A.Y. 2023/2024

Hedonic Price Method-
Assessing the Impact of the Jin-Yi-Dong Rail Transit Line on Housing Prices in Yiwu City

Economic Assessment of Urban Transformations

Group 10
Yuxin Chen 214994
Fan Chen 221459
Xin Zhang 220716

Professors: Caragliu Andrea Antonio; Dellisanti Roberto
Management of Built Environment

1. Introduction

The Jin-Yi-Dong Rail Transit Line is a suburban rapid transit line in Zhejiang Province, China, connecting three neighbouring county-level cities of Zhejiang: Jinhua, Yiwu and Dongyang. Serving as both a metropolitan railway and an urban metro, the line is 107.17 km long with 31 stations and was completely commissioned in 2023. The total investment cost of the project is approximately RMB 34 billion (equivalent to 5 billion $). The aim of the project is to reduce the urban-rural development gap, promote the integration of metropolitan areas and drive industrial integration along the line.

The positive impact on the rural areas, especially the junction area, is self-evident. Nevertheless, the economic benefits for urban plots have been somewhat overlooked and controversial, this paper is going to focus on the impact of the line within Yiwu on urban property prices.

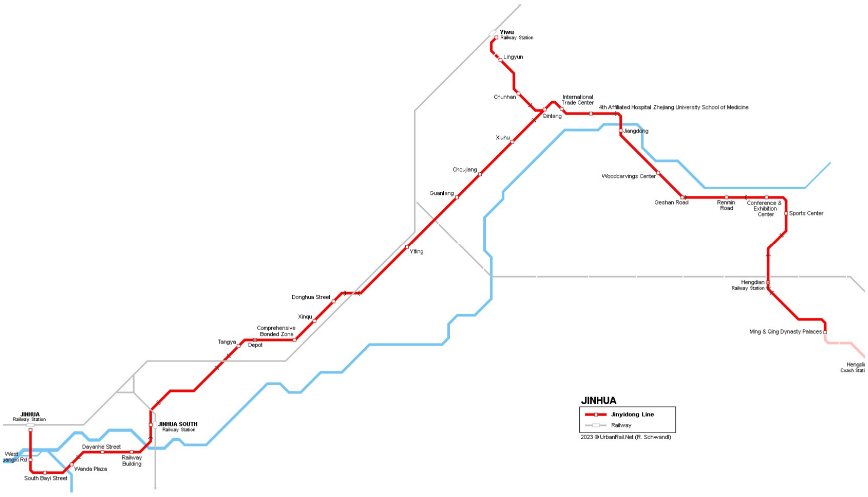


Fig 1. Jin-Yi-Dong Line map

2. Methodology

Here we do the Hedonic Price Method (HPM). Because it provides a robust framework allowing us to decompose the price of residential properties into various contributing factors, and well-suited for analysing how them attributes to a residential property. The HPM allows us to isolate and measure the effect of the rail infrastructure on housing prices, controlling for other influencing factors.

2.1. Data Collection and Preparation

We will gather data on residential property transactions in Yiwu City from China's property trading websites[1], focusing on the on samples within the urban area. The main urban area of Yiwu is divided into seven districts. Twenty samples were randomly selected from each district. The data will include the following variables:

General data
District
Total residential sale prices
Size *(mq)*

Data	**Variables**
Housing price per square meter *(price_sqm)*	in RMB
Presence of bedrooms *(bed)*	numbers
Presence of balcony *(balc)*	yes=1; no=0
Level of floor *(floor)*	low=1; medium=2, high=3
Presence of elevator *(elevator)*	yes=1; no=0
Housing type *(type)*	Second hand=0; new=1
Presence of decoration *(decoration)*	yes=1; no=0
Quality level of school *(qual_sch)*	low=1; medium=2, high=3
Distance to the school *(dis_sch)*	kilometres
Distance to the cbd *(dis_sch)*	kilometres
Distance to the railway station *(rail_dist)*	kilometres
Distance to the nearest new metro station *(metro_dist)*	kilometres

Fig 2. All available data for the model

For checking the data sheet, please click here.

2.2. Model Specification

To align the model closer to a normal distribution, we will use a hedonic pricing model in the form of a double-log linear regression. The dependent variable is the logarithm of the property price, which helps to normalize the distribution and interpret the coefficients as elasticities. The independent variables also undergo logarithmic transformation (note: dummy variables will not be transformed), allowing us to capture the percentage change in property prices resulting from a percentage change in each attribute. The model can be specified as follows:

$$\ln(P)=\ln\beta_0+\beta_1\ln(mq)+\beta_2\ln(be)+balcony+\beta_3\ln(fl)+elevator+type+decoration+\beta_4\ln(sd)+\beta_5\ln(sq)+\beta_6\ln(cd)+\beta_7\ln(rd)+\beta_8\ln(md)+\varepsilon$$

where:

P: the price of property

(*): * represents a vector of property-specific characteristics (e.g., size, floor),

β_0: the intercept

[1] https://yiwu.anjuke.com/

β_1, β_2, β_3…: the coefficients to be estimated

ε: the error term

3. Application of HPM

3.1. Construct the initial linear regression model.

Our dataset comprised 140 observations and 12 variables. After handling missing values and convert categorical variables to dummy variables, we constructed our initial regression model. The dependent variable was the price per square meter, while the independent variables included m^2, bed, balc, elevator, and other factors.

The initial regression model revealed a large range of residuals, indicating potential outliers or high variability in predictions. This variability suggested that further refinement of the model was necessary.

```
> summary(reg1)

Call:
lm(formula = price_sqm ~ mq + bed + balc + floor + elevator +
    type + decoration + qual_sch + dis_sch + cbd_dist + rail_dist +
    metro_dist, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-17035.9  -2881.4    -42.8   2535.8  20517.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12873.244   6142.625   2.096  0.03809 *
mq              5.617     27.107   0.207  0.83616
bed          -229.999   1231.626  -0.187  0.85216
balc         1638.598   1424.851   1.150  0.25230
floor         125.638    622.338   0.202  0.84033
elevator     6987.379   2187.715   3.194  0.00177 **
type        -2400.718   1592.014  -1.508  0.13404
decoration    600.635   1211.584   0.496  0.62093
qual_sch     2804.317    999.352   2.806  0.00580 **
dis_sch       385.213    313.788   1.228  0.22186
cbd_dist    -1321.715    195.599  -6.757 4.56e-10 ***
rail_dist     309.930    132.592   2.337  0.02098 *
metro_dist   -291.393    311.460  -0.936  0.35127
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5514 on 127 degrees of freedom
Multiple R-squared:  0.5812,    Adjusted R-squared:  0.5416
F-statistic: 14.69 on 12 and 127 DF,  p-value: < 2.2e-16
```

Fig 3. Summary of initial regression model (reg1)

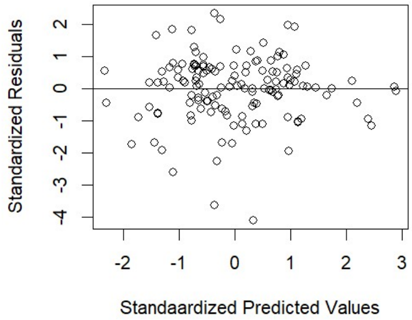
3.2. Interpretation and validation

Outliers were removed and the regression model was reconstructed using the cleaned data. Subsequently, the Variance Inflation Factor (VIF) was employed to ascertain the absence of significant multicollinearity among the independent variables. The analysis demonstrated that the VIF values were within the acceptable limits.

Then, the model fit was evaluated using a standardized residuals plot, which facilitated an understanding of the distribution and variance of residuals. Furthermore, we conducted a pair plot matrix analysis, which indicated the potential for non-linear relationships between the variables. This insight highlighted the necessity to consider non-linearity in the model.

```
> vif(reg1a)
       mq       bed      balc     floor  elevator      type decoration  qual_sch   dis_sch  cbd_dist rail_dist metro_dist
 2.069308  2.056992  1.236692  1.087932  1.187238  1.307828  1.411958  2.619977  1.414216  2.239666  1.412400  1.643123
```

Pic 4. VIF of reg1



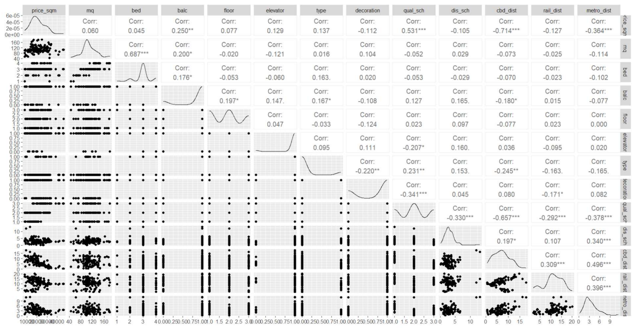


Fig 5. Standardized predicted values of reg1. *Fig 6. Pair plot matrix. of reg1.*

3.3. Log-Transformed Regression Model

To address the non-linear relationships identified, we transformed our variables using logarithms. This log transformation significantly improved the model fit and reduced residual variability, making our predictions more reliable.

```
> vif(reg2)
       mq_log        bed_log           balc      floor_log       elevator 
     2.425077       2.362973       1.214000       1.075304       1.194740 
         type     decoration   qual_sch_log    dis_sch_log   cbd_dist_log 
     1.303263       1.331732       2.520059       1.282531       1.896417 
 rail_dist_log metro_dist_log 
     1.419088       1.391321 
```

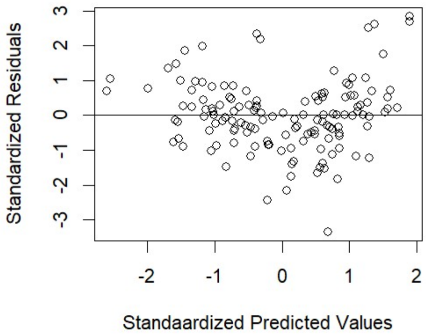


Fig 7. VIF of reg2. *Fig 8. Standardized predicted values of reg2.*

3.4. Result Analysis

The final step involved analysing the results of our log-transformed regression model.

```
Call:
lm(formula = price_sqm_log ~ mq_log + bed_log + balc + floor_log + 
    elevator + type + decoration + qual_sch_log + dis_sch_log + 
    cbd_dist_log + rail_dist_log + metro_dist_log, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.94342 -0.11251  0.02092  0.14134  0.54232 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     8.902615   0.585143  15.214  < 2e-16 ***
mq_log          0.185563   0.135865   1.366  0.17444    
bed_log         0.029186   0.147581   0.198  0.84355    
balc            0.140713   0.061702   2.281  0.02426 *  
floor_log      -0.002367   0.048051  -0.049  0.96079    
elevator        0.279297   0.095929   2.911  0.00426 ** 
type           -0.182960   0.069499  -2.633  0.00953 ** 
decoration     -0.009839   0.051725  -0.190  0.84945    
qual_sch_log    0.175182   0.077116   2.272  0.02480 *  
dis_sch_log    -0.019951   0.038690  -0.516  0.60700    
cbd_dist_log   -0.317721   0.041306  -7.692 3.62e-12 ***
rail_dist_log   0.123553   0.046369   2.665  0.00872 ** 
metro_dist_log -0.070296   0.040550  -1.734  0.08544 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.241 on 126 degrees of freedom
Multiple R-squared:  0.6214,	Adjusted R-squared:  0.5854 
F-statistic: 17.24 on 12 and 126 DF,  p-value: < 2.2e-16
```

Fig 9. Summary of log-transformed regression model(reg2).

We identified significant predictors that influenced the price per square meter:

i. Significant Positive Predictors: Balcony (balc), elevator (elevator), quality of school (qual_sch_log), and distance to rail station (rail_dist_log).
ii. Significant Negative Predictors: Type of property (type), distance to central business district (cbd_dist_log), and distance to metro station (metro_dist_log).
iii. Non-Significant Predictors: Area in square meters (mq_log), number of bedrooms (bed_log), floor level (floor_log), decoration, and distance to school (dis_sch_log).

A notable finding was the impact of metro distance. Our analysis showed that a 1% increase in distance to the metro station was associated with a 7.03% decrease in price per square meter. This finding was marginally significant with a p-value of 0.08544.

4. Conclusion

From the above statistical analysis of the HPM, we can conclude that a negative coefficient for the variable of distance to metro stations, implying a statistically weak correlation, which in turn reveals that the Jin-Yi-Dong Rail Transit Line has a positive impact on property prices to a certain extent.

Looking further into the matter, we can say that the economic benefits of the intercity rail project in terms of the value added to land in the urban area of Yiwu are very limited. However, the project was originally constructed with the intention of promoting the intensive development and integration of neighbouring areas, as well as improving accessibility and convenience of the public transportation. Therefore, its overall impact, including its non-economic value, is worthy of in-depth investigation.

This analysis underscores the importance of considering multiple factors and applying robust statistical methods to understand the complex dynamics influencing real estate markets. By leveraging these insights, stakeholders can make more informed decisions, ultimately contributing to more sustainable urban development.