

Application of Multiple Equal Linear Regression Model on real estate batch valuation

Abstract:

Linear regression method is one of the methods for assessing real estate batch. The defects of simple linear regression method are analyzed concretely in this paper, by selecting Foshan city real estate trading information for samples with empirical method. Equal linear regression is applied to analyze the housing prices in Foshan, by comparing the effect of linear regression and standard regression analysis of data, observing different regression models, analyzing the impact of pings number, room number, hall number, number of bathrooms and the age of the building variables on housing prices, to improve the linear regression method. The results show that if using the standard square regression to analyze, there may be undervalued and overvalued, which cannot correctly reflect what is the real factors that influence the estate price. This paper found that the number of rooms, bathrooms, and the age of the building will be undervalued in high-priced estate if using the standard linear regression to estimate there, namely in the high-priced estate, these variables have significant impact on housing prices.

Keywords: Batch valuation; Property valuation; Analysis of regression; Multiple Equal linear regression

1、 Introduction

Since China began to implement the reform of the housing system in 1998, the price of property in China has continued to soar. In recent years, the state has made greater efforts to regulate and control the real estate industry. The number of policies promulgated has gradually increased. After years of state regulation of the housing market, housing prices have eased somewhat. In today's volatile housing market, it is important to use a proper valuation method to make a reasonable valuation of the property. In view of this, this paper will study on the variables such as the number of pings, the number of rooms, and provide some advice for property valuation.

In the past, most of the researchers used a least-squares regression to perform regression analysis on the data when evaluating the real estate volume(Zhao Yuanyuan,2008;Liu Bing). In addition, the quantile regression model proposed by foreign scholars (1978) has also been widely used in the study of real estate bulk assessment. For example, domestic scholars Zhang Yiwen and Jiang Yinghui used the quantile regression model to study the effect of commodity housing's own property on real estate prices in atypical residential buildings in 1997, verifying that the standard regression model cannot accurately estimate the distribution of property prices for various characteristic attributes. The conclusion of the difference, and thus put forward the method of using the quantile regression model to evaluate the real estate. However, it is complex and difficult to understand, and it is difficult to study.

Therefore, in this paper, we use another method called Multiple Equal Part Linear Regression Model (EPLRM), in which the analysis data are divided into several equal parts to build the linear regression model. This way, the model trend of each equal part can be independently observed. It can be compared with the general linear regression. By using this method, we can learn that whether there will be undervalued or overvalued if using least square regression analysis and find out the real factors that influence the real estate price.

2、Study Method

2.1 Equal part linear regression model

In this paper, we adopt the Multiple Equal Part Linear Regression Model (EPLRM) which proposed by Professor Pan(2017). It can divide the analysis data into several equal parts and build the linear regression model respectively. In this way, the trend of each equal parts can be independently observed. Also, it can be compared with the standard linear regression.

Assume y is a continuous dependent variable dependent on x . The analysis data is divided into several equal parts and built fit linear regression model for the different equal parts. These three linear regression equations can be expressed as:

$$y_i^{\tau} = \beta_0^{\tau} + \beta_1^{\tau} x_i^{\tau} + \varepsilon_i^{\tau}$$

Others such as standard estimation:

$$\hat{\beta}_0^{\tau} = \frac{\sum_{i=1}^n (x_i^{\tau} - \bar{x}^{\tau})(y_i^{\tau} - \bar{y}^{\tau})}{\sum_{i=1}^n (x_i^{\tau} - \bar{x}^{\tau})^2}$$

$$\hat{\beta}_1^{\tau} = \bar{y}^{\tau} - \bar{\beta}_0^{\tau} \bar{x}^{\tau}$$

Coefficient of determination and confidence interval:

$$(\hat{\beta}_i^{\tau} - t_{\alpha/2} \times s_{\hat{\beta}_i^{\tau}}, \hat{\beta}_i^{\tau} + t_{\alpha/2} \times s_{\hat{\beta}_i^{\tau}}), i=0,1$$

Symbol “ t ” is added at certain points in the equation.

52

2.2 Variable description and data source

In the case, this paper uses the house price to carry out factor analysis. The number of rooms (X1), the number of rooms (X2), the number of halls (X3), the number of bathrooms (X4) and the age of the building (X5) are taken as the study objects.

The study takes 550 Foshan's real estate data to analyze in R program. And these data are came from China major real estate websites such as HomeLink Real Estate, Jia Jia Shun real estat and so on.

Table 1 below is the sample statistics for Foshan.

Table 1 Foshan sample description statistics

Variable	mean	S.D	min	max
Y	226.16	264.69	25.00	2980.0
X1	127.55	93.28	34.5	767.79
X2	3.20	0.97	1.00	8.00
X3	1.98	0.42	0.00	4.00
X4	1.92	0.90	1.00	7.00
X5	5.40	4.64	1.00	26.00

3. Case study

3.1 Variable descriptive statistics and correlation analysis

The empirical analysis takes house price (Y) as the dependent variable, and takes the standard regression analysis of the number of pings (X1), the number of rooms (X2), the number of halls (X3), the number of bathrooms (X4) and the age of the building (X5). Table 2 is the narrative statistics of standard Regression for the sample data of Foshan, and our study mainly focuses on the impact of property characteristics on house prices.

Table 2 The Standard regression coefficient and constant

	系 数
CONSTANT	-48.442820
X1	2.698195

X2	-7.678986
X3	-1.849502
X4	-17.236979
X5	-1.512460

70

71 From Table 2, we can see that in the linear regression model, the number of rooms (X2), number of
72 bathrooms (X4), the age of the building (X5) and house prices (Y) are negatively correlated. There was a
73 slight negative correlation between the age of the building (X5) and house price (Y), a significant negative
74 correlation between room number (X2), and a very significant negative correlation between bathroom
75 number (X4) and house prices.

76

77 3.2 Multiple equal part linear regression model analysis

78 Take the sum of the data collected from Foshan as a sample, take house price (Y) as the dependent
79 variable, number of pings (X1), number of rooms (X2), number of halls (X3), number of bathrooms (X4), the
80 age of the building (X5) as independent variable. We divide the data into three parts for multiple equal linear
81 regression.

82 **Table 3 Multiple Equal linear regression sample description statistics**

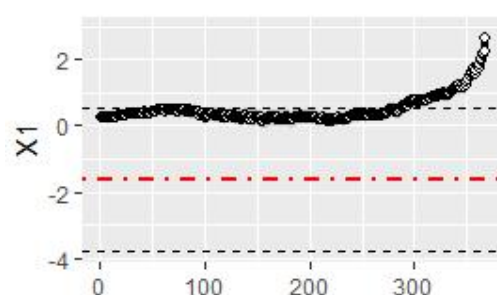
	LRM			EPLRM $\tau=1$			EPLRM $\tau=2$			EPLRM $\tau=3$		
	R ² =0.79			R ² =0.83			R ² =0.80			R ² =0.67		
Stat.	Conf.	T	Sig.	Conf.	T	Sig.	Conf.	T	Sig.	Conf.	T	Sig.
X1	2.70	32.31	***	2.88	20.00	***	2.47	17.67	***	1.98	11.92	***
X2	-7.68	-0.91	-	0.69	0.04	-	9.62	0.75	-	-9.04	-0.73	-
X3	-1.85	-0.12	-	-79.86	-2.65	***	40.30	1.78	*	32.36	1.50	-
X4	-17.24	-1.98	**	-1.29	-0.07	-	-44.85	-3.25	***	3.82	0.29	-
X5	-1.51	-1.34	-	-0.94	-0.33	-	-10.48	-1.41	-	3.11	0.59	-

83 note: *indicates 10% significance level; **indicates 5% significance level; ***indicates 1% significance level.

84 From Table 3, it is found that the number of pings (X1) and the number of halls (X3) are more
85 significant in the first equal part, the number of douche (X4) is more significant in the second equal part.

86 The following figures show the trajectories of the equal part linear regression's coefficients and the
87 confidence intervals. The red dashed line is the standard linear regression line, and the upper and lower
88 straight horizontal lines are the confidence intervals of the standard linear regression. The irregular black line
89 is the equal part linear regression line. The upper and lower gray areas are the confidence intervals for
90 equally divided linear regression.

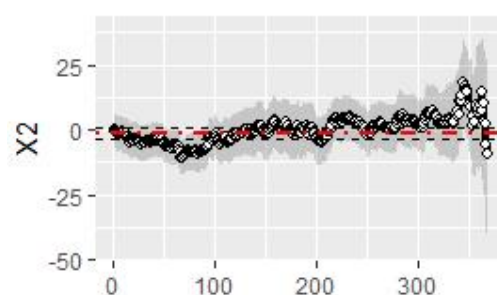
91



92

93 **Fig.1 the trajectories of the multiple equal part linear regression's coefficients and the confidence**
94 **intervals(X1)**

95 From Figure 1, it can be seen that the equal part linear regression lines of the number of pings (X1) in
96 Foshan is located above the confidence interval of the standard linear regression. By comparing the data
97 analysis results of the equal part linear regression and the least-squares regression, we find that in the third
98 equal part of the data, there is a significant difference between the two regression models. In other words, if
99 we use standard linear regression to estimate high-priced estate, there will be underestimation. This paper
100 believes that buyers in Foshan have higher potential demand for houses with a larger number of pings. The
101 reason may be that Foshan City has a large area per capita and its geographical location attracts investors
102 from Guangzhou.



103

Fig.2 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X2)

From Figure 2, it can be seen that the equal part linear regression line of the number of house rooms in second and third equal parts are above the confidence interval of the standard linear regression. This paper believes that in the middle and high-priced estate, the number of rooms has a greater impact on house prices.

It can be concluded that the impact of the number of housing rooms in Foshan on house prices is generally underestimated at high prices. This paper believes that the reasons may be as follows: First, the introduction of the second-born policy, the family's permanent population has increased, the demand for the number of rooms has become larger; Second, the improvement in economic standards has led to an increase in the quality of life requirements, and thus the use of the room increased.

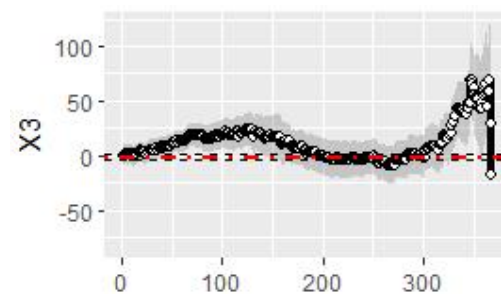


Fig.3 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X3)

It can be seen from Figure 3 that the equal part linear regression line at the second and third parts in Foshan is located above the confidence interval of the standard linear regression. The difference between the two regression models is relatively large, that is, the degree of significance is high and there is an underestimation in the middle and high-priced estate .

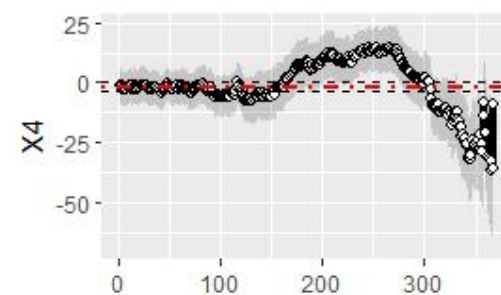


Fig.4 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X4)

From Figure 4, it can be seen that the equal part linear regression line of the number of bathrooms in the second-part is located above the confidence interval of the standard linear regression. It means that in the middle-priced estate, the number of bathrooms has a greater impact on house prices.

In addition, the equal part linear regression line of the number of bathrooms in the second part is located below the confidence interval of the standard linear regression which indicated that in high-priced estate, the number of bathroom may seem not too important.

The difference between the two regression models is larger, that is, the degree of significance is higher and there is an underestimated situation. Due to the cancellation of China's family planning policy, people in middle class tend to possess a house with more bathrooms.

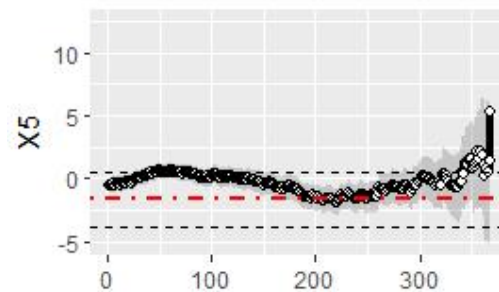


Fig.5 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X5)

From Figure 5, it can be seen that the equal part linear regression line in the third part is located above the confidence interval of the standard linear regression. So, if we use the standard linear regression to estimate the price, the impact on high-priced estate may be underestimated. It can be concluded that the age of the building have been underestimated when house prices are at high points. This paper believes that the reasons may be as follows: First, old buildings are mostly concentrated in the old town or in the city center. The location is superior and the transportation is convenient. The second is the old building style and historical value that make buyers prefer to owning.

Table 4 F Examination of Difference between Different Equal Parts

Variable	$\tau_1 - \tau_2$		$\tau_1 - \tau_3$		$\tau_2 - \tau_3$	
	F_value	Sig.	F_value	Sig.	F_value	Sig.
Constant	0.2998	***	0.0043	***	0.0142	***
X1	0.3082	***	0.0858	**	0.2782	***
X2	0.9138	-	0.4323	***	0.4730	***
X3	1.0941	-	0.8169	***	0.7467	*

X4	1.0793	-	0.1689	***	0.1565	***
X5	1.2227	-	2.0893	***	1.7088	***

note: *indicates 10% significance level; **indicates 5% significance level; ***indicates 1% significance level.

Table 5 shows the difference examination between groups carried out for original data in this paper. F examination is carried out respectively for data of each equal part. Table 5 shows that in the multiple equal part linear regression model, there is significant difference between the first equal part and second equal part of X1 and X2. And there is significant difference between the second equal part and the third equal part of all the variable. Therefore, further detailed analysis is required.

Table 5 Multiple equal linear regression coefficient difference F test

Variable	$\tau_1 - \tau_2$		$\tau_1 - \tau_3$		$\tau_2 - \tau_3$	
	F_value	Sig.	F_value	Sig.	F_value	Sig.
Constant	1.231977279	-	1.129733847	-	0.91700867	-
X1	15.34492744	***	1.601296238	*	0.104353458	***
X2	0.97687879	-	7.37439739	***	7.548937967	***
X3	9.846182681	***	20.16674692	***	2.048179236	***
X4	0.259003936	***	5.387015415	***	20.79897121	***
X5	0.432462742	**	0.07672594	***	0.177416302	***

note: *indicates 10% significance level; **indicates 5% significance level; ***indicates 1% significance level.

Among the three coefficients of linear regression, there is a significant difference in the number of pings, halls, bathrooms, and building age in the first and second equal parts. There are significant differences between the second and the third equal, first and third equal except the constant term. This result is consistent with the expectation of the linear regression model of equal parts. These differences can be determined with the linear regression equations for each parts.

4. Conclusion and suggestions

In this paper, a new method to explore the accuracy of the batch value is applied. Unlike the standard linear regression which can not probe deeply into the significance of some extreme value data implied by the

real data. Multiple equal part linear regression model can divide the analysis data into several equal parts and build the linear regression model respectively. In this way, the trend of each equal parts can be independently observed. In terms of research samples, 550 estate information of Foshan is collected by the website of China estate company and used. The empirical analysis found that:

In Foshan, the number of housing pings, rooms, and halls and the building age variable, if estimated by the standard regression, the impact on the house price in the high price part will generally be underestimated. Therefore, this paper suggests that real estate developers should always pay attention to the national policy when building new properties, analyze their impact on housing demand in time for national policies, and make timely adjustments. When setting up a high-priced apartment, developers should consider increasing the number of rooms and developing new uses for the room. At the same time, we consider that location is an important factor affecting the age of the building, so this paper suggests that real estate developers can reasonably use the location advantage of the old city or downtown.

In addition, the analysis method of the multiple equal part linear regression linear regression model used in this paper is seldom used as the research method. It is used for the first time in the real estate appraisal. Therefore, this method can be extended to the property valuation in other provinces. It can also be extended to other large appraisal models, such as real estate tax assessment and inventory valuation, which can effectively avoid data distortion caused by extreme data values.

References

- Wen-Tsao Pan (2017). 'A Newer Equal Part Linear Regression Model: A Case Study of the influence of Educational Input on Gross National Income.' *EURASIA Journal of Mathematics Science and Technology Education*, 13 (8):5766-5768.
- Koenker, R.W. and Bassett, G. (1978). 'Regression Quantile', *Econometrics*, Vol.46, No.1, pp.33-50.
- Li-Jun Deng, Wen-Tsao Pan, Zu-Chang Zhong, Xue-Ying Chen (2017). 'The Influence of Education Expenditures on Industrial Upgrading: Empirical Analysis via Equal Part Linear Regression Model'. *EURASIA Journal of Mathematics Science and Technology Education*, Accept.
- Koenker, R. and K. Hallock (2001). 'Quantile Regression', *Journal of Economic Perspectives*, 15, 143-156.
- Cade, B. and B. Noon (2003). A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment*, 1, 412-420.
- Koenker, R (2005). Quantile Regression, Econometric Society Monograph Series, Cambridge University Press. Errata list.

Zhang Yiwen, Jiang Yinghui, Zhang Jinbiao. Application of Component Regression in Large-scale Valuation Models - Improvement of Atypical Housing Evaluation [J]. *Metropolis and Planning*, 2009, 36, (3): 281-304.

- 194 Lin Qiuzhen. Discussion on the Rent Mode of the Robust Residential Housing: An Analysis of the
195 Abnormal Point [J]. Chinese Journal of Housing, 1996, 4:51-72.
- 196 Lin Zujia, Ma Yujun. Application of Mass Estimate Method of Characteristic Equations in Real Estate
197 Market in Taiwan [J]. Journal of Housing Science, 2007, 16, (2): 1-22
- 198 Li Zhe. Research and empirical analysis of real estate batch valuation model [J]. China Real Estate Appraisal
199 and Brokerage, 2017, 1: 21-27.
- 200 Zhao Tiequan. Analysis of the Influencing Factors of Housing Price in Hangzhou Based on VAR [D].
201 Zhejiang: Zhejiang Sci-Tech University, 2009.
- 202 Zhao Yuanyuan. Analysis on the Influencing Factors of Commercial Housing Prices in Hebei Province [D].
203 Hebei: Hebei Agricultural University, 2008.
- 204 Li Hengkai, Xu Qixing, Wang Xiuli. Research and application of price evaluation model for real estate
205 feature [J]. Science of Surveying and Mapping, 2011, 36(6): 262-264.
- 206 Wu Jiannan, Ma Wei. "Estimating Extreme Behavior Model: Quantile Regression Method and Its
207 Implementation and Application [J]. Mathematical Statistics and Management, 2006, 25(5): 536-543.
- 208 Wei Xiahai, Yu Lingbiao. Empirical Study on the Differences between Regular Employment and Informal
209 Employment in Urban China: Based on the Discovery of Quantile Regression and Decomposition [J].
210 Quantitative Economic and Technical Economics Research, 2012(1): 78-90.