An Analysis of Factors Affecting the Housing Market of San Jose Using MLR Model

Bo Yang[1,a,*]

[1]*Mathematics Department, University of California - Santa Cruz (95064), Santa Cruz*
a. byang48@ucsc.edu
*corresponding author

Abstract: This article seeks to explain the unstable curve of housing prices in San Jose by analyzing the data of factors that may affect the housing market. By previous research conducted by other scholars, we believe there were five major factors which cause the housing prices to fluctuate, which are construction costs, total population, median income, mortgage rates and crime rates. Introducing multiple linear regression (MLR) as an approach to analyze and based on the monthly data of five-years period, from 2015 to 2019, we find an increase of mortgage rates and construction costs can lead to an increase in housing prices and other factors may not affect the housing prices as expected. However, due to lacking monthly data of certain factors like total population, we use other information associated with the factors to find an approximation of the real monthly data. Hence, while the conclusion provides certain explanations of housing prices for the buyers, it is built on many assumptions and may not perfectly fit the real behaviour of the housing market of San Jose.

Keywords: multiple linear regression, housing market, San Jose

1.   Introduction

In the 21st century, as modernization and urbanization thrive all over the world, homes as a special kind of good in metropolises become valuable due to high inflow of population. Such long lasting unmatching between consumers and goods resulted in a gradual increase in housing prices. Especially in recent years, the growing trend of housing prices in metropolises is incredible. It's being said that buying a house is the best investment. There's little risk of losing values and a great chance of winning multiple times the inputs. However, since the 2008 subprime mortgage crisis, the housing market has become one of the most unstable markets. As the input of housing investing is numerous, individuals may be unable to take the consequences of any investment failure. In addition to the worries of fortune seekers, the trend of future housing prices also becomes a concern for a large group of people who seek more affordable housing, since purchasing a home is going to cost them years of earnings. Therefore, an understanding of housing prices is needed. San Jose is the greatest city in northern California, which is also the location of Silicon Valley. The soaring local technology companies had attracted thousands if not millions of people to move in which subsequently brought up the housing prices which is many's concern. This paper will focus on the housing market of San Jose and analyze its behaviour by recapitulating the factors affecting the housing market, collecting the associated data of San Jose, and applying Multiple Linear Regression

(MLR) to build a statistical model of housing prices based on evaluated factors. Within this research, we may help the buyers gain more knowledge of the behaviour of the local housing market from multiple factors to make a relatively better decision.

San Jose is the largest city of northern California with a population of nearly a million and an area of 179.97 square miles, which indicates the local housing market is huge. Considering houses are high-value, durable and immobile consumption goods, the major affecting factors can be generalized into four categories: supply, demand, the economic environment, and the characteristics of the goods.

The supply includes the factors that are directly associated with the production of homes. As the cost of the product, which is the homes, rises, the selling prices will rise accordingly. According to Alessio and Galo's research on the relationship between construction prices and housing prices, in the United States, from 1987 to 2007, construction prices and housing prices both had an increase of around 70 percent. Moreover, in the United States, the trends of these two prices have a strong connection, as the construction prices kept tracking the housing prices from 1987 to 2001 [1]. Therefore, construction prices are one affecting factor of housing prices in the United States.

The demand is associated with buyers and potential buyers. The population, ages, races and relationship status act as deciding roles in the need of housing. Males, elders, asians and married people exhibit a greater housing demand. Also, a larger population will have the same effect on the demand for housing [2]. In addition to the size of population, the general income also positively affects local demand for housing [3, 4]. For most people, the majority of their income is their salary. A rise in a regional average salary and household income indicates an increase in housing purchasing power, which consequently leads to a rise in housing demand and then in housing prices [2]. Research further shows that education levels are positively associated with the amount of income, and a larger proportion of educated people in the total population leads to higher local housing prices [5-7]. Considering growing wealth disparity and for the purpose of simplicity, this paper will choose the median number of income and the size of population as affecting factors.

The economical environment also has an effect on the potential buyers. The implementation of monetary policies results in the shift in interest rates. And interest rates are another affecting factor [3]. Under a high interest rate, the cost of loans increases, which affects the housing demand as most buyers loan for houses, and people are more likely to deposit their money in banks than investing [2]. Research also shows an increased mortgage rate indeed leads to a drop in housing prices [8]. Hence, we may conclude that interest rates are also one of the factors affecting housing prices.

Since homes are immobile goods, their accessory characteristics weigh a lot in the buyers' consideration. For example, the location of a home decides buyers' future transportation convenience. Such accessory characteristics will be reflected on housing prices. Local crime rate is one of the characteristics. A high crime rate in a community causes more move-outs and less move-ins. As a result, the local housing prices drop as the demand drops [9, 10]. As mentioned before, transportation convenience is also a factor affecting housing prices. Time spent on commuting and traveling to central business districts (CBD) is negatively correlated to housing prices [6, 7, 11-13]. Since San Jose is located in the bay area which is the area with one of the highest housing prices among the world, most people who live here tend to work here, or they would move out for lower housing costs. The CBD of San Jose is its downtown which is at the center of San Jose, then commuting time would be similar for homes in different locations. So this paper will only focus on crime rates as a factor affecting the housing prices in San Jose.

From the previous research, this paper selects the construction prices, the size of population, the regional median income, the interest rates, and the crime rates as major factors affecting the housing prices of San Jose. When the construction prices, the size of population, and the regional median

income increase, the housing prices increase. When the interest rates and the crime rates increase, the housing prices drop.

2.   Methodology

This paper uses Multiple Linear Regression (MLR) to analyze data of the housing prices and major factors affecting the housing prices of San Jose. The statistical model contains five explanatory variables where each variable represents a major affecting factor and one dependent variable which is the median housing price. The formula is as follows.

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 \tag{1}$$

where Y represents Housing Prices, X_1 represents Construction Prices, X_2 represents Total Population, X_3 represents Median Income, X_4 represents Average 30-Year Mortgage Rates, X_5 represents Crime Rates and a_i represents Constant regression coefficients.

We will then introduce data of a five year period, including housing prices provided by Zillow, construction prices by Terner Center, California Construction Cost Index (CCCI) by California Government, population and median income by United States Census Bureau, mortgage rates by Freddie Mac, and crime rates by San Jose Police Department (SJPD). As it takes a great amount of samples to output correlations between the dependent variable which is housing prices and explanatory variables which are factors affecting the housing prices, we use monthly data to compute as early annual data will be hard to trace. Due to data availability, collected construction prices were of LIHTC housing which was carried out by the Tax Reform Act of 1986 for affordable housing. In the found data, there will be three factors that only have annual data, which are exact construction costs, total population, and median income. For factors that only have annual data, we deal with them in two approaches, since they could have collinearity between each other if they do not change over 12 months. The first approach is to assume they had a stable trend during the years, so they do not change when presenting data month by month. For example, Total Population will stay 1025980 in every month of 2015. The second approach is to reasonably select data for each month to prevent collinearity, so that monthly data fluctuate while they can be converted into the annual data, and vice versa. For the exact construction costs, we found a substitute with monthly data that indicates the construction costs, which is CCCI. Monthly data are used to compute the MLR model but will not be shown in this paper to shorten the length. Monthly housing prices and income are both median numbers, as average can be pushed higher or lower by an extreme case.

Table 1: Annual data of each factor.

YEAR	HP[a](USD)	CP[b](USD)	TP[c]	MI[d](USD)	MR[e](%)	CR[f]
2015	775,000	420,000	1,025,980	32,400	3.85	334.5
2016	850,000	490,000	1,030,242	34,100	3.65	373
2017	920,000	460,000	1,032,335	36,100	3.99	400.4
2018	1,100,000	480,000	1,028,020	38,400	4.54	422.7
2019	1,020,000	485,000	1,021,795	41,000	3.94	437.3

[a] Housing Prices
[b] Construction Prices (Construction Costs)
[c] Total Population
[d] Median Income
[e] Mortgage Rates
[f] Crime Rates

3. Results and Discussion

3.1. First Approach

Using constant construction prices, total population and crime rates for each month from 2015 to 2019, we have the following model.

$$Y = 92626245.504 - 2.943 * X_1 - 86.981 * X_2 - 4930.659 * X_3 + 57214.230 * X_4 + 420602.637 * X_5 \tag{2}$$

$$Y = -0.636 * X_1 - 2.665 * X_2 - 10.585 * X_3 + 0.174 * X_4 + 10.848 * X_5 \tag{3}$$

Table 2: Details of the model 2 and 3.

	b^a	σ^b	Standardized b^c	p^d	VIF^e
Constant	92626245.5	14890358.87	-	0	-
CP	-2.943	0.608	-0.636	0	19.538
TP	-86.981	13.873	-2.665	0	204.184
MI	-4930.659	816.89	-10.585	0	3475.01
MR	57214.23	17381.016	0.174	0.002	3.151
CR	420602.638	65749.771	10.848	0	3249.35
$R^{2\,f}$	0.952		p=0.000		

These variables are defined the same in other tables.
[a] Regression Coefficients
[b] Standard Deviation
[c] Regression Coefficients After Data Standardization (Used To Remove The Constant a_0)
[d] Indicates Whether The Correlation Between The Factors And The Housing Prices Is Significant
[e] Indicates The Existence of Collinearity In The Model
[f] Indicates How Matched The Predicted Housing Prices Is Compared To The Actual Housing Prices(%)

From table 2, we observe that the model can explain 95.2% of behavior of the housing prices and construction prices, total population and median income have a negative effect on the housing prices while mortgage rates and crime rates have a positive effect. However, there exists collinearity between Construction Prices, Median Income, Total Population and Crime Rates as their VIF values are greater than 10. So from this model, we will not be able to conclude any result. Found that Median Income has the greatest impact on other factors, we removed it to generate another model as follows.

$$Y = 2992857.331 + 0.457 * X_1 - 3.498 * X_2 + 133370.273 * X_4 + 23968.367 * X_5 \tag{4}$$

$$Y = 0.099 * X_1 - 0.107 * X_2 + 0.405 * X_4 + 0.618 * X_5 \tag{5}$$

Table 3: Details of model 4 and 5.

	b	σ	Standardized b	p	VIF
Constant	2992857.331	1403607.28	-	0.037	-
CP	0.457	0.294	0.099	0.125	2.773
TP	-3.498	1.384	-0.107	0.014	1.236
MR	133370.273	15328.681	0.405	0	1.49
CR	23968.367	2832.977	0.618	0	3.669
R^2	0.92		p=0.000		

As VIF values are less than 5, there doesn't exist collinearity between factors. However, the p-value of Construction Prices is greater than 0.05, so Construction Prices do not appear to be

significantly affecting Housing Prices. From the regression analysis, we may conclude that Mortgage Rates and Crime Rates have a positively correlated to Housing Prices, and Total Population has a negative impact on Housing Prices during 2015 to 2019 in San Jose.

3.2. Second Approach

We substitute Construction Prices with CCCI, and compute Crime Rates of each month using the number of Use of Force Incidents. As claimed before, the majority of the income is salary which means Median Income is a relatively more stable factor, and Total Population is also a more or less stable factor, so we leave these two factors unchanged during each month. The monthly Crime Rates are based on the number of Use of Force Incidents provided by SJPD per month. We assume the number of use of force incidents are proportional to the number of total violent crimes. Therefore, the more the use of force incidents happen in a month, we may expect the more violent crime happens that month. This approach may help prevent collinearity in the model, while losing a little accuracy.

$$Y = -2157848.175 + 163.08 * X_1 + 0.953 * X_2 + 138.774 * X_3 + 136709.862 * X_4 + 226.294 * X_5 \tag{6}$$

$$Y = 0.371 * X_1 + 0.029 * X_2 + 0.298 * X_3 + 0.415 * X_4 + 0.097 * X_5 \tag{7}$$

Table 4: Details of model 6 and 7.

	b	σ	Standardized b	p	VIF
Constant	-2157848.175	1666761.371	-	0.201	-
CP	163.08	64.939	0.371	0.015	15.165
TP	0.953	1.678	0.029	0.572	1.842
MI	138.774	76.106	0.298	0.074	18.598
MR	136709.862	14337.126	0.415	0	1.322
CR	226.294	129.783	0.097	0.087	2.165
R^2	0.922		p=0.000		

The collinearity still exists in the model, and it's between CCCI and Median Income as their VIF values are greater than 10. This reveals that Median Income could be correlated with CCCI which reflects construction costs. Unfortunately, it could also be caused by monthly data scarcity of Median Income. To build a more reliable model, again, we remove Median Income and generate a model without collinearity.

$$Y = -579344.608 + 271.007 * X_1 - 0.897 * X_2 + 145666.521 * X_4 + 257.711 * X_5 \tag{8}$$

$$Y = 0.616 * X_1 - 0.028 * X_2 + 0.443 * X_4 + 0.111 * X_5 \tag{9}$$

Table 5: Details of model 8 and 9.

	b	σ	Standardized b	p	VIF
Constant	-579344.608	370321.197	-	0	-
CCCI	271.007	27.275	0.616	0	2.567
TP	-0.897	1.364	-0.028	0.513	1.168
MR	145666.521	13751.164	0.443	0	1.167
CR	257.711	131.325	0.111	0.055	2.127
R^2	0.918		p=0.000		

The model shows 91.8% behavior of Housing Prices can be explained by the above four factors and Total Population and Crime Rates do not have significant effects on Housing Prices while CCCI and Mortgage Rates are both positively correlated to Housing Prices.

From both approaches, we see that the effects of Mortgage Rates on Housing Prices contradicts the previous research which says as interest rates grow, housing prices drop. Using the first approach, we also have Crime Rates and Total Population influencing Housing Prices. However, due to lack of their monthly data, the monthly data of these two factors remains constant during a year, which means the result may not be able to reflect the truth. Using the second approach, we find CCCI which reflects Construction Prices also has a positive effect on Housing Prices which comply with the research. So, from the provided information, we may conclude that Construction Prices and Mortgage Rates have a positive influence on Housing Prices.

4. Conclusion

Five factors affecting housing prices of San Jose are evaluated in this paper based on monthly data from 2015 to 2019, including construction prices, total population, median income, mortgage rates and crime rates. According to the MLR analysis, as construction costs and mortgage rates grow, housing prices grow. The correlations between other factors and housing prices cannot be concluded from the analysis.

Certainly, this research overlooks factors affecting the housing market other than the listed five which causes a loss of accuracy in the conclusion. In addition, the model is based on a large amount of assumptions on data which could further lead to a deviation from reality. Future research will collect a more detailed data set of each factor to generate a more precise model of the behavior of housing prices. And another approach other than MLR may be chosen for modeling, as it would not be simple to add many factors in MLR analysis.

References

[1] Moro, Alessio, and Galo Nuño 2012 Does total-factor productivity drive housing prices? A growth-accounting exercise for four countries. Economics Letters 115 2: 221-224.

[2] Wei-Shong Lin, et al. 2014 Effects of Socioeconomic Factors on Regional Housing Prices in the USA. International Journal of Housing Markets and Analysis, vol. 7, no. 1, pp. 30-41. ProQuest, https://www.proquest.com/scholarly-journals/effects-socioeconomic-factors-on-regional-housing/docview/1508425885/se-2, doi:https://doi.org/10.1108/IJHMA-11-2012-0056.

[3] Jud, G. Donald, and Daniel T. Winkler 2002 The dynamics of metropolitan housing prices." The journal of real estate research 23.1: 29-46.

[4] Potepan, Michael J 1996 Explaining intermetropolitan variation in housing prices, rents and land prices. Real Estate Economics 24.2: 219-245.

[5] Goodman, Allen C., and Masahiro Kawai 1982 Permanent income, hedonic prices, and demand for housing: New evidence. Journal of Urban Economics 12.2: 214-237.

[6] Haurin, Donald R., and David Brasington. 1996 School quality and real house prices: Inter-and intrametropolitan effects. Journal of Housing economics 5.4: 351-368.

[7] Goodman, Allen C 1977 A comparison of block group and census tract data in a hedonic housing price model." Land Economics 53.4: 483-487.

[8] Reichert, Alan K 1990 The impact of interest rates, income, and employment upon regional housing prices. The Journal of Real Estate Finance and Economics 3.4: 373-391.

[9] Rizzo, Mario J 1979 The effect of crime on residential rents and property values. The American Economist 23.1: 16-21.

[10] Katzman, Martin T 1980 The contribution of crime to urban decline. Urban Studies 17.3: 277-286.

[11] Katz, Lawrence, and Kenneth T. Rosen 1987 The interjurisdictional effects of growth controls on housing prices. The Journal of Law and Economics 30.1: 149-160.

[12] Shimizu, Chihiro, Kiyohiko G. Nishimura, and Tsutomu Watanabe 2010 Housing prices in Tokyo: A comparison of hedonic and repeat sales measures. Jahrbücher für Nationalökonomie und Statistik 230.6: 792-813.

[13] Ham, Charlotte, et al 2012 Accounting for heterogeneity of public lands in hedonic property models. Land Economics 88.3: 444-456.