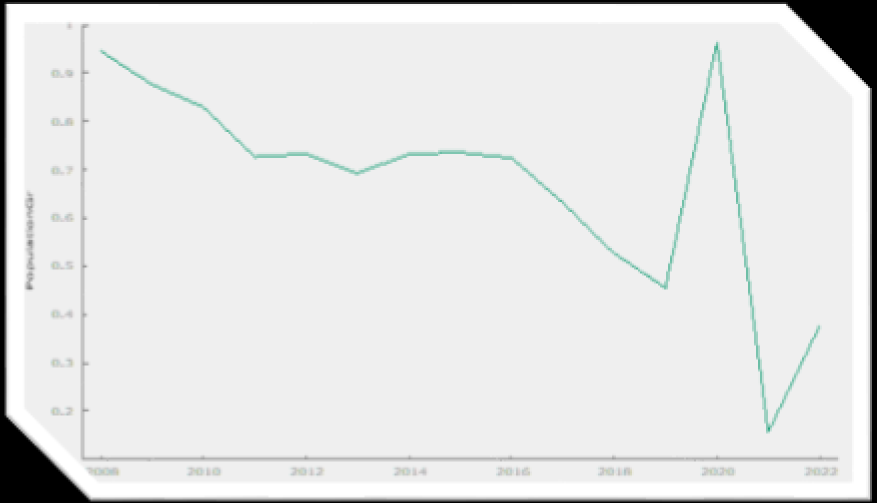
Mortgage rate



GDP pet capita squared



Population growth



INTRODUCTION

HERE WE ARE USING THE METHOD OF ORDINARY LEAST SQUARES WITH HOUSING PRICE AS THE REGRESSAND AND GDP PER CAPITA , GDP PER CAPITA SQUARED, MORTGAGE RATES AND POPULATION GROWTH AS EXPLANATORY VARIABLES OR REGRESSORS.

WE NOW PROVIDE A BRIEF EXPLANATION OF THE EXPLANATORY VARIABLES

- Mortgage Rates: Lower mortgage rates generally make it more affordable to buy a house, which can lead to increased demand and higher housing prices. Conversely, higher mortgage rates can dampen demand and put downward pressure on prices.
- GDP per capita: GDP per capita is a measure of a country's economic output per person. A rising GDP per capita can indicate a stronger economy and potentially higher household incomes, which can fuel demand for housing and drive up prices.
- Population Growth: Population growth can lead to increased demand for housing, especially in areas with limited land availability. This can put upward pressure on housing prices.

REGRESSION RESULTS

-

DEPENDENT VARIABLE: HOUSING PRICES

METHOD: ORDINARY LEAST SQUARES

INCLUDED OBSERVATIONS: 15 (2008-2022)

Variables	Cofficient	Std.Error	t-Statistic	P-Value
Const	-1085.33	172.071	-6.307	5.78e-05 ***
MR	-26.4155	9.31669	2.835	0.0162 **
PG	22.6134	49.3673	0.4581	0.6558
GDP	0.0230781	0.00280332	8.232	4.97e-06 ***
Mean dependent var	392.4173	S.D. dependent var		85.63291
Sum squared resid	11578.76	S.E. of regression		32.44402
R-Squared	0.887215	Adjusted R-Squared		0.856455
F(4, 10)	26.80389	P-Value (F)		0.000023
Log-Likelihood	-71.15066	Akaike criterion		150.3013
Schwarz criterion	153.1335	Mannan-Quinn		150.2711
rho	0.590298	Durbin-Watson		0.834852

ANOVA TABLE
(ANALYSIS OF VARIANCE)

SOURCE OF VARIATION	SUM OF SQUARES	D.O.F	MEAN SUM OF SQUARES
Regression	91083.2	3	30361.1
Residual	11578.8	11	1052.61
Total	102662	14	7333

R^2 = 91083.2 / 102662 = 0.887215

F(4, 10) = 30361.1 / 1052.61 = 28.8435 [P-VALUE 1.64E-005]

INTERPRETATION OF THE RESULTS

➢ <u>ESTIMATED MODEL:</u>

HPI^ = -1085.33 + 26.4*MR + 22.6*PG + 0.0231*GDP

(1) WHEN MR INCREASES BY ONE UNIT THEN MEAN PREDICTED HPI DECREASES BY 26.4 UNITS, KEEPING ALL OTHER VARIABLES CONSTANT.

(2) WHEN PG INCREASES BY ONE UNIT THEN MEAN PREDICTED HPI INCREASES BY 22.6 UNITS , KEEPING ALL OTHER VARIABLES CONSTANT.

(3) WHEN GDP INCREASES BY ONE UNIT THEN MEAN PREDICTED HPI INCREASES BY APPROX. .0231 UNITS. CETERIS PARIBUS.

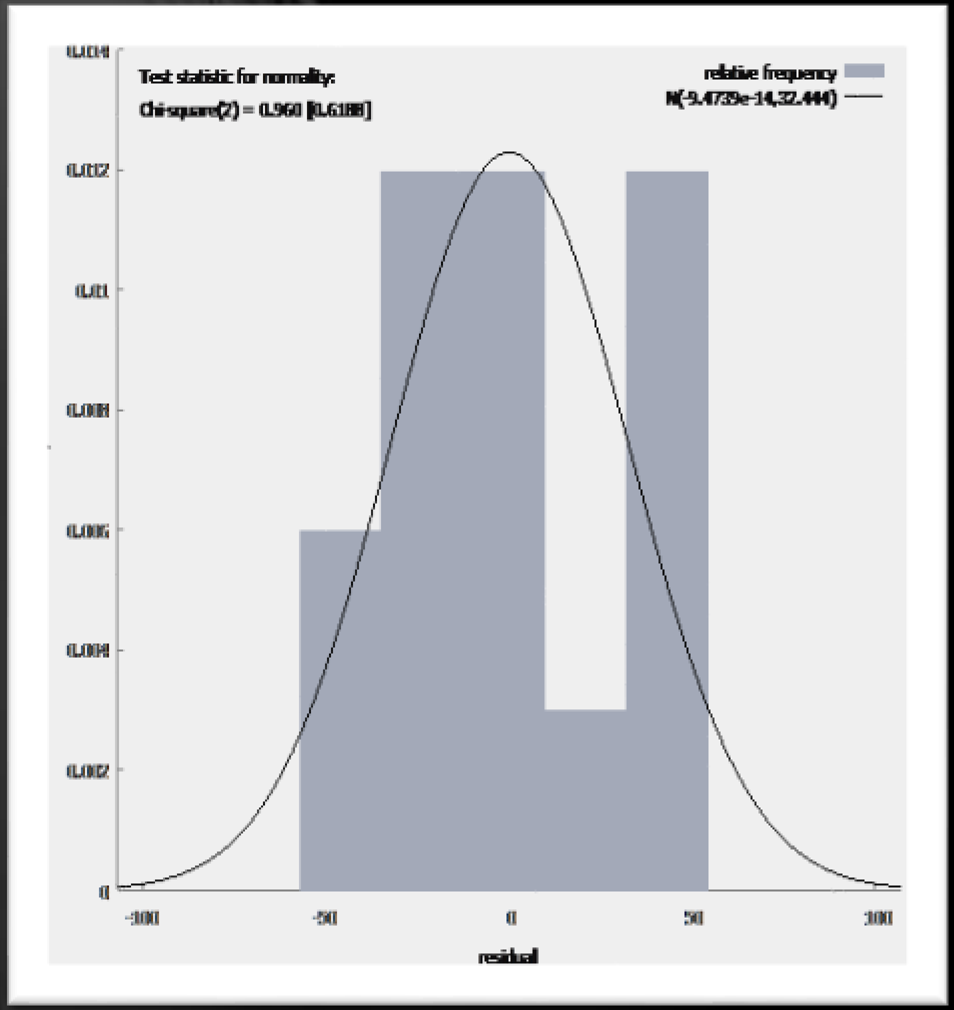
(4) THE HIGH R-SQUARE REPRESENTS THAT 88.72% OF THE VARIATION IN HPI IS EXPLAINED BY THE VARIATION IN MR, PG, GDP . IN OTHER WORDS THE MODEL HAS A GOOD FIT.

(5)THE INDEPENDENT VARIABLES MR , GDP ARE SIGNIFICANT BUT THE REGRESSOR PG COMES OUT TO BE INSIGNIFICANT AS CAN BE SEEN FROM RESPECTIVE T-STATISTICS AND THE ASSCOIATED PROBABLITIES

<u>WE WILL NOW CHECK IF OUR OLS ESTIMATES ARE VALID OR NOT BY CHECKING IF SOME OF THE STANDARD CLRM ASSUMPTIONS ARE VIOLATED. FOR THIS WE CARRY OUT A FEW TESTING EXERCISES</u>.

JARQUE-BERA NORMALITY TEST OF THE RESIDUALS

- THE TEST CHECKS WHETHER THE RESIDUALS OR ERROR TERM ARE NORMALLY DISTRIBUTED OR NOT.



HO: RESIDUALS ARE NORMALLY DISTRIBUTED

HA: RESIDUALS ARE NOT NORMALLY DISTRIBUTED

JB(TEST STATISTIC)=0.960 < C.V.(AT 5% L.O.S)= 5.991

HENCE , ON THE BASIS OF SAMPLE EVIDENCE WE REJECT HO ,

SO RESIDUALS FOLLOW NORMAL DISTRIBUTION.

RESTRICTED LEAST SQUARES

- LET'S SEE THE CONSEQUENCES OF OMITTING RELEVANT VARIABLES FROM THE REGRESSION MODEL.
- *RESTRICTION SET* HO: RESTRICTIONS ARE VALID

1: B[MORTGAGE] = 0 HA: RESTRICTIONS ARE NOT VALID

- TEST STATISTIC: ROBUST F(1, 11) = 8.03887, WITH P-VALUE = 0.016218
- ASSUME(L.O.S=5%), NOW C.V.=F(2,11)= 3.98
- SINCE F-VALUE IS HIGHLY SIGNIFICANT, ALSO THE P-VALUE IS TOO SMALL I.E HIGHER THE EVIDENCE AGAINST THE NULL ,SO ON THE BASIS OF SAMPLE EVIDENCE WE REJECT HO
  - SO WE REJECT THE RESTRICTED REGRESSION
  - THE MORTGAGE RATES ,POPULATION GROWTH & GDP TOGHETHER SIGNIFICANTLY IMPACTING THE HOUSING PRICE INDEX IN UNRESTRICTED REGRESSION MODEL.

DETECTION OF MULTICOLLINEARITY

In order to check whether or not there is any problem of correlation among the explanatory variables whether or not our model suffers from the problem of multicollinearity, we use the Variance Inflating Factor (VIF) where,

$$VIF = 1/1-R_J\text{^}2$$

We regress each explanatory variable on the others referred to as Auxiliary Regressions. the squared multiple correlation coefficients obtained from each of these regressions. If the VIF corresponding to any auxiliary regression is greater than 10 we say that at the model has regression is greater a severe multicollinearity problem and VIF between 2 and 10 implies moderate multicollinearity. The general solution suggested is to drop the corresponding regressor.

DETECTION OF MULTICOLLINEARITY

WE HAVE OBTAINED THE FOLLOWING RESULTS

REGRESSORS	VIF
MR	1.119
PG	2.382
GDP	2.470

From the results obtained above we see that in our model the VIF of MR, PG & G D P are LESS THAN 10.Thus we conclude that the model is free from multicollinearity or in other words, the explanatory variables are uncorrelated.

WHITE'S HETEROSKEDASTICITY TEST

- The White test is a statistical test used in regression analysis to check for heteroscedasticity. Heteroscedasticity refers to the violation of an assumption in linear regression that the variance of the errors is constant across all levels of the independent variables. In simpler terms, it checks if the spread of the residuals (errors) in your regression model is consistent or not.

White's test for heteroskedasticity
OLS, using observations 2008-2022 (T = 15)
Dependent variable: uhat^2

Variables	Cofficient	Std.Error	t-Statistic	P-Value
CONST	-1.03125E+06	942935	-1.094	0.3240
MR	59399.7	41165.4	1.443	0.2086
PG	345537	328602	1.052	0.3412
GDP	27.4822	26.3317	1.044	0.3444
SQ_MR	1078.92	1933.61	0.5580	0.6009
X2_X3	-20212.2	18370.7	-1.100	0.3214
X2_X4	-0.929608	0.760582	-1.222	0.2761
SQ_PG	-14874.3	25656.2	-0.5798	0.5872
X3_X4	-4.28773	4.02813	-1.064	0.3358
SQ_GDP	-0.000179915	0.000181229	-0.9927	0.3664

RESULTS

HO: R^2= 0

HA: R^2 NOT EQUAL TO ZERO

TEST STATISTIC: T = 11.351686,

P- VALUE = 0.252371

ASSUME (L.0.S=5%), SO C.V.=16.919

HENCE C.V. > T(11.351686,) ,SO ON THE BASIS OF

SAMPLE EVIDENCE WE DO NOT REJECT HO, ALSO THE P-VALUE IS LARGE (P-VALUE=0.252371),HENCE THERE IS NO HETEROSKEDASTICITY.

RAMSEY RESET TEST

The Ramsey RESET test is a statistical test used in linear regression to assess whether the model is correctly specified. In simpler terms, it checks if the chosen linear model is the most appropriate way to represent the relationship between the variables. The test works by adding higher-order terms (squares, cubes, etc.) of the fitted values from the initial regression into a new regression equation. If the original model was well-specified (linearity holds), these added terms should not be statistically significant. If they are significant, it suggests that a non-linear relationship exists between the variables, and the linear model may not be the best fit.

Auxiliary regression for RESET specification test
OLS, using observations 2008-2022 (T = 15)
Dependent variable : HPI

	Coefficient	Std. error	t-ratio	P-value	
Const.	2570.40	1197.99	2.146	0.0575	*
MR	-65.4224	30.8868	-2.118	0.0602	*
GDP	-0.0422354	0.0213023	-1.983	0.0755	*
PG	15.8639	46.0481	0.3445	0.7376	
FITTED^2	0.00349888	0.00113225	3.090	0.0114	**

- THIS TEST CAN BE DONE BY TAKING THE NO OF FITTED ITEMS AS 1

- Test statistic : F = 9.549434
- With p-value = P(F(1,10) > 9.54943) = 0.0114

RESULTS – RAMSAY RESET TEST

Ho : THE MODEL IS CORRECTLY SPECIFIED
HA : THE MODEL IS MIS-SPECIFIED .

P-VALUE=0.0114
ASSUME (L.O.S = 5%) , SO C.V = 4.97 , F = 9.549439

SINCE F(TEST – STATISTIC) > CV, ALSO THE P-VALUE IS TOO SMALL INDICATES HIGHER EVIDENCE AGAINST THE NULL,SO WE REJECT HO

SINCE THE , F- VALUE IS MUCH HIGH THEN C.V, SO ON BASIS OF SAMPLE EVIDENCE WE REJECT HO , HENCE THE MODEL IS MISPECIFIED OR SIGNIFICANT .

AUTOCORRELATION

1. DURBIN-WATSON TEST: The Durbin-Watson test is a statistical test used in regression analysis to detect autocorrelation in the residuals, or errors, of the model. Autocorrelation means that the mistakes from one observation are correlated with the errors from previous observations. This can invalidate the regression analysis results, making it difficult to draw accurate conclusions about the relationships between the variables. The Durbin-Watson statistic takes a value between 0 and 4.A value close to 2 indicates no autocorrelation, while values closer to 0 or 4 suggest positive or negative autocorrelation.

RESULTS-DURBIN WATSON TEST

Our Computed d-value is 0.834852.
Ho: No Positive autocorrelation
Ha: No Negate autocorrelation
→ Computed lower limit (dl) = 0.591
upper limit (du) = 101.465
these limits values are taken from the Durbin- Watson table
K = No of regressors
→Since the computed d of 0.834852 lies b/w the lower (du) & upper (dl) limit.
So we can conclude that the autocorrelation exists, since it lies in the zone of indecision or region of ignorance.

REMEDIAL MEASURE TO CORRECT AUTOCORRLEATIOM

USE A GENERALIZED LEAST SQUARES (GLS) ESTIMATION:

- IF THE AUTOCORRELATION PATTERN IS KNOWN, THE GLS ESTIMATION METHOD CAN BE USED TO CORRECT FOR AUTOCORRELATION. THIS METHOD INVOLVES TRANSFORMING THE ORIGINAL REGRESSION MODEL TO REMOVE THE AUTOCORRELATION AND THEN APPLYING ORDINARY LEAST SQUARES (OLS) TO THE TRANSFORMED.
- MODEL: PRAIS-WINSTEN, USING OBSERVATIONS 2008-2022 (T = 15)
  DEPENDENT VARIABLE: HPI
  RHO = 0.916827

VARIABLES	Coefficient	Std. error	t-ratio	P-value
Const.	-832.327	361.768	-2.301	0.0420 **
MR	12.8144	8.65774	1.480	0.1669
PG	44.2121	34.2178	1.292	0.2228 ***
GDP	0.0198956	0.00600036	3.136	0.0069

NOW OUR TRANSFORMED MODEL AFTER RUNNING (GLS)-AUTO REGRESSIVE(1)

MODEL: HPI^ = -832 + 12.8*MR + 44.2*PG + 0.0199*GDP

HERE WE ASSUMED THAT THE ERROR TERM FOLLOWS THE AR(1) SCHEME

NOW THE MODEL OBTAINED ABOVE IS FREE FROM THE PROBLEM OF AUTOCORRELATION

THE OLS ESTIMATORS ARE NOW BLUE (BEST LINMEAR UNBIAS ESTIMATORS) INSTEAD OF LUE, IT WILL NO LONGER LOOSE ITS EFFICIENCY

THERE WILL NO LONGER BE THE ISSUES OF INSIGNIFICANT T-RATIOS OR INVALID HYPOITHESIS TESTING

THANKYOU