Analysing the Differentiation of Housing Price Influence Based on PCA Clustering-Entropy Weight Method: A Case Study in Sichuan

Xuedong Liang1, Qian Lu2 and Qian Lu2*

1Business School, Sichuan University, Chengdu, Sichuan, 610065, China
2Business School, Sichuan University, Chengdu, Sichuan, 610065, China
3Corresponding author’s e-mail: 2018225025133@stu.scu.edu.cn

Abstract. The immobility of real estate products and the regional nature of the consumption of real estate products make the real estate industry have obvious regional characteristics. The current policy adjustments in the real estate market emphasize differentiation, that is, the regulation methods of each city are becoming more and more different, not only because of the policy of the city, but also the policy of the district. For exploring the basis of differential regulation, the article takes Sichuan Province of China as an example. Through the selection of four first-level indicators and thirteen secondary indicators, 21 cities (districts) are divided into four types of areas. Then the entropy method is used to obtain the factors that affect the housing prices in each category. Based on these, it is suggested to improve the real estate market related policies from the three aspects: promoting the construction of the housing rental market, guiding the rational expectations of buyers and correctly guiding the flow of people.

1. Introduction
Housing values play a critical role in the stability of national economies and financial markets [1]. In recent years, the development and distribution of real estate market in various regions have become more and more differentiated [2]. In 2018, the national real estate regulation and control exceeded 400 times [3]. However, the national housing prices are still showing an unbalanced and unhealthy development trend. Therefore, how to classify and formulate effective control policies for different regional characteristics has certain practical significance. Scholars have made relevant research on this issue. The following paper will be divided into two parts, the first one is the regional difference study of real estate price fluctuations, and the second is the research on the influencing factors of real estate prices.

1.1. Research on regional differences of real estate price fluctuation
Many studies have shown that regional real estate prices are very different. Wenkang Wu proposed that there are obvious regional differences in the scale of industrial development, real estate market supply and consumer demand in China's regional real estate industry [4]. H J Yang used the spatial autocorrelation index combined with the spatial variogram to analyse the spatial structure characteristics and driving mechanism of Xi'an housing prices [5]. The results show that residential prices show significant spatial autocorrelation and the spatial variability of residential prices continues to increase. The result indicated that house prices are influenced by economic variables in other regions. Liang and Gao pointed out that the regional imbalance of their fluctuations was significant, and then furthered the
reasons for the regional differences in housing price fluctuations through the panel data model [6]. The study suggests that the real estate market should adopt a localized regulatory policy to avoid a one-size-fits-all situation in the past.

1.2. Research on the factors affecting real estate prices
Scholars have also conducted corresponding research on the factors affecting housing prices. N Adela found through empirical research that the main reasons for excessive housing prices are interest rates, unemployment rates, large numbers of immigrants, and household population and income [7]. Joseph Gyourko and others accurately record land value, construction costs, and overall house prices, and then create some supply and demand indicators to see if price growth can be reasonably interpreted to reflect local market fundamentals [8].

The above research shows that there are some researches on regional differences in housing price fluctuations. In addition, there are many factors affecting housing prices, but the importance of various factors on housing prices is significantly different. China has a vast territory, and the level of consumption and economic development among different regions are quite different. Therefore, the factors affecting the rise in housing prices among various consumption regions are analyzed, and the corresponding policies are formulated according to the factors affecting each consumption interval. Taking Sichuan Province as an example, the article firstly divided the 21 cities (districts) in the province and then conducted panel data analysis.

2. Materials and Methods

2.1. Entropy Weight Method
The entropy weight method is a method of determining the standard by using information entropy as the weight. Entropy is a thermodynamics concept first introduced into information theory by Claude Elwood Shannon, in which entropy is used to represent the measure of disorder in a system [9]. The entropy weight method is based on the improvement of information entropy definition in Shannon information theory. Information entropy:

\[ H(x) = -\sum_{i=1}^{n} p(x_i) \ln p(x_i) \]

2.2. Principal component analysis
Principal component analysis is a common method for data dimensionality reduction. It converts more variables into fewer linearly unrelated components through linear combination. Generally, several principal components with a contribution rate greater than or equal to 85% are extracted to explain the original variable.

If the original data constitutes a matrix \( X_{n \times m} \), the column vectors \( X_1, X_2, ..., X_m \) represent the respective variables, the coefficient matrix \( A_{n \times m} \), and the principal component is finally expressed as:

\[
\begin{align*}
F1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1m}X_m \\
F2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2m}X_m \\
&\quad \cdots \\
Fm &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mm}X_m
\end{align*}
\]

2.3. Systematic cluster analysis
Cluster analysis is a common method for object classification. Generally, the Euclidean distance is used to calculate the distance between each sample, and the distances are similar. The paper clusters the data after PCA to complete the classification. The distance between the variables:
3. Selection of Indicators and data sources

3.1. Evaluation Index System Construction

This study intends to use economic factors, demand factors, supply factors and social factors to identify factors affecting regional differences in real estate market prices in specific regions, and analyses them through panel data models. Regional economic development is directly proportional to real estate prices. Economic indicators are mainly composed of regional GDP, per capita GDP, real estate investment and general public budget. Economically speaking, the fluctuation of real estate price level is the result of the combination of supply and demand of real estate. Among them, demand factors include per capita disposable income and housing sales area. Factors affecting real estate supply are land availability, completed housing costs, and house prices. In addition to economic, supply and demand factors, there is also a very important part of social factors, including urbanization rate, population growth potential, marriage rate and divorce rate. The urbanization rate and the population growth potential are both demographic factors.

Table 1. Index system

| Primary indicator | Secondary indicators                      |
|-------------------|-------------------------------------------|
| Economic Factors  | GDP-F1                                    |
|                   | Real GDP per capital-F2                   |
|                   | Real estate investment-F3                  |
|                   | General public budget revenue-F4          |
| Demand Factors    | Per capita disposable income-F5           |
|                   | House sales area-F6                       |
| Supply Factors    | Land availability-F7                      |
|                   | Completed house cost-F8                   |
|                   | House price-F9                            |
| Social Factors    | Urbanization rate-F10                     |
|                   | Population growth potential-F11           |
|                   | Marriage rate-F12                         |
|                   | Divorce rate-F13                          |

3.2. Data resources

All data used in this paper are from the Sichuan Statistical Yearbook for the period 2010-2016 and the National Bureau of Statistics.

4. Analysis Process

4.1. Principal component analysis

This paper deals with 13 evaluation indicators. Since the evaluation indicators have different dimensions, the first step is to perform dimensionless processing through linear normalization. Then, the correlation matrix of the indicator data is used to calculate the eigenvalues and eigenvectors and the contribution rate of the principal components. The cumulative contribution rate is used to determine the number of components to calculate the principal component. And the expressions of the comprehensive indicators F of each city in 2010 are obtained:

\[ F = 8.91919358 \times F1 + 0.73276165 \times F2 + 1.26722657 \times F3 \]

The cumulative variance contribution rates of the three main components F1, F2 and F3 from 2011 to 2016 are all greater than 85%, so the situation in 2010-2016 is as follows:
### Table 2. Result of Principal component analysis.

| City           | 2010       | 2011       | 2012       | 2013       | 2014       | 2015       | 2016       |
|----------------|------------|------------|------------|------------|------------|------------|------------|
| Chengdu        | 110.762    | 104.343    | 98.66972   | 103.2247   | 104.8271   | 101.4138   | 102.0687   |
| Zigong         | -2.27196   | -1.9262    | -5.28205   | -4.03764   | -5.82745   | -4.62844   | -5.4928    |
| Panzhihua      | 9.247335   | 8.963522   | 8.669976   | 10.18803   | -0.43399   | 2.772214   | 3.686295   |
| Luzhou         | -0.51535   | -1.00081   | -2.9415    | -0.99099   | -3.01119   | 6.356099   | -0.33031   |
| Deyang         | 5.031853   | 5.662388   | 6.734165   | 6.187878   | 0.782086   | -2.27724   | -3.85237   |
| Mianyang       | 6.770394   | 7.026283   | 6.949791   | 4.626891   | 5.395742   | 3.607407   | 1.277166   |
| Guangyuan      | -15.3253   | -13.8741   | -14.0337   | -15.1493   | -10.9018   | -11.9513   | -12.4398   |
| Suining        | -11.875    | -10.702    | -8.24413   | -6.96836   | -5.84299   | -6.39328   | -5.65746   |
| Neijiang       | -5.92565   | -4.67588   | -6.46086   | -5.62086   | -4.66046   | -4.55311   | -10.2362   |
| Leshan         | 2.802917   | 3.24077    | 1.074701   | 0.039987   | -3.26102   | -0.40124   | -3.91702   |
| Nanchong       | -7.07093   | -7.28273   | -4.87291   | -4.52481   | 0.548623   | -1.57571   | -4.71164   |
| Meishan        | -3.73934   | -4.02772   | -3.5001    | -5.17462   | -4.67821   | -1.56095   | -1.23407   |
| Yibin          | 1.756468   | 1.109608   | 1.484675   | 1.58253    | 5.031533   | -0.98194   | 2.569448   |
| Guang'an       | -13.0206   | -12.231    | -9.46907   | -8.59938   | -9.36982   | -6.23367   | -6.05931   |
| Dazhou         | -6.52246   | -11.3927   | -8.21809   | -6.54194   | -4.88299   | -4.02815   | -6.10156   |
| Ya'an          | -9.14236   | -7.74131   | -9.86453   | -9.13322   | -7.9801   | -19.5979   | -8.51673   |
| Bazhong        | -22.2237   | -20.9665   | -16.6002   | -18.0556   | -12.2951   | -14.2465   | -12.7439   |
| Ziyang         | -7.47287   | -3.05309   | -1.01136   | -3.13227   | 0.641916   | -0.3782   | -7.41206   |
| Aba Prefecture | -18.6151   | -16.6801   | -16.5253   | -18.5159   | -23.6519   | -17.0393   | -12.9916   |
| Ganzi Prefecture| -14.9127   | -18.0804   | -21.5925   | -20.6082   | -20.9341   | -16.699   | -16.3352   |
| Liangshan Prefecture| 2.256977 | 3.288865 | 5.033292 | 1.202247 | 0.504964 | -1.60359 | 8.431119 |

### 4.2. Systematic cluster analysis

After the principal component is reduced in dimension, then the cities is clustered according to the obtained comprehensive indicator panel data, as follows:

![Figure1. Result of System cluster.](image)

### 4.3. Entropy weight and average calculation for each type

| City          | X1       | X2       | X3       | X4       | X5       | X6       | X7       |
|---------------|----------|----------|----------|----------|----------|----------|----------|
| Chengdu       | 0.06450  | 0.06192  | 0.06140  | 0.06459  | 0.05866  | 0.13749  | 0.04952  |
| Zigong        | 0.07393  | 0.06945  | 0.06209  | 0.08421  | 0.07573  | 0.05299  | 0.10523  |
| Neijiang      | 0.06459  | 0.06178  | 0.09814  | 0.08036  | 0.08880  | 0.05891  | 0.05736  |
| Nanchong      | 0.06377  | 0.06303  | 0.05351  | 0.07667  | 0.07264  | 0.04384  | 0.09326  |
| Meishan       | 0.06130  | 0.05995  | 0.08514  | 0.06645  | 0.06522  | 0.05640  | 0.11546  |
| Guangan       | 0.05826  | 0.05307  | 0.12720  | 0.06159  | 0.06081  | 0.03999  | 0.06929  |
| Dazhou        | 0.06418  | 0.06153  | 0.05446  | 0.08797  | 0.08346  | 0.05112  | 0.08520  |
IOP Conf. Series: Materials Science and Engineering 780 (2020) 072054 doi:10.1088/1757-899X/780/7/072054

5. Conclusions and policy implication

(1) Promote the construction of the housing rental market

China's housing leasing market is developing at a slower pace, which is closely related to China's current policy of selling new commercial housing and not renting, and people's awareness of renting

4.4. Result Analysis:

The real estate market of 21 cities and districts in Sichuan Province is divided into four categories through the cluster analysis method. The results are as follows: Chengdu is a category; Dazhou city, Zigong city, Ya'an city, Guang'an city, Meishan city, Nanchong city, Neijiang city, Zigong city, Suining city as a class; the Aba Prefecture, Bazhong city, Guangyuan city and Ganzi prefecture are one class; Panzhihua, Deyang city, Liangshan city, Mianyang city, Yibin city, Luzhou city and Leshan city as a class.
housing is weak. To solve this problem, we need to construct a good and orderly housing leasing system from the two fundamental and critical perspectives of policy system and market governance. First of all, we must strengthen macro management; establish a scientific housing rental market management system on a macro level; optimize supervision and management environment, and promote the development of the housing rental market to a standardized development with efficient and high-quality housing rental management services and housing rental market system. Secondly, we must change our concept and gradually transform from “re-sale light rent” to “rental and sale”.

(2) Guide buyers to rational expectations

People's expectations of future housing prices will also affect real estate price volatility. Therefore, it is necessary to correctly guide investors and consumers' expectations of real estate prices, and to report misrepresentation of consumers who are misleading consumers in the market environment, and to prevent the risk of bubbles in the real estate market. Therefore, government departments also need to manage bad behaviors that disrupt the real estate market.

(3) Correctly guide population movement

The population is also an important factor affecting the price change of real estate. The denser the population, the higher the average real estate price. In order to achieve a balanced distribution of population in various regions, local governments should incorporate the coordinated development of the regional economy into the overall planning, dynamically adjust the industrial structure, and tighten the ecological environment of the backward regions to attract more population inflows and rationally adjust the price of regional real estate.

References

[1] Fan, Y., Yang, Z., Yavas, A. (2019) Understanding real estate price dynamics: The case of housing prices in five major cities of China. Journal of Housing Economics, 43: 37–55.
[2] Pijnenburg, K. (2017) The spatial dimension of us house prices. Urban Stud. 54:466–481.
[3] Wang, J.H., (2018) Real estate financing under financial regulation and real estate control. http://www.zhonglun.com/Content/2018/10-09/1133151274.html
[4] Wu, W.K. (2005) Research on the development law of regional real estate industry in China [D]. Jilin University.
[5] Yang, H.J., Gao, E.L., Yang, X.Z., (2017) Study on the Evolution of the Spatial Structure of Commercial Housing Price in Xi'an City. Science of Surveying and Mapping. 42:86–92.
[6] Liang, Y.F., (2007) Empirical Analysis of Regional Differences in China's Real Estate Price Volatility. Economic Research. 08:133-142.
[7] Adela Nistor, Diana Reianu. (2017) Determinants of housing prices: evidence from Ontario cities. International Journal of Housing Markets and Analysis, 11: 541-556.
[8] Jing Wu, Joseph Gyourko, Yongheng Deng. (2016) Evaluating the risk of Chinese housing markets: What we know and what we need to know. Journal of China Economic Review. 39: 91–114.
[9] Wu, G., Duan, K. (2017) Integrated Sustainability Assessment of Public Rental Housing Community Based on a Hybrid Method of AHP-Entropy Weight and Cloud Model. Sustainability. 603.