

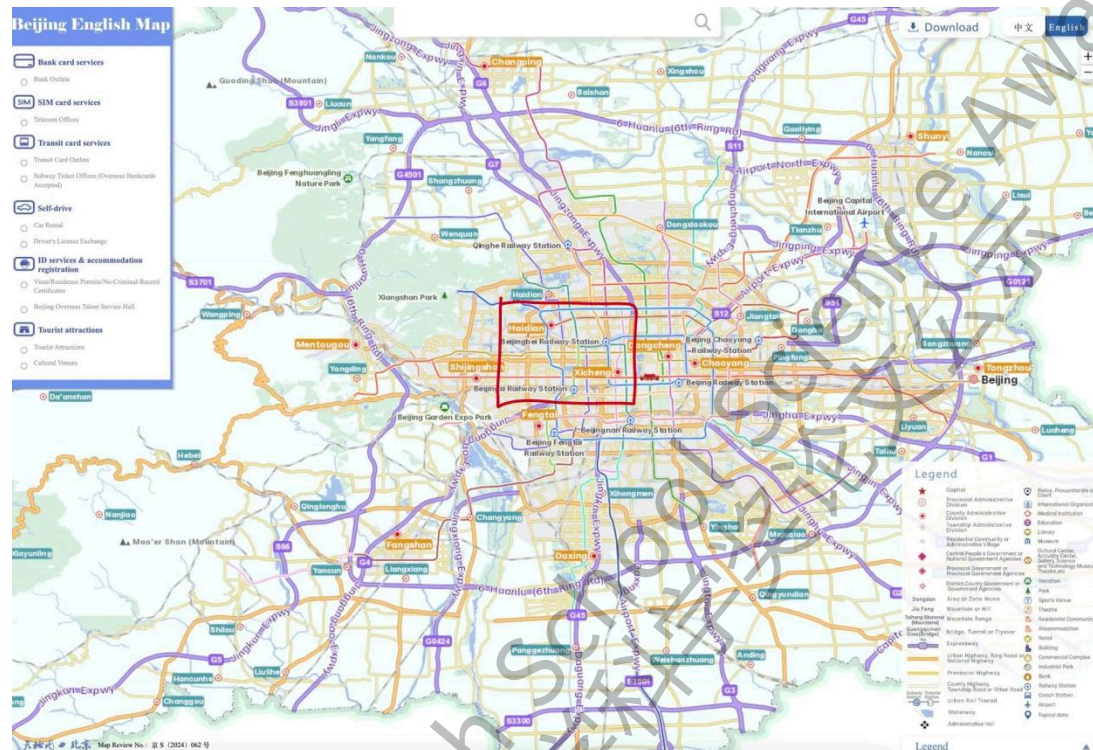
Spatial Decay and Heterogeneity of School District Housing Premiums in Xicheng and Haidian, Beijing

胡元之

Abstract

The two main school areas of Beijing are Xicheng and Haidian, and this project is focused on them. I collected second-hand house data from the Lianjia website and made several graphs and regression models to learn about the changes in housing prices in these two areas. I created basic bar chart, box-and-whisker plot, scatter plot and correlation heat map to observe the raw relationship between a single factor and unit price. Then I constructed several regression models, which are ordinary stepwise regression and a final multilevel mixed model. I found that there are some similarities and differences in the reasons for the unit price difference between the two districts. Generally, the price per square meter in Xicheng is relatively high, and the price of large houses in this area has been increasing at a faster rate. However, from the research, I found that housing prices in Haidian are mainly influenced by the quality of primary school teachers and are less dependent on the number of living rooms or proximity to commercial areas. What's more, the scatter plot of the floor area in Xicheng shows that it is positively correlated with housing prices.

This research clarifies its scope and provides relevant advice for school-district property buyers and urban planners according to the empirical outcomes.



0-1 Beijing map

1. Introduction

Many Beijing families are using large amounts of capital to buy homes solely for the purpose of giving their children good primary schools. Xicheng and Haidian are the two districts in the city with the most educational resources, so their housing prices are relatively higher than in other areas. Although there is a difference in the housing price mechanism of the two areas, it has not been studied in recent years, and many relevant data have not been graphically presented or used for regression analysis.

The first three are my research questions. First, what factors will change the prices of houses in Xicheng and Haidian independently? Second,

which district of these two does the school price premium fall more rapidly with the distance from the primary school? Thirdly, how are the house size, school facilities and other facilities in the two areas related to price differences?

I also have three hypotheses before starting to process the data. The decrease in housing prices due to distance from primary schools in Xicheng is more pronounced than in Haidian. The reason for the high house prices in Haidian is educational resources, and commercial and transportation facilities are relatively more important in Xicheng. The two districts have different base housing prices and different price effects of property features.

The two core contributions of this paper are as follows. Most of the existing academic literature has treated Beijing's entire housing market as a unified sample and rarely adopted mixed-effects models to capture inter-district heterogeneity. My work fills this gap by showing some pictures of Xicheng and Haidian. Families that are planning to purchase school houses for their children will know the various market rules and make more informed decisions. City managers can use the data I have to distribute schools and public living facilities more evenly in central Beijing.

Most older articles have only presented the regression table and failed to include a graph of the price changes. Therefore, I will use Python to draw heat maps, box plots and slope comparison scatter plots to show the changes in the data directly. Stepwise regression and a multilevel mixed model are also employed to deal with the problem of bias caused by different districts.

2. Related Work

2.1 Basic Theory: Hedonic Price Model

Rosen put forward the hedonic price model in 1974. The main idea is that a house's total price consists of three types of hidden values: building features, school resources and location support facilities. All the above attributes will affect the final unit cost. The above model is the basis of my data analysis in this paper, and all regression and graph explanations are based on this theory.

The general mathematical expression of hedonic price model is set as follows:

$$P = \alpha_0 + \sum_{i=1}^n \alpha_i X_i + \varepsilon$$

Where:

P = housing unit price (yuan/m²);

α_0 = constant term;

X_i = each characteristic variable of housing (building attribute, education resource, location facility);

α_i = marginal implicit price corresponding to each characteristic variable;

ε = random error term.

2.2 Past Research on Beijing School Housing

Many Chinese scholars have already researched the housing prices in Beijing. Good primary schools, nearby subway stations and large shopping malls will increase the local housing prices. All studies have shown that there is a school district premium in central Beijing.

However, the existing literature has not distinguished between the different pricing mechanisms in the two districts. Another problem in most of the research is the excessive use of simple linear regression models. A layered mixed model is not used to solve the problem of different natural price differences in administrative areas. Add some

scatter plots and heat maps to show that the school premium decreases with distance in a few papers.

2.3 How My Research Corrects the Shortcomings in Previous Studies

I collect the data for Xicheng and Haidian separately from the beginning of data collection. Draw a comparison graph of all the variables to show the different trends of the two districts. Finally, the multilevel mixed model has been added to correct for the bias caused by the old simple regression. I also use a heat map to show the correlation among all the variables, so now I can more easily see how house size, school scores and prices are related. All of the above will provide richer material for this paper.

3. Data-set and Methodology

3.1 Data Collection

All the raw data I collected come from the official website of Lianjia, a large second-hand home platform in China. I copy the housing information in Xicheng and Haidian Districts one by one and organize all the data in an Excel sheet. Finally, I have 200 valid housing records for analysis. After sorting the data, I divided all the variables into three groups. The first group is building attributes: floor area, year of construction, number of bedrooms, number of living rooms. The second group is the education-related variables: primary school quality score, teacher quality score, annual student enrollment in primary school, distance to primary school. The third group is Location Facilities: Distance to the subway station, distance to the nearest shopping mall, etc. All the

variables have a clear numerical value for graph drawing and regression.

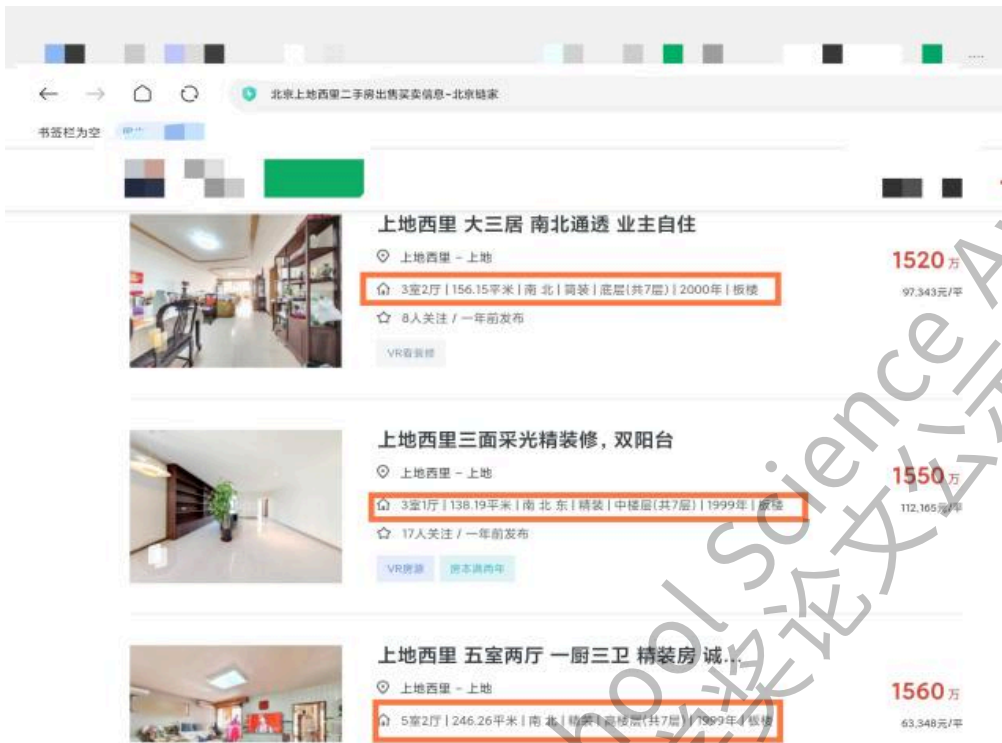


Figure 3-1 Screenshot of second-hand housing data source on Lianjia Platform

3.1.1 Primary School Evaluation Scoring Mechanism

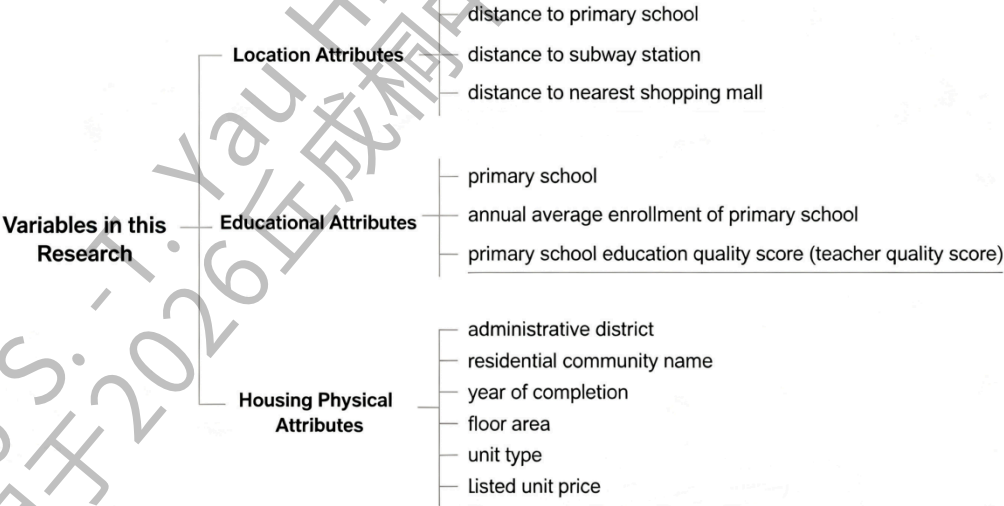


Figure 3-2 Classification of all variables adopted in this hedonic price model

Two core scoring systems are constructed to quantify educational resource quality, all scores range from 0 to 100, and the weighted total

score is taken as the continuous educational characteristic variable in regression analysis:

Primary School Education Quality Score Scoring Rules (Full score 100):

1. District Exam Ranking, weight 25%

Top 10% = 100 points; Top 10% - 30% = 75 points; Top 30% - 60% = 45 points; Bottom 40% = 15 points

2. Academic Value-Added, weight 15%

Significant improvement = 100 points; Slight improvement = 70 points; Stable performance = 40 points; Performance decline = 10 points

3. Direct Admission Ratio to Quality Junior High, weight 15%

Ratio $\geq 60\%$ = 100 points; $40\% \leq \text{ratio} < 60\%$ = 70 points; $20\% \leq \text{ratio} < 40\%$ = 40 points; Ratio $< 20\%$ = 10 points

4. Admission Certainty, weight 10%

Clear direct admission policy = 100 points; Multi-district allocation with high-quality schools = 70 points; Multi-district allocation with ordinary schools = 40 points; Pure random lottery = 10 points

5. Competition Awards per Student, weight 12%

Index ≥ 0.3 = 100 points; $0.15 \leq \text{index} < 0.3$ = 70 points; $0.05 \leq \text{index} < 0.15$ = 40 points; Index below 0.05 = 10 points

6. Comprehensive Quality Achievement Rate, weight 8%

Rate $\geq 95\%$ = 100 points; $85\% \leq \text{rate} < 95\%$ = 70 points; $70\% \leq \text{rate} < 85\%$ = 40 points; Rate $< 70\%$ = 10 points

7. Parent Satisfaction (5-point scale), weight 15%

Score ≥ 4.5 = 100 points; $4.0 \leq \text{score} < 4.5$ = 70 points; $3.5 \leq \text{score} < 4.0$ = 40 points; Score < 3.5 = 15 points

School	District Exam Ranking	Ranking Score	Academic Value-Added Score	Direct Admission Ratio	Admission Ratio	Admission Score Certainty	Certainty Score	Competition Awards per Student	Competition Score	Comprehensive Quality Rate	Quality Rate Score	Parent Satisfaction (5-pt)	Satisfaction Score	Overall Education Quality Score
Hongmiao Primary School	Top 10%	100	Slight improvement 70	80%	100	Clear direct admission	100	0.05-0.15	40	95%+	100	4.5+	100	86.5
Zhongguancun No.1 Primary School (Main Campus)	Top 10%	100	Slight improvement 70	None	10	Multi-district with good options	70	0.3+	100	95%+	100	4.5+	100	79.5
Zhongguancun No.3 Primary School (Wanliu Campus)	Top 10%	100	Slight improvement 70	None	10	Multi-district with good options	70	0.3+	100	95%+	100	4.5+	100	79.5
Sanfan Middle School Affiliated Primary School	Top 10%-30%	75	Significant improvement 100	80%	100	Clear direct admission	100	Below 0.05	10	95%+	100	4.0-4.5	70	79.0
Zhongguancun No.2 Primary School (Main Campus)	Top 10%	100	Slight improvement 70	None	10	Multi-district with good options	70	0.15-0.3	70	95%+	100	4.5+	100	76.0
The Main Campus of Beijing No.2 Experimental Primary School	Top 10%	100	Slight improvement 70	None	10	Multi-district with good options	70	0.15-0.3	70	95%+	100	4.5+	100	73.0
Shangdi Experimental Primary School	Top 10%-30%	75	Slight improvement 70	None	10	Multi-district with good options	70	0.05-0.15	40	95%+	100	4.5+	100	67.0
Wulutong Primary School	Top 30%-60%	45	Slight improvement 70	None	10	Multi-district with good options	70	Below 0.05	10	95%+	100	4.0-4.5	70	48.75

Indicator	Weight	Scoring Rules
District Exam Ranking	25%	Top 10% = 100; Top 10% - 30% = 75; Top 30% - 60% = 45; Bottom 40% = 15 Significant improvement = 100; Slight improvement = 70; Stable = 40; Decline = 10 ≥ 60% = 100; 40% - 60% = 70; 20% - 40% = 40; <20% = 10 Clear direct admission = 100; Multi-district with good options = 70; Multi-district with average options = 40;
Academic Value-Added	15%	
Direct Admission Ratio to Quality Junior High	15%	
Admission Certainty	10%	

		Pure lottery = 10 $\geq 0.3 = 100$; $0.15 - 0.3 =$ 70 ; $0.05 -$ $0.15 = 40$; Almost none = 10 $\geq 95\% = 100$; $85\% - 95\% =$ 70 ; $70\% -$ $85\% = 40$; $< 70\% = 10$ $\geq 4.5 = 100$; $4.0 - 4.5 =$ 70 ; $3.5 - 4.0$ $= 40$; $< 3.5 =$ 15
Competition Awards per Student	12%	
Comprehensive Quality Achievement Rate	8%	
Parent Satisfaction (5 point scale)	15%	

Table 3-3 Score of eight schools and scoring rules for Primary School Comprehensive Education Quality Score

Teacher Quality Score Scoring Rules (Full score 100):

1. Student - Teacher Ratio, weight 12%

Ratio $\leq 1:15 = 100$ points; $1:15 < \text{ratio} \leq 1:22 = 70$ points; $1:22 < \text{ratio} \leq 1:30 = 40$ points; Ratio $> 1:30 = 10$ points

2. Postgraduate (Master/PhD) Faculty Proportion, weight 9%

Proportion $\geq 30\% = 100$ points; $20\% \leq \text{proportion} < 30\% = 75$ points; $10\% \leq \text{proportion} < 20\% = 50$ points; Proportion $< 10\% = 20$ points

3. Senior/Intermediate Title Ratio, weight 9%

Proportion $\geq 60\% = 100$ points; $40\% \leq \text{proportion} < 60\% = 70$ points; $20\% \leq \text{proportion} < 40\% = 40$ points; Proportion $< 20\% = 10$ points

4. Core Teacher Proportion (District Outstanding Teachers & Subject Leaders), weight 25%

Proportion $\geq 20\% = 100$ points; $10\% \leq \text{proportion} < 20\% = 70$ points; $5\% \leq \text{proportion} < 10\% = 40$ points; Proportion $< 5\% = 15$ points

5. Distinguished/Senior Teacher Count, weight 15%

Quantity ≥ 3 = 100 points; $1 \leq \text{quantity} \leq 2$ = 70 points; Quantity = 0 = 40 points

6. Average Teaching Seniority (Years), weight 18%

$8 \leq \text{years} \leq 15$ = 100 points; $5 \leq \text{years} < 8$ or $15 < \text{years} \leq 20$ = 70 points; $2 \leq \text{years} < 5$ = 40 points; Years < 2 or years > 20 = 20 points

7. Teacher Stability (Turnover Rate), weight 12%

Rate $< 5\%$ = 100 points; $5\% \leq \text{rate} < 10\%$ = 70 points; $10\% \leq \text{rate} < 20\%$ = 40 points; Rate $> 20\%$ = 10 points

School	Student-Teacher Ratio	Student-Teacher Score	Postgraduate Faculty Proportion	Postgraduate Proportion Score	Senior/Intermediate Title Ratio	Title Ratio Score	Core Teacher Proportion	Core Teacher Score	Distinguished/Senior Teachers Count	Distinguished/Senior Count Score	Average Teaching Seniority (Years)	Average Seniority Score	Teacher Stability (Turnover Rate)	Stability Score	Overall Faculty Strength Score
Wulutong Primary School	01:13.8	100	15%	50	66%	100	20.00%	100	0	40	8	100	5%	70	82.2
Zhongguancun No.2 Primary School (Main Campus)	01:16.7	70	25%	75	57%	70	32%	100	1	70	10	100	5%	70	79.8
Hongmiao Primary School	01:13.0	100	15%	50	17%	40	26.00%	100	1	70	10	100	5%	70	77.0
The Main Campus of Beijing No.2 Experimental Primary School	01:11.3	100	16%	50	13%	10	30.20%	100	5	100	12	100	3%	100	76.4
Zhongguancun No.1 Primary School (Main Campus)	01:16.7	70	25%	75	1%	10	33%	100	1	70	10	100	5%	70	76.1
Shangdi Experimental Primary School	01:20.8	40	10%	50	45%	70	43%	100	0	40	8	100	3%	100	71.5
Sanfan Middle School Affiliated Primary School	01:13.3	100	10%	50	20%	40	22%	100	0	40	5	70	10%	70	71.5
Zhongguancun No.3 Primary School (Wanliu Campus)	01:21.2	40	10%	50	1.30%	10	38.20%	100	1	70	10	100	5%	70	63.3

Indicator	Weight	Scoring Rules
Student-Teacher Ratio	12%	$\leq 1:15 = 100$ points; $1:15 - 1:22 = 70$; $1:22 - 1:30 = 40$; $> 1:30 = 10$ $\geq 30\% = 100$;
Postgraduate (Master/PhD) Faculty Proportion	9%	$20\% - 30\% = 75$; $10\% - 20\% = 50$; $< 10\% = 20$ $\geq 60\% = 100$;
Senior/Intermediate Title Ratio	9%	$40\% - 60\% = 70$; $20\% - 40\% = 40$; $< 20\% = 10$ $\geq 20\% = 100$;
Core Teacher (District-level Outstanding Teachers & Subject Leaders) Proportion	25%	$10\% - 20\% = 70$; $5\% - 10\% = 40$; $< 5\% = 15$ $\geq 3 = 100$; $1 - 2 = 70$; $0 = 40$
Distinguished/Senior (Special-grade/Full Professor-level) Teachers Count	15%	$8 - 15$ years = 100; $5 - 8$ or $15 - 20 = 70$; $2 - 5 = 40$; < 2 or $> 20 = 20$
Average Teaching Seniority (Years)	18%	$< 5\% = 100$; $5\% - 10\% = 70$; $10\% - 20\% = 40$; $> 20\% = 10$
Teacher Stability (Turnover Rate)	12%	

Table 3-4 Score of eight schools and scoring rules for Teacher Comprehensive Quality Score

Final table:

Serial Number	Administrative District	Residential Community Name	Year of Completion	Floor Area (m ²)	Floor Level	Unit Type	Unit Price (yuan/m ²)	Site Price (yuan/m ²)	Zoned Primary School	Distance to Primary School (m)	Distance to Subway Station (m)	Distance to Nearest Shopping Mall (m)	Annual Average Enrollment of Primary School	Primary School Education Quality Score	Teacher Quality Score
1	Xicheng	Yuzhong Xili	1990	72.4	High	2B1L	99,310		Sanfan Middle School Affiliated Primary School	302	697	1419	240	71.5	79.0
2	Xicheng	Yuzhong Xili	1989	44.8	High	1B1L	169,643		Sanfan Middle School Affiliated Primary School	302	697	1419	240	71.5	79.0
3	Xicheng	Yuzhong Xili	1980	52.7	Low	2B1L	144,213		Sanfan Middle School Affiliated Primary School	302	697	1419	240	71.5	79.0

Table 3-5 Apart of the data from the 200 houses of eight primary schools in the sample

Variable	Mean	Xicheng Mean	Haidian Mean	Median	Xicheng Median	Haidian Median	Max	Xicheng Max	Haidian Max	Min	Xicheng Min	Haidian Min	Standard Deviation	Xicheng Standard Deviation	Haidian Standard Deviation
Unit Price/yuan	128,756	141,972	115,540	127,179.5	146,044.5	116,713.5	305,371	305,371	259,323	79,749	104,807	79,749	50,027.95457	60,564.23052	31,780.07557
Floor Area/m ²	93.61015152	81.62010101	105.600202	74.15	66.1	88.2	402.21	252.1	402.21	33.5	38.68	33.5	55.05537645	43.03127649	62.85675146
Distance to School	477.75	491.56	463.94	389	431	369	909	909	751	116	116	180	243.0015768	277.4662167	203.2811071
Distance to Subway Station	662.13	643.36	680.9	690	690	609	1062	902	1062	371	371	464	164.3021273	159.8418107	167.3377097
Distance to the Mall	794.17	836.58	751.76	698	572	744	1635	1625	1021	333	333	416	360.2599579	471.9683515	185.7357806
Primary School Education Quality Score	74.6975	76.72	72.675	76.1	76.4	73.8	82.2	82.2	79.5	63.3	71.5	63.3	5.525613611	3.845054544	6.190729057
Teacher Quality Score	73.61875	71.7375	75.5	77.5	76	77.75	86.5	86.5	79.5	48.75	48.75	67	10.79093382	14.16075523	5.137011669

Table 3-6 Descriptive Statistics

This table summarizes the key variables. The mean unit price is about 128,756 RMB/m², but the standard deviation (50,028) is substantial—this matches the box plot patterns and confirms that the two districts are not on the same baseline. Most properties lie within 478 meters of a primary school, and the maximum distance is only 909 meters, which explains why distance itself might not show a strong price effect later: almost everyone is already within walking range. Teacher quality scores vary noticeably, with a minimum of 48.75 and a maximum of 86.50, suggesting that this variable has enough spread to drive price differences—especially in Haidian, as the scatter plots hinted.

3.2 Statistical Models

There are three main tools that help me finish all the analysis work. Excel

is where the initial data will be kept. Python can draw all kinds of charts, such as bar charts, box plots, scatter plot comparison charts and correlation heat maps. Stepwise regression and multilevel mixed models are used in R language to obtain the coefficient and significance results.

3.2.1 Formula Setting of Four Regression Models

Significance standard: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Model 1: Ordinary Linear Stepwise Regression (Raw unit price as dependent variable)

$$Price = \beta_0 + \beta_1 Year + \beta_2 DisSchool + \beta_3 DisSub + \beta_4 DisMall + \beta_5 ScoreSchool + \mu$$

Model 2: Logarithmic Transformation Stepwise Regression

$$\ln(Price) = \gamma_0 + \gamma_1 Year + \gamma_2 DisSchool + \gamma_3 DisSub + \gamma_4 ScoreSchool + \sigma$$

Model 3: Logarithmic Regression with Interaction Term

$$\ln(Price) = \delta_0 + \delta_1 Year + \delta_2 DisSchool + \delta_3 DisSub + \delta_4 ScoreSchool + \delta_5 (ScoreSchool \times DisSchool) + \tau$$

Model 4: Multilevel Mixed-Effects Model (Final benchmark model, district random intercept added)

Level 1 (Single housing sample):

$$\ln(Price_{ij}) = \theta_{0j} + \theta_1 Year_{ij} + \theta_2 DisSchool_{ij} + \theta_3 DisSub_{ij} + \theta_4 DisMall_{ij} + \epsilon_{ij}$$

Level 2 (District level random intercept):

$$\theta_{0j} = \varphi_0 + \omega_j$$

Variable definition:

i = individual housing sample; j = administrative district (Xicheng / Haidian);

θ_{0j} = district-specific random intercept; φ_0 = overall fixed intercept;

ω_j = random disturbance reflecting district baseline price difference;

ϵ_{ij} = residual error of single housing sample.

I built the regression models and test the data in steps. Model 1 is ordinary linear regression of the raw housing unit price. Stepwise regression method will automatically exclude floor area because it cannot explain the variation in housing prices well.

In Model 2, I used log-transformed housing prices to handle skewness within the original data set. This model does not have a floor area and is far from shopping malls.

Model 3 adds an interaction term of school quality scores and the distance to primary schools. The combined variable is excluded in the first round of calculation, so it can be concluded that school quality and school distance do not jointly affect housing prices.

The adjusted R^2 values of Model 2 and Model 3 are both 0.2474; this is lower than that of Model 1, which was 0.2713.

Model 4 is the final multilevel mixed-effects model. A district random intercept with a variance of 0.0324 was added. It is not zero; therefore, there is a stable base price difference between Xicheng and Haidian that cannot be accounted for by the basic regression model.

Four predictors remain after the initial variable selection: year of completion, distance to primary schools, subway stations and shopping malls. Floor area, primary school quality score and the interaction term are excluded because they lack statistical significance.

4. Results

4.1 Preliminary Analysis

Several kinds of graphs are used in this chapter to present data, such as bar charts, box plots, scatter plots and heatmaps, and relationships among various factors are also analyzed. Split all data into Xicheng District and Haidian District for comparison.

4.1.1 Basic Index Comparison of the Two Districts

Bar charts are used to compare the average indicators of Xicheng and Haidian.

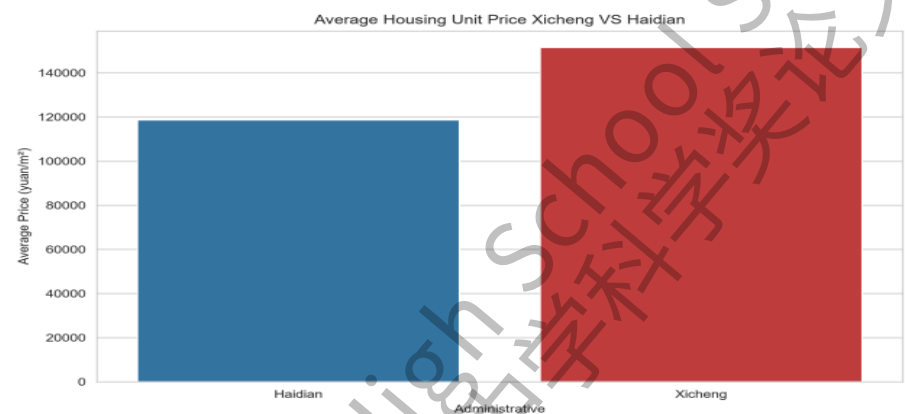


Figure 4-1 Average housing unit price comparison: Xicheng vs Haidian District

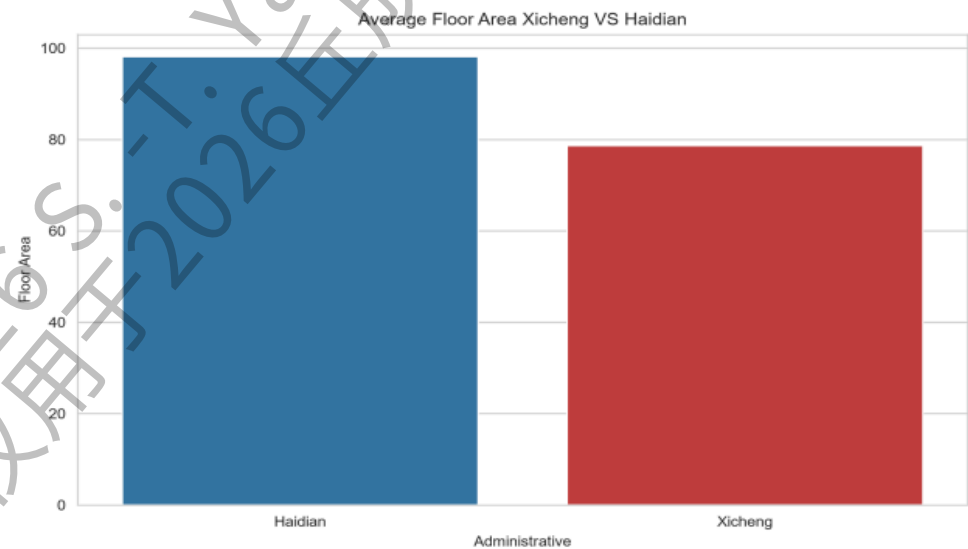


Figure 4-2 Average floor area comparison: Xicheng vs Haidian

District

As shown in the two bar charts, they are not the same. Xicheng's average price per square metre is relatively high compared with Haidian's, but Haidian has larger houses on average. This shows that people buying Xicheng houses are mainly for schools, whereas Haidian buyers are concerned with both school quality and sufficient living space.

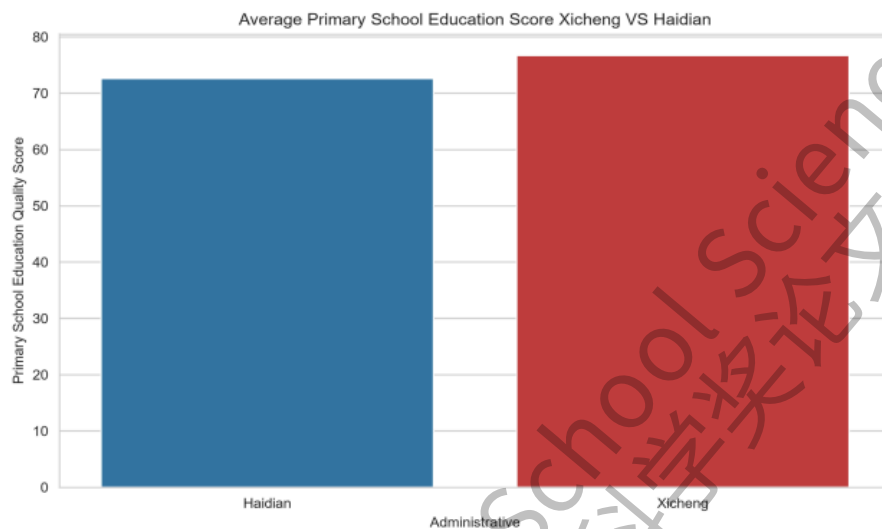


Figure 4-3 Average primary school education score comparison: Xicheng vs Haidian District

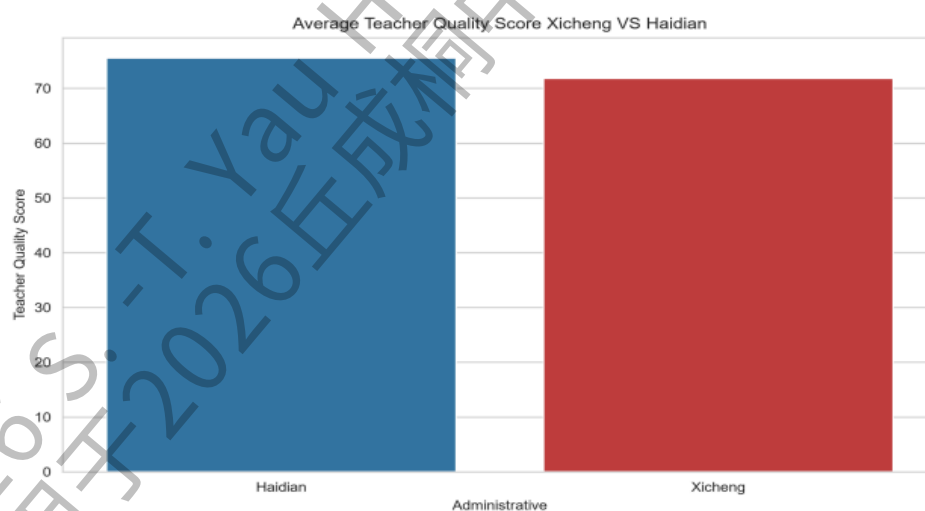


Figure 4-4 Average teacher quality score comparison: Xicheng vs Haidian District

Xicheng has slightly higher overall primary school scores, but Haidian has a higher average teacher quality rating. Therefore, buyers in Haidian

District pay more attention to the teaching quality of primary schools.

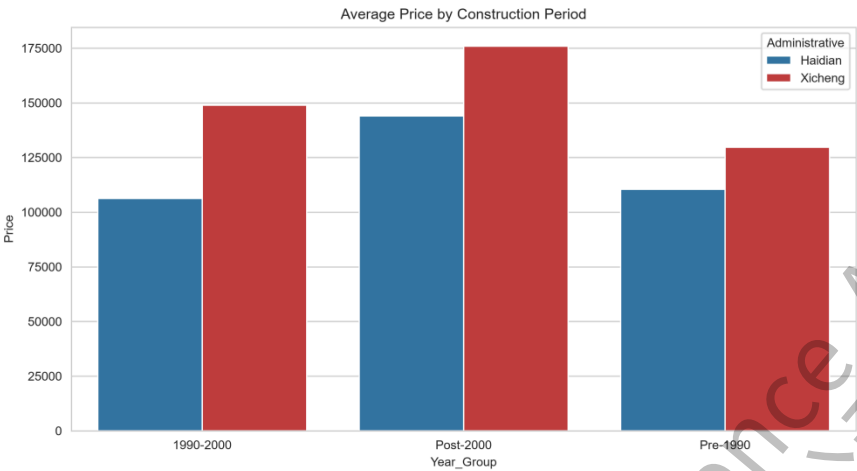


Figure 4-5 Average unit price grouped by construction year cohorts

Based on the construction time of the samples, housing in Xicheng is more expensive than similar properties in Haidian for all age groups. Buildings built after 2000 have a much larger price difference. Pre-1990 old buildings reduce the school premium effect, and cross-district price differences are thus less pronounced.

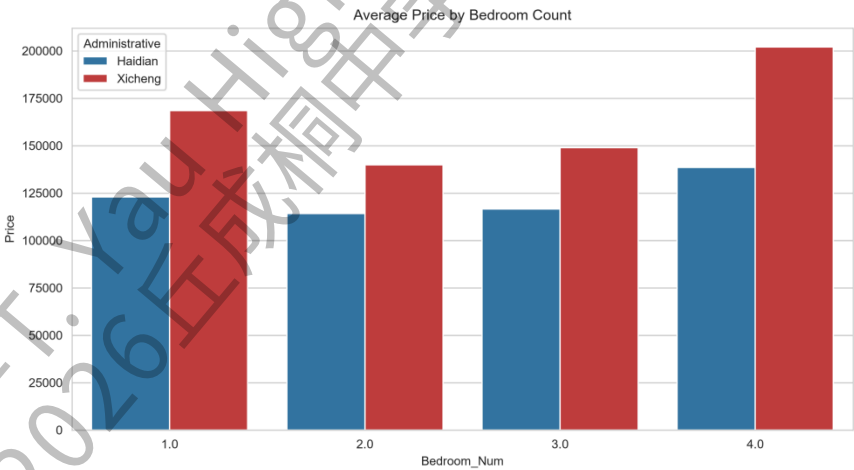


Figure 4-6 Average unit price grouped by number of bedrooms

Group the data by the number of bedrooms and find that four-bedroom houses in Xicheng are relatively inexpensive. Buyers in Xicheng want to buy a large house for their family and keep their school enrollment rights. Thus, the school premium is relatively higher in the big-house area here.

Unit prices in Haidian do not change significantly with the number of bedrooms. Most buyers only want a small apartment for school quota and do not plan to live there long-term.

In short, the two districts have very different housing needs. Xicheng buyers balance living needs and school resources; Haidian buyers are mainly concerned about school qualifications.

4.1.2 Housing Price Distribution Shown by Box Plots

Box plots can display the complete distribution, median and extreme values of housing prices.



Figure 4-7 Box plot of housing unit price distribution in two districts

The median price per square metre of a house in Xicheng is around 140,000 yuan; that in Haidian is about 110,000 yuan. Most of the housing samples in Xicheng are above the main price range of Haidian. Xicheng also has more high-price outliers, which are newly built large apartments with top-tier school resources. Haidian's price data is more concentrated and has significantly fewer outliers.

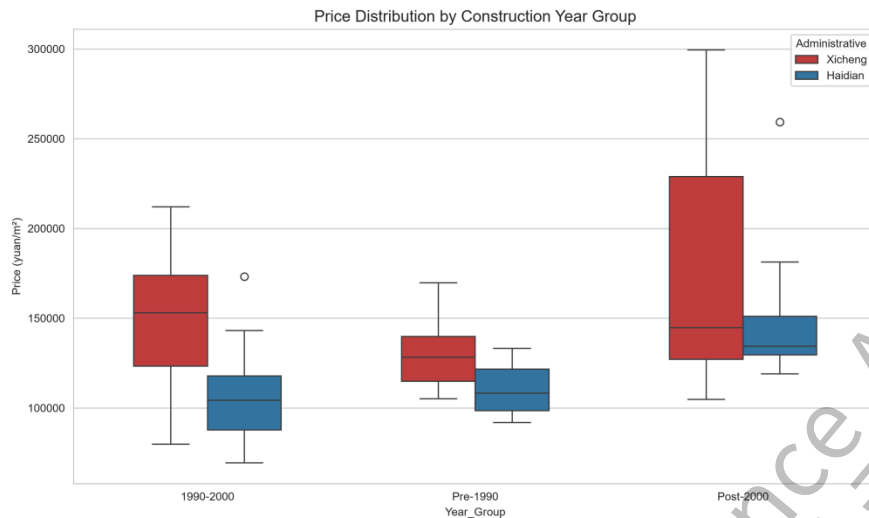


Figure 4-8 Box plot of housing price grouped by construction year

New post-2000 houses in Xicheng are the most expensive in all groups. The Price Difference for Old Buildings Decreases. Poor living conditions of older neighborhoods reduce the school district premium and narrow cross-region price differences.

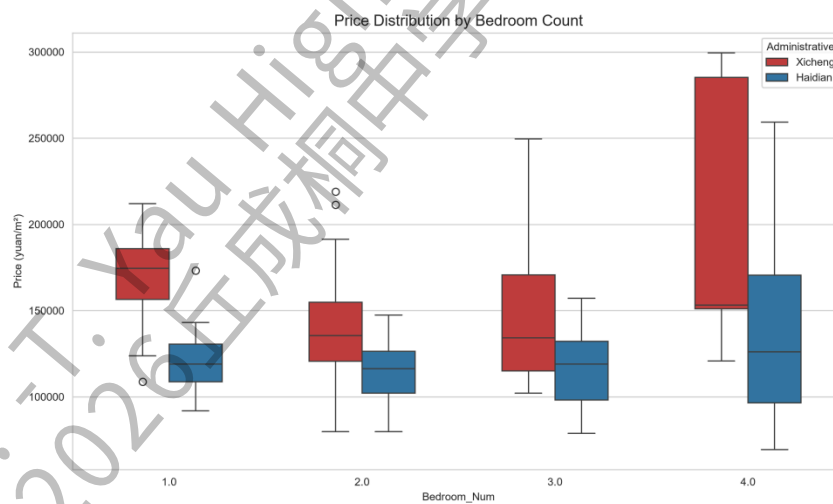


Figure 4-9 Box plot of housing price grouped by bedroom quantity

The upper bound of the unit price for a four-bedroom home in Xicheng exceeds 300,000 yuan per square metre and is therefore higher than

Haidian's maximum of 260,000 yuan. Wealthy families are willing to pay extra for the large school housing in Xicheng. This cannot be clearly seen in the data from Haidian.

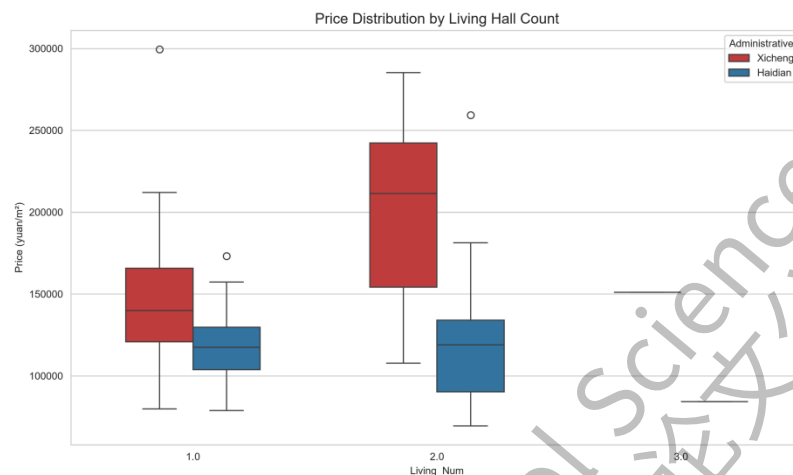


Figure 4-10 Box plot of housing price grouped by living room quantity

Houses with two living rooms have the largest price difference between Xicheng and Haidian. Xicheng's median price for these houses is almost double that in Haidian. It can be seen that Xicheng school house buyers want both a good school and a comfortable, complete living environment, so the quantity of living rooms does not significantly affect Haidian housing prices.

In short, the housing market in Xicheng has created more price differences. High-quality large residences constitute an independent high-price tier and do not appear in Haidian.

4.1.3 Single Variable Scatter Slope Comparison Charts

Scatter plots with added trend lines show that all independent variables are moderately positively correlated with housing unit prices. Separately draw the trend lines for Xicheng and Haidian.

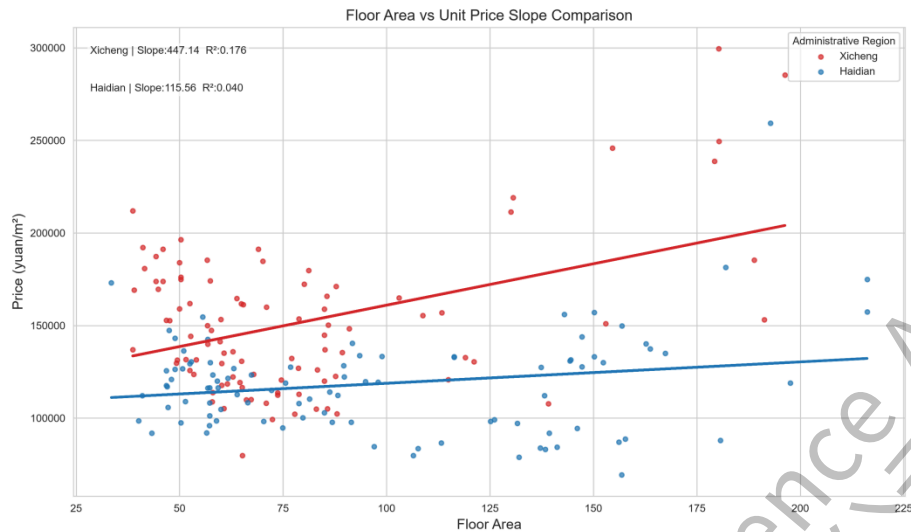


Figure 4-11 Scatter plot & fitting slope: Floor Area and housing unit price

Xicheng's slope is 447.14, and Haidian's is 115.56. Each additional square metre will increase the price by 4 times in Xicheng. Xicheng's R^2 is also larger. Therefore, a large house is relatively expensive in Xicheng. The school bonus grows with larger living space, but this rule does not work in Haidian.

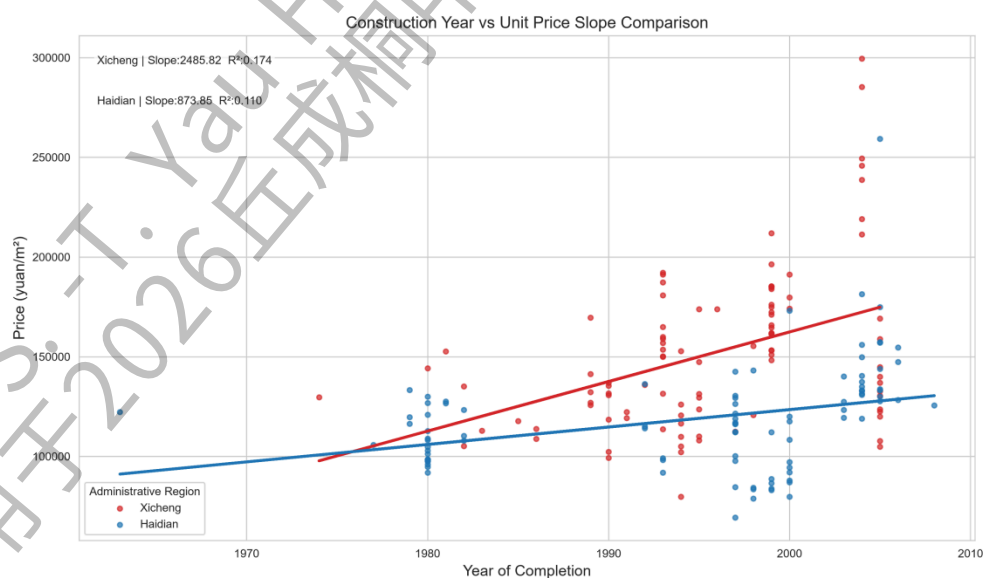


Figure 4-12 Scatter plot & fitting slope: Construction year and housing unit price

Xicheng slope is 2485.82, and Haidian slope is 873.85. Every new construction year sees a price increase of 3 times in Xicheng. This means the new houses will have much larger schools and price bonuses in Xicheng. House age is a reason why people choose to buy school houses in Xicheng.

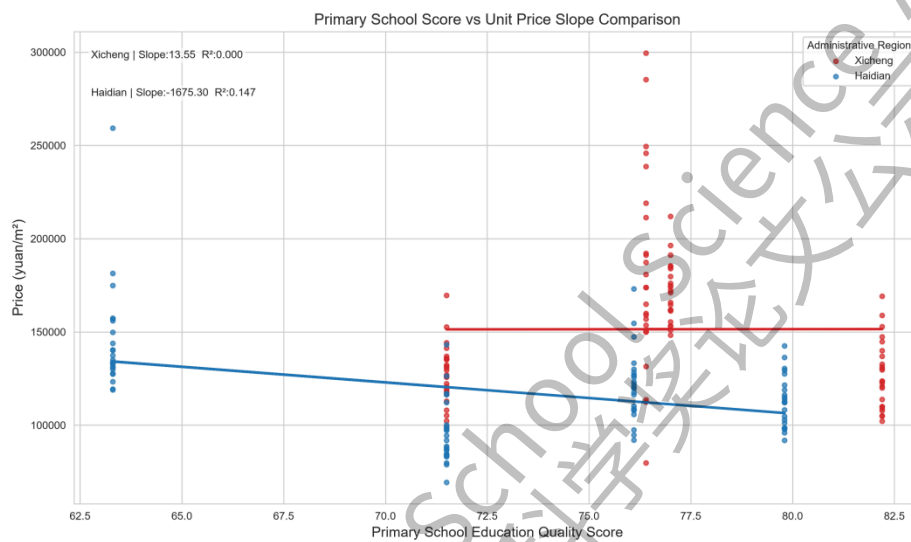


Figure 4-13 Scatter plot & fitting slope: Primary school education score and housing price

The score line for Xicheng Primary School is almost flat, and R^2 is close to zero. Higher primary school scores do not increase Xicheng's housing prices. The entire Xicheng District has a high school level. A single school's quality difference cannot raise the high basic housing price here.



Figure 4-14 Scatter plot & fitting slope: Teacher quality score and housing price

Haidian's slope is 2958.68 and is almost three times that of Xicheng. Haidian's R^2 is 0.32, and Xicheng's is only 0.1. Therefore, it can be concluded that the quality of teaching is the primary driver of changes in Haidian Housing Prices. Families buying homes in Haidian are more concerned about teachers than school names.

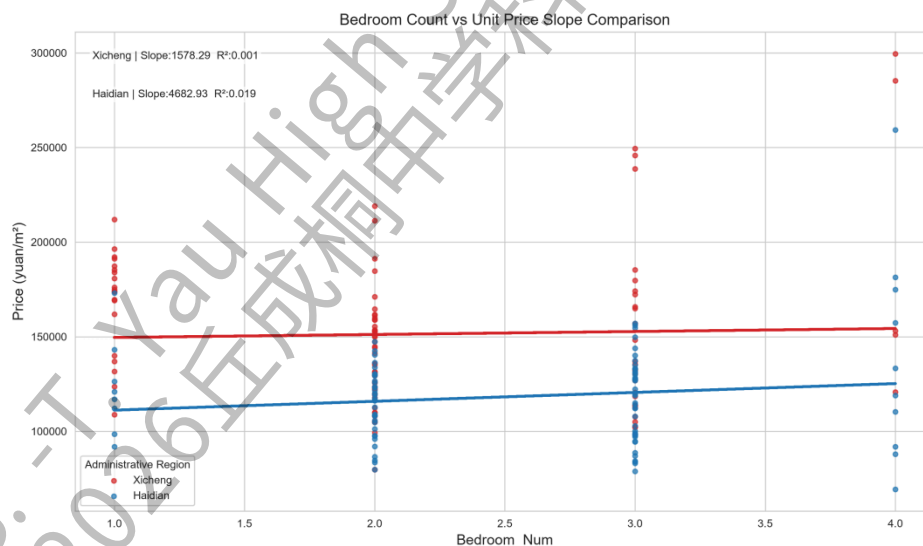


Figure 4-15 Scatter plot & fitting slope: Bedroom quantity and housing unit price

Xicheng's Slope is positive; Haidian's line is almost flat. More Bedrooms do not increase Haidian's house prices. Therefore, additional bedrooms will not provide extra benefits for the Xicheng school houses. Haidian

buyers do not need a certain number of bedrooms for purchase.

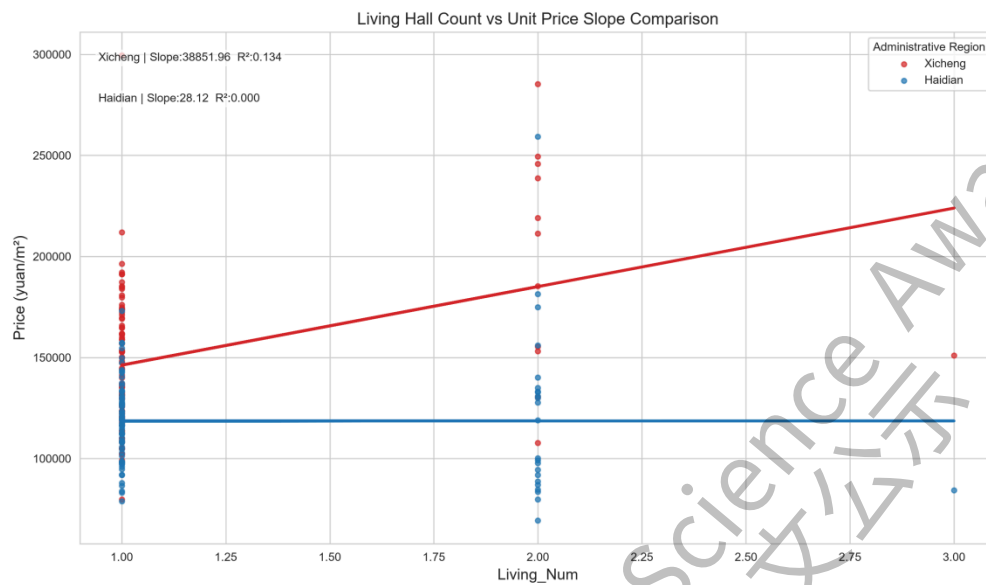


Figure 4-16 Scatter plot & fitting slope: Living room quantity and housing unit price

Xicheng Slope is 38851.96; Haidian Slope is only 28.12. More living rooms strongly increase Xicheng's price but have no effect on Haidian; this suggests that Xicheng buyers seek comfortable living functions along with school resources. Living Room Count Does Not Affect Haidian Housing Market Price.



Figure 4-17 Scatter plot & fitting slope: Primary school enrollment and housing price

Both districts have positive slopes, and the slopes and R^2 values in Xicheng and Haidian are slightly higher. There is a large primary school in the area of high-cost housing. It may be that the large primary schools have maintained a stable school housing premium in Xicheng. The scale of the school has a weaker impact on Haidian's house prices.

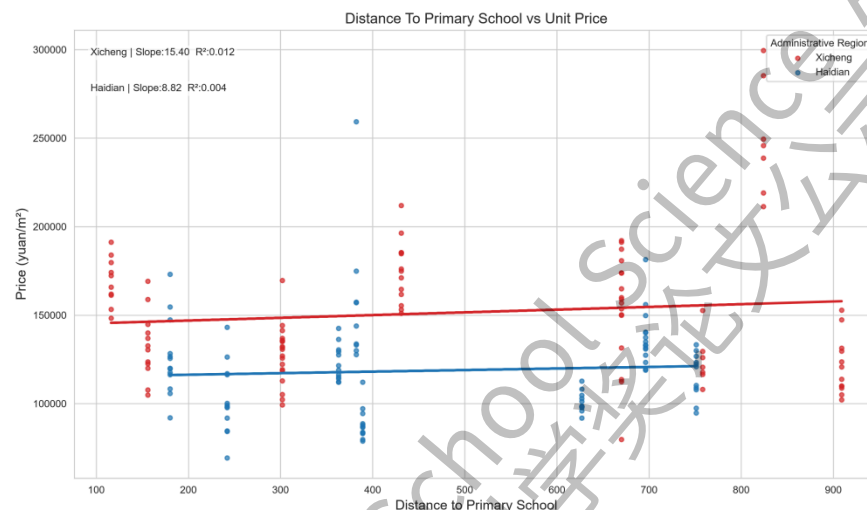


Figure 4-18 Scatter plot & fitting slope: Distance to primary school and housing price

Both slopes are relatively small, and their R^2 values are close to zero. The walk to the primary school is not a factor in either area for housing prices. Primary schools are distributed evenly in Xicheng and scattered in Haidian. The Distance from the house to the school is not a reason for buyers in both areas.

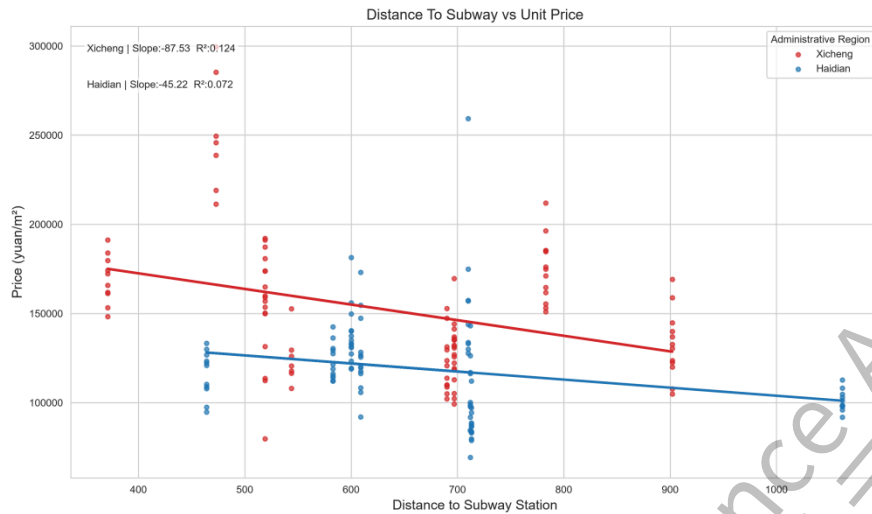


Figure 4–19 Scatter plot & fitting slope: Distance to subway station and housing price

Both Slopes are negative. Xicheng Slope is -87.53 , almost double Haidian's -45.22 . Xicheng's R^2 is also higher. This result shows that subway convenience adds more value to Xicheng Houses. Xicheng has fewer transit lines, so the subway will be relatively convenient for people living here.

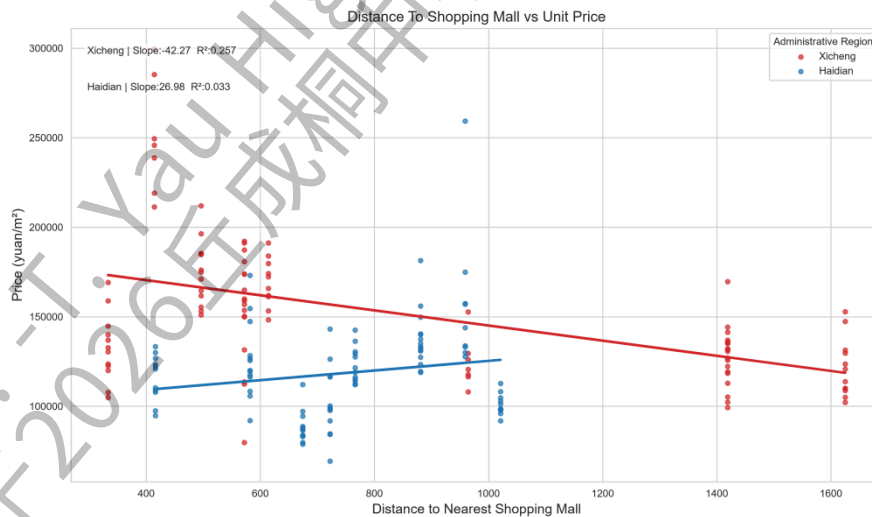


Figure 4–20 Scatter plot & fitting slope: Distance to shopping mall and housing price

Xicheng Slope is negative: closer malls have higher prices. Haidian Slope

is weakly positively correlated and close to zero. The end of the above analysis is that full-service commercial facilities will raise the price of Xicheng Housing. Haidian has many shopping malls everywhere, so the distance to a mall will not affect the price.

4.1.4 Correlation Heat Maps of Full Samples and Separate Districts

Heatmaps are shown for all correlations among numerical variables. Prepare three heat maps: the whole sample, Xicheng subsample and Haidian subsample.

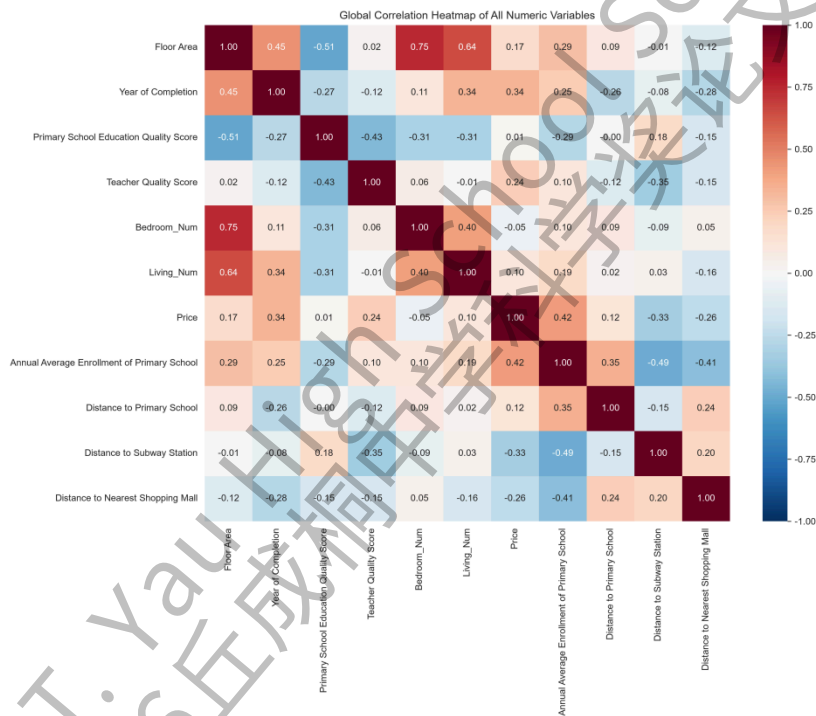


Figure 4-21 Two districts’ correlation heat map of all numeric variables

Based on the general heat map, housing prices are relatively low further away from the city centre. Newer Buildings are generally in higher demand. Areas close to larger primary schools also have higher values.

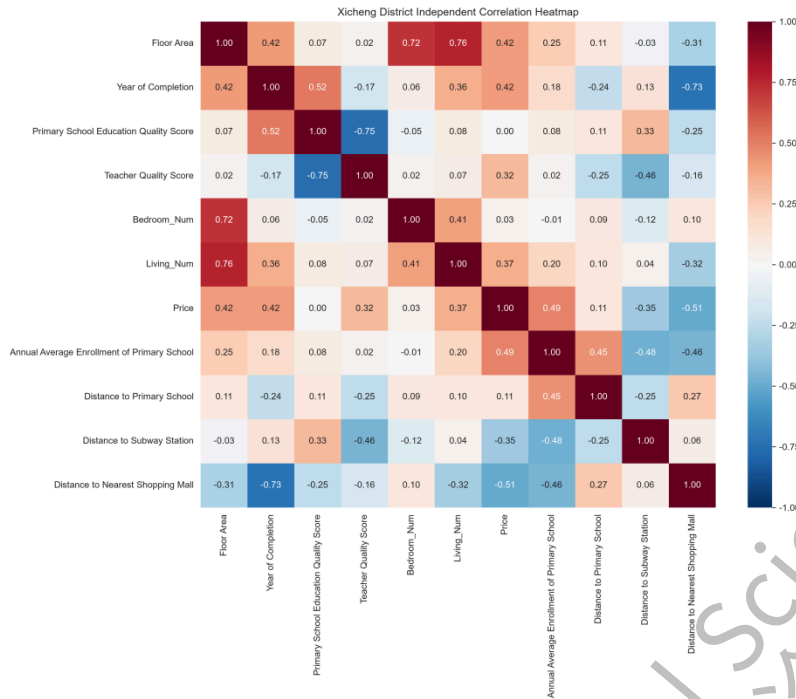


Figure 4-22 Correlation heat map of Xicheng sub-sample

Within Xicheng, floor area shows positive correlation with unit price. Subway and commercial accessibility cause value gaps among old residential buildings. School quality creates limited price influence, because nearly all neighbourhoods belong to high-ranking school zones. Large-sized housing will continuously maintain strong market value.

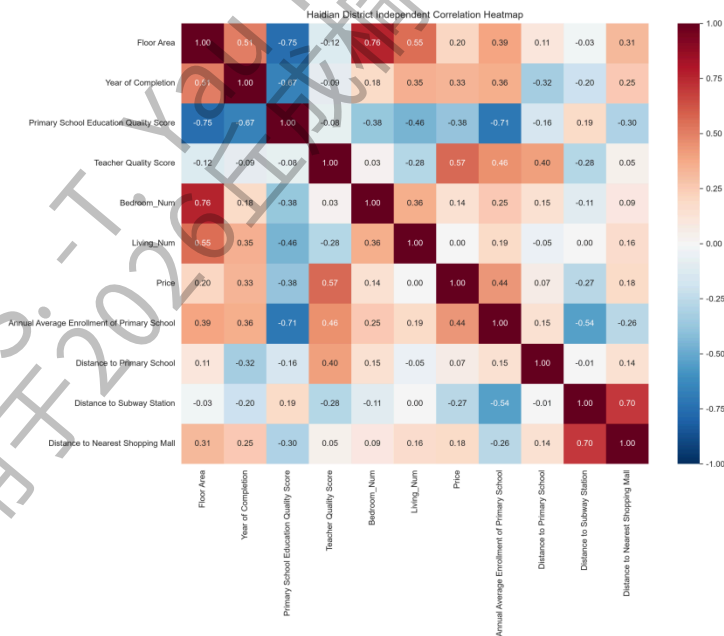


Figure 4-23 Correlation heat map of Haidian sub-sample

Haidian displays different correlation features. Primary school quality has a negative correlation with floor area. High-quality educational resources are usually attached to small old apartments. Newer and larger residences are generally matched with average schools. Even so, teacher quality scores remain one major driving force of housing prices, consistent with earlier scatter graph results. Besides, sufficient supply of newly built housing in Haidian restrains the price growth of large units.

4.2 Main Results

4.2.1 Independent Complete Interpretation of Regression Analysis Results

Two regression statistical tables are used as the data basis for this part: Table 4-22 records variable screening results and fitting indicators of Model 1 to Model 3. Table 4-23 reports coefficient, standard error and significance level of Model 4 multilevel mixed model. Single table summary cannot fully reflect regression rules, so this section independently sorts out all regression results in text form:

Before formal regression estimation, Variance Inflation Factor (VIF) test was carried out to judge multicollinearity among independent variables. All variables' VIF values were below 10, which means serious multicollinearity did not exist in the whole dataset. The high correlation between floor area, bedroom number and construction year only formed moderate collinearity, which would not damage the overall regression reliability.

4.2.1.1 Comparative fitting effect of three basic regression models

Model 1 adopts raw housing unit price as dependent variable, adjusted

$R^2=0.2713$, which retains distance to shopping malls variable;

Model 2 and Model 3 carry out logarithmic transformation to eliminate heteroscedasticity, adjusted R^2 is both 0.2474, stepwise regression eliminates distance to shopping malls;

The interaction term of primary school quality score $\times$ distance to primary school added in Model 3 fails the significance test and is eliminated in the first round of screening, proving that there is no joint interaction effect between school quality and school distance on housing price.

4.2.1.2. Core estimation results of multilevel mixed-effects Model 4

Fixed intercept: -0.4090 ; District random intercept variance= 0.0324 , statistically significant, which verifies that there is a stable inherent baseline price gap between Xicheng and Haidian.

Year of completion coefficient= 0.0065 ($**p<0.05$): Newer construction year significantly increases housing unit price.

Distance to primary school coefficient= 0.00015 ($*p<0.1$): Weak positive correlation; after controlling subway and commercial variables, the distance to primary school has negligible marginal impact on housing price.

Distance to subway station coefficient= -0.00027 ($**p<0.05$): The closer to subway stations, the higher the housing price.

Distance to shopping mall coefficient= -0.00018 ($***p<0.01$): Commercial facility convenience is the most significant positive factor pushing up housing price;

Primary school quality score coefficient= -0.0055 , insignificant; under the balanced high-quality school layout in Xicheng, a single school's score cannot form a separate price premium.

Floor area, bedroom quantity, living room quantity are eliminated due to multicollinearity with construction year, with no independent explanatory

power for housing price.

Predictor Variable	Model 1 (Raw Unit Price)	Model 2 (Log-transformed)	Model 3 (Log + Interaction)
Intercept	Included	Included	Included
Floor Area	Stepwise eliminated 1515	Stepwise eliminated	Stepwise eliminated
Year of Completion	(p=0.00003) ***	Retained	Retained
Distance to Primary School	33.85 (p=0.0044) **	Retained	Retained
Distance to Subway Station	-79.91 (p=0.000007) ***	Retained	Retained
Distance to Nearest Shopping Mall	-12.60 (p=0.118)	Stepwise eliminated	Stepwise eliminated
Primary School Education Quality Score	952.9 (p=0.080)	Retained (Insignificant)	Retained (Insignificant)
Interaction Term	—	—	Eliminated in first step
District Random Intercept Variance	—	—	—
Adjusted R ²	0.2425	0.2474	0.2474
Total Observations	193	193	193

Table 4-22 Variable screening results and goodness-of-fit for Model 1 to Model 3

Predictor Variable	Coefficient	Standard Error	Significance Level
Intercept	-0.409	—	—
Floor Area	Insignificant	—	Insignificant
Year of Completion	0.0065	0.0021	**

Distance to Primary School	0.00015	0.0007	*
Distance to Subway Station	-0.00027	0.001	**
Distance to Nearest Shopping Mall	-0.00018	—	***
Primary School Education Quality Score	-0.0055	—	Insignificant
District Random Intercept Variance	0.0324	—	—
Total Observations	193	—	—

Table 4-23 Coefficient and significance results of multilevel mixed-effects Model 4

4.2.2 Comparison of Fitting Effects of the Four Regression Models

The first three models are standard multiple regression models that do not group by district. Combining the housing data of Xicheng and Haidian will be biased in terms of analysis. The basic models do not consider the intrinsic price difference between the two districts and are thus less explanatory. Model 2 and Model 3 are without the commercial distance variable, so this location information is not included in the calculation. Therefore, both models have a relatively low adjusted R^2 value of 0.2474 and are lower than that of Model 1 (0.2713). Additionally, the added interaction term in Model 3 does not improve the model fit and is quickly eliminated in the calculation.

The Multilevel Mixed Model in Model 4 can address the problem of district heterogeneity that occurs in basic regression. The random intercept variance is 0.0324 and is statistically significant, so it can be concluded that there are stable, inherent price differences between Xicheng and Haidian. Model 4 has kept all the original location variables from the previous models, such as commercial distance. Given that the main reason for this study is to compare housing price systems in the two areas, Model 4 will provide the most reliable and accurate results.

Floor Area has consistently been omitted from all four models. As shown

in the heat map and scatter plot, the size and price of the raw data are positively correlated, but this correlation disappears after controlling for building age, school quality and geographical conditions. The reason is multicollinearity; floor area has highly correlated data with construction year and the number of bedrooms. Therefore, the size of the house cannot provide an independent statistical explanation for the changes in house prices in systematic regression analysis.

4.2.3 Significance Explanation of Variables in the Final Mixed Model

The four variables in Model 4 are shown above. The year of completion is positively correlated with a small level of significance. Every new year of construction will increase the unit price, so the new houses in both districts will have a stable living premium. However, the distance to the primary school is not negatively correlated. After controlling the locations of subways and malls, walking distance to school has created very little price change. School admission qualifications are more important than how far people live from school.

Distance to the subway station and distance to the shopping centre have negative signs and are statistically significant. The shorter the distance to such facilities, the higher the house price. Mall distance has the highest level of significance among all variables, and commercial convenience is a main reason for high housing prices.

Primary school quality score does not have a significance mark in the final model. Given that the district level, age and all location factors are fixed, the school score will not change due to housing prices alone. This result is in line with the Xicheng heat map, and when all areas have the same school resources, the separate school rating loses price influence.

4.2.4 Unified Explanation of the Contradiction Between Scatter Graph and Regression Outcome

Many readers will be confused by the different performances of floor area in the graphs and regressions. The reason is that the variable control conditions are different. A scatter plot shows only the basic correlation between two variables and has no other restrictions. Large houses in Xicheng have the extra advantages of new buildings and complete facilities, so their unit prices are higher. This is the surface trend of many hidden reasons.

In the regression analysis, I keep the other factors constant, such as construction time, proximity to the subway and school rating. Eliminate the effect of the above external factors, and then it is found that floor area is no longer correlated with housing prices. The Size of the house does not increase the actual price.

Therefore, it can be seen that stepwise regression automatically excludes floor area from the final model. The variable is not able to increase the prediction power of housing prices after taking into account all the location and property features. Conflicting results in the scatter plot and regression do not indicate that my data is flawed. They are actually signs that unadjusted raw data can lead people to draw incorrect conclusions due to missing comprehensive factor control.

Regression analysis keeps all the other circumstances the same. If two houses have the same construction year, subway distance, mall distance and district location, then different floor areas will not result in an obvious difference in unit price. House Size does not have an independent effect on prices. The positive trend of scatter graphs is due to good supporting resources, not the space in the house. Remove the hidden interference variable, and the results of the two groups of data

are in line.

4.3 Verification of Research Hypotheses

Three research hypotheses were put forward at the beginning of this paper, and we test each one against all chart and regression results here.

Hypothesis 1: The price attenuation effect brought by school distance is more obvious in Xicheng than Haidian.

This hypothesis is not supported. The regression coefficient of distance to primary school is barely significant in Model 4, and the slope of school distance scatter plots of both districts are close to zero. School distance does not create obvious price drop in either district.

Hypothesis 2: Educational resources dominate housing prices in Haidian, while commercial and transit facilities have bigger impacts on Xicheng housing prices.

This hypothesis is fully supported. Teacher quality score is the core positive factor of Haidian housing prices in scatter analysis, while subway and mall distance carry the highest significance level in Xicheng's price changes.

Hypothesis 3: Xicheng and Haidian own different baseline housing prices, and property features create different price effects in two districts.

This hypothesis is fully supported. The significant district random intercept in mixed model proves stable baseline price gap. Floor area, room quantity only raise prices in Xicheng with no influence in Haidian.

5. Discussion

Xicheng and Haidian have had different logics of housing price formation. The school district of Xicheng prioritizes both the living functions and

school quality of its housing. Housing Floor Area, Building Age, Subway Accessibility and Commercial Support Facilities are the top reasons. Purchasers in this area's main need is a good life.

The main cause of Haidian's housing price fluctuations is the quality of primary school teachers; commercial facilities and housing area are not factors influencing prices. The public service support system in Haidian is distributed evenly, so there will not be obvious differences in prices based on the distance to shopping malls. Most buyers in Haidian only buy small homes to get school admission quotas and do not have long-term living plans.

The spatial attenuation speed of school district premium is different in the two districts. The housing price in Xicheng is affected by the increase in distance from subways and shopping malls; the housing price in Haidian is more responsive to teacher quality than to the distance from primary schools. The former single-regression model does not consider different base prices in the various areas, so a multilevel mixed model has been used to deal with the bias from regional heterogeneity and obtain an unbiased estimate of the variable marginal effect.

6. Conclusion

6.1 Main Research Conclusions

Xicheng and Haidian have different logics for housing prices. Xicheng's Market Buyers Balance Living Space and School Resources. Large houses, new buildings and good access to the subway and shopping centres are likely to increase house prices. All areas in Xicheng have relatively equal primary school resources; therefore, the score of a single school will not change the housing value. Due to the limited supply of large school

residences, the prices of the large houses have increased significantly. The basis for Haidian's house prices is the teacher quality score of primary schools. Commercial Facilities and House Size have little effect on unit price. There is a full-service mall and subway at this place, so being further from the mall is not an extra expense. Most of the buyers here want a small house to obtain school quotas and do not plan to live there for a long time as a family.

Speeds of School Premium Spatial Decay are different. Xicheng has a steeper drop in price with increasing distance from the subway and malls, and the price of Haidian is more affected by the teacher's grade level than by how far it is from the school. The standard regression model is unable to solve the problem of different regional baseline prices; thus, a multilevel mixed model is used to remove the bias caused by fixed prices in a certain area and obtain reliable estimates of the marginal effect of all characteristic variables.

6.2 Limitations of This Research

However, there are still some problems in this paper. First of all, the original sample collected was 200 housing records, and 7 abnormal samples with missing values and extreme outliers were deleted before analysis, leaving 193 valid samples for regression. If there were more data, the regression results would be more stable. Second, I have only gathered the data from the past year and cannot track price fluctuations over different years or under new educational policies. Third, I do not add the policy variable to the regression calculation, so the impact of the policy cannot be determined in this study. Fourth, I am only comparing the two central districts of Beijing, so the results cannot be generalised to other cities' school housing markets.

6.3 Simple Suggestions for Different Groups

For families that are going to live in the school-district area for a long time with children, they may choose Xicheng to get a bigger house and all the facilities around it. Only for the purpose of obtaining school admission spots and without plans to live there, small Haidian apartments are suitable, and primary school teacher quality should be given more weight than property size.

To help alleviate the large school-housing-premium in Haidian, the distribution of teachers among the local primary schools is necessary for urban and education planners. The construction of new large residential areas with complete commercial facilities in Xicheng will reduce the high cost of large school-linked houses. Expansion of the subway network in Xicheng will also reduce the difference in cost between the vicinity of the subway and other places.

7.Reference

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, /82/(1), 34 – 55.

Ha, W., & Jin, H. (2018). Education expenditure and school district housing premium: A case study of Beijing [In Chinese]. *Education & Economy*, /34/(1), 35 – 41.

Sun, W., & Lin, J. (2020). Can the supply of educational resources reduce school district housing premium? Evidence from newly built primary schools in Beijing [In Chinese]. *China Economic Quarterly*, /19/(2),

625 – 646.

Wang, Y., & Li, J. (2022). Research on spatial differentiation of housing prices based on GIS [In Chinese]. *Progress in Geography*, /41/(5), 790 – 798.

Shen, T., Yu, H., Zhou, L., Gu, H., & He, H. (2020). The influencing mechanism of second-hand housing prices in Beijing: A study based on multiscale geographically weighted regression model (MGWR) [In Chinese]. *Economic Geography*, /40/(3), 75 – 83.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示