2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics

Study of Residential Price Index Based on Quantile Regression Model — A Case Study on Hangzhou, China

Haizhen Wen*
Zhejiang University
Department of Civil Engineering
Hangzhou 310058, China
e-mail: wenhaizhen@zju.edu.cn

Yao Liu
Zhejiang University
Department of Civil Engineering
Hangzhou 310058, China
e-mail: liuyao.aaaa@qq.com

***Abstract*—With the rapid development of the real estate market, housing prices at different price levels are significantly different from each other. Therefore, this paper selects the second-hand housing transaction data from 2007 to 2015 in Hangzhou, uses the hedonic price model, and introduces the quantile regression to compile the first quantile house price index in Hangzhou. Then we compare it with Ordinary Least Squares estimation to verify its feasibility and necessity. The empirical results show that the variation trend of the price index at different quantiles is the same, but the fluctuation range is significantly different. The rate of increase in housing prices from the high to low prices decline. This study can help the agencies compiling price index to expand their thinking, and provide a more scientific basis for government policy making, real estate agencies and buyers' decision-making.**

Keywords-hedonic price model; real estate price index; quantile regression model; residential property prices

I. INTRODUCTION

In the past decades, the real estate market has ushered in a period of vigorous development and plays an important role in economic development. The rise and fall of the real estate industry has a direct impact on the industrial structure and domestic demand, as well as the smooth operation of a country's economy, financial security and people's livelihood. As the main component of the real estate market, housing accounts for about 80% of real estate sales. House price as an important economic indicator provides important real estate market information for policy formulation, real estate agencies and buyers, so the compilation of real estate price index is of practical significance. Since 2004, policies have been introduced successively, and real estate information systems have been established in major cities. At the same time, the government and civil organizations have continued to improve the compilation of real estate price indexes to improve the accuracy of their monitoring [1]. However, different methods and models are used to construct price index, market information reflected by price indices is different. How to choose a price index compilation method that is suitable for the current situation of the real estate market in China has been the focus of attention of all sectors of society [2].

"NBS 70 Cities Index" and "NBS Average Price Index" have many problems. The full sample data of the NBS Average Price Index is obtained by surveying and sampling the commercial housing in the market after clustering and classification, resulting in a strong subjectivity in data sample selection. The NBS 70 Cities Index is calculated using a chained Laplace index method, without considering the particularity of real estate, and the impact of non-market factors is greater [3] .Therefore, many scholars have proposed different improvement methods based on the existing price index, such as two-phase chain update method, hedonic price index method, repeat sales regression method, machine learning method, etc., but these methods focus on the overall distribution of house prices, and do not distinguish between houses at different price levels. Therefore, many scholars have begun to use quantile regression to study houses with different price levels, and found that the characteristic variables have different degrees of impact on houses with different price levels. The research in this article will use Hangzhou quarterly second-hand housing transaction data from 2007 to 2015 to construct a hedonic price model to empirically study the prices of residential properties, while using quantile regression models to compare the prices of residential properties at different quantiles. Finally, we construct the first quarterly quantile price index of second-hand housing in Hangzhou from 2007 to 2015.

II. LITERATURE REVIEW

With the continuous improvement of the compilation of the price index system, the price index plays an important role in reflecting market trends, guiding the behavior of relevant decision-makers, and providing basic information. The development status of the real estate market in different countries is different, so the method of compiling the price index is also different. The research on price index at home and abroad mainly focuses on two aspects [4]. on the one hand, Research on the source of housing price data [5], on the other hand, selection of estimation methods [6,7]. The compilation methods of price indexes commonly used at home and abroad are the repeated sales regression model and the hedonic price model. This article will focus on the discussion of estimation methods.

The repeat sales regression model was first proposed in 1963 by Bailey, Muth, and Nourse [8]. It was used to independently compile real estate price indexes after 1980. Case and Shiller proposed a weighted repeat sales method for compiling price indexes in 1987 [9]. This method may depend on the time interval between sales. In practical applications, the real estate price index compiled by the

978-1-7281-6024-5/20/$31.00 ©2020 IEEE

repeated sales regression model requires a lot of transaction data [3]. Therefore, it is used more in countries which the real estate market is more mature, because its mortgage lenders can provide a large number of complete data [10]. But Chinese real estate market is relatively backward, and it is difficult to collect data on repeated transactions of houses. There are few cases of using the repeated transaction model to compile house price indexes. According to the complex status quo of Chinese real estate market, it is more appropriate to compile the house price index with a hedonic price model [2].

The hedonic price model originated in 1930s and 1940s, It is a common method for studying price changes in the real estate market. At present, the price index of most domestic real estate is based on the least square method (OLS) to perform linear regression to establish the price index. To some extent, it reflects the change in the implied average price, but ignores the long-term changes in the price index of houses with different prices. Least squares estimation is based on conditional mean values and susceptible to extreme values, especially it does not reflect the full distribution of house prices [11]. Quantile regression is able to show the change of the implied price of the explanatory variable at different price levels [13]. Therefore, a small number of scholars at home and abroad have begun to study the price index based on the quantile regression method.

Coulson and McMillen used the quantile regression model to construct price indices for three major cities in Chicago, compared the price indices of the three cities' 0.1, 0.5, and 0.8 quantiles, and then used a filtering model to study dynamic interaction between indicators. Residential supply only has a greater impact on high price housing [7]. McMillen used a hedonic price model to perform quantile regression on all home sales data in Cook County, Illinois from 2000 to 2011, and explored the impact of time factors on different quality homes in the same geographical location. The empirical results show that there is a significant difference in price index between different quality houses and locations. The original method of estimating the price index simplifies the influence of time and space [13]. Shi, Wang, and Gu used the quantile regression model to construct new house and second-hand house price index in Shanghai, then compared it with OLS regression. Quantile regression effectively distinguishes the price trends and fluctuations of different quality houses. The price index of new houses fluctuates more obviously than the price index of second-hand house [14]. Zhang lei and Yi Yiming constructed the quantile house price index for Beijing and compared it with OLS estimates. It was found that the quantile price index reflected more comprehensive information [3].

Therefore, using the quantile price index can reflect the full distribution of house prices and the change in characteristic prices at the quantile, which is more comprehensive than the price index reflection constructed by OLS regression. This article selects the second-hand housing data in Hangzhou from 2007 to 2015, we construct the first residential quantile price index in Hangzhou.

III. DATE AND MODEL

A. Data Sources

This paper selects the six main urban areas of Hangzhou as the empirical research areas, and the residential data comes from Hangzhou Real Estate Administration. The homes studied in this article are second-hand ordinary homes, so they do not include townhouses and villas. We collected the transaction data of second-hand housing in Hangzhou from January 2007 to December 2015, then eliminated some missing and duplicate data, and finally got 14,717 valid data samples.

B. Variable Selection

Based on literature analysis at home and abroad, the average transaction price of the residential district was selected as the explanatory variable. We used the hedonic price model and selected three attributes. There are one building characteristics, three location attributes, and six neighborhood attributes.

We added 36 quarterly dummy variables D to the hedonic price model, with the first quarter of 2007 as the base period.

C. Model

Through constant attempts and comparisons of functional forms, and reference to other scholars' research on the real estate price index. this article adopts a logarithmic form of hedonic price model.

$$ln\, p_{it} = \alpha + \beta' ln\, X_{it} + \gamma' ln\, Y_{it} + \delta' Z_{it} + \varphi' D_{it} + \varepsilon_{it} \quad (1)$$

where p_{it} is the sales price of the residential house i at time t; X_{it} is the building characteristics such as building age; Y_{it} includes location attributes; Z_{it} is neighborhood attributes such as natural environment; D_{it} represents a series of quarterly dummy variables; ε_{it} is error term; φ captures the price index from the second quarter to the t quarter(*Index*= *exp* (φ_t)),time 1 is used as the base period.

Using the quantile regression method to study the characteristic variables of residential prices, the quantile regression model constructed is as follows:

$$ln\, p_{it} = \alpha_\tau + \beta_\tau' ln\, X_{it} + \gamma_\tau' ln\, Y_{it} + \delta_\tau' Z_{it} + \varphi_\tau' D_{it} + \varepsilon_{\tau it} \quad (2)$$

where τ represents the quantile(e.g., τ =0.1, 0.2…, 0.9), Quantile regression studies the effects of explanatory variables on the position, size, and shape of the dependent variable distribution. The least square method uses symmetric regression, which needs to meet the assumptions of normality and homogeneity of variance.

TABLE I. DEFINITION OF VARIABLES

Variable	Definition	Expected Sign
Building age	The age of the building (year, transaction years minus actual built years	-
Distance to Wulin Square	Straight-line distance between the community center and Wulin Square (km)	-
Distance to West	Straight-line distance between the community	-

Lake	center and the coast of West Lake (km)	
Distance to Qianjiang New City	Straight-line distance between the community center and the Qianjiang New City (km)	-
Natural environment	Comprehensive score of the surrounding natural environment, divided into 5 levels: good (5 points), good (4 points), fair (3 points), poor (2 points), poor (1 point)	+
Cultural atmosphere	=1，If there is a university within 1km of the neighborhood；0, otherwise	+
Living facilities	Is there a hospital, park, bank, farmer's market and supermarket within 1km of the community center (1 point for each item, 5 points in total)	+
Sports facilities	The comprehensive score of the number and quality of community sports facilities is divided into 6 levels: good (5 points), good (4 points), fair (3 points), poor (2 points), poor (1 point), very Poor (0 points)	+
Education facilities	Is there a kindergarten, elementary school, junior high school or high school within 1km of the neighborhood (1 point for each item, 4 points in total)	+
Property management	The service level of residential properties is divided into 5 levels: good (5 points), good (4 points), fair (3 points), poor (2 points), poor (1 point)	+
D1 – D36	=1, if the house is sold in this quarter; 0, otherwise	?

IV. RESULT ANALYSIS

Regression was performed using a quantile regression model. In order to verify the quantile effect of the house, the change of the characteristic variable with the quantile is plotted in Fig. 1.

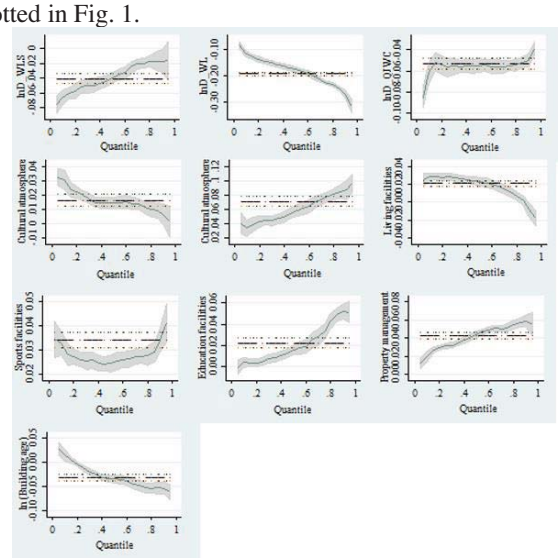

Figure 1. Coefficient Estimates by Quantile

A. Quantile Regression Coefficient Analysis

The coefficients for OLS and quantile estimates show:

1) Sign of the OLS regression coefficient in Table II is consistent with the expectation, and the significance of the characteristic variables is 1%, indicating that the 10 characteristic variables have a significant impact on the house. The R^2 of the model reached 0.7595, indicating that the model has a good fit and there are fewer unexplained variables.

2) Analysis of OLS estimation coefficients. As can be seen from Table II, residential prices decrease with the increase of distance to Wulin Square, distance to West Lake, and distance to Qianjiang New City. The comparison of the coefficients of the three regional characteristic variables found that people are more inclined to choose a house closer to the West Lake. With the rise of natural environment, cultural atmosphere, living facilities, sports facilities, educational facilities, and property management levels, housing prices have gradually risen.

3) The Pseudo R^2 of quantile regression is approximately in the range of 45% to 60%. Almost all characteristic variables in the model are significant. The coefficients of different quantile regression are quite different from those estimated by OLS, and there is also a significant difference in the impact of the same characteristic variable on different price levels of houses. It can be seen that OLS estimation cannot accurately reflect the impact of characteristic variables on housing at different price levels, and quantile regression can reflect the full distribution of characteristic variables in housing prices.

4) The location characteristics of residences have obvious differences in their impact on different price levels. The housing price decreases with the distance to Wulin Square. When the quantile of housing prices changes from a low quantile to a high quantile, the coefficient of distance to Wulin Square also changes from large to small, and the coefficient of distance to West Lake changes from small to large. It can be seen that the buyers of high-priced homes have higher requirements for the distance from their homes to the West Lake District than those of low-priced homes, and pay more attention to the living environment.

5) The neighborhood attributes of houses have different impacts on houses at different price levels. From the 0.1 quantile to the 0.9 quantile, the coefficients of natural environment, living facilities, and building age have shown a downward trend, and the coefficients of cultural atmosphere, educational facilities, and property management have shown an upward trend. It shows that high-priced buyers pay more attention to the cultural atmosphere, educational support and property management level than low-priced buyers when choosing a house.

B. Analysis of Index Results for Quantile Housing Price

The time dummy variables in the above results are all significant. According to the calculation formula of the house price index, the quarterly price index of Hangzhou's housing from 2007 to 2015 is calculated (see Fig. 2).

1) The comparison between the price index obtained by OLS regression and the price index obtained by quantile regression shows that the fluctuation trend of the curve is basically the same. Before 2010, the four curves almost coincided, and after 2010, the distance between the four curves gradually increased. From a numerical point of view, after 2010, the second-hand house price index calculated by using the quantile regression to obtain, the 0.8 quantile was the highest, and the second-hand house price index calculated at quantile 0.2 was the lowest. The price index at

quantile 0.5 is close to the fluctuation range of the OLS price index.

2) As can be seen from Fig. 2, the price index of second-hand housing in Hangzhou from 2007 to the second quarter of 2008 basically showed a slight upward trend. After the third quarter of 2008, there was a slight decline, and there was an obvious trough period, but this trough period was short-lived, and the residential price index rebounded sharply from the first quarter of 2009. Prior to the second quarter of 2010, the second-hand housing price index continued to rise, and has since entered a gradual period. The price index has maintained a gentle fluctuation, and it declined slightly in the third quarter of 2013.

3) Comparing the price index curve in Fig. 2 with the "NBS Average Price Index", it is found that the OLS estimated price index, quantile regression price index and China Real Estate Index show very similar trends, but the three house price index fluctuations are all different. The largest fluctuation is the quantile price index, and the most volatile is the China Housing Price Index curve. The quantile house price index is more in line with the changing circumstances of the real estate market.

Figure 2. Quantile House Price Index

TABLE II. RESULTS OF OLS AND QUANTILE REGRESSION

	OLS	*QR_10*	*QR_20*	*QR_30*	*QR_40*	*QR_50*	*QR_60*	*QR_70*	*QR_80*	*QR_90*
Ln (Distance to Wulin Square)	-0.0406***	-0.0665***	-0.0578***	-0.0508***	-0.0467***	-0.0393***	-0.0329***	-0.0212***	-0.0177***	-0.0172***
	0.0034	0.0059	0.0042	0.0037	0.0034	0.0036	0.0038	0.0038	0.0045	0.0062
ln (Distance to West Lake)	-0.1910***	-0.1160***	-0.1370***	-0.1510***	-0.1620***	-0.1750***	-0.1900***	-0.2150***	-0.2330***	-0.2680***
	0.0034	0.0059	0.0042	0.0037	0.0034	0.0036	0.0038	0.0038	0.0045	0.0062
ln (Distance to Qianjiang New City)	-0.0527***	-0.0617***	-0.0546***	-0.0555***	-0.0550***	-0.0551***	-0.0558***	-0.0564***	-0.0545***	-0.0498***
	0.0027	0.0046	0.0033	0.0029	0.0027	0.0028	0.003	0.003	0.0035	0.0048
Natural environment	0.01650***	0.0306***	0.0219***	0.0176***	0.0146***	0.0147***	0.0152***	0.0139***	0.0101***	0.0066*
	0.0021	0.0036	0.0026	0.0023	0.0021	0.0022	0.0023	0.0023	0.0027	0.0038
Cultural atmosphere	0.0709***	0.0336***	0.0434***	0.0444***	0.0496***	0.0589***	0.0653***	0.0738***	0.0810***	0.0876***
	0.0034	0.0059	0.0042	0.0038	0.0034	0.0036	0.0039	0.0038	0.0045	0.0062
Living facilities	0.0209***	0.0287***	0.0275***	0.0272***	0.0250***	0.0223***	0.0201***	0.0141***	0.0068***	-0.0081***
	0.0017	0.003	0.0021	0.0019	0.0017	0.0018	0.0019	0.0019	0.0023	0.0031
Sports facilities	0.0343***	0.0327***	0.0273***	0.0253***	0.0245***	0.0244***	0.0260***	0.0274***	0.0278***	0.0340***
	0.0016	0.0027	0.0019	0.0017	0.0016	0.0017	0.0018	0.0018	0.0021	0.0029
Education facilities	0.0220***	0.0036	0.0033	0.0070***	0.0107***	0.0158***	0.0189***	0.0278***	0.0423***	0.0522***
	0.0024	0.0042	0.003	0.0026	0.0024	0.0025	0.0027	0.0027	0.0031	0.0043
Property management	0.0427***	0.0196***	0.0291***	0.0322***	0.0380***	0.0435***	0.0492***	0.0504***	0.0544***	0.0587***
	0.0017	0.0029	0.0021	0.0019	0.0017	0.0018	0.0019	0.0019	0.0022	0.0031
ln (Building age)	-0.0315***	0.0158***	-0.0043	-0.0199***	-0.0308***	-0.0347***	-0.0391***	-0.0480***	-0.0532***	-0.0531***
	0.0033	0.0057	0.004	0.0036	0.0033	0.0034	0.0037	0.0037	0.0043	0.0059
_cons	9.157***	8.887***	9.017***	9.098***	9.1470***	9.169***	9.193***	9.271***	9.333***	9.468***
	0.0204	0.0353	0.025	0.0222	0.0204	0.0213	0.023	0.0228	0.0267	0.0368
R^2 or pseudo R^2	0.7595	0.581	0.5971	0.5881	0.5661	0.5419	0.5193	0.501	0.4823	0.4567

***, **, and * represent the 1%, 5%, and 10% significance levels, respectively.

V. CONCLUSION

This article selects the second-hand housing transaction data of Hangzhou from 2007 to 2015. Based on the hedonic price model, it uses the OLS and quantile regression methods to construct the Hangzhou residential price index. The empirical results show that the characteristic variables have different impacts on different quality homes; the housing price index changes at different quantiles have the same fluctuation trend, and the fluctuation ranges are different; the quantile house price index and the OLS housing price index and "NBS Average Price Index" Compared, we find that the quantile-regressive residential price index is more suitable for the changing real estate market. The above empirical results fully demonstrate the superiority and applicability of using the quantile to construct a price index.

The research results of this article have practical guiding significance and can provide scientific decision-making for government policy formulation, real estate agencies and home buyers. Here are some suggestions and deficiencies for

the construction of house price index. The housing price index at different levels should be constructed separately; the government's regulatory policies can be based on the impact of different housing price. Statistical departments should supplement the required data and provide technical support to further expand data sources and control data quality.

ACKNOWLEDGEMENT

This study is supported by the Natural Science Foundation of China (71974169).

REFERENCES

[1] J. Wu, H.Y.Liu, and Y.N.Ma, "the main compilation method and choice of housing price index,"Construction Economy, vol.7, pp.2730, 2007. doi:10.14181/j.cnki.1002-851x.2007.07.007.

[2] J.Wu, Y.H.Deng,and H.Y. Liu, "House price index construction in the nascent housing market: The case of China,"The Journal of Real Estate Finance and Economics ,vol.48, pp.522-545, 2014.

[3] L.Zhang, Y.M.Yi," Quantile house price indices in Beijing,"Regional Science and Urban Economics,vol.63, pp.85-96, 2017.

[4] C.Shimizu, K. G.Nishimura, and T. Watanabe," House prices at different stages of the buying/selling process,"Regional Science and Urban Economics,vol.59, pp.37-53, 2016.

[5] E.Anenberg, S.Laufer, " A more timely house price index,"Review of Economics and Statistics,vol.99, pp.722-734, 2017.

[6] T.Bollerslev, A.J.Patton,and W.Wang,"Daily house price indices: Construction, modeling, and longer-run predictions, " Journal of Applied Econometrics ,vol.31, pp.1005-1025, 2016.

[7] N.E.Coulson , D.P. McMillen, "The dynamics of intraurban quantile house price indexes,"Urban Study,vol.44, pp.1517–1537, 2007.

[8] J.Wu, M.J.Bailey, R. F. Muth, and H.O. Nourse,"A Regression Method of Real Estate Price Index Construction,"Publications of the American Statistical Association,vol.58, pp.933-942, 1963.

[9] K.E.Case, R.J.Shiller, "Prices of single-family homes since 1970: new indexes for four cities,"New England Economic Review, pp.4556, September 1987.

[10] Y.Deng, D.P.Mcmillen, and T.F.Sing, "Matching indices for thinlytraded commercial real estate in Singapore,"Regional Science and Urban Economics, vol.47,pp.86-98, 2014.

[11] H.G.Kim, K.C. Hung, and S.Y.Park,"Determinants of Housing Prices in Hong Kong: A Box-Cox Quantile Regression Approach," The Journal of Real Estate Finance and Economics, vol.50, pp.270287, 2015.

[12] R.T.Mora-Garcia, M.F.Cespedes-Lopez, V.R. Perez-Sanchez, P. Marti, and J.C.Perez-Sanchez ," Determinants of the price of housing in the province of Alicante (Spain):Analysis using quantile regression,"Sustainability ,vol.11, pp.437, January 2019.

[13] D.P.McMillen,"Local quantile house price indices,"Journal of Urban Economics,2013.

[14] W.Shi, H.W.Wang,and Q.D.Gu,"Constructing Housing Price Index Using Quantile Regression Feature Price Model," Statistics and Decision, vol.4,pp.9-12,2015.

[15] A.Elliot, L.Steven, "A more timely house price index,"Review of Economics and Statistics,vol.99, pp.722-734, 2017.

[16] H.Fang, Q.L.Gu, W.Xiong, and Li. A. Zhou, "Demystifying the Chinese housing boom,"NBER Macroeconomics Annual,vol.30,pp. 105-166. Novermber 2016.