**Citation:** LI Wenxi, BIAN Changzhi, XU Qi, DONG Tianyi, ZHAO Yixin. The Nonlinear Impact and Spatial Characteristics of TOD Built Environment on Land Value: A Case Study of Beijing[J], Urban Transport of China, 2024 (6).

# The Nonlinear Impact and Spatial Characteristics of TOD Built Environment on Land Value: A Case Study of Beijing

LI Wenxi[1,2], BIAN Changzhi[2], XU Qi[3], DONG Tianyi[4], ZHAO Yixin[2]

*1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China;*
*2. China Academy of Urban Planning & Design, Beijing 100037, China;*
*3. Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China;*
*4. School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China*

**Abstract:** The improvement of land value brought by urban rail transit can be attributed to three aspects: the premium of adjacent transit stations, the premium of accessibility improvement, and the construction premium around the station. However, research on the impact of the latter two factors on the improvement of lane value remains insufficient. In response, integrating multi-source urban big data, this paper proposes the “5D” indicator system and its calculation methodology for the Transit-Oriented Development (TOD) built environment of urban rail transit. Relaxing the assumption of the linear relationship between housing price and its influencing factors in exiting research, this paper employs a nonlinear model for feature pricing based on the XGBoost machine learning algorithm and the local interpretation method SHAP to analyze the nonlinear impact of TOD built environment on the promotion of land value and its spatial distribution characteristics. Case studies in Beijing demonstrate that the TOD built environment around rail stations significantly influences the land value, accounting for 64.30%, and the relative importance of transit accessibility holds is 17.66%, which is more important than station proximity. The impact of TOD built environment on land value has an obvious nonlinear relationship and threshold effect. The spatial distribution characteristics of the nonlinear impact indicate that station land development strategies should be tailored to specific regions, according to local conditions, the stations in the peripheral areas of cities should pay attention to the accumulation of resources, and the stations in the central urban area should focus on the improvement of travel quality and environmental quality. **DOI:** 10.13813/j.cn11-5141/u.2024.0036-en

**Keywords:** urban rail transit; built environment; land value uplift; accessibility; machine learning; threshold effect; Beijing

## 0 Introduction

The rapid urbanization in China has a great rigid demand for urban rail transit, but it faces difficulties such as the shortage of construction funds and insufficient operating income. Usually, the construction fund for urban rail transit mainly comes from investment means such as financial appropriations of the government and low-interest commercial loans. However, with the surge in funding demand for urban rail transit construction, without other diversified investment and financing channels, there is no other way but to increase the tax. In addition, urban rail transit generally faces the problem of revenue and expenditure balance in operation and the financial sustainability cannot be maintained by relying solely on the ticket revenue. It may lead the traveler to indirectly share the operating cost due to the increase in the ticket price or reduction of the service quality, which is unfair.

As a typical public good, urban rail transit has significant positive externalities, which are mainly reflected in the value-added effect on land and property along the line. However, due to the separation of investment and income entities, urban rail transit construction and operation enterprises, as the main investors, receive less income from external benefits. Therefore, it is imperative to fully utilize external benefits to support the construction of urban rail transit. In response to this trend, scholars have begun to study various sustainable investment and financing methods for urban rail transit, including a series of land value capture tools, aimed at effectively realizing the increase in the land value brought about by urban rail transit investment.

The Transit-Oriented Development (TOD) model guided by public transportation is an inevitable choice for the integrated development between urban rail transit and cities [1]. Urban rail transit and surrounding real estate development can interact positively, achieving a win-win situation. World-class metropolises such as Tokyo, London, and Hong Kong have actively developed urban rail transit, promoting its integration with urban development and achieving remarkable success. China and other developing countries have followed the TOD model and the construction of urban rail transit is booming [2–4]. Therefore, it is necessary to

**Received:** 2024-06-29

quantitatively analyze the impact of urban rail transit on the land value under the TOD model, thus providing a reliable basis for decision-making using tools such as land value capture.

According to the theory of urban land economics [5–7], transportation investment brings about an improvement in accessibility, which reduces the general transportation cost for the commuter, and promotes the premium value of land and property around the stations while creating location advantages. The increase in the land value brought by urban rail transit can be divided into three parts: (1) The transportation proximity premium brought by being close to urban rail transit stations in space to obtain travel services; (2) The accessibility improvement generated by the reduction in the travel cost drives the development of the supporting facility, thereby increasing the value of surrounding land; (3) The construction of the transportation infrastructure brings the new development opportunity, which leads to construction premiums around the stations.

In terms of transportation, a large amount of theoretical research and practical experience have shown that the construction of urban rail transit will be capitalized as the land value [8–11], and there is a proximity effect of increasing the land value around stations. This transportation proximity premium varies according to different research scenarios [12–14]. In the meantime, accessibility, as a key element in the built environment of TOD, is a key research object, and proximity and accessibility together constitute multi-scale proximity of public transportation in space and time. Previous studies often use proximity as a representation of accessibility [11, 13, 15–17], but considering spatial proximity alone is not sufficient to highlight the interaction between transportation and land use. Some scholars have improved based on the gravity model [18] and spatial syntax [19], but have not considered the impact of multiple transportation modes on the accessibility calculation.

There is generally less research on TOD urban construction compared to transportation. The construction premium originates from the control of the land use nature and development by urban planning and transportation planning, such as adjusting industrial land to residential land or increasing the development intensity. However, in existing research, this construction premium is usually manifested as the value-added effect of the housing neighborhood environmental factor, rather than focusing on the vicinity of the station [19–21]. Research focusing on TOD often divides the built environment variables of the urban construction dimension based on the “3D” principle (density, design, and diversity), and concludes that TOD land use characteristics may be an important factor affecting the premium effect of urban rail transit [8, 16], such as land mixed use [20, 22], high-density development [20], and a good pedestrian environment [16, 20, 23], which can lead to an expansion of the land premium.

Scholars in China and abroad have achieved rich research results in the field of urban rail transit and land value, but there are still directions to be improved and deepened. On the one hand, the mechanism by which accessibility enhances the land value deserves further research. The intrinsic connection between urban rail transit and construction premium stems from the accessibility enhancement driven by the reduced travel cost [24], which is a necessary condition for urban development and renewal, and its calculation method needs to be optimized. On the other hand, research on the construction premium is not yet sufficient. Existing studies often assume a linear relationship between the TOD urban construction factor and the land value and the non-linear effect and its spatial distribution characteristics still need to be further explored.

In summary, this article focused on the following two issues for research: 1) Which of the availability (proximity to the station) and accessibility of urban rail transit has a more significant impact on land value uplift? 2) What is the impact of the TOD built environment, especially the urban design factor, on land value uplift and its spatial distribution characteristics?

Given the crucial role of the housing price in reflecting the land value, this article used the housing price as the research object. According to the “5D” (density, design, diversity, proximity, and accessibility) development concept of the TOD built environment [25], taking Beijing as an example, this study integrated multi-source urban big data to calculate the cumulative opportunity accessibility of public transportation and private cars, as well as other TOD built environment variables, constituting the influencing factor of the housing price together with the housing price control variable. This article relaxed the assumption of a linear relationship between the housing price and its influencing factors, used the machine learning model eXtreme Gradient Boosting (XGBoost) to study the quantitative relationship between the two in the framework of a nonlinear model of feature pricing, and employed the local interpretation method Shapley Additive Explanation (SHAP) to analyze the nonlinearity of the influencing factor, exploring the impact of the TOD built environment on the land value and its spatial distribution characteristics.

## 1 Research area and data explanation

Beijing consists of 16 districts with a total area of 16 410 km$^2$. At the end of 2022, the permanent population of Beijing was 21.843 million, with nearly 80% of residents living within the Sixth Ring Road. To meet the growing demand for housing, real estate developers have purchased a large amount of land construction rights from the government. With the continuous improvement of the urban rail transit system, Beijing has experienced explosive growth in housing construction, reshaping the urban landscape. Taking into account factors such as urban layout and data availability, this article selected the area within the Sixth Ring Road of Beijing as the research area, and the main data used includes:

### 1) Second-hand housing transactions

Compared to first-hand houses mainly located in the suburbs, second-hand housing transactions involve a wider range of property types and have a wider coverage. Since 2008, second-hand houses have become the mainstay of housing transactions in Beijing and the transaction ratio even reached over 95% from 2011 to 2017 [26]. Therefore, this article obtained a panel dataset of second-hand housing transactions in Beijing from Lianjia Network using a Python web crawler (January 2011 to March 2016), including structural features such as the house age, floor number, and area. The dataset was collected at the community level and averaged over time to form cross-sectional data, including 274 172 transactions from 5 296 communities.

### 2) Urban rail transit system

Eighteen subway lines opened in 2016 were selected, with a total length of 555 km. Based on the reasonable walking radius, the Guidelines for Planning and Design of Urban Rail Transit Areas issued by the Ministry of Housing and Urban-Rural Development in 2015, and the “10-minute walking circle”, a circular area with a radius of 800 m was adopted as the impact range of the subway station, namely, the station buffer zone.

### 3) Built environment of the station

Based on the open platform of the Gaode Map, Point of Interest (POI) data within the influence range of subway stations were obtained. Through the open-source data platform of the Baidu Map, OpenStreetMap, and Mapbox websites, the Area of Interest (AOI), 3D vectors of buildings, urban roads, and isochronous circles data were obtained. The retrieval time range for the above data was March 2020. The population was sourced from the 100 m × 100 m grid data in 2019 in China on the WorldPop website. The shared bicycle data was sourced from Mobike’s booking data during the morning and evening peak hours (7:00 to 9:00 and 17:00 to 19:00) on weekdays from May 10–18, 2017.

### 4) Trip information

The trip planning service based on the Gaode Map adopted the shortest time strategy to obtain the full trip chain information of private cars and public transportation between any two 1 km × 1 km grid centers within the Sixth Ring Road. The search time range was during the morning rush hour (7:00 to 9:00) of working days in November 2019, and its information included the name and latitude of the starting/ending station, travel time, etc., which were used to calculate the accessibility.

## 2 Built environment variables and calculation method

The built environment characteristics along the urban rail transit line are characterized by the “5D” development concept of TOD. According to the obtained data, the definitions of built environment variables and housing price control variables in each dimension are shown in Tab. 1.

Accessibility is calculated using the cumulative opportunity, which means the total number of opportunities obtained through a certain mode of transportation within a certain period, and its results have the advantages of simplicity and high interpretability. Based on trip planning of the Gaode Map and POI data, this paper considered the full trip chain cost of multiple transportation modes and calculated the employment accessibility for employment-related POIs, and the accessibility calculation formula is:

$$A_i^z = \sum O_j f\left(t_{ij}^z\right), \tag{1}$$

where, $A_i^z$ is the accessibility, which refers to the cumulative opportunity of starting point $i$ under the $z$-th mode of transportation; $z$ is the mode of transportation, $z$ = 1 means public transportation, and $z$ = 2 means a private car; $O_j$ is the number of POIs in the $j$-th grid, with 6 types of POIs, including employment, school, medical, shopping, leisure and entertainment, and scenic spot (only employment type POIs are used to calculate the employment accessibility); $t^z_{ij}$ is the travel time/min from grid $i$ to $j$ using the $z$-th mode of transportation under trip planning. When $t^z_{ij} \le T$, $f(t^z_{ij}) = 1$; otherwise, it is 0, where $T$ is the average commuting time in Beijing (47 min).

To further explore the impact of the difference in accessibility between public transportation and the private car on the land value, this article defined two variables: relative accessibility and relative employment accessibility, to reflect the differences in overall and employment accessibility between the two modes of transportation.

## 3 Analysis model of land value and TOD built environment variables

### 3.1 Feature pricing model

The feature pricing model aims to estimate the implied price of a property using the regression calculation method based on its feature set, and has become a powerful tool for revealing the value of facilities in the land and real estate market. For example, when we estimate the premium value of urban rail transit, it is often assumed to have a linear relationship with the influencing factor. However, the impact of the TOD built environment on the land value exhibits nonlinear characteristics. This article used the nonlinear model of feature pricing as the research framework to estimate housing prices, and the calculation formula is:

$$Price = f(S, L, N, T, P, A) \,, \qquad (2)$$

where *Price* is the housing price, in yuan; *f* is in the form of a nonlinear function, and the influencing factors include the structural factor *S*, location factor *L*, and neighborhood factor *N* commonly used in house pricing research, as well as proximity *T*, the TOD urban construction factor *P*, and accessibility *A*.

**Tab. 1** Descriptive statistics of TOD built environment and housing price index

| Classification | | | Variable | Variable description | Mean | SD |
|---|---|---|---|---|---|---|
| Dependent variable | | | Housing price/(CNY 10 000/m$^2$) | Housing price/m$^2$ | 3.92 | 1.62 |
| Variables of TOD built environment | Proximity | | Proximity of subway station/km | The straight-line distance to the nearest subway station | 0.95 | 0.89 |
| | | | Proximity of bus stop/m | The straight-line distance to the nearest bus stop | 197.19 | 105.59 |
| | TOD urban construction | Density | Population density/(10 000 persons·m$^{-2}$) | Population within the station buffer zone | 1.98 | 0.62 |
| | | | Employment density/(10 000 jobs·m$^{-2}$) | Number of job opportunities within the station buffer zone | 1.03 | 0.62 |
| | | | Building density/% | The ratio of the total floor area of buildings on the AOI within the station buffer zone to the total area of the AOI | 21 | 6 |
| | | | Plot ratio | Average building plot ratio within the station buffer zone | 3.37 | 1.53 |
| | | | Compactness | The ratio of the building plot ratio within a radius of 500 m to that within a radius of 500–800 m at the station | 1.05 | 0.41 |
| | | Diversity | Mixing degree of land use | $E_L = -\sum_i^n P_i \ln P_i / \ln(n)$, where $E_L$ is the land use mixing degree of the station; $P_i$ is the proportion of the number of the *i*-th type of POI based on the land use type to that of the total types in the station buffer zone; *n* is the number of the land use type 1) | 0.48 | 0.04 |
| | | | Mixing degree of economic industry | $E_E = -\sum_j^m P_j \ln P_j / \ln(m)$, where $E_g$ is the mixing degree of the economic industry in the stations; $P_j$ is the proportion of the number of the *j*-th type of POI based on the land use type to that of the total types in the station buffer zone; *m* is the number of the economic industrial type 2) | 0.75 | 0.06 |
| | | Design | Non-motorized traffic index | The ratio of the 10 min walking time circle centered around the station to the buffer zone area of the station | 0.42 | 0.13 |
| | | | Pedestrian road density/(km·km$^{-2}$) | The ratio of the length of walkable roads within the station buffer zone to the station buffer zone | 13.13 | 4.57 |
| | | | Public bus stop density/(stops·km$^{-2}$) | The ratio of the number of bus stops within the station buffer zone to the station buffer zone | 5.80 | 2.36 |
| | | | Cycling shuttle ratio/% | The proportion of cumulative shared bicycle orders within a 200 m radius of the station during peak hours to the total number of orders | 0.30 | 0.27 |
| | Accessibility | | Public transportation accessibility | Public transportation accessibility of the grid where the community is located (1 km × 1 km) | 19 414 | 12 477 |
| | | | Relative accessibility | The ratio of public transportation accessibility to private car accessibility in residential areas | 0.15 | 0.07 |
| | | | Relative employment accessibility | The ratio of public transportation employment accessibility to private car employment accessibility in residential areas | 0.25 | 0.12 |
| Housing price control variable | Structure | | Area/m$^2$ | Housing area | 86.59 | 45.59 |
| | | | Bedroom number | Number of bedrooms included in a house | 2.11 | 0.57 |
| | | | Living room number | Number of living rooms included in a house | 1.18 | 0.38 |
| | | | Orientation | House facing south = 1; otherwise = 0 | 0.78 | 0.26 |
| | | | Building story/floor | The total number of floors in the building where the house is located | 12.00 | 6.76 |
| | | | Floor location | House located below 2/3 of the total building floors = 1; otherwise = 0 | 0.62 | 0.19 |
| | | | Taxes | High transaction tax paid for purchasing it = 1; otherwise = 0 | 0.62 | 0.23 |
| | | | Housing age | Years since construction | 18.69 | 10.09 |
| | Location | | Distance from important city centers/km | The straight-line distance between the house and the nearest important city center 3) | 5.21 | 3.25 |
| | | | Trading volume/time | The total amount of housing transactions in the same community | 60.63 | 94.07 |
| | Neighborhood | | Shopping point | There are commercial facilities within 1 000 m = 1; otherwise = 0 | 0.82 | 0.39 |
| | | | School | There are key primary schools within 1 000 m = 1; otherwise = 0 | 0.24 | 0.43 |

1) *n* is a collection of POI types classified according to the Urban Land Classification and Standard for Planning Construction Land (GB 50137-2011), *n* = 8; 2) The classification standard for *m* is Classification of National Economic Industries (GB/T 4754-2017), *m* = 20; 3) The important urban centers include the regional center (Tiananmen Square in the center of Beijing, People's Government of Beijing City in the sub center) and the business centers (CBD, Zhongguancun, Financial Street, Shangdi, Wangjing, Yizhuang, and Fengtai Science Park).

### 3.2 Nonlinear model

The machine learning model XGBoost is an ensemble tree model improved based on the gradient boosting technology. The core of its solution is to use the residual value of the ($d$ – 1)-th round of the decision tree to fit the $d$-th round of the basic decision tree, and calculate the result of all decision trees after reaching the iteration number. It has good modeling ability for the nonlinear characteristics of various types of data. Therefore, a feature pricing nonlinear model based on XGBoost was adopted to study the nonlinear impact of the TOD built environment on house pricing. The calculation formula is

$$\hat{y}_x^{(d)} = \sum_{p=1}^{d} f_p(m_x) = \hat{y}_x^{(d-1)} + f_d(m_x) \text{ ,} \quad (3)$$

where $\hat{y}_x^{(d)}$ is the predicted housing price of sample $x$ after the $d$-th iteration; $p$ is the number of base models; $f_p(m_x)$ is the model of the $p$-th decision tree; $m_x$ represents the house pricing control variable and TOD built environment variable determined in Tab. 1; $\hat{y}_x^{(d-1)}$ is the predicted housing price of the previous ($d$ – 1)-th decision tree set; $f_d(m_x)$ is the model of the $d$-th decision tree.

### 3.3 Model explanation

#### 3.3.1 SHAP local interpretation method

This article used the local interpretation method SHAP to explain the XGBoost model. SHAP draws on the Shapley value in the cooperative game theory to explain the prediction of a specific instance by calculating the marginal contribution of a certain feature to the model's prediction. Compared to conventional importance algorithms that are limited to obtaining results within the entire population (such as machine learning's built-in feature importance ranking), SHAP can obtain results on each instance, thus reflecting the spatial feature of contribution. The calculation method for the Shapley value is

$$\varphi_k(f) = \sum_{S \subseteq N \setminus \{k\}} \frac{|S|!\left(K-|S|-1\right)!}{|K|!}\left[f\left(S \cup \{k\}\right) - f(S)\right] \text{ ,} \quad (4)$$

where $\varphi_k(f)$ is the Shapley value of feature $k$, reflecting the impact of feature $k$ on an individual prediction and the average value of overall predictions; $N$ is the set of all features in the training set; $N\setminus\{k\}$ is the set of permutations and combinations of all features except for feature $k$; $S$ is one of the combinations of $N\setminus\{k\}$; $K$ is the total number of features/piece; $f(S\cup\{k\})$ is the predicted value containing feature $k$; $f(S)$ is the predicted value without feature $k$.

SHAP extends the concept of the Shapley value and constructs an explanatory model of additivity, and considers all features as "contributors". The calculated SHAP value emphasizes the contribution of each feature in the individual sample to improving the overall predictive ability of the model, that is, the importance of the feature. If the SHAP value is positive, it contributes positively to the model prediction, and vice versa, which can be used to analyze the importance, direction of influence, and interaction effects of the feature.

#### 3.3.2 Importance of features

The importance of the explanatory variable is the most important explanatory tool for the machine learning model, and the larger its value, the higher its corresponding importance. The global feature importance can be obtained by calculating the average absolute value of all SHAP values corresponding to each feature, that is, one feature corresponds to one global feature importance, thus ranking the feature contribution. The calculation method is

$$I_k = \sum_{x=1}^{M}\left|\varphi_k^{(x)}\right|/M \text{ ,} \quad (5)$$

where $I_k$ is the global importance of feature $k$; $M$ is the number of samples; $\varphi_k^{(x)}$ is the SHAP value of feature $k$ in a single sample $x$.

#### 3.3.3 SHAP partial dependence plot

The SHAP partial dependency plot is an important tool for interpreting the prediction results of the machine learning model. It intuitively reflects the relationship between the dependent variable and the explanatory variable through a scatter plot of SHAP values, combining the intuitiveness of the partial dependency plot with the interpretability of the SHAP value. This article used the SHAP partial dependency plot to reveal the local pattern of the impact of the TOD built environment on the land value.

## 4 Results analysis

Taking Beijing as a case study, this paper used the XGBoost toolbox in Python to solve the regression gradient boosting tree ensemble model. The dataset was divided into a training set (70%) and a testing set (30%), and the optimal combination of hyperparameters was calculated using grid search and five-fold cross validation. The final model consisted of 800 decision trees with a maximum depth of 5, with a learning rate of 0.1 and an explanatory power of 84.2%.

### 4.1 Relative importance of explanatory variable

The global feature importance of each feature was calculated using Eq. (5), which was normalized to form the relative importance (%), with a sum of 100%. We took the average relative importance of the variables under each classification and sorted them, with the results shown in Table 2. The ranking results show that the impact of TOD built environment variables on the land value reaches 64.30% and accessibility is the most important variable, followed by density, design, and diversity. Specifically, the relative importance of public transportation accessibility ranks first, reaching 17.66%, followed by the population density and walkable road density, which are 8.76% and 8.04%, respectively, and then the building density of 6.73%. In addition, the relative importance of proximity to the subway station (1.80%) is much lower than that of accessibility, indicating that the improvement of the land value requires an increase in accessibility rather than just proximity to transportation facilities. Compared to the "last kilometer" problem, the ability to use public transportation to obtain services is more important, that is, the convenience to obtain various types of opportunities such as the service facility, commercial center, and workplace, is more important. The relative importance of relative employment accessibility ranks tenth, indicating that the competitiveness of public transportation and private cars during commuting is also an important variable affecting land value. The proportion of shuttle by bike ranks 11th, indicating that the "urban rail transit + shared cycling" model plays a significant role in urban transportation, but its influence is limited and can only serve as a supplement to other important influencing factors.

In addition, the ranking of the impact of different variables on the land value varies depending on the location. The top 5 variables in terms of importance are listed by the administrative region (see Fig. 1). The population density ranks in the top two in Daxing District, Fangshan District, Changping District, Tongzhou District, and Shunyi District, while its influence is not high in Dongcheng District, Xicheng District, and Chaoyang District. It indicates that population aggregation at suburban subway stations has a more significant effect on improving the land value compared to the central urban area; The development intensity around stations in Dongcheng District, Xicheng District, and Chaoyang District has a significant impact on the land value, while in Fengtai District, Haidian District, Changping District, and Tongzhou District, we need to pay special attention to optimizing the non-motorized transportation environment of stations.

## 4.2 Threshold effect and spatial distribution characteristics of nonlinear effects

### 4.2.1 TOD urban construction

The top three representative variables in the TOD urban construction dimension were selected for specific analysis: population density, walkable road density, and building density, ranking third (8.76%), fourth (8.04%), and fifth (6.73%), respectively, which belong to the density and design categories. Fig. 2 shows the SHAP partial dependency plots of three variables, where the blue scatter points represent the SHAP values; the dark blue solid lines represent polynomial fitting curves, and gray bar charts represent the numerical distribution of sample frequencies. It can be seen that there is a significant nonlinear relationship and threshold effect between the three variables of the TOD urban construction dimension and land value.

The changes in the curves of Figs. 2a and 2b can be roughly divided into three stages: In the early stage, they show a step-like steady climb, and in the later stage, they tend to stabilize, where the variable value does not have a significant impact on the land value when it reaches about the threshold of 19 000 persons·km$^{-2}$ and 16 km·km$^{-2}$, respectively. The results indicate that there are upper and lower limits to the impact of TOD built environment on land value uplift. In the early stage of land development, human aggregation is needed as a support, and moderate population aggregation around the station increases the land value, but beyond the threshold, the value-added effect decreases significantly.

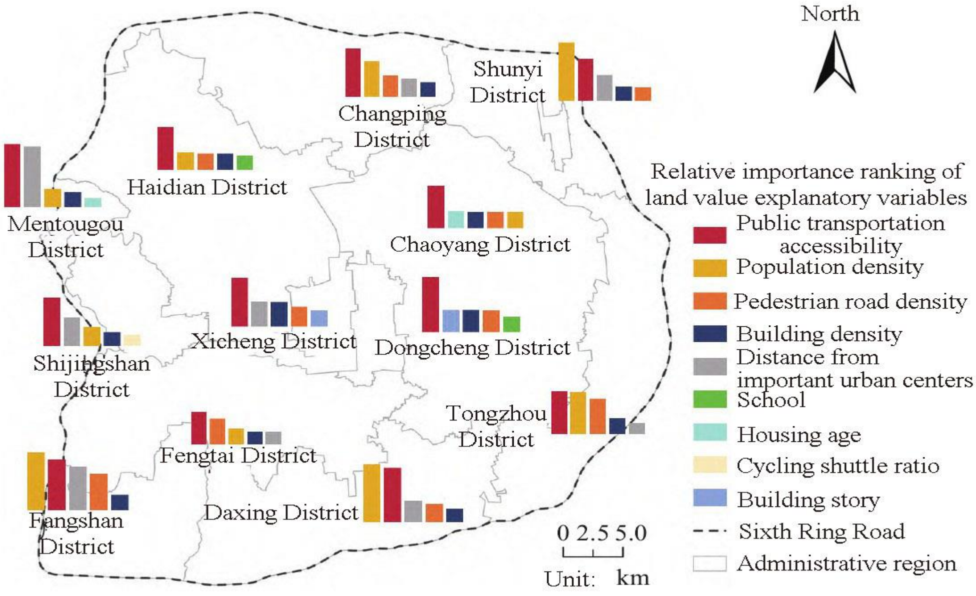



**Fig. 1** Top five relative importance of land value explanatory variables in various regions of Beijing

**Tab. 2** Relative importance ranking of explanatory variables

<table>
<tr><th colspan="3">Classification</th><th>Variable</th><th>Relative importance /%</th><th>Individual ranking</th><th>Sum of relative importance /%</th><th>Mean/%</th><th>Ranking of the type</th></tr>
<tr><td rowspan="16">TOD built environment variable</td><td rowspan="2" colspan="2">Proximity</td><td>Proximity to subway stations</td><td>1.80</td><td>21</td><td rowspan="2">3.36</td><td rowspan="2">1.68</td><td rowspan="2">8</td></tr>
<tr><td>Proximity to bus stops</td><td>1.56</td><td>24</td></tr>
<tr><td rowspan="11">TOD urban construction</td><td rowspan="5">Density</td><td>Population density</td><td>8.76</td><td>2</td><td rowspan="5">21.16</td><td rowspan="5">4.23</td><td rowspan="5">3</td></tr>
<tr><td>Employment density</td><td>2.02</td><td>17</td></tr>
<tr><td>Building density</td><td>6.73</td><td>4</td></tr>
<tr><td>Plot ratio</td><td>1.82</td><td>20</td></tr>
<tr><td>Compactness</td><td>1.83</td><td>19</td></tr>
<tr><td rowspan="2">Diversity</td><td>Land use mixing</td><td>2.21</td><td>13</td><td rowspan="2">3.80</td><td rowspan="2">1.90</td><td rowspan="2">7</td></tr>
<tr><td>Economic industry mixing</td><td>1.59</td><td>23</td></tr>
<tr><td rowspan="4">Design</td><td>Non-motorized traffic index</td><td>1.78</td><td>22</td><td rowspan="4">13.87</td><td rowspan="4">3.47</td><td rowspan="4">4</td></tr>
<tr><td>Pedestrian road density</td><td>8.04</td><td>3</td></tr>
<tr><td>Bus stop density</td><td>1.23</td><td>27</td></tr>
<tr><td>Cycling shuttle ratio</td><td>2.82</td><td>11</td></tr>
<tr><td rowspan="3" colspan="2">Accessibility</td><td>Public transportation accessibility</td><td>17.66</td><td>1</td><td rowspan="3">22.11</td><td rowspan="3">7.37</td><td rowspan="3">1</td></tr>
<tr><td>Relative accessibility</td><td>1.34</td><td>25</td></tr>
<tr><td>Relative employment accessibility</td><td>3.11</td><td>10</td></tr>
<tr><td rowspan="12">Housing price control variable</td><td rowspan="8" colspan="2">Structure</td><td>Area</td><td>3.48</td><td>9</td><td rowspan="8">21.91</td><td rowspan="8">2.74</td><td rowspan="8">5</td></tr>
<tr><td>Bedroom number</td><td>1.94</td><td>18</td></tr>
<tr><td>Living room number</td><td>2.15</td><td>16</td></tr>
<tr><td>Orientation</td><td>2.20</td><td>14</td></tr>
<tr><td>Building story</td><td>3.59</td><td>8</td></tr>
<tr><td>Floor location</td><td>1.31</td><td>26</td></tr>
<tr><td>Taxes</td><td>2.16</td><td>15</td></tr>
<tr><td>Housing age</td><td>5.08</td><td>6</td></tr>
<tr><td rowspan="2" colspan="2">Location</td><td>Distance from important urban centers</td><td>6.72</td><td>5</td><td rowspan="2">9.48</td><td rowspan="2">4.74</td><td rowspan="2">2</td></tr>
<tr><td>Trading volume</td><td>2.75</td><td>12</td></tr>
<tr><td rowspan="2" colspan="2">Neighborhood</td><td>Shopping point</td><td>0.06</td><td>28</td><td rowspan="2">4.30</td><td rowspan="2">2.15</td><td rowspan="2">6</td></tr>
<tr><td>School</td><td>4.24</td><td>7</td></tr>
</table>

In the early stage of Fig. 2c, when the building density is less than 20%, the SHAP value increases very slowly. After it reaches about 23%, the SHAP value becomes positive and the increase significantly rises, for the development of station areas above this value has reached a certain level, and multifunctional buildings have formed, which is conducive to gathering the passenger flow and facilitating production activities around the subway station, thereby promoting the increase of the surrounding land value.

The impact of the TOD built environment on the land value may vary due to the location, therefore further mapping the SHAP value of the above variables to space for visual comparative analysis (see Fig. 3).

In Fig. 3a, the population density in the central area of the city has a significant positive driving effect on the land value, but its distribution is relatively scattered, reflecting to some extent the multi-center development pattern of Beijing. Fig. 3b shows that the pedestrian supporting facilities around subway stations in the north are more complete than those in the south, especially in the southeast of Haidian District, but those in the peripheral areas of the city are generally poor, such as in Qinghe, Fengtai Science Park and Muxiyuan, focus should be on strengthening the pedestrian connectivity design. Fig. 3c illustrates the suppression of the land value in the Tongzhou area due to the low building density, and attention should be paid to the development of buildings along subways Line 1 and the Batong Line. The low SHAP value in the northern and southern parts of Chaoyang District is due to that stations developed with medium density cannot meet the living needs and commercial activities of residents in the area.

Therefore, special attention should be paid to the gathering of resources towards stations in the peripheral areas of a city to enhance the vitality of land use around stations, while stations in the central urban areas should focus on improving pedestrian and environmental quality.

### 4.2.2 Dimensions of proximity and accessibility

As important features representing TOD transportation, proximity and accessibility constitute multi-scale public transportation proximity in space and time, corresponding to the availability and accessibility of public transportation, which reflect the convenience and quality of obtaining public transportation services respectively, playing an important role in urban development and the land market.

The relative importance of public transportation accessibility ranks first (17.66%), and its SHAP partial dependence plot is shown in Fig. 4a. Public transportation accessibility has significant threshold effects around 10 000 and 30 000, respectively. This trend indicates that the improvement of public transportation accessibility has a positive impact on the land value, and there are upper and lower limits, with a positive impact after about 15 000; ultimately, its positive impact on house pricing tends to stabilize.

The proximity of the station is reflected in the availability of public transportation services in space and the relative importance of subway station proximity ranks 21st (1.80%). Fig. 4b shows a significantly negative correlation between it and the land value within the sample aggregation range of 0–0.2 km. If it exceeds about 800 m, it has a negative impact on the increase in the land value, indicating that within 800 m is the optimal distance to leverage the advantages of nearby subway stations.

The spatial distribution of SHAP values for accessibility and proximity variables was analyzed (Fig. 5). It can be seen that the overall SHAP value of public transportation accessibility shows a decreasing trend with positive values in the city center and negative values in the city periphery, similar to the distribution of feature values. Specifically, the advantage of public transportation accessibility and its promotion of the land value increase in a circular manner towards the city center, and the accessibility of public transportation in the peripheral areas of the city, especially in Shijingshan, Daxing, and the west side of Tongzhou, urgently needs to be strengthened. The positive SHAP values for the proximity of subway stations are mainly concentrated along the subway lines, but negative SHAP values appear in areas with a high concentration of subway stations in the city center, indicating that multiple modes of transportation coexist in this area, and besides the subway, there are many other modes of transportation to choose from, or they are affected by other negative factors such as subway noise and congestion.

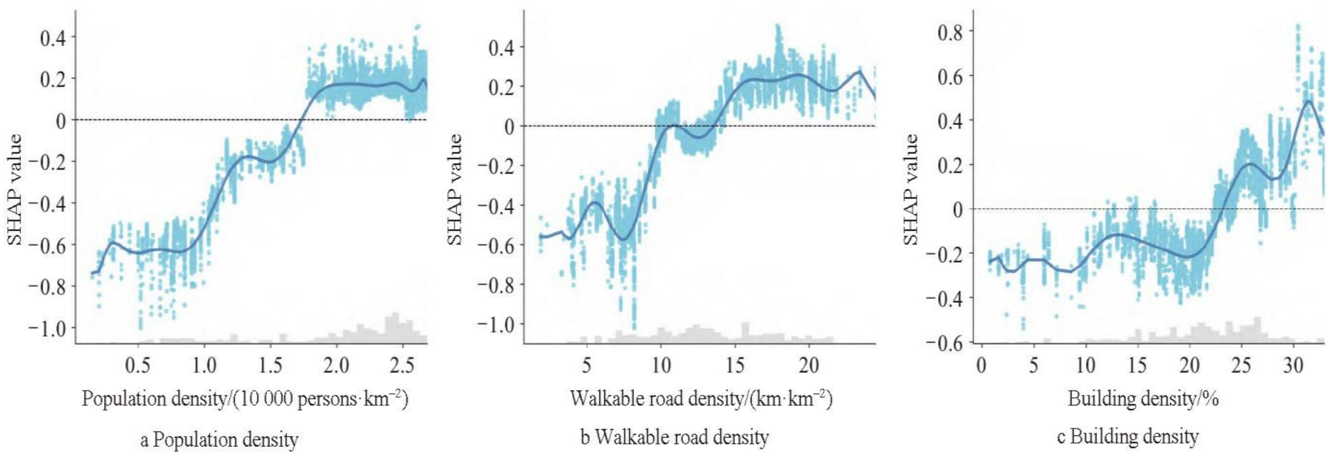



**Fig. 2** SHAP partial dependence graph of TOD urban construction dimensions variables

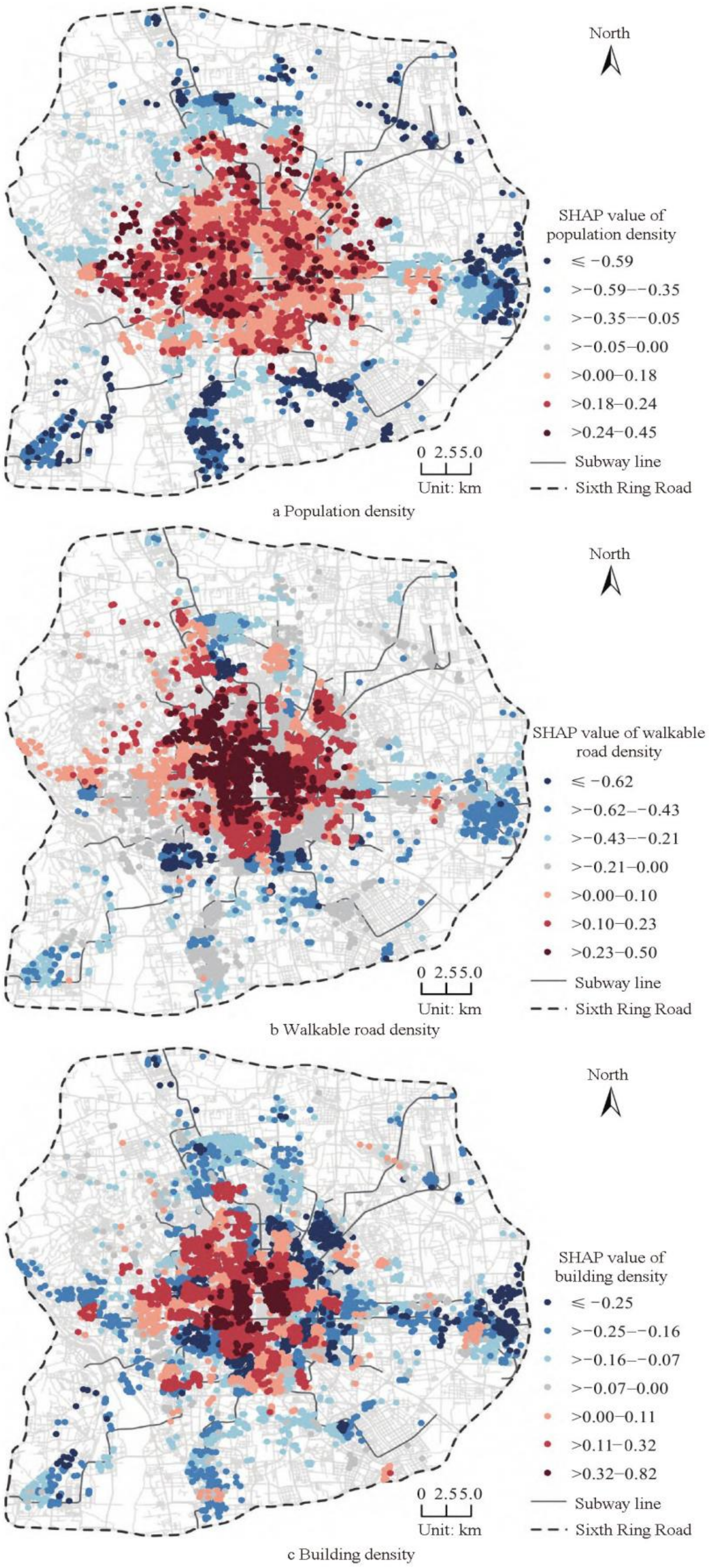



**Fig. 3** Spatial distribution of SHAP value of TOD urban construction dimensions variables

The above results show that there is a significant difference in the impact of accessibility and proximity on the land value and accessibility is more important. It indicates that in the entire trip chain, compared to the “last kilometer”, people pay more attention to the comprehensive “point-to-point” travel path, that is, the convenience of obtaining various opportunities such as service facilities, commercial centers, and workplaces, is more important.

In addition, the results indicate that the interaction between transportation facilities of the subway station and urban construction plays an important role in the impact of the TOD built environment on the land value. Accessibility comprehensively considers factors such as travel time and cost, mainly focusing on the convenience of the entire trip chain of public transportation. However, more importantly, it measures the degree of organic integration between public transportation and land use planning, emphasizing the interconnection between subway stations, bus stations, bicycle parking spots, and pedestrian paths. Compared to emphasizing the proximity of stations to the public transportation facility, the importance of accessibility lies in its more comprehensive consideration of people’s travel needs and the integrated transportation network.

Therefore, accessibility should be the main factor in judging the convenience of public transportation, rather than just considering station proximity. Meanwhile, in urban development, attention should be paid to the close integration of transportation and land use planning, and a balance should be sought based on comprehensive consideration of these two factors, making the entire travel path more convenient and improving the travel efficiency and quality of life of urban residents.

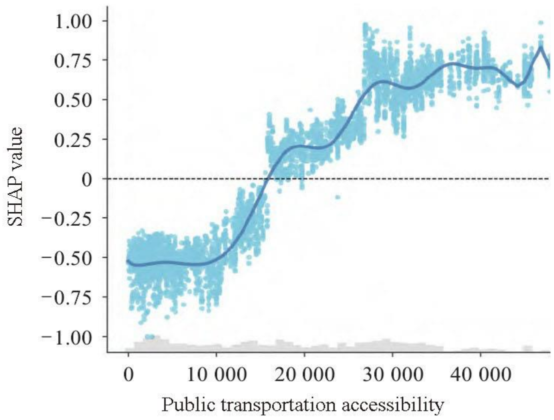
## 5 Conclusions

This article constructed a nonlinear feature pricing model based on the machine learning model XGBoost, and used the local interpretation analysis method SHAP to reveal the nonlinear impact of the TOD built environment on the land value and its spatial distribution characteristics. The results show that:

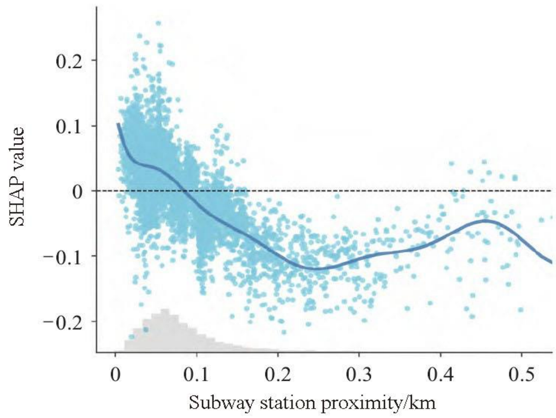
1) The non-linear impact of the TOD construction environment on the land value in the subway station is significant, with a degree of impact of 64.30%, wherein the key factor is the accessibility of public transportation, with a relative importance of 17.66%. The dimensions of transportation and urban construction in TOD are both very important factors affecting the land value, and the high importance of accessibility indicates that the key direction of urban development and renewal should focus on the organic combination of transportation and land use planning.

2) The density and design classification variables are relatively important in the dimension of TOD urban construction, and attention should be paid to the development of land around stations and transportation connections such as walking and cycling. Variables with a high relative importance ranking have significant nonlinear relationships and threshold effects with the land value, producing a positive impact. The threshold and spatial distribution characteristics of the influence of various variables provide upper and lower limit references for the development of TOD in stations, which can provide targeted guidance to policymakers, homebuyers, and investors.

3) In the dimension of transportation, proximity, and accessibility constitute the multi-scale proximity of public transportation in space and time, which has opposite effects on the housing price, and both of them have threshold effects. Within a distance of 800 m, the advantages of nearby subway stations can be maximized. In contrast, the relative importance of accessibility is significant, indicating that people pay more attention to the comprehensive “point-to-point” travel path compared to the “last kilometer”. In terms of transportation, public transportation remains a key focus for residents; the competitiveness between public transportation and private cars in commuting, and the distribution of employment opportunities are key factors affecting the land value.

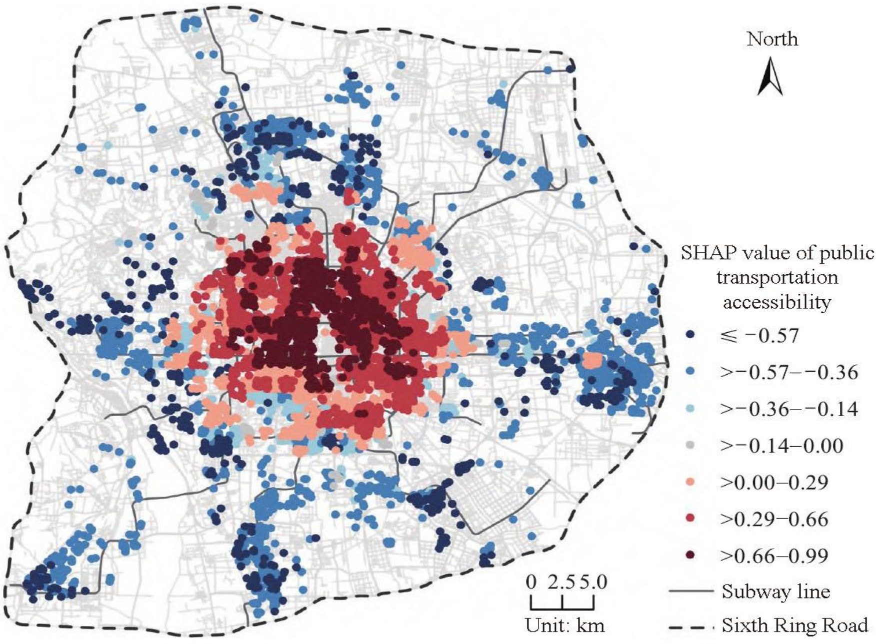



a Public transportation accessibility

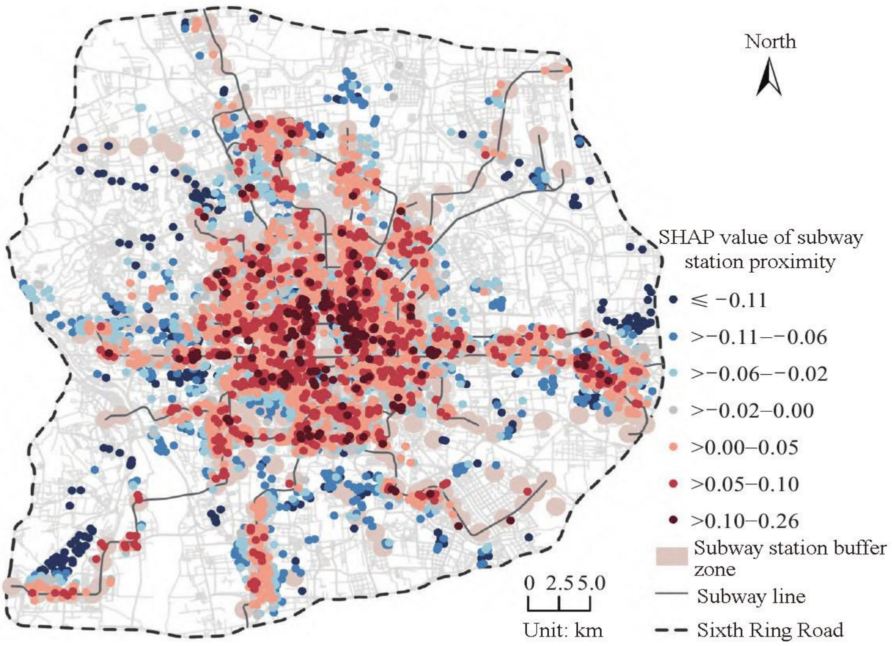



b Subway station proximity

**Fig. 4** SHAP partial dependence graph of accessibility and proximity variables




a Public transportation accessibility




b Subway station proximity

**Fig. 5** Spatial distribution of SHAP values of accessibility and proximity variables

4) The impact of the TOD built environment on the land value exhibits spatial heterogeneity and varying degrees of spatial aggregation. Special attention should be paid to the gathering of resources towards the subway station in peripheral areas of the city, and stations in central urban areas should focus on improving the travel quality and environmental quality. The development of station land in different regions should adopt differentiated development strategies tailored to local conditions, or make appropriate development adjustments based on the surrounding development plan and population needs.

## References

[1] CERVERO R. The transit metropolis: a global inquiry[M]. Washington DC: Island Press,1998.
[2] CERVERO R, DAY J. Suburbanization and Transit-Oriented Development in China[J]. Transport policy, 2008, 15(5): 315–323.
[3] WANG X, TONG D, GAO J, et al. The reshaping of land development density through rail transit: the stories of central areas vs suburbs in Shenzhen, China[J]. Cities, 2019, 89: 35–45.
[4] CAO X J, DING C, YANG J. CAO X J, DING C, YANG J. Global synthesis of transport and land use planning[M]//Urban transport and land use planning: a synthesis of global knowledge: Vol. 9[M]. London: Academic Press, 2022: 1–8.

[5] ALONSO W. Location and land use: toward a general theory of land rent[M]. Cambridge: Harvard University Press, 1964.
[6] MUTH R F. Cities and housing: the spatial pattern of urban residential land use[M]. Chicago: University of Chicago Press, 1969.
[7] MILLS E S. Studies in the structure of the urban economy[M]. Baltimore: Johns Hopkins Press, 1972.
[8] BARTHOLOMEW K, EWING R. Hedonic price effects of pedestrian-and Transit-Oriented Development[J]. Journal of planning literature,2011, 26(1): 18–34.
[9] HIGGINS C D, KANAROGLOU P S. Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg[J]. Transport reviews, 2011, 36(5): 610–634.
[10] XIONG Y F, ZHANG A L, LIU M B. Spatial-temporal effect and heterogeneity analysis of the impact of rail transit on housing prices in Wuhan: from the perspective of network structure and scale[J]. China land science, 2022, 36(12): 47–57. (in Chinese)
[11] XU T, TAO J. Research on the urban transit access premium variety: theory, phenomenon and mechanism[J]. Modern urban research, 2020(9): 116–123. (in Chinese)
[12] MOHAMMAD S, GRAHAM D, MELO P, et al. A meta-analysis of the impact of rail projects on land and property values[J]. Transportation research part A: policy and practice, 2013, 50: 158–170.
[13] SEO K, GOLUB A, KUBY M J. Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona[J]. Journal of transport geography, 2014, 41: 53–62.
[14] ZHANG S J, XU Q, JIA S P, et al. Spatial and temporal effects of new urban rail transit lines on residential property value uplift[J]. Journal of transportation systems engineering and information technology, 2021, 21(4): 54–62. (in Chinese)
[15] YANG L C, CHAU K W, SZETO W Y, et al. Accessibility to transit, by transit, and property prices: spatially varying relationships[J]. Transportation research part D: transport and environment, 2020, 85: 102387.
[16] SU S, ZHANG J, HE S, et al. Unraveling the impact of TOD on housing rental prices and implications on spatial planning: a comparative analysis of five Chinese megacities[J]. Habitat international, 2021, 107: 102309.
[17] WANG S T, ZHENG S Q, FENG J. Spatial accessibility of housing to public services and its impact on housing price: a case study of Beijing's inner city[J]. Progress in geography, 2007(6): 78–85. (in Chinese)
[18] HIGGINS C D, KANAROGLOU P S. Rapid transit, Transit-Oriented Development, and the contextual sensitivity of land value uplift in Toronto[J]. Urban studies, 2018, 55(10): 2197–2225.
[19] JIN T, CHENG L, LIU Z, et al. Nonlinear public transit accessibility effects on housing prices: heterogeneity across price segments[J]. Transport policy, 2022, 117: 48–59.
[20] LI J, HUANG H. Effects of Transit-Oriented Development (TOD) on housing prices: a case study in Wuhan, China[J]. Research in transportation economics, 2020, 100813: 80.
[21] GU Y Z, ZHENG S Q. The impacts of rail transit on property values and land development intensity: the case of No.13 line in Beijing[J]. Acta geographica sinica, 2010, 65(2): 213–223. (in Chinese)
[22] JIANG Y, GU P Q, CAO Z J, et al. Impact of Transit-Oriented Development on residential property values around urban rail stations[J]. Transportation research record: journal of the transportation research board, 2020,2674: 362–372.
[23] DUNCAN M. The impact of Transit-Oriented Development on housing prices in San Diego, CA[J]. Urban studies, 2011, 48: 101–127.
[24] LIN X, NIU B, LIU W, et al. Land premium effects of urban rail transit and the associated policy insights for TOD: a case of Ningbo, China[J]. Urban rail transit. 2022, 8(3–4): 157–166.
[25] EWING R, CERVERO R. Travel and the built environment: a synthesis[J]. Transportation research record: journal of the transportation research board, 2001, 1780: 87–114.
[26] SHEN T Y, YU H C, ZHOU L, et al. On hedonic price of second-hand houses in Beijing based on multi-scale geographically weighted regression: scale law of spatial heterogeneity[J]. Economic geography, 2020, 40(3): 75–83. (in Chinese)