

Significance Level of Factors Effecting House Price

Jinhe Liu *

Concord college, Shropshire, SY5 7PF, United Kingdom

* Corresponding Author Email: 2437902@concordcollege.org.uk

Abstract. This paper is about analyzing dataset recorded between 1975 and 2021 about the house prices in United State and uses multiple linear regression to plot linear equation to help to predict the future house prices. The data includes some essential factors, population for example, that influences house prices index both directly and indirectly and did a further analysis to find the correlation between these factors and house prices index, and the correlation between each influencing factors. These factors all have difference significance level comparing to house price index, making it worth to find out even deeper the most influencing factors that affect house price index so the most important factors can be deduced. In the result, real disposable income is being concluded as the most important factors effecting the house price index among all other factors. In addition, the author also uses determinant coefficient to further confirm the accuracy of the multiple linear regression, making sure linear regression is a valid method to analyze the significance of all influencing factors.

Keywords: Linear regression; house price; determinant coefficient.

1. Introduction

Real estate has a fundamental role in the economy. It effects the entire market in various ways. In real life conditions, house price is affected by a lot of factors, including societal, economical and policy influences [1]. For instance, population density in an area will significantly affect the local house price. Also, a well environment also has some impacts on the local house prices [2]. By renovating the street landscape, or having a better transportation network around houses, all of these will cause the house prices to rise. In addition, some special geographical area, places that has great cultural values for instance, may also cause the house prices around to add value.

Despite the societal factors, economical influences also cause the house price to fluctuate a lot. Land, being the most related commodity associated with real estate, has a huge impact on the house prices. Land is a necessary resource that is required to construct houses. Because of this, the supply of available lands become extremely important in determination of house prices [3]. When lands are in a great scarce and the demands are higher than the supply, the house price are forced to push up so the consumers that are willing to pay a higher price gets it. In addition, interest rate also plays an important role in the house price determination. Generally, the rise in interest rate will cause the house price to fall, displaying an inverse relationship between interest rate and house price [4]. This is because the increase in interest rate will leads to a rise in the cost of purchasing houses, causing consumers unable and less willing to purchase houses. This will lead to a shrink in demand for houses, which will pressure the firms to lower house prices to attract consumers. On the other hand, increase in interest rate also significantly raise the cost of constructing houses for firms, causing supply to decrease. But the house price in this situation is still uncertain [4].

Furthermore, the policy set by government in each country also directly influences the direction of house price. A suitable policy is usually being understand by allowing the house price to have little fluctuation and allows citizens to be able to afford houses by giving subsidies in special conditions. Yuan Jing et al. discussed the ideal housing policy by analyzing the economy structure and welfare loss in extreme house price fluctuation in China [5]. By increasing the transparency of the policy, letting the citizens to be more certain on the future expectation of house prices, this is a type of expected policy direction that government will generally make. Apart from this, house prices are also related with monetary policy. Monetary policy is a series of policy that generally effects the aggregate demand in the economy. Some of the strategy being used in monetary policy includes setting interest

rate in central bank [6]. Although different articles describe the effectiveness of altering interest rate on the house price to be either significant or indifferent, interest still shows an influence both in short run and long run. Not only monetary policy, fiscal and some special policies also aim to regulate the fluctuation of policy [7]. Purchase restriction is one of it. This policy helps to prevent a rapid house price growth, but also sacrifice the volume of demand in the market. Among the direct effect on house prices of these macroeconomics policies, they mostly effect the house prices indirectly in a long term. The main effect of the macroeconomics is to promote a sustainable economic growth with a low unemployment level, allowing everyone's income level to rise, which allows them to be able to purchase big ticket product, such as real estate, in a long run [8].

Overall, house price is an important support for the entire economy that each government would aims to stabilize it. This is because house price will affect the demand and purchase of houses, which determine the future economic growth [9]. Most purchase of houses in developed countries is through mortgage. In an economic boom, the demand for houses will rises due to an improved purchasing power of citizens, causing a shortage in supply. This will then encourage producers to construct more houses, which creates more job opportunities, allowing more people to be employed, which will have a very positive effect on the future economy.

In this paper, the author uses linear regression method to predict the future directions of house prices [10]. As mentioned above, government aims for a minimal fluctuation in house prices in most cases. Thus, during the modelling, a stable and smooth prediction is favored. However, in some extreme cases that goes out of government controls such as epidemics for example, an extreme fluctuation will appear, which is very challenging to predict what will it brings to the house price levels.

2. Methods

2.1. Data Source

In this paper, the data is founded on the website named "Kaggle". Inside the data set, it contains the house price index of United States between 1975 and 2021, including the other related index and data such as population and real disposable income.

2.2. Variables Selection

In the original data set, there are a vast amount of data which is not all necessary in predicting housing price. Thus, an abstraction is being done and left out with the only variables needed for carrying on linear regression and is displayed in table 1:

Table 1. Definition of each column name variables and data range

Variables	Meaning	Data Range
House_Price_Index	The house price present in percent form	>0
Consumer_Price_Index	The price of basket of good presented in percent form	>0
Population (in thousand)	Number of people living in the country	>0
Unemployment_Rate	Percentage of people that does not have a job but actively looking for a job within labor force	>0
Mortgage_Rate	Interest rate when people borrow loans	>0
Real_Disposable_Income	Actual purchasing power of income	>0

The author retains the house price index to represent the changes in house price during these years and some other variables that affect the house price index so that it lowers the complexity of the analyzing and having a more comprehensive analyzation, comparing only the changes in house price of each variable and the relations of each variable. With this, finding the variables that have the strongest relationship with the house prices can also be analyzed.

2.3. Method Introduction

In this paper, simple linear regression is being used as the main tool in predicting the future house price of Aberdeenshire. The equation of a unary linear regression is being expressed like this:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Inside the equation, β_0 and β_1 is an unknown constant, x is the time, y is the house price and ε is a random factor that influences the housing prices. Next step, to find out the β_0 and β_1 , Least square method is used to obtain the values of unknown constants:

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x} \quad (2)$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (3)$$

Where inside the equations, $\bar{x}$ and $\bar{y}$ means the average of the observed value x and y in the data set respectively, and $\widehat{\beta}_0$ and $\widehat{\beta}_1$ is the estimated value of the linear regression line predicted out with the data set. Furthermore, to make the calculation simpler, the author also uses SPSSAU as a tool to help analyze each different factor.

3. Results and Discussion

3.1. Correlation Analysis

First, a Pearson correlation is calculated to first see the significance of each variables comparing house price index, so the author can further abstract out the variables that does not have a strong correlation with house price and use the variables that have a significant effect on house prices to predict the future house prices.

Table 2. Pearson correlation

Variables		House_Price_Index
Consumer_Price_Index	correlation coefficient	0.969**
	p value	0.000
Population(in thousands)	correlation coefficient	-0.616**
	p value	0.000
Unemployment_Rate	correlation coefficient	-0.340*
	p value	0.020
Mortgage_Rate	correlation coefficient	-0.835**
	p value	0.000
Real_Disposable_Income	correlation coefficient	0.987**
	p value	0.000
* p<0.05 ** p<0.01		

From Table 2, all the p value are smaller than 0.05 and only the p value for unemployment rate exceeds 0.01. Thus, the author can conclude that all the variables listed above have an important significance in compare with house prices. Furthermore, consumer price index and real disposable income have strong positive correlations with house price index, while population, unemployment rate and mortgage rate have a negative correlation with house price index.

Next step, despite the relationship between house price index and other influencing factors, the correlation between each influencing factor can also be investigated. From Table 3, the correlation between unemployment rate and population is the only two factors that doesn't have strong significance between these two factors, will the rest of the factors all have an important significance which p value is smaller than 0.05. In addition, Consumer price index and real disposable income has a strong positive correlation while having a relatively weaker negative correlation with population,

unemployment and mortgage rate. By comparing the data for population and other factors, population has a weak positive correlation with unemployment rate and mortgage rate but a weak negative correlation with real disposable income. Furthermore, Unemployment rate has a weak positive correlation with mortgage rate but a weak negative correlation with real disposable income. Lastly, mortgage rate has a strong negative correlation with real disposable income.

Table 3. Correlation between influencing factors

	Consumer price index	Population	Unemployment rate	Mortgage Rate	Real disposable income
Consumer_Price_Index	1				
Population (in thousand)	-0.551**	1			
Unemployment_Rate	-0.307*	0.068	1		
Mortgage_Rate	-0.853*	0.423**	0.316*	1	
Real_Disposable_Income	0.989**	-0.597**	-0.335*	-0.874**	1

* $p < 0.05$ ** $p < 0.01$

After investigating the correlations between each of the factor, scatter diagram is used to visualize the relationship between the house price index and influencing factors more clearly.

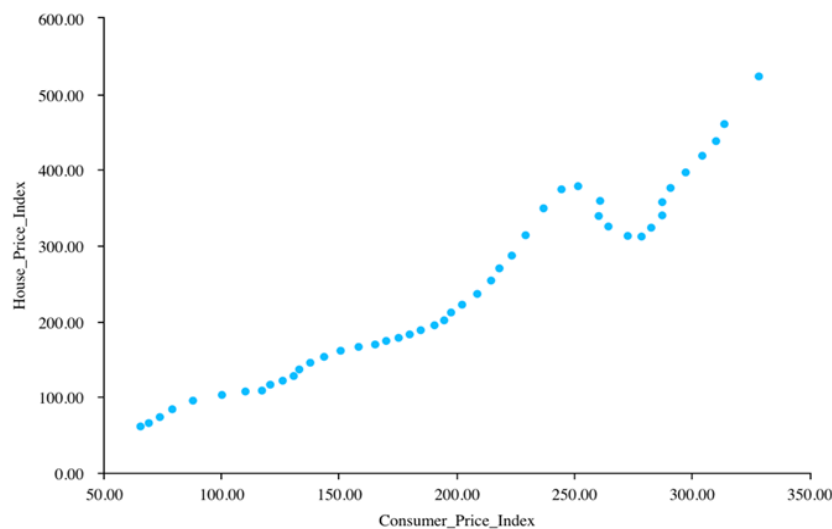


Fig. 1 Relation between house price index and consumer price index

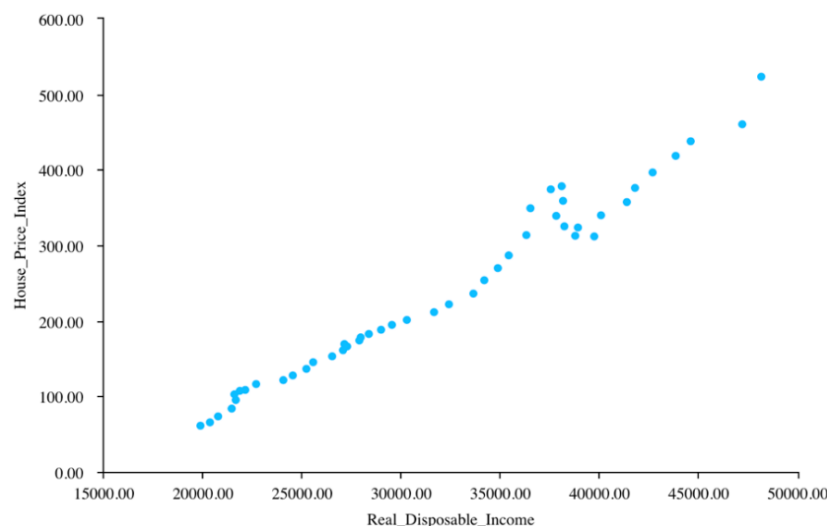


Fig. 2 Relation between house price index and real disposable income

After plotting Figure 1 and Figure 2, the author observes that the trend of the real disposable income between house price index and the trend of the consumer price index is highly similar, meaning that these two influencing factors may be highly repeated, thus, leaving one variable during the linear regression are also viable.

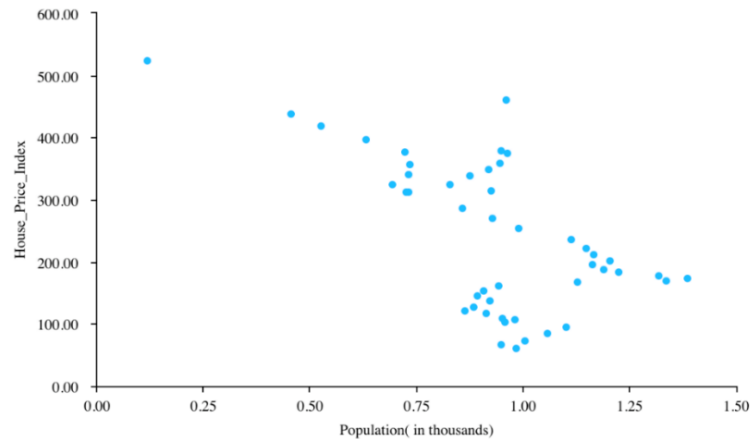


Fig. 3 Relation between house price index and population

From Figure 3, the house price index is showing a downward flow as the population increase. However, the data point after 0.75 is too spreading, making the relationship between population and house price index being ambiguous.

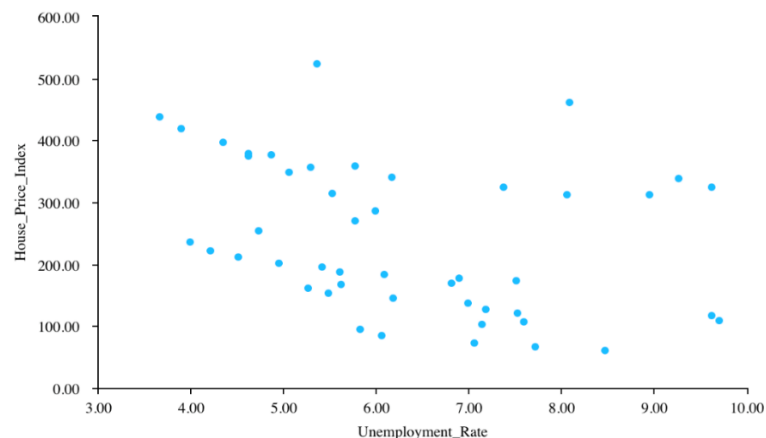


Fig. 4 Relation between house price index and unemployment rate

From Figure 4, the general trend is even harder to identify when comparing to Figure 3, making it hard to analyze the actual relationship between house price index and unemployment rate.

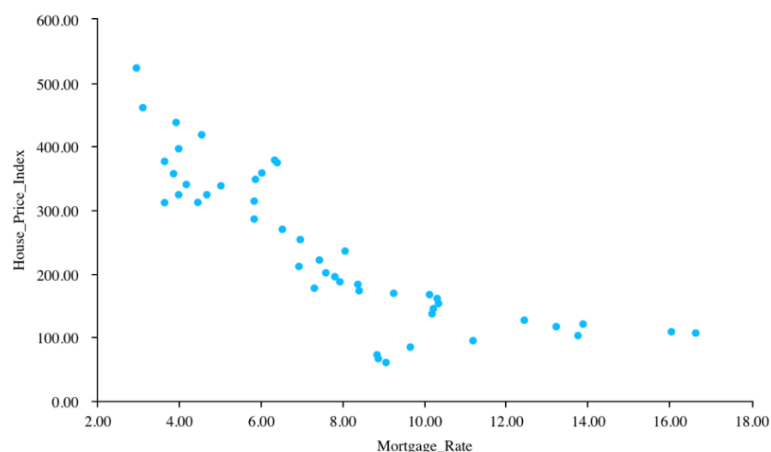


Fig. 5 Relation between house price index and mortgage rate

From Figure 5, it is more clearly that the house price index is decreasing when mortgage rate increase, showing a negative relationship between two of them. Generally, the data is all compacted together in each mortgage rate, proving that the correlation house price index and mortgage rate is very strong.

3.2. Model Results

After plotting all the diagrams, unemployment rate seems to be having the weakest correlation with house price index, thus, the author decides to remove this factor when doing the final linear regression.

Carrying on, to determine the most significance factor that effect the house price index, a multiple linear regression is calculated to further compare the relationship between them and the values is recorded as follow:

From table 4, the significance level of these 4 factors is gradually decreasing starting from real disposable income, mortgage rate, consumer price index and population. By combining the result from Pearson correlation, the reliability of this sequence of significance level can be confirmed correct.

Table 4. Linear regression analysis

	Regression coefficient	Standard error	p value	Multicollinearity tolerance
Constant	-382.390	63.656	0.000**	-
Consumer_Price_Index	-0.561	0.262	0.038*	0.018
Population (in thousand)	-0.091	16.183	0.996	0.515
Mortgage_Rate	4.616	1.735	0.011*	0.207
Real_Disposable_Income	0.022	0.003	0.000**	0.014

* p<0.05 ** p<0.01

3.3. Check for Accuracy

To check for the validity of the method in predicting the future house price, a determinant coefficient is calculated to check the accuracy of the model. To find the determinant coefficient, the average $\bar{y}$ of all the observed value y_i is first calculated:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

Where n is the total number of observed values recorded in the data set. Next, the total sum of squares (TSS) is calculated by first squaring the difference between $\bar{y}$ and each observed value y_i , and then added all together, which can also be expressed as following equations:

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

In addition, residual sum of squares (RSS) is also needed, which is finding the sum of the square of the difference between estimated value $\hat{y}$ and each observed value y_i :

$$RSS = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (6)$$

After finding all the necessary information needed to calculated determinant coefficient R2, the last step is to divide RSS by TSS, and let one subtract the value retrieved from above step and get the final value R2 equals to 0.979. This is a very highly accurate result because the R2 value is very close to 1, proving that the degree of fitting is extremely high.

4. Conclusion

In conclusion, the author discovers the relationship between a few factors relating to the house price index and uses multiple linear regression to identify the most related factors influencing house prices, which is real disposable income. In addition, this result has a high accuracy because of having a high determinant coefficient. This study is important in the way that it can be helped to understand the most significant reason behind the changes in house prices, finding the main reason about some social problem that relates to real estates. House prices can reveal a lot of economic status by understanding the factors correlated to house prices, it can be easier for government to fix the problem of high house prices by directly fixing the root of the factors that cause the abnormal change in house prices. Furthermore, an analyze on significance of each factor can help people to predict a more accurate future house price, because the calculation can now weigh the most significance higher to differ the influence level of each factor during the prediction. Lastly, this result is limited due to the lack of other variables that effects the house price. To further improve this analysis, more factors can be added into this research so the analysis can be more concise and comprehensive, giving a better result in predicting future house prices.

References

- [1] Zhou Y. Research on factors influencing housing price through supply and demand mechanism. *Enterprise Technology Development*, 2009, 28(02): 59-60.
- [2] Jin B, et al. Analysis of spatial variation of street landscape greening and influencing factors: an example from Fuzhou city, China. *Sci Rep*, 2023, 13: 21767.
- [3] Zhang Lei, Liu Kexin, Huang J G. Land supply and housing prices: The case of Hong Kong. *Journal of China Textile University (English Edition)*, 1997, 14(02): 74-81.
- [4] Sun Binyi, Yu X N. A study on the mixed effects of interest rate and purchase restriction policies on housing price changes. *Journal of east China normal university (philosophy and social sciences edition)*, 2018, 50(6): 92-99 + 175.
- [5] Yuan Jing, Liu Lei, Liu Jieqiong. Welfare cost effect of housing price fluctuation and optimal housing price policy in China. *Journal of shandong college of industry and commerce*, 2017, 31(6): 42-51.
- [6] Xu Zhong, Zhang Xuechun, Zou Chuanwei. The Interactions among Housing Price, Inflation and Monetary Policy in China: Facts and Implications. *China Economist*, 2013, 8(01): 112-123.
- [7] Wenwen Zhou, Mengyao Chen, Yang Gao, Ruilin Feng. Research on the Impact of Market Concern for Real Estate Policy on Housing Prices: Evidence from Internet Search and Hedonic Price Theory. *Computer Modeling in Engineering & Sciences*, 2022, (06): 1635-1652.
- [8] Jin Bei. The Right Policy Target for China's Real Estate Regulation. *China Economist*, 2014, 9(02): 1.
- [9] Brzezicka J, et al. The missing asymmetry in the Polish house price cycle: an analysis of the behaviour of house prices in 17 major cities. *J Hous and the Built Environ*, 2022, 37, 1029–1056.
- [10] Li Yumao, He Tao, Liu D. Theory and application of unitary linear regression method. *Journal of chifeng institute (natural science edition)*, 2017, 33(15): 1-2.