Factors Affect the House Price

Minhong Shao

Qingdao, China

Email: 804156271@qq.com

November, 2021

#### Abstract

The object of this paper is to find whether house age and distance between house and nearest metro station could affect house price of unit area. A sample of 414 houses in New Taipei City, Taiwan is selected. By using multiple linear regression, there is a linear relationship between house price and house age, and distance. By creating scatter plots of dependent variables versus independent variables, the graphs show that the linear regression model might not be the best model to explain the relationships of house price of unit area and the two explanatory variables.

#### Introduction

Our lives can’t be separated by food, clothes, transportation, and house, all of them are our daily necessities. All of the are our daily necessities, but only food and cloths have relative cheap and stable price. Although the price of the car is very high, it costs very little if taking public transports like bus and metro. However, house is also a necessity, but its price increases year by year. House and apartment are also commodities, so they also follow supply and demand curve. As more and more people would like to buy a house, the price will increase. If fewer people buy houses, house price will fall. The house bubble in recent years in the United States is related to the demand and supply curve. In 2008, United States had encountered the biggest financial crisis, which was caused by house bubble. Supply and demand curve shows that the more people buy a house, the higher the house price will be. With the price peaking, not everyone who wants to buy a house can afford it, so loans had become the popular choice of buyers. However, at that time, loan institutions did not consider the repayment ability of buyers, so they lent them money, but buyers did not have the ability to repay, which led to the subprime mortgage crisis. Finally, the loan institutions didn’t have enough money to provide buyers to purchase a house, so the number of buyers decreased, which led the house price decreased either. That has eased the subprime mortgage crisis.

House prices has been a hot issue for a long time. House price greatly impacts our daily lives and our decisions. More and more people would like to rent a house or short-term apartment like Airbnb instead of purchasing one because of the price. In Beijing, we can rent a house for about 200 years with the money we buy it, it is obvious that we could not live that longer. House price is one of the reasons that people would like to rent a house instead of buying one. Also, with the continuous rise of house prices, it is difficult for people with low wages to purchase a house with full amount, and they have to borrow money from loan institutions. If people owe moneys, then their happiness will be greatly affected. In fact, the rising house prices will also lead to the fertility rate decline. Young people have no time and money to raise children, because they need to spend most of their money on housing and car loans. Although the house price now is higher than before, more and more people want to buy a house because they think owning a house will bring them a sense of security. Therefore, I think it’s necessary for us to discuss the factors that could affect house price.

Many people have already discussed the factors affecting house prices. In Dennis R. Capozza, Patric H. Hendershott, Charlotte Mack, and Christopher J. Mayer’s paper [1], they use data set of 62 metro areas from 1979 to 1995 in United States, and they find that high house price can occur in high real construction cost areas. Boston, New York, San Francisco are such places. Stijn Van Nieuwerburgh and Pierre-Olivier Weill use dynamic equilibrium model of the housing market to show that the house price in a certain area is correlated with the per capita income in that area [2]. Owen Lamont and Jeremy C. Stein use city level data to show that price is correlated with loan-to-value ratios [3], they find areas that have higher loan-to-value ratio have relative higher house price. However, only when people have higher salary, loan institution will lend them higher loans than people who have less income. Hence, we could use Owen Lamont and Jeremy C. Stein’s result to derive Stijn Van Nieuwerburgh and Pierre-Olivier’s result. Melissa A. Boyle and Katherine A. Kiel find that the environment goods like, air quality, water quality, and distance from toxic or potentially toxic sites could also affect the house price [4]. Gang-Zhi Fan, Seow Eng Ong and Hian Chye Koh use decision tree approach to figure out that in Singapore, some characteristics of the house like floor area, model type, floor level, and recreational facilities could also affect the house price [5]. Bradford Case, Henry O. Pollakowski, and Susan M. Wachter use a different perspective to analyze the factors that affecting housing price [6]. In their article, they use four America counties’ house data to find that the more frequently the house is traded, the higher estimated price of the house. In Bernardina Algieri’s article, she found that not only income, long run interest rates, stock prices and inflation also can affect house price, but latent component could also play a significant role in explaining real house prices [7]. The latent component reflects factors that are not observed specifically, like structure changes and changing preferences. Atharva Chouthai, Mohammed Athar Rangila, Sanved Amate, Prayag Adhikari, and Vijay Kukre use machine learning to figure out that material used for construction, number of bedrooms, and number of garages are also related with house price [8]. Chinese house prices are much more complex than those of any other countries. Compared with other countries, there are more factors affecting Chinese house price. Policy is one of the important factors affecting Chinese house price. Huayi Yu uses Chinese house price to figure out that the real estate policy and land policy of China could also lead house price difference [9]. Last, Martijn Droes and Alex van de Minne use almost 200 years of historical data on house price from Amsterdam to figure out that house price changes over time [10], they argue that house price depends on economic state of affairs in each different era.

#### Data Source

The data set I used to analyze factors that affecting house price is Real\_estate.csv, the full name is Real estate valuation data set. This data set is created by Prof. I-Cheng Yeh from Tamkang University, Taiwan in 18th August, 2018. Data set can be retrieved from [Real estate price prediction | Kaggle](https://www.kaggle.com/quantbruce/real-estate-price-prediction). The data set could be derived from Kaggle. Kaggle is an online community of data scientists and machine learning practitioner. It allows users to find and publish data sets, explore and build models in a web-based data-science environment, and enter competitions to solve data science challenges. The market historical data set of real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan. This data set includes transaction date, house age, distance to the nearest MRT station, number of convenience stores, latitude, longitude and price of 414 instances for New Taipei City of Taiwan.

#### Unit

I use year to measure the age of houses; use meters to measure the distance between houses and the nearest metro station；use integer to measure the number of convenience stores in the living circle on foot; use degrees to measure latitude and longitude of houses in the geographic coordinate. House price of unit area is our dependent variable, and I use 10000 New Taiwan Dollar per Ping to measure it, where Ping is a local unit, 1 Ping is about 3.3 meter squared.

#### Variables and Description

Table 1.1 Description of Independent and Dependent

Variables Description Unit Min and Max Value

House price House price of unit area 10000 New Taiwan dollars/Ping 7.6 and 117.5

House age House age Year 0 and 43.8

Distance Distance between houses and nearest metro station Meter 23.38 and 6488.02

I choose house age and distance as my independent variables, and price of houses as my dependent variable. The house price shows 1 Ping (about 3.3 squared meter) could cost how many New Taiwan dollars in 10000 dollars. The house age represents the year from it was constructed to 2018.

#### Model and Analysis

In this section, I will use Real\_estate.csv data set to discuss how house age and distance between house and nearest metro station could affect the house price in Taiwan. I will use multiple linear regression for analyzing. Multiple linear regression, which is also known as multiple regression or just MLR, is a technique used in statistics. The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables. Multiple linear regression is used extensively in econometrics and financial inference.

The general formula of multiple linear regression is

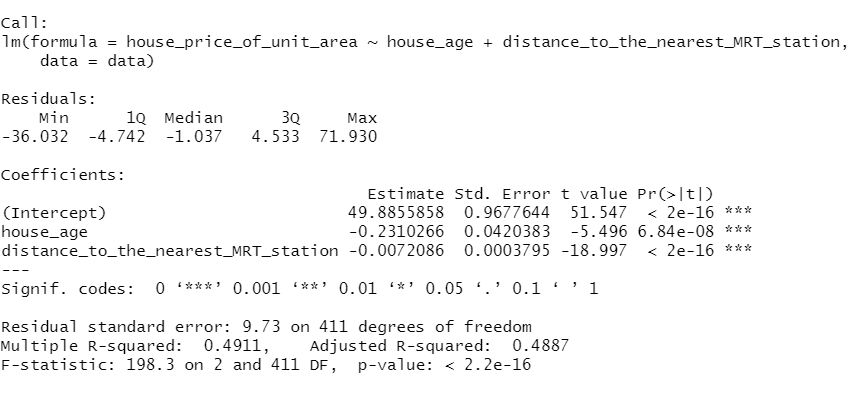
Y = β0 + β1 X1 + β2 X2.

X1 and X2 are our independent variables, Y is our dependent variable. β0 is the intercept, it represents the value of dependent variable when all of the independent variables are equal to zero.

β1 is the coefficient respects to independent variable X1 , the dependent variable Y will change β1 if X1 increases one unit when setting other variables fixed. β2 is also the coefficient but respects to independent variable X2 , the dependent variable Y will change β2 if X2 increases one unit when setting other variables fixed. In our research, Y is house price in New Taiwan Dollars per unit Ping, X1 is the age of the house, and X2 is the distance between the house and its nearest metro station.

#### Result

Figure 1.1 Multiple Linear Regression between House Price and House Age, Distance between House and its Nearest Metro Stations.



I use RStudio (R language) to create the summary of the model. As figure 1.1 shows, a multiple linear regression was calculated to predict house price based on house age and the distance between house and nearest MRT station. A significant regression equation was found (F (2,411) = 198.3, p<.0001), with an R2 of 0.4911. Participants’ predicted house price is equal to 49.89 - 0.231\*house age - 0.00721\*distance, where house age is measured as how long does the house exist in years, and distance is measured as the distance between house and nearest metro station in meter. Object of measurement decrease 0.231 times ten thousand New Taiwan dollars for house age increases one year, and decrease -0.00721 times ten thousand New Taiwan dollars for the distance increases one meter. Both house age and distance were significant predictors of house price. The coefficient of intercept is 49.89, it means that when the house age is zero, and the distance between house and metro stations is zero meters, the price of the house is 49.89 times ten thousand New Taiwan dollars. However, there is no house that zero-meter closes to the metro stations, so the intercept does not have actual meaning.

For house age, the null hypothesis is that there is no linear relationship between house age and house price, and the alternative hypothesis is that there is a linear relationship between house age and house price. Since the p-value of coefficient of house age is 6.48 x 10-8, which is much smaller than 0.0001, we have statistically significant evidence to argue against the null hypothesis. Hence, we can say that there is a linear relationship between house age and house price. For distance to the nearest MRT station, the null hypothesis is that there is no linear relationship between distance and house price, and the alternative hypothesis is that there is a linear relationship between house age and house price. Since the p-value of coefficient of the distance between house and nearest metro station is much less than 0.0001, we have statistically significant evidence to argue against the null hypothesis. Hence, we can say that there is a linear relationship between distance of house and the nearest metro station and house price.

#### Discussion

From the model, it is not hard to see that if people would like to purchase a low-price house, then they need to choose an old house or houses far away metro stations. However, although the house price per unit may be cheap, it is very inconvenient for people living in it. Since the old houses may have many small problems, such as aging heating pipes, and the design itself may also have defects. If people have private cars, living away from the metro station may not affect their daily lives, but for people without a private car, living away from the metro station means that they will spend more time on going to the metro station than people who live near the station. Also, the model has some limitations. First, I am not confident that the model will work in all areas not just New Taipei City. For houses that only have one or two metro lines or no metro lines nearby, it is possible that there is no linear relationship between house price and the distance between house and nearest metro station. Another example is that coastal areas in eastern China were once colonized by Germany, and during colonial period, many German houses were constructed. Since they have unique architecture style, and very rare, their house prices are still high even though they have been built for more than a century.

#### Conclusion

This paper is based on the house price in New Taipei City in 2018, and I am going to find whether there is a linear relationship between house age and house price per Ping and whether there is a linear relationship between distance (distance between houses and nearest metro station) and house price per Ping. I find out that if house age and distance between houses and nearest metro station increases, then the house price per Ping in New Taipei City will decrease. There is a linear relationship between house age and house price per ping, and linear relationship between distance and house price per ping.

Although multiple linear regression model has already proved that there is a linear relationship between house price per ping and distance, and house price per ping and house age, may there are other models could better explain the relationship between house price, distance and house age than multiple linear regression model. We first need to see the distribution between independent variables and dependent variables.

Figure 2.1 House Prices of Unit Area vs House Age

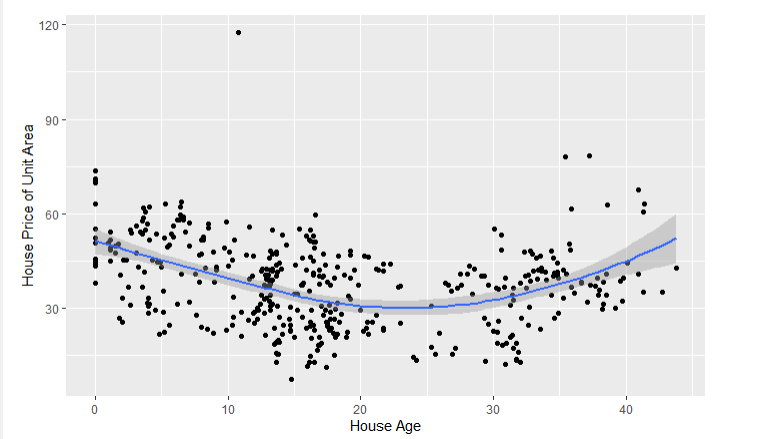
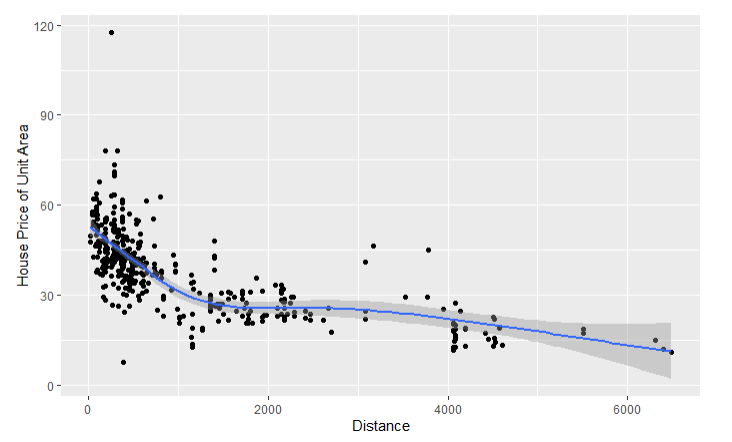


Figure 2.2 House Prices of Unit Area vs Distance between Houses and nearest MRT station.



Both Figure 2.1 and Figure 2.2 are created by RStudio. Figure 2.1 is a scatter plot of house price per unit area over age of house with a smooth line. By observing the scatter plot, I find that the smooth line is like a parabola, so there might be a quadratic relationship between price and age. When the house age is between 0 and 22, the house price will decrease as house age increases. However, if age is greater than 22, the house price of unit area will increase as house age increases. Therefore, quadratic model might be better to explain the relationship of house price per unit area and house age than linear regression.

Figure 2.2 is a scatter plot of house price per unit area over distance between house and nearest metro station. By observing the smooth line, I find that the house price will decrease as distance between house and metro station increases, if distance is less than 1500 meters. However, when the distance is between 1500 meters and 3000 meters, the house price of unit area almost remains constant no matter distance increases or decreases. If distance is greater than 3000 meters, the house price of unit area will decrease as distance increases. Hence, linear regression model is not the best model to explain the relationship between house price of unit area and distance.

Therefore, by analyzing Figure 2.1 and Figure 2.2, I find out that multiple linear regression model may not be the best model to explain the relationship of house price of unit area and house age, and house price of unit area and distance between house and nearest metro station.

#### Reference

[1] Dennis R. Capozza, Patric H. Hendershott, Charlotte Mack, and Christopher J. Mayer, DETERMINANTS OF REAL HOUSE PRICE DYNAMICS, 2002, which is available at

<https://www.nber.org/system/files/working_papers/w9262/w9262.pdf>

[2] Stijn Van Nieuwerburgh, and Pierre-Olivier, 2010, Why Has House Price Dispersion Gone Up? which is available at

[Why Has House Price Dispersion Gone Up? | The Review of Economic Studies | Oxford Academic (oup.com)](https://academic.oup.com/restud/article/77/4/1567/1645685?login=true)

# [3] Owen Lamont and Jeremy C. Stein, 2000, Leverage and House-Price Dynamics in U.S. Cities, which is available at

# [aeg6195.tmp (nber.org)](https://www.nber.org/system/files/working_papers/w5961/w5961.pdf)

[4] Melissa A. Boyle and Katherine A. Kiel, A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities (2001), which is available at

[A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities on JSTOR](https://www.jstor.org/stable/44103421)

[5] Gang-Zhi Fan, Seow Eng Ong and Hian Chye Koh, Determinants of House Price: A Decision Tree Approach, 2006, which is available at

[Determinants of House Price: A Decision Tree Approach - Gang-Zhi Fan, Seow Eng Ong, Hian Chye Koh, 2006 (sagepub.com)](https://journals.sagepub.com/doi/abs/10.1080/00420980600990928)

[6] Bradford Case, Henry O. Pollakowski, and Susan M. Wachter, Frequency of Transaction and House Price Modeling, 1997, which is available at

[Frequency of Transaction and House Price Modeling | SpringerLink](https://link.springer.com/article/10.1023/A:1007736521741)

[7] Bernardina Algieri, House Price Determinants: Fundamentals and Underlying Factors, 2013, which is available at

[House Price Determinants: Fundamentals and Underlying Factors | SpringerLink](https://link.springer.com/article/10.1057/ces.2013.3)

[8] Atharva Chouthai, Mohammed Athar Rangila, Sanved Amate, Prayag Adhikari, and Vijay Kukre, House Price Prediction Using Machine Learning, 2019, which is available at

[IRJET-V6I3115120190910-55587-hlw9rw-with-cover-page-v2.pdf (d1wqtxts1xzle7.cloudfront.net)](https://d1wqtxts1xzle7.cloudfront.net/60555584/IRJET-V6I3115120190910-55587-hlw9rw-with-cover-page-v2.pdf?Expires=1635401500&Signature=EjNZpofZ3mmsjibQSxK6zHpmelDFwt96lr-E6ImmFovjVdVuSn1tVUrVkTeiiRxoI9kdYRP0Jj3KV~oY4HfwOdUakL75pLrheonnbvCh54MovBe-EybhK6KyDHMrrj1Z~wBPVrHaPIhzSIZD3IrwGklFpsp45GVWX4hJG46A6GyddUCJ9WVYod~KltbUfDbymNron7tbbt-PR0v5orQhU1GJ7YmEXoy3fC3JehoKOX3ty9xClgnOnz01cjZ5IQxfMWa4IkoV6~DepgMQRaS7aRcuu1P~ZovB~8142H2Jt~LuNk1BSJ7rKCzfMnLJ~SG4rpxzMXK9q2g6cH9qdCQjRw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)

[9] Huayi Yu, China’s House Price: Affected by Economic Fundamentals or Real Estate Policy? 2010, which is available at

<https://brill.com/view/journals/fecc/5/1/article-p25_2.xml>

[10] Martijn Droes and Alex van de Minne, Do the Determinants of House Prices Change over Time? Evidence from 200 Year of Transactions Data, 2016, which is available at

[Do the Determinants of House Prices Change over Time? Evidence from 200 Years of Transactions Data (repec.org)](https://ideas.repec.org/p/arz/wpaper/eres2016_227.html)