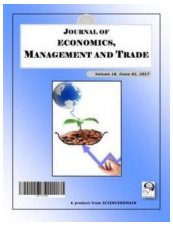

Journal of Economics, Management and Trade

21(5): 1-7, 2018; Article no.JEMT.41833
ISSN: 2456-9216
(Past name: British Journal of Economics, Management & Trade, Past ISSN: 2278-098X)

Application of Multiple Equal Linear Regression Model on Real Estate Batch Valuation

Hui-ling Yang[1*] and Jun-sheng Chen[1]

[1]School of Business, Guangdong University of Foreign Studies, Guangzhou, China.

Authors' contributions

This work was carried out in collaboration between both authors. Author HIY designed the study, wrote the protocol, wrote the first draft of the manuscript and revised the manuscript. Author JSC performed the statistical analysis and managed the analyses of the study. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JEMT/2018/41833
Editor(s):
(1) Stefano Bresciani, Professor, Department of Management, University of Turin, Italy.
Reviewers:
(1) Darmesh Krishanan, Management and Science University, Malaysia.
(2) Kadri Cemil Akyüz, Karadeniz Technical University, Turkey.
(3) N. V. Krishna Prasad, Gandhi Institute of Technology and Management, India.
(4) R. Shenbagavalli, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/24889

Original Research Article

Received 19th March 2018
Accepted 24th May 2018
Published 30th May 2018

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.

**Corresponding author: E-mail: augustyanghl@gmail.com;*

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) [1,2,3]. In addition, the quantile regression model proposed by foreign scholars (Koenker and Bassett,1978) [4,5,6] has also been widely used in the study of real estate batch valuation. For example, domestic scholars Zhang Yiwen and Jiang Yinghui [7] 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 [8,9].

Therefore, in this paper, we use another method called Multiple Equal Part Linear Regression Model (EPLRM), is was proposed by Professor Pan(2017). It can divide analysis data into several equal parts to build the linear regression model. This way, the model trend of each equal part can be independently observed. In addition, it can also be compared with the general linear regression [10,11,12].

Professor Pan has used this method to analyze the impact of the input of fixed assets stock and input of human capital stock of China government on GDP. By collecting a group of sample data of 21 years, he draws a conclusion that insufficient input of fixed assets stock will result in a noticeable decrease of GDP and suggested that government should input fixed assets stock to considerably facilitate the growth of GDP in China [13].

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. 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.

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:

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

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

2.2 Variable Description and Data Source

In the case, this paper uses the house price to carry out factor analysis. 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) are taken as the study objects [14,15,16,17].

The study takes 550 Foshan’s real estate data to analyze in R program. And these data come from the database of China major real estate such as

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

HomeLink Real Estate, JiaJiaShun Real Estate and so on.

Table 1 is the sample statistics for Foshan.

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 focus on the impact of property characteristics on house prices [18,19].

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

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

3.2 Multiple Equal Part Linear Regression Model Analysis

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

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.

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 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

Table 3. Multiple equal linear regression sample description statistics

	LRM R^2=0.79			EPLRM τ=1 R^2=0.83			EPLRM τ=2 R^2=0.80			EPLRM τ=3 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*

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
```

Table 5. Multiple equal linear regression adjusted sample statistics statement

	LRM R^2=0.80			EPLRM τ=1 R^2=0.75			EPLRM τ=2 R^2=0.74			EPLRM τ=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	-

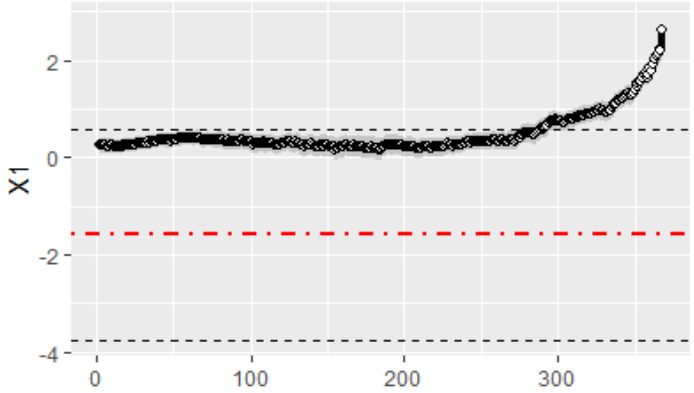

Fig. 1. The trajectories of the multiple equal part linear regression's coefficients and the confidence intervals(X1)

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.

From Fig. 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 in the third part. 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.

From Fig. 2, 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 third part is located blow 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.

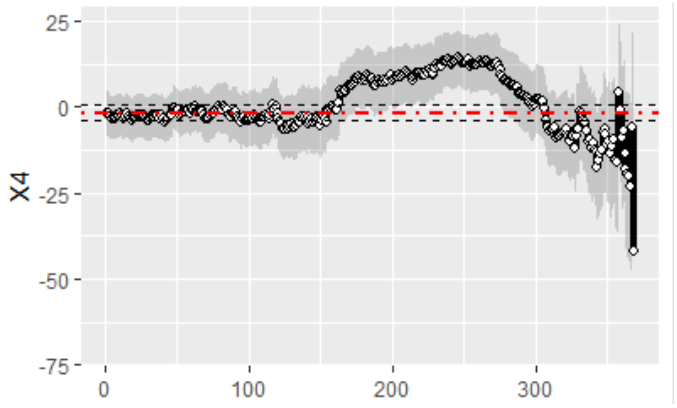

Fig. 2. The trajectories of the equal part linear regression's coefficients and the confidence intervals(X4)

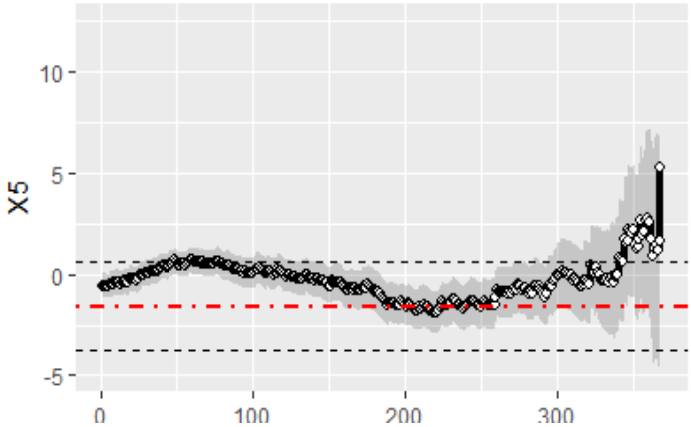

Fig. 3. the trajectories of the equal part linear regression's coefficients and the confidence intervals(X5)

Table 6. F Examination of difference between different equal parts

Variable	$\tau1-\tau2$		$\tau1-\tau3$		$\tau2-\tau3$	
	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 7. Multiple equal linear regression coefficient difference F test

Variable	$\tau1-\tau2$		$\tau1-\tau3$		$\tau2-\tau3$	
	F_value	Sig.	F_value	Sig.	F_value	Sig.
Constant	0.40664796	***	0.003586985	***	0.00882086	***
X1	0.835935484	-	0.028267616	***	0.033815548	***
X4	1.102663294	-	0.431273277	***	0.391119646	***
X5	1.76231003	***	0.747164948	*	0.423969072	***

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

From Fig. 3, 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 6 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 shows that among the three coefficients of linear regression, there is a significant difference there is a 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

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 [18,19,3].

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Zhao YY. Analysis on the Influencing factors of commercial housing prices in Hebei Province. Hebei: Hebei Agricultural University; 2008.
2. Lin QZ. Discussion on the rent mode of the robust residential housing: An analysis of the abnormal point. Chinese Journal of Housing. 1996;4:51-72.
3. Zhang YW, Jiang YH, Zhang JB. Application of component regression in large-scale valuation models improvement of atypical housing evaluation. Metropolis and Planning. 2009;36(3):281-304.
4. Koenke RW, Bassett G. Regression quantile. Econometrics. 1978;46(1):33-50.
5. Koenker RW, Hallock K. Quantile regression. Journal of Economic Perspectives. 2001;15: 143-156.
6. Cade B, Noon B. a gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment. 2003;1412-420.
7. Zhang YW, Jiang YH, Zhang JB. Application of component regression in large-scale valuation models improvement of atypical housing evaluation. Metropolis and Planning. 2009;36(3):281-304.
8. Wu JN, Ma W. Estimating extreme behavior model: quantile regression method and its implementation and application. Mathematical Statistics and Management. 2006;25(5):536-543.
9. Wei XH, Yu LB. Empirical study on the differences between regular employment and informal employment in urban China: Based on the discovery of quantile regression and decomposition. Quantitative Economic and Technical Economics Research. 2012;1:78-90.
10. Deng LJ, Pan WC, Zhong ZC, Chen XY. 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; 2017.
11. Zhong ZC, Chen XY, Yang HY, Wang G, Pan WC. Theoretical and case study on multiple equal part linear regression. 2017 2nd International Conference On Computational Modeling, Simulation Ang Applied Mathematics; 2017.
12. Shao YM, Li XW, Yang TT, Ma FX, Pan WC, Luo SH, Chen XY. 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; 2018.
13. Pan WC. 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. 2017;13(8):5766-5768.
14. Lin ZJ, Ma YJ. Application of mass estimate method of characteristic equations in real estate market in Taiwan. Journal of Housing Science. 2007;16(2):1-22.
15. Zhao TQ. Analysis of the influencing factors of housing price in Hangzhou based on VAR. Zhejiang. Zhejiang Sci-Tech University; 2009.
16. Li HK, Xu QX, Wang XL. Research and application of price evaluation model for real estate feature. Science of Surveying and Mapping. 2011;6(6):262-264.
17. Huang FF, Hu HC. Analysis of the influence factors of China's tourism income-based on principal component regression. Statistics and Management. 2015;2:46-48.
18. Li HK, Xu QX, Wang XL. Research and application of price evaluation model for real estate feature. Science of Surveying and Mapping. 2011;6(6):262-264.
19. Xiong YL. The main component of linear regression coefficients of pathological model: Ridge regression estimate. Shuxue Xuexi Yu Yanjiu. 2014;9:121-121.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/24889