

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 in the high-priced estate if using the standard linear regression to estimate there, the number of rooms and the age of the building will be undervalued, and the number of bathrooms will be overvalued.

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 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). In addition, the quantile regression model proposed by foreign scholars (Koenker,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.

33

34 Therefore, in this paper, we use another method called Multiple Equal Part Linear Regression Model
35 (EPLRM), is was propose Professor Pan(2017). It can divide analysis data into several equal parts to build
36 the linear regression model. This way, the model trend of each equal part can be independently observed. In
37 addition, it can also be compared with the general linear regression.

38 Professor Pan has used this method to analyze the impact of the input of fixed assets stock and input of
39 human capital stock of China government on GDP. By collecting a group of sample data of 21 years, he
40 draws a conclusion that insufficient input of fixed assets stock will result in a noticeable decrease of GDP
41 and suggested that government should input fixed assets stock to considerably facilitate the growth of GDP
42 in China.

43 2、 Study Method

44 2.1 Equal part linear regression model

45 In this paper, we adopt the Multiple Equal Part Linear Regression Model (EPLRM) which proposed by
46 Professor Pan(2017). It can divide the analysis data into several equal parts and build the linear regression
47 model respectively. In this way, the trend of each equal parts can be independently observed. Also, it can be
48 compared with the standard linear regression. By using this method, we can learn that whether there will be
49 undervalued or overvalued if using least square regression analysis and find out the real factors that influence
50 the real estate price.

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

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

54 Others such as standard estimation:

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

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

55 Coefficient of determination and confidence interval:

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

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

58

59 2.2 Variable description and data source

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

63 The study takes 550 Foshan's real estate data to analyze in R program. And these data come from the
64 database of China major real estate such as HomeLink Real Estate, JiaJiaShun Real Estate and so on.

65 Table 1 below is the sample statistics for Foshan.

66 **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

67

68 3. Case study

69 3.1 Variable descriptive statistics and correlation analysis

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

75

76 **Table 2 The Standard regression coefficient and constant**

	COEFFICIENTS
CONSTANT	-48.442820
X1	2.698195
X2	-7.678986
X3	-1.849502
X4	-17.236979
X5	-1.512460

77

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

83

84 3.2 Multiple equal part linear regression model analysis

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

89 **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	-

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

From Table 3, it is found that the number of pings (X1) and the number of halls (X3) are more significant in the first equal part, the number of bathrooms (X4) is more significant in the second equal part. In order to find out the variables that really affect house price (Y), we need to analyze the data further.

Table 4 Stepwise regression analysis of the variables

```
d=read.csv(file='testdata.csv',header=T)#读取数据
attach(d)
lm.model=lm(Y~.,data=d)#建立多元线性回归方程
summary(lm.model)#查看多元线性回归结果
step.model=step(lm.model)#逐步回归分析
summary(step.model)#查看逐步回归结果
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -67.74919   13.87775  -4.882 1.38e-06 ***
X1             2.65372    0.07275  36.479 < 2e-16 ***
X4            -18.43418    7.47430  -2.466  0.014 *
X5             -1.58648    1.10487  -1.436  0.152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 120 on 545 degrees of freedom
Multiple R-squared:  0.7957,    Adjusted R-squared:  0.7946
F-statistic: 707.5 on 3 and 545 DF,  p-value: < 2.2e-16
```

So this paper use stepwise regression to ensure that the resulting set of explanatory variables is optimal. Each variable is introduced into the model one by one, and an F-test is performed after each explanatory variable is introduced. The t-test is performed for the selected explanatory variable one by one. If the originally introduced explanatory variable is no longer significant due to the introduction of explanatory variables, and it will be delete. And from the result of Table 4, number of rooms (X2) and number of halls (X3) are found that not significant in the model. Therefore, we excluded the interference of X2 and X3 and mainly focus on the influence of other variables on Y.

Table 5 Multiple Equal linear regression adjusted sample statistics statement

	LRM $R^2=0.80$			EPLRM $\tau=1$ $R^2=0.75$			EPLRM $\tau=2$ $R^2=0.74$			EPLRM $\tau=3$ $R^2=0.76$		
Stat.	Conf.	T	Sig.	Conf.	T	Sig.	Conf.	T	Sig.	Conf.	T	Sig.
X1	2.70	36.5	***	0.25	3.34	***	0.81	1.60	-	2.61	20.00	***
X4	-18.40	-2.5	**	-1.82	-0.56	-	8.27	1.84	*	-42.06	-2.89	***
X5	-1.59	-1.44	-	-0.58	-0.90	*	-1.13	-2.15	**	5.22	1.42	-

Table 5 is the sample statistics statement after adjusted. From the above table, it is found that in the first equal part, the number of pings (X1) and the age of the building (X5) are more significant. In the third equal part, the number of pings (X1) and the number of bathrooms (X4) are more significant.

Next, this paper will focus on the number of pings (X1), the number of bathrooms (X4) and the age of the building (X5), as well as their impact on house price (Y).

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

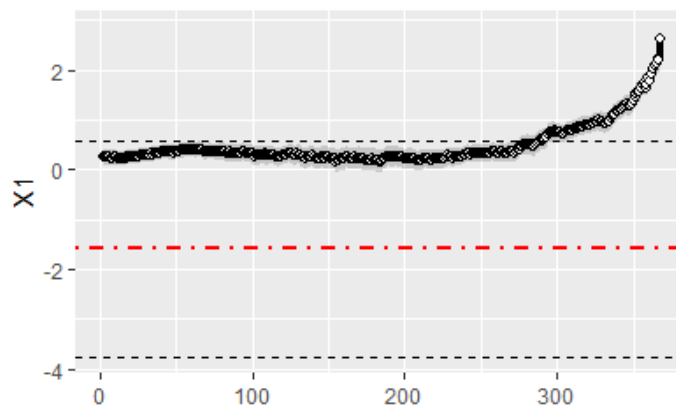


Fig.1 the trajectories of the multiple equal part linear regression's coefficients and the confidence intervals(X1)

From Figure 1, it can be seen that the equal part linear regression lines of the number of pings (X1) in

Foshan is located above the confidence interval of the standard linear regression. By comparing the data analysis results of the equal part linear regression and the least-squares regression, we find that in the third equal part of the data, there is a significant difference between the two regression models. In other words, if we use standard linear regression to estimate high-priced estate, there will be underestimation. This paper believes that buyers in Foshan have higher potential demand for houses with a larger number of pings.

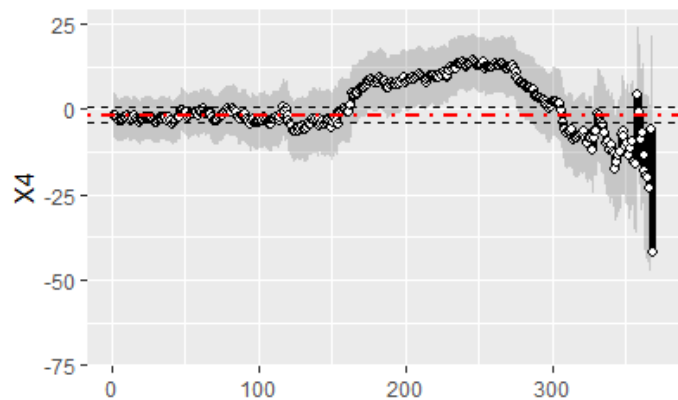


Fig.4 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X4)

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.

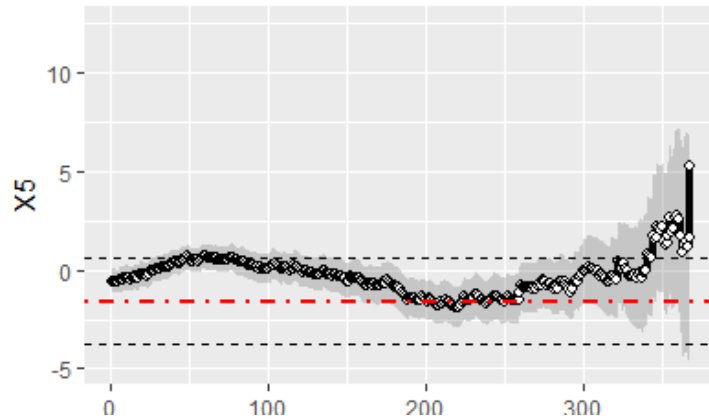


Fig.5 the trajectories of the equal part linear regression's coefficients and the confidence intervals(X5)

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 has 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 6 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	0.4066	***	0.0035	***	0.0088	***
X1	0.8359	-	0.02826	***	0.0338	***
X4	0.7722	*	0.2126	***	0.2753	***
X5	1.1266	***	2.0184	***	1.7916	***

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 6 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 7 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	0.40664796	***	0.00358698 5	***	0.00882086	***
X1	0.835935484	-	0.02826761 6	***	0.03381554 8	***
X4	1.102663294	-	0.43127327 7	***	0.39111964 6	***
X5	1.76231003	***	0.74716494 8	*	0.42396907 2	***

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 constant and building age in the first and second equal parts. There are significant difference between first and the third equal part, first and third equal part. 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. And by this, the The empirical analysis found that:

In Foshan, the number of housing pings 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. It is

suggested that high-priced estate requires more pings and better location. As for the number of bathrooms, it is overestimated in the high-priced estate, and it is suggested that real estate developers should always pay attention on customers' demand and design more suitable units.

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. In addition, this method can be extended to the property valuation in other districts. 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

Pan, W.C. (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*, vol.13, No.8, pp.5766-5768.

Koenker, R.W. and Bassett, G. (1978). 'Regression Quantile', *Econometrics*, Vol.46, No.1, pp.33-50.

Deng, L.J., Pan, W.C., Zhong, Z.C. and Chen, X.Y. (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.W. and Hallock, K. (2001). 'Quantile Regression', *Journal of Economic Perspectives*, vol.15, pp.143-156.

Cade, B. and Noon, B. (2003). 'A Gentle Introduction to Quantile Regression for Ecologists', *Frontiers in Ecology and the Environment*, vol.1, pp.412-420.

Zhong, Z.C., Chen, X.Y., Yang, H.Y., Wang, G. and Pan, W.C. (2017). 'Theoretical and Case Study on Multiple Equal Part Linear Regression.' *2017 2nd International Conference On Computational Modeling, Simulation and Applied Mathematics*.

Shao, Y.M., Li, X.W., Yang, T.T., Ma, F.X., Pan, W.C., Luo, S.H., Chen, X.Y. (2018). 'The Equal Regression Analysis of Total Social Consumable and Fixed Assets and Freight Turnover of China.' *Proceedings of the 2018 International Conference on Modern Management and Innovation*.

Huang, F.F. and Hu H.C. (2015). 'Analysis of the Influence Factors of China's Tourism Income-Based on Principal Component Regression', *Statistics and Management*, Vol.2, pp.46-48.

Xiong, Y.L. (2014). 'The Main Component of Linear Regression Coefficients of Pathological Model: Ridge Regression Estimate', *Shuxue Xuexi Yu Yanjiu*, Vol.9, pp.121-121.

Zhao, B.F., Hou, Y.G. and Wen, Z.J. (2011). 'An Improved Multiple Linear Regression Model of Air Conditioning Load Prediction', *Refrigeration Air Conditioning & Electric Power Machinery*, Vol.32, No.4, pp.49-51.

Zhang, Y.W., Jiang Y.H., Zhang, J.B. (2009). 'Application of Component Regression in

208 Large-scale Valuation Models - Improvement of Atypical Housing Evaluation'
209 Metropolis and Planning, vol.36,No.3, pp.281-304.

210 Lin, Q.Z. (1996). 'Discussion on the Rent Mode of the Robust Residential Housing: An
211 Analysis of the Abnormal Point', *Chinese Journal of Housing*, vol.4, pp.51-72.

212 Lin, Z.J., Ma, Y.j. (2007). 'Application of Mass Estimate Method of Characteristic
213 Equations in Real Estate Market in Taiwan', *Journal of Housing Science*, vol.16,
214 No.2, pp.1-22.

215 Li, Z. (2017). 'Research and empirical analysis of real estate batch valuation model',
216 *China Real Estate Appraisal and Brokerage*, vol.1, pp.21-27.

217 Zhao, T.q.(2009). 'Analysis of the Influencing Factors of Housing Price in Hangzhou
218 Based on VAR', Zhejiang: Zhejiang Sci-Tech University.

219 Zhao, Y.Y. (2008). 'Analysis on the Influencing Factors of Commercial Housing Prices in
220 Hebei Province', Hebei: Hebei Agricultural University.

221 Li, H.K., Xu, Q.X., Wang, X.L. (2011). 'Research and application of price evaluation
222 model for real estate feature' *Science of Surveying and Mapping*, vol.6, No.6,
223 pp.262-264.

224 Wu, J.N., Ma, W. (2006). 'Estimating Extreme Behavior Model: Quantile Regression
225 Method and Its Implementation and Application', *Mathematical Statistics and
226 Management*, vol.25, No.5, pp.536-543.

227 Wei, X.H., Yu, L.B. (2012). 'Empirical Study on the Differences between Regular
228 Employment and Informal Employment in Urban China: Based on the Discovery of
229 Quantile Regression and Decomposition', *Quantitative Economic and Technical
230 Economics Research*, vol.1, pp.78-90.