

The Impact of Household Structure and Urban Agglomeration on Real Estate Prices in California

Hua Zhou*

College of Zhuoyue Honors, Hangzhou Dianzi University, Zhejiang, China

*Corresponding author: 19141335@hdu.edu.cn

Abstract. The purpose of this paper is to study the impact of regional economic and demographic factors on real estate prices and provide a theoretical basis for policy makers. The first section examines the economic traits of California in the 1990s and reveals a significant urbanization agglomeration impact. Through the use of linear regression, the data analysis model is further developed to examine the reasons for regional variances in home prices. The frequency of numerous families, small family sizes, and the concentration of big cities are proven to have a significant impact on housing costs.

Keywords: housing price; family structure; Urban agglomeration

1. Introduction

Real estate is an important part of the regional economy. Its development level depends on the local economic development, and in turn affects the local economic development. It promotes and restricts each other with the macro-economic development. At the same time, due to the immobility of real estate products and the regional characteristics of consumption, the real estate industry has strong regional differences. By studying the development of the local real estate industry, the public can have an insight into all aspects of the local economy.

The early literatures mainly studied the fluctuation of real estate prices through the theory of economic growth, while the agglomeration effect and regional problems of metropolis did not attract the attention of researchers. According to research by Downs (1993), the fundamental variations in market circumstances are what drive the variations in real estate price fluctuations [1]. Kuznets (1952) and Grebler et al (1956) believe that the long cycle of real estate is mainly affected by the urbanization process and the fluctuation of population birth rate [2,3]. Maisel (1960), Easterlin (1966), Jaffee & Rosen (1979), Smith (1984), explains the long real estate cycle in terms of demographic changes and household structure [4-7].

From the perspective of asset allocation, Mayo (1981) found that when the income crossed a certain threshold, individuals were more concerned with the quality of homes than the quantity, which caused the price of homes in desirable regions to rise more quickly than those in less desirable places [8,9]. Horioka (1988) believed that while analyzing the housing price, family income should be taken into account rather than individual income. He discovered in his article that the relative demand for non-housing assets increases with family size, and the effectiveness of raising the housing price should be diminished accordingly [10]. The results of Horioka (1988) are still debatable, nevertheless. It is more plausible that family size determines housing demand, with a smaller family size having a larger upward pressure on housing costs.

In addition, Clayto (1996) found through his research on real estate prices in Vancouver that the fluctuations of housing prices were mainly determined by the basic value of houses, and the influence of rational expectations on the fluctuations of housing prices could be ignored [11].

Recently, some scholars have also focused on the study of agglomeration effects and regional issues. Chen et al (2004) made a comparative study of four Asian cities, Hong Kong, Singapore, Taipei and Tokyo, and found that the high volatility of wealth in Hong Kong, Singapore and Taipei was closely related to the high local economic growth rate [12]. Liang & Gao (2007) studied the regional differences of housing price fluctuations in 28 provinces in China from 1990 to 2006. They found that economic growth had a greater impact on housing price fluctuations in the central region,

but had a smaller impact on the eastern and western regions. The scale of real estate credit has a large impact on the east and west, but a small impact on the central part [13].

In order to study the endogenous link between real estate prices and the degree of local social and economic development from a micro viewpoint using a linear regression model, and to address the time-lag impact between real estate prices and violent crimes.

2. Data Analysis

The shortage of land resources is getting more acute as urbanization progresses, and the impact of inflation makes it impossible to reverse the trend of growing real estate prices.

As one of the most developed states in the United States, California has complex economic activities and is a good sample to study the influencing factors of real estate prices. This paper selects the 1990 California census data as the research model.

In 1990, the population of California was relatively evenly distributed and mostly clustered in coastal areas. As shown in figure 1, three major urban clusters were formed with Los Angeles-San Diego, Sacramento, and San Francisco-San Jose as the centers, and the population density of giant city centers reached the peak, showing significant urbanization effect.

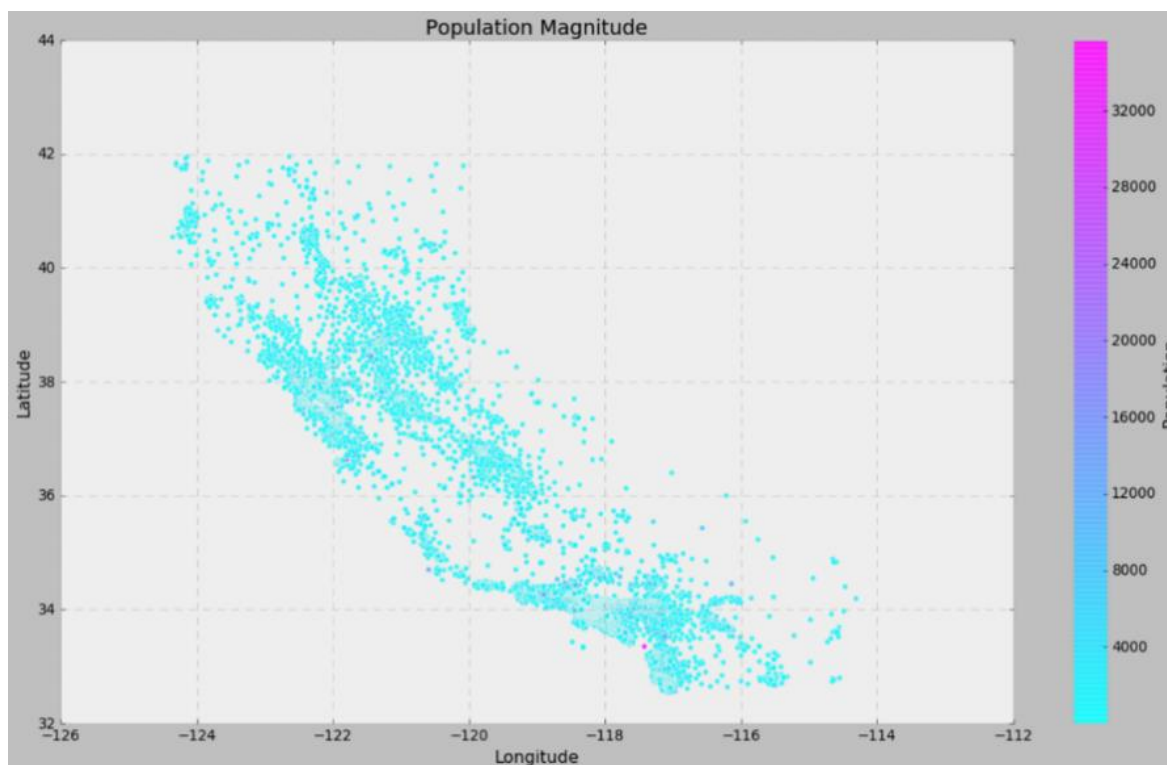


Fig 1. Two or more references

And the median income map, which also fits the characteristics described by the population map. However, in figure 2, the differences are even more striking, with SAN Francisco-San Jose and Los Angeles-San Diego having strong siphoning effects, but Sacramento's distance from the coast is much lower.

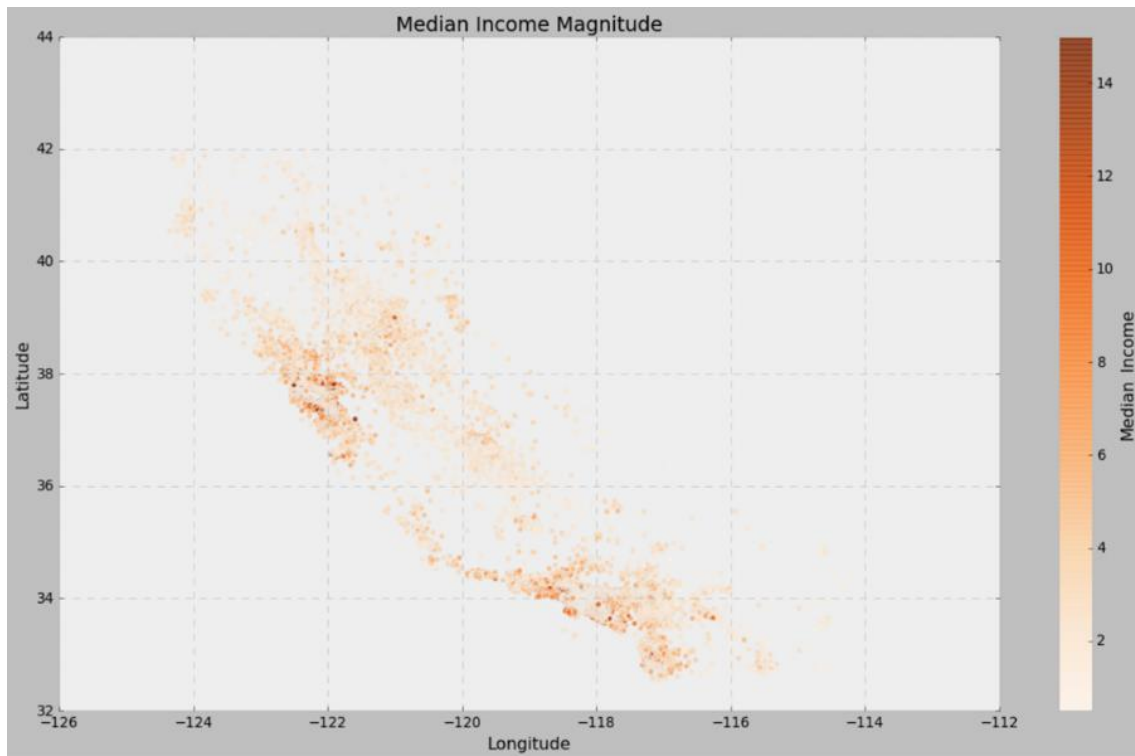


Fig 2. Median income Magnitude

Previous studies have shown that, in general, there is a positive correlation between per capita income and the average price and supply of local real estate. Based on this, we propose the hypothesis that housing supply will produce agglomeration effect in economically developed areas.

This paper takes LA, San Diego, San Francis and San Jose as major cities, calculates the distance of each house by latitude and longitude, and selects the minimum value as the variable 'Distance_to_nearest_major_city'. The results, shown in Figure 3, are as expected.

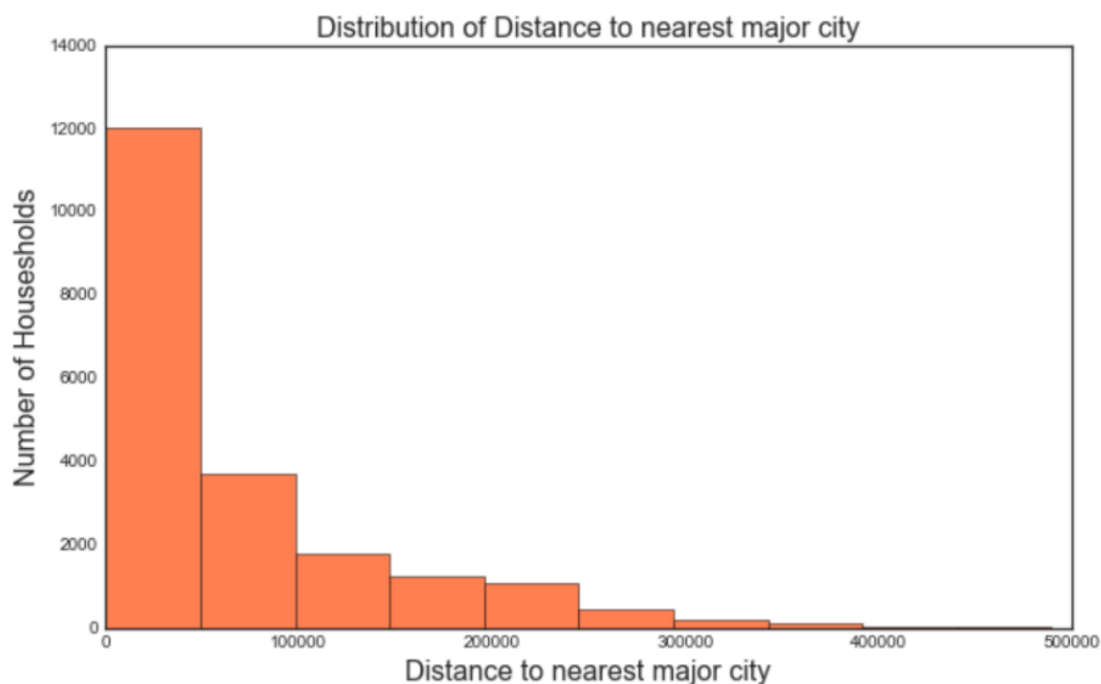


Fig 3. Distribution of Distance to nearest major city

Similarly, a new variable "distance to coast" is introduced, and the results are shown in the following table. The distribution of houses is shown in figure4, similar to Figure 3, so similar conclusions can be drawn.

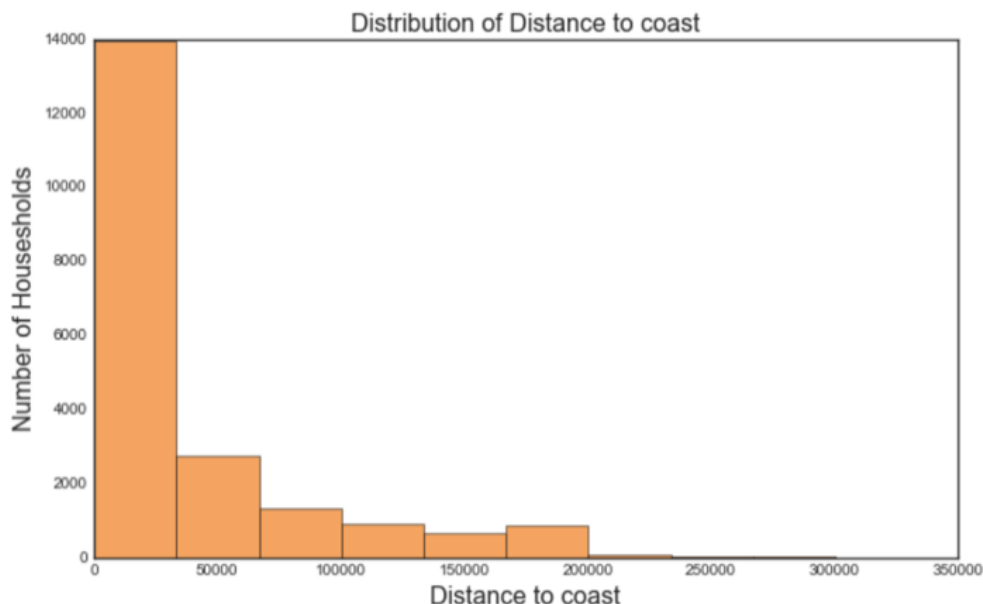


Fig 4. Distribution of Distance to coast

Figure 5 and Figure 6 show the median distribution of real estate price and median distribution of housing age respectively. Graphically, they have strong similarity, but also have differences. From the point of view of the median house price, within the range of 0-500K, the figure roughly presents a normal distribution, with the median being 206K. However, after the range exceeds 500K, the price rises abnormally, indicating that there are a large number of high net worth real estate in this region.

As for Figure 5, there are two peaks, respectively 16 years and 35 years, indicating that there are two peak periods of construction industry development in this area, corresponding to two large-scale urbanization processes, and a large number of buildings with an age of more than 50 years remain.

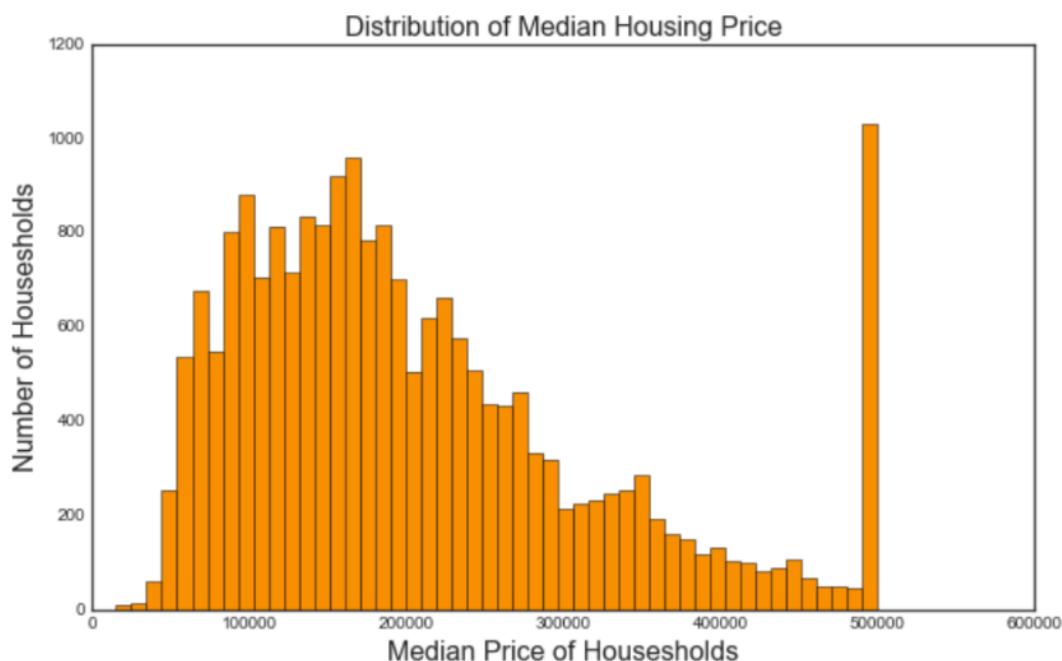


Fig 5. Distribution of Median Housing Price

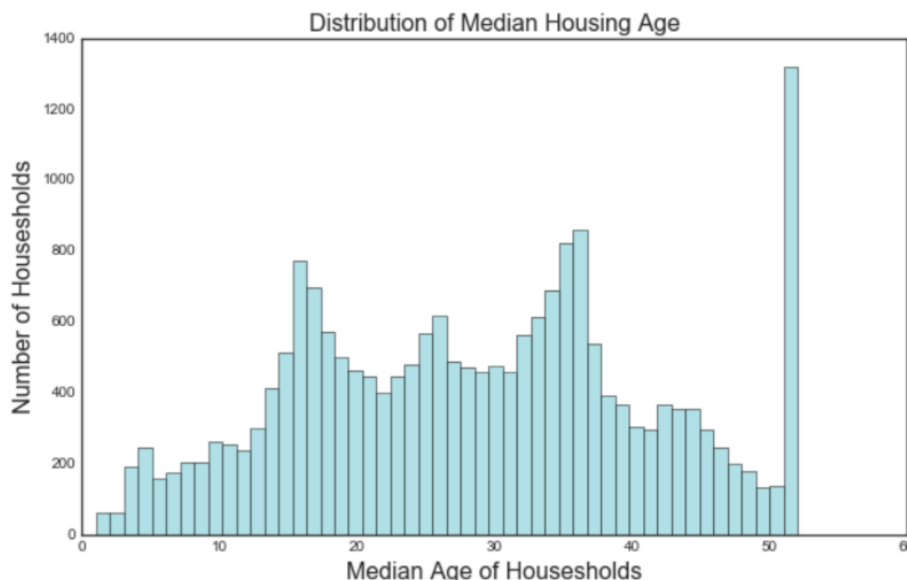


Fig 6. Distribution of Median Housing Age

Next, this paper introduces four new variables. Since the purchase of real estate usually takes the family as the basic unit, we use the family as the basic unit to calculate the average number of rooms in the family, the average number of bedrooms in the family and the average number of people in the family. The ratio of house price to income is obtained by dividing the average house price by the average wage. As summarized in table 1, it can be intuitively concluded that the average number of rooms in the family is 5.4, the average number of bedrooms is 1.1, and the average price-to-income ratio is 5.65. Although the above three values have some special values, they have less STD and less impact on the overall model. As for the index of family size, STD is relatively large, and the difference between the outlier and the 75% quantile is extremely significant. In addition, according to the distribution chart of the median housing price in Figure 7, it indicates that the outlier is somehow related to the high value of the region.

Table 1. Regional Family Structure

	Room_per_Households	Beds_per_Households	Pops_per_Households	Years_to_Buy
mean	5.429	1.097	3.071	5.65
std	2.474	0.474	10.386	3.24
min	0.846	0.333	0.692	0.36
25%	4.441	1.006	2.430	3.86
50%	5.229	1.049	2.818	4.95
75%	6.052	1.100	3.282	6.59
max	141.909	34.067	1243.333	100.00

3. Modelling

The model (1) was determined to be where the house prices are taken in the logarithmic form to reduce the multicollinearity problem.

$$Y = \alpha + \sum_{i=1}^{14} \beta_i * X_i + u \quad (1)$$

Where, Y is the independent variable, in this paper is chosen to be house price. X_i is the the dependent variables, including that X_1 is Median Income; X_2 is Median Age of the house; X_3 is Total Rooms; X_4 is Total Bedrooms; X_5 is Population in the block; X_6 is Households in the block;

X_7 is latitude of the block; X_8 is longitude of the block; X_9 is Distance to the coast; X_{10} is Distance to the nearest major city; X_{11} is Room per Households; X_{12} is Beds per Households; X_{13} is Pops per Households; X_{14} is Years to Buy the house. α and β are intercept and residual, respectively

After the model is established, the overall data set is divided into training data set and testing data set, and 20% of the original data set is selected as the training set. The results are shown in the Table 2.

Table 2. Model 1

Variables	Values	Variables	Values
α	-1509028.37	X_8	-16554.57
X_1	45465.23	X_9	-0.02
X_2	698.83	X_{10}	-0.02
X_3	2.49	X_{11}	187.56
X_4	11.49	X_{12}	3460.35
X_5	-21.07	X_{13}	-120.38
X_6	46.69	X_{14}	1.79
X_7	-15977.57		

After the model is established, the overall data set is divided into training data set and testing data set, and 20% of the original data set is selected as the training set. The results are shown in the figure below. After testing, the accuracy of the model on the training set is 84.62% and the accuracy on the test set is 83.16%.

In order to avoid spurious regression phenomenon, the robustness of the model is tested in this paper, and the first-order lag term of Y is selected to repeat the test. The accuracy of the new model on the training set is 84.14%, and the accuracy on the test set is 82.78%, which proves that the conclusion is robust.

4. Collinearity problem

In the research process of collinearity problem, since X_{11} and X_{12} are derived variables, the correlation between various variables is calculated considering the multicollinearity problem, and the results are shown in the Figure 8.

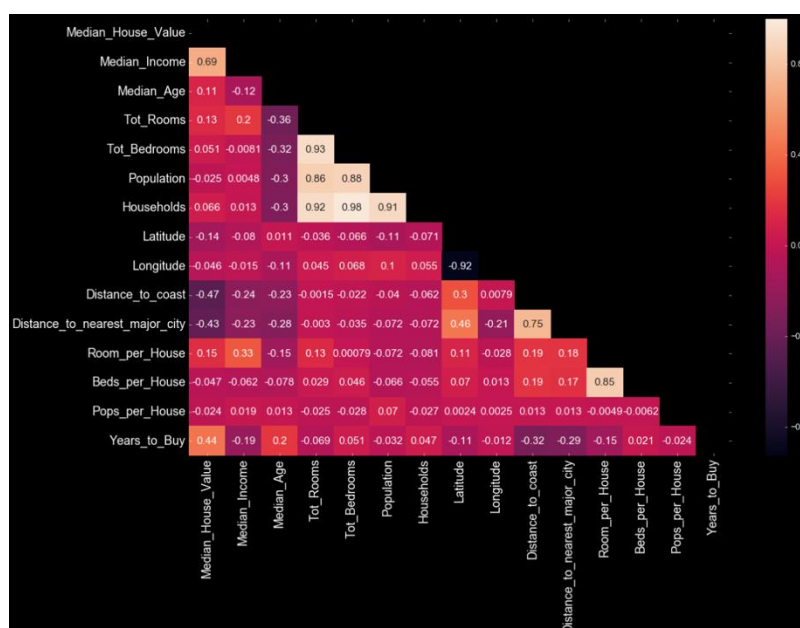


Fig 7. Coefficient

In order to eliminate the influence of repeated factors, a new model, model (2) was constructed, its independent variables remained unchanged, X3 and X4 were deleted as dependent variables, and the model was constructed in the same way. The results are shown in the following figure.

Table 3. Model 2

Variables	Values	Variables	Values
α	-1518999.59	X_9	-0.02
X_1	45244.72	X_{10}	-0.02
X_2	689.62	X_{11}	1790.45
X_5	-23.89	X_{12}	-2102.46
X_6	79.61	X_{13}	-7.35
X_7	--15948.55	X_{14}	1.80
X_8	-16612.65		

After testing, the accuracy of the model on the training set is 84.43%, and the accuracy on the test set is 83.84%.

The stability test was also carried out by the first-order lag term. The accuracy of the model for the training set was 83.63%, and the accuracy of the model for the test set was 84.62%, which proved that the conclusion was robust.

In this way, delete the redundant part in the previous text.

5. Discussion

5.1 Page Numbers

In this paper, the economic development of California in 1990 was described qualitatively, and it was found that there were strong regional differences in population and income. Most of the population was clustered in the megacities, and the income level was decreasing from the megacities.

At the same time, the housing supply also showed obvious regional differences, and the total housing supply decreased from the giant cities and coastlines to the outside. Housing supply is mainly divided into two types, rigid type and luxury type. The target users of rigid housing are ordinary citizens, and the average price of housing is 206K; The target of luxury housing is high net worth group, the average price of housing starts at 500K.

Through model analysis, the following phenomena can be found: Home values are highly positively correlated with the local area's per capita income and the age of the home. Generally speaking, high net worth homes tend to be concentrated in urban centers that have been developed for a long time and are highly urbanized, where more high-value jobs can be provided and the per capita income of the area is higher. At the same time, because of the relatively early development time, the housing age is relatively high.

In model one, house value is positively correlated with the number of rooms and bedrooms; But in model two, house value is positively related to the number of rooms per household and negatively related to the number of bedrooms. The explanation here is that for high-net-worth families, they are willing to pay more premium for the improvement of life quality and hope that the house can have more space, which leads to the increase of the total number of rooms and bedrooms. When the family is considered as the basic unit, the number of bedrooms in a single-room house is more about the family planning, while high-net-worth families usually have low fertility intention, which leads to their lack of too many bedrooms.

Housing values are negatively correlated with the number of people in the area, positively correlated with the number of households, and negatively correlated with the average number of people in the household. For high-net-worth families, they want a more comfortable environment when buying a house. They often choose areas with less population density, and they have a lower

fertility intention and a lower average household population. For families, security is an important factor. Households tend to cluster in safer neighborhoods, and the property tax generated by high net worth homes can maintain relatively better neighborhoods.

The housing value has a weak negative correlation with the distance to the megacities and the coast. The main reason for this is that in the old city center, due to the relatively early development time, the aging of infrastructure and housing quality has resulted in deurbanization. Although there are still many high-value houses in the city center, many high-net worth families choose the newly developed suburban apartments.

Home values are positively related to the housing-wage ratio. Literature References

6. Conclusion

This paper qualitatively describes the regional differences in various regions of California, and analyzes and discusses the reasons for the regional differences in housing prices from the micro perspective through linear regression model, and draws the following conclusions:

Urbanization and the agglomeration effect of scenic spots have a phased impact on housing prices. In the early years, as modernization progressed, only the wealthiest citizens were able to buy the most expensive houses in the center. In a later period, as the central area becomes unlivable due to aging infrastructure and environmental pollution, wealthy families will move to newly constructed upscale neighborhoods.

Regional population structure and family structure have strong correlation with housing price. In areas with high housing prices, there are more families, less population and less average population.

At the same time, this paper also has corresponding limitations. Due to the relatively difficult data acquisition, this paper only uses the real estate data of California in 1990, which is small in scale, has certain contingency and regional characteristics, and does not consider the fluctuation of time length.

References

- [1] Downs A. Real Estate and Long-Wave Cycles. National Real Estate Investor, 1993, June.
- [2] Kuznets Simon. Long Term Changes in National Income of the United States Since 1870. Review of Income and Wealth, 1952, 2(1): 29-241.
- [3] Grebler Leo, David Blank, Louis Winnick. Capital Formation in Residential Real Estate: Trends and Prospects. Journal of Finance, 1956, 24(1): 626.
- [4] Maisel, Sherman J. Changes in the Rate and Components of Household Formation. Journal of American Statistical Association, 1960, 55(2): 268-283.
- [5] Easterlin, Richard A. Economic-Demographic Interactions and Long Swings in Economic Growth. America Economic Review, 1966, 56: 1063-1104.
- [6] Jaffee, Dwight Mand T. Rosen Kenneth. Mortgage Credit Availability and Residential Construction. Brookings Papers of Economic Activity, 1979, 2: 333-376.
- [7] Smith, Lawrence B. Household Headship Rates, Household Formation and Housing Demand in Canada. Land Economics, 1984, 60(2): 180-188.
- [8] Mayo, Stephen K. Theory and Estimation in the Economics of Housing Demand. Journal of Urban Economics, 1981, 10: 95-116.
- [9] Horioka C. Y. Tenure Choice and Housing Demand in Japan. Journal of Urban Economics. Journal of Urban Economics, 1988, 24(3): 289-303.
- [10] Clayton J. Rational Expectations, Market Fundamentals and Housing Price Volatility. Journal of Real Estate Economics, 1996, 24(4): 441-470.
- [11] Chen M., Kawaguchi Y., Patel K. An Analysis of The Trends and Cyclical Behaviours of House Prices in The Asian Markets. Journal of Property Investment & Finance, 2004, 22(1): 55-75.
- [12] Liang Y., Gao T. An Empirical Analysis of Regional Differences in Real Estate Price Volatility in China. Journal of Economic Research, 2007, 8.