Exploring the Spatial Discrete Heterogeneity of Housing Prices in Beijing, China, Based on Regionally Geographically Weighted Regression Affected by Education

Zengzheng Wang \(^1,2\), Fuhao Zhang \(^1,2,\ast\) and Yangyang Zhao \(^2\)

1. Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610000, China
2. Research Centre of Geospatial Big Data Application, Chinese Academy of Surveying and Mapping, Beijing 100830, China

\ast\ Correspondence: zhangfh@casm.ac.cn

Abstract: Spatial heterogeneity analysis of housing prices, in general, is crucial for maintaining high-quality economic development in China, especially in the post-COVID-19 pandemic context. Previous studies have attempted to explain the associated geographical evolution by studying the spatial non-stationary continuous heterogeneity; however, they ignored the spatial discrete heterogeneity caused by natural or policy factors, such as education, economy, and population. Therefore, in this study, we take Beijing as an example and consider educational factors in order to propose an improved local regression algorithm called the regionally geographically weighted regression affected by education (E-RGWR), which can effectively address spatial non-stationary discrete heterogeneity caused by education factors. Our empirical study indicates that the \(R^2\) and \(R^2_{\text{adj}}\) values of E-RGWR are 0.8644 and 0.8642, which are 10.98\% and 11.01\% higher than those of GWR, and 3.26\% and 3.27\% higher than those of RGWR, respectively. In addition, through an analysis of related variables, the quantitative impacts of greening rate, distance to market, distance to hospitals, and construction time on housing prices in Beijing are found to present significant spatial discrete heterogeneity, and a positive relationship between school districts and housing prices was also observed. The obtained evaluation results indicate that E-RGWR can explain the spatial instability of housing prices in Beijing and the spatial discrete heterogeneity caused by education factors. Finally, based on the estimation results of the E-RGWR model, regarding housing prices in Beijing, we analyze the relationships between enrollment policy, real estate sales policy, and housing prices, E-RGWR can provide policy makers with more refined evidence to understand the nature of the centralized change relationship of Beijing’s housing price data in a well-defined manner. The government should not only carry out macro-control, but also implement precise policies for different regions, refine social governance, promote education equity, and boost the economy.

Keywords: GWR; RGWR; spatial discrete heterogeneity; house price valuation

1. Introduction

In recent years, in addition to the obvious spatial non-stationary features, the housing price changes in China’s major cities are also directly affected by the macro-economy and policies [1−3]. China’s macro-policy of “housing is for living in, not for speculation” and new housing policies for school districts in Beijing, Shanghai, and other places have led to a slowdown in the growth rate of housing prices in large cities, which had risen rapidly in recent years. It has been confirmed that the Coronavirus Disease 2019 (COVID-19) pandemic has posed significant impacts on the global economy [4], which has led to varying degrees of stagnation or even recession in the development of the global economy [5−8]. As the real estate market is one of the most active markets in China and plays a vital role in the Chinese economy, exploring the spatial heterogeneity of housing prices is
considered to be crucial for revealing major issues in real estate market development, understanding effective strategies of economic macro-control, and promoting the high-quality development of internal economics in the post-COVID-19 pandemic era [9–12].

The hedonic pricing model is a widely-used approach to estimate the quantitative relationship between housing characteristics, the surrounding environment, geographical location, and housing prices [13]. In 1972, Rosen proposed the hedonic model [14], which aims to measure real estate prices with respect to many environmental factors. In the hedonic model, the structural attributes of the house determine its basic functions and have a direct effect on its price. The area of the house, floor area ratio, greening rate, number of bathrooms, construction time, and other factors are usually considered as house structure attributes [15–17]. Over time, more and more independent variables have been taken into account and more statistical measures have been added to test the validity of the model. Further studies have partially incorporated the surrounding environment of the house [18], reflecting the accessibility of amenities and the socioeconomic status of communities. The influence of supermarkets, shopping centers, primary schools, and bus stations contribute greatly to explaining housing prices [19,20]. Measures of these factors are able to estimate the positive or negative correlation between each independent variable and house price more accurately.

However, the hedonic pricing model is a global stationary model that constantly encounters problems when dealing with spatial heterogeneity (i.e., the same independent variable has different effects on house prices in different regions) [2]. To account for spatial heterogeneity, spatial statistics have changed from global matching to local models. Brunsdon has proposed geographic weighted regression (GWR) [21–23], which is built on the premise of Tobler’s famous first law of geography—that “everything is related to everything else, but near things are more related than distant things” [24]—and the coefficients fully take into account the influence of adjacent data points. Since GWR has been used for real-estate research, the non-stationary nature and spatial heterogeneity of housing prices have been widely revealed. Subsequently, an increasing number of studies have explored the non-stationary nature and spatial heterogeneity of housing prices from different perspectives [1–3,25–32]. For example, Yu et al. have employed GWR to examine the spatial dependence and heterogeneity of housing-market dynamics in the city of Milwaukee [29]. Bitter et al. have compared two approaches to examine spatial heterogeneity in housing attribute prices within the Tucson, Arizona housing market: the spatial expansion method and GWR [30]. They found that GWR could outperform the spatial expansion method in terms of explanatory power and predictive accuracy. Yang et al. have used travel distance, instead of the original European distance, to estimate house prices [25]. Recently, Li and Wei have used GWR to examine the local geographies of the housing value bust (2008–2012) and boom (2012–2016) since the financial crisis in Salt Lake County, Utah [31]. Wang et al. have used geographically neural network weighted regression (GNNWR) to improve the accuracy of real estate evaluation [2]. Li et al. have used the GWR model to investigate the spatial impacts of COVID-19 on housing price changes in the U.S. real estate market [32].

The ability of GWR to express spatial heterogeneity is limited. The specification of spatial heterogeneity can be classified into continuous heterogeneity and discrete heterogeneity [33]. Existing GWR models and extended models can detect spatially continuous heterogeneity, to a certain extent, in practical applications by means of bandwidth optimization [34]; however, they cannot reveal the discrete heterogeneity [35]. In fact, housing prices not only have the characteristics of spatial non-stationary, but also spatial discrete heterogeneity, which is particularly obvious in China, where real estate prices are directly affected by the macro-economy and policies. The selling prices of houses near high-quality schools are affected by the neighborhood-level attribute space [36–39]. In the current stage of compulsory education, most cities have the concept of “school district” in China, but, unlike other countries, residents can send their children to schools in school districts only when they buy houses in school districts, while renters are excluded. At the same time,
high-quality educational resources in Chinese cities are very scarce. For example, the number of high-quality primary schools in Beijing, the capital, accounts for less than 10% of all primary schools. There is a large gap between the quality of high-quality primary schools and ordinary primary schools [40]. Enrollment in primary and secondary schools is strictly in accordance with the policy of “nearby enrollment” to where the child’s registered residence is located. Therefore, many parents are willing to pay high housing prices for high-quality schools. As a result, the current prices of housing in the school district are much higher than those in adjacent regions [41]. Spatial discretization causes proximity-based heterogeneity. Few studies have simultaneously investigated both spatial continuous and discrete heterogeneity in depth. Our research aims to fill this gap and to provide valuable reference for the healthy development of the real estate market.

Beijing is a representative city in the study of housing prices in China. Many studies in the literature have found that, in Beijing, education is the main reason for the locally discrete nature of housing prices. The price of second-hand school district housing is significantly higher than that of neighboring second-hand non-school district housing [40,42,43]. The objective of this study is to expand the traditional GWR model to detect the spatial non-stationary characteristics and spatial discrete heterogeneity of housing prices simultaneously in the research area, considering the spatial discrete nature caused by specific educational factors. This study seeks to contribute to the literature on the topic in the following three ways. First, we propose the education-affected regionally geographically weighted regression (E-RGWR) model, which uses regional education factors as variables to reflect regional differences. Second, we examine and compare the effects of GWR, RGWR, and E-RGWR in determining modeling housing prices by means of a case study in Beijing city, China. Third, through the expanded GWR model (E-RGWR), the project can more accurately identify the differences in housing prices and their influencing factors between different regions, providing support for policy makers to understand the influence of different factors on China’s housing price data in the post-COVID-19 pan-demic era in order to achieve accurate analysis and precise policy implementation.

The remainder of this paper is structured as follows. In Section 2, we briefly describe the study area, data sources, and methods, using the E-RGWR approach. In Section 3, we use the GWR, RGWR, and E-RGWR models to empirically compare the case study results and analyze the influencing factors. Finally, the discussion and conclusions of the paper are reported in Sections 4 and 5, respectively.

2. Data Sources and Methods

2.1. Study Area and Data

We selected Beijing as the study area. Located in northern China, Beijing is the capital, political center, and cultural center of China. Beijing is one of the most populated cities in China, having many primary schools. This large number of primary schools comes with a diversity of educational quality and social reputation. Moreover, unlike other cities such as Suzhou, where high-quality primary schools are often located together and close to each other, high-quality primary schools in Beijing are relatively far away from each other, and there are typically ordinary primary schools located in between. Beijing consists of sixteen administrative regions, and, in this paper, we restrict the analysis to the main urban districts covered by the subway in the central area of Beijing, including Dongcheng, Xicheng, Chaoyang, Fengtai, Shijingshan, Haidian, Mentougou, Fangshan, Tongzhou, Shunyi, Changping, and Daxing. The listed sale prices of residential buildings in Beijing are used as the characteristic price data to construct the characteristic price model and carry out the experimental analysis.

2.1.1. Housing Hedonic Price Data

We restrict the scope of the analysis in this paper to the main urban districts covered by the subway in the central area of Beijing, with the sampling points mainly including 7282 communities in the main urban areas within the sixth ring road. We obtained the
average listing price (yuan/square meter) for each community in April 2022, as well as information on characteristics of selected residential districts that may affect house prices. Their geographic locations are shown in Figure 1. Specifically, characteristic information including construction time, floor area ratio, greening rate, and property management fee were collected, and the characteristic price data of houses were obtained from Anjuke (https://www.anjuke.com/; last accessed on 22 April 2022). In addition, the coordinate data of each community were obtained by calling the Gaode map API.

Figure 1. Distribution map of sampling points in Beijing.

2.1.2. Location POI and Base Map Data

We collected data on points of interest such as bus stations, subway stations, hospitals, parks, shopping malls, universities, firefighting, and public security in Beijing. The point of interest data were obtained from Gaode (https://www.gaode.com/; last accessed on 26 April 2022). In addition, data on the administrative areas, main roads, and water of Beijing came from China Map Publishing House.

2.1.3. Middle and Primary School Data

We obtained the statistical reports of the 2021–2022 school year for each middle and primary school in each district of Beijing from the official website of the Beijing Municipal Education Commission (http://jw.beijing.gov.cn/; last accessed on 26 April 2022), and obtained the basic data of each middle and primary school, including the number and name of schools. The coordinate data corresponding to each school were obtained by calling the
According to the data provided by the platform, there are 837 primary schools and 667 middle schools in Beijing [44].

2.2. Research Methodology

2.2.1. Geographically Weighted Regression (GWR)

Geographically weighted regression is based on the first law of geography [21–23] and is used to explore the non-stationary nature of space, change the regression coefficient from global to local, embed the geographical location in the regression parameters, and change the weights of adjacent points according to different distances in the regression framework. The model is expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \epsilon_i, \quad i = 1, 2, \ldots, n$$

where $\beta_k(u_i, v_i)x_{ik}$ is the $k$-th regression parameter at the $i$-th sampling point. We used the local weighted least squares approach, and the regression parameter estimation $\hat{\beta}_i$ at point $i$ is given by Formula (2):

$$\hat{\beta}_i = (X'W_iX)^{-1}X'W_iy$$

The estimated values in the matrix $\hat{y}$ can be expressed as:

$$\hat{y}_i = X_i\hat{\beta}_i = X_i(X'W_iX)^{-1}X'W_iy$$

The spatial weight matrix $W_i$ is expressed as:

$$W_i \triangleq \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_j \end{bmatrix}$$

The spatial weight matrix is used to express different understandings of spatial relationships between data by selecting different spatial kernel functions. The weight kernels usually involve Gaussian, bi-square, tri-cube, and exponential functions. These functions can relatively simply express the complex relationship between spatial proximity (e.g., spatial distance) and spatial non-stationarity (i.e., spatial weight) [2]. The most-used kernel function is a function based on Gaussian distance attenuation, and its function form is as follows:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{b^2}\right)$$

where $d_{ij}$ is the distance between points $i$ and $j$ and $b$ is called the bandwidth. There are two methods to choose from: fixed bandwidth and adaptive bandwidth. With a fixed bandwidth, the distance is constant, but the number of neighbors is variable. With an adaptive bandwidth, the distance will change, but the number of neighbors will remain unchanged.

2.2.2. Regionally and Geographically Weighted Regression (RGWR)

RGWR is an extension of GWR used to explore spatial non-stationary and spatially discrete heterogeneity by adding regional dummy variables to GWR [35]. The model can be expressed as:

$$y_i = R\beta_0(u_i, v_i) + \sum_{k=1}^{p} R\beta_k(u_i, v_i)x_{ik} + \epsilon_i, \quad i = 1, 2, \ldots, n$$
The spatial weight matrix $\mathbf{RW}_i$ is expressed as:

$$
\mathbf{RW}_i \triangleq \begin{bmatrix}
rw_1 & 0 & \cdots & 0 \\
0 & rw_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & rw_j \\
\end{bmatrix}
$$

(7)

The RGWR Gaussian kernel function is:

$$
rw_{ij} = \delta \exp\left(-\frac{(d_{ij}/b)^2}{2}\right)
$$

(8)

where $\delta$ is an introduced dummy variable: the regional influence factor. Meanwhile, $i$ and $j$ are located in the same region, $\delta = 1$, which is the traditional GWR; however, when $i$ and $j$ are located in different regions, $\delta = 0$, it means that $rw_{ij} = 0$, indicating the RGWR.

In addition, the bi-square kernel function is also a commonly-used weight calculation method for GWR. RGWR of the bi-square kernel function is given as:

$$
rw_{ij} = \begin{cases} 
\delta \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 & d_{ij} \leq b \\
0 & d_{ij} > b 
\end{cases}
$$

(9)

2.2.3. Regionally Geographically Weighted Regression Affected by Education (E-RGWR)

E-RGWR replaces the regional dummy variable in the RGWR method with the education measurement factor of the spatial region and introduces the regional education impact factor in order to accurately construct a model of spatial discrete heterogeneity caused by educational factors between regions. The model can be expressed as:

$$
y_i = E\beta_0(u, v_i) + \sum_{k=1}^{p} E\beta_k(u, v_i)x_{ik} + \epsilon_i \quad i = 1, 2, \ldots, n
$$

(10)

The E-RGWR spatial weight matrix $\mathbf{ERW}_i$ is expressed as:

$$
\mathbf{ERW}_i \triangleq \begin{bmatrix}
erw_1 & 0 & \cdots & 0 \\
0 & erw_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & erw_j \\
\end{bmatrix}
$$

(11)

The E-RGWR Gaussian kernel function $erw_1$ is:

$$
erw_{ij} = er_{ij} \ast w_{ij} = er_{ij} \ast \exp\left(-\frac{d_{ij}^2}{b^2}\right)
$$

(12)

The E-RGWR bi-square kernel function is given as:

$$
erw_{ij} = \begin{cases} 
er_{ij} \ast \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 & d_{ij} \leq b \\
0 & d_{ij} > b 
\end{cases}
$$

(13)

where $er_{ij}$ denotes the regional education impact factors: when $i$ and $j$ are located in the same region, $er_{ij} = 1$; when $i$ and $j$ are located in different regions, the regional impact factors are calculated in different regions according to the Y-value, as depicted in Figure 2.
In the case of this study, we propose the measurement method of regional education impact factors according to the spatial weight of regional economic distance proposed by Lin et al. [45]:

$$er_{ij} = 1 - \frac{|er_i - er_j|}{er_i + er_j} \quad (14)$$

$$er_i = \frac{q_i}{q_{all}}, \quad er_j = \frac{q_j}{q_{all}} \quad (15)$$

where $q_i$ is the number of high-quality educational resources in the region where $i$ is located and $q_{all}$ is the number of all educational resources in the region where $i$ is located. In this study, as previously stated, educational resources refer specifically to public primary schools. When $i$ and $j$ are in the same region, $|er_i - er_j| = 0$, which gives the traditional GWR; when $i$ and $j$ are located in different regions, the E-RGWR is taken.

2.2.4. Model Evaluation of Performance

We consider the following metrics to evaluate the performance of the model. First, for bandwidth selection, we adopt the Akaike information criterion, corrected (AICc) [22], defined as:

$$AICc = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n\left[\frac{n + tr(S)}{n - 2 - tr(S)}\right] \quad (16)$$

Other measures of model performance [46] include root mean square error (RMSE), mean square error (MSE), coefficient of determination ($R^2$), and adjusted $R^2$ ($R_{adj}^2$). Their definitions are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (17)$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (19)$$

$$R_{adj}^2 = 1 - \left[\frac{n - 1}{n - tr(S) - 1}(1 - R^2)\right] \quad (20)$$

Figure 2. Calculation diagram of E-RGWR regional education impact factors.
where $\bar{y}$ is the average of the observed values and $tr(S)$ is the effective number of parameters.

2.3. Data Preprocessing

2.3.1. Coordinates (Latitude and Longitude)

The longitude and latitude of the study area range from 115.925 to 116.7638 and from 39.62064 to 40.29134, respectively. For the neighborhoods, schools, and other POI points used in this study, we called the Gaode map API to obtain spatial coordinate data. In addition, all latitude and longitude coordinates used in this study were converted to the WGS 1984 50N coordinate system after projection conversion.

2.3.2. Structure Covariates

The structural characteristics of each community are described by the three covariates of greening rate, floor area ratio, and property fee. If the property fee was given as an interval range, the average value of upper and lower limits was considered. The greening rate, floor area ratio, and property fee were all logarithmically transformed into LNGR, LNFAR, and LNPF, respectively.

2.3.3. Temporal Covariates

The temporal covariates were the construction year (NA). We recorded the earliest construction year as base 1 and increased it year by year. If the construction year was given as an interval value, the construction completion time was considered.

2.3.4. Neighborhood Covariate

We calculated the distance of each residential area to the nearest subway station, hospital, park, market, primary school, and middle school, and transformed these into LNS, LNH, LNP, LNM, LNPS, and LNMS, respectively, by taking the logarithm. An overview of the variables involved in housing prices is given in Table 1.

Table 1. Variables used to predict housing prices in Beijing, China.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LnPrice</td>
<td>Price of the house</td>
<td>-0.8144</td>
<td>1.9288</td>
<td>3.3426</td>
</tr>
<tr>
<td>RC</td>
<td>Corresponding administrative regionalization code</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LNFAR</td>
<td>Log of the FAR</td>
<td>-1.2040</td>
<td>0.5498</td>
<td>2.2148</td>
</tr>
<tr>
<td>LNGR</td>
<td>Log of the green ratio</td>
<td>-2.9957</td>
<td>3.3834</td>
<td>4.3820</td>
</tr>
<tr>
<td>LNPF</td>
<td>Log of the property fees</td>
<td>-4.6052</td>
<td>0.1729</td>
<td>2.8904</td>
</tr>
<tr>
<td>NA</td>
<td>Normalized building age</td>
<td>0</td>
<td>76.3704</td>
<td>97</td>
</tr>
</tbody>
</table>

2.3.5. School Quality

As China has adopted the enrollment policy of the nearest entrance, primary schools and junior high schools in most cities use the school districts designated by each school as the basis for enrollment. However, relevant evidence has shown that, although this policy of dividing enrollment by school districts is generally adopted by primary schools in Beijing, it is not strictly followed in the enrollment of junior high schools [47]. Furthermore, the quality of a primary school plays a fundamental role in later academic achievement [43].
Therefore, this study focuses on the quality of primary schools rather than other educational stages, as in existing studies.

In order to promote educational equity and reduce inter-school differences, Beijing’s key school system was abolished around 2000. At the same time, Beijing does not publicly release detailed school quality data such as per capita student expenditure, teacher-student ratio, and standardized test scores. Nevertheless, manpower, material resources, policies, reputation accumulated over a long period of time, and other aspects make the “key primary schools” identified in the early days still considered as a group of schools with the best education quality in Beijing, which is still of great reference value. The public’s judgment of the quality of primary schools is also mainly based on resources such as the title of “Key Primary School” and word-of-mouth in major forums. Therefore, for this study, we combined the existing list of key primary schools in the city, the housing websites of various university districts, parent forums, and other resources to designate a total of 28 primary schools in the urban area of Beijing as indicators of quality education.

2.4. Analytical Framework Design and Implementation

The analytical framework for modeling housing prices is shown in Figure 3. The analytical framework is divided into five parts. First, is selection of the optimal bandwidth; following which, the GWR, RGWR, and E-RGWR models are used for modeling and analysis. Finally, the parameter estimation and comparison of different models, including the estimation of regression coefficients, fitting values, and evaluation measures.

Set the bandwidth value range
Build the GWR model
Building a spatial kernel function
Build the RGWR model
Establishing a spatial weight function based on the regionalization
Establishing a regional factor affected by education between point i and j
Establishing a Spatial Weight Matrix E-RW
Building a spatial kernel function
Build the E-RGWR model
Building a spatial kernel function
Calculate Regression Coefficients $\beta_i$
Calculate Spatial Weight Matrix $W_i$
Calculate the fitted value $y$
Compute and compare Model Evaluation Metrics $R^2, R^2_{adj}, MSE, RMSE, \ldots$
GWR
RGWR
E-RGWR
Figure 3. Analytical framework of the modeling methods used.
The data include independent variables, dependent variables, spatial location variables, education quality variables, and alternative bandwidth.

The pseudo-code of E-RGWR is presented in Algorithm 1.

Algorithm 1: Pseudo-code describing the E-RGWR model.

Algorithm: E-RGWR
INPUT: explanatory variables X
spatial coordinates
dependent variable Y
PROCESS
Find the optimal spatial bandwidth
For $i \in \{1, 2, \ldots, n\}$ do
construct the spatial kernel weight $W_{ij}$ between $(u_i, v_i)$ and $(u_j, v_j)$
calculate the AICc value
obtain the optimal spatial bandwidth
end for
Establish a regional factor affected by education $er_{ij}$ between $(u_i, v_i)$ and $(u_j, v_j)$
Construct the spatial kernel weight affected by regional education $E$-RW$_{ij}$ between $(u_i, v_i)$ and $(u_j, v_j)$
Calculate the fitted value $E$-R$_Y$
Calculate $R^2$, $R^2_{adj}$, MSE, and RMSE

2.5. Parameter Setting

According to the algorithm process described in Section 2.4. of this paper, we used the GWR, RGWR, and E-RGWR methods to establish characteristic price models. First, the AICc method was used to determine the optimal bandwidth of GWR, where the fixed type was 10,012.4 and the adaptive type was 900. Then, the characteristic price models of GWR, RGWR, and E-RGWR were established using the optimal bandwidth, and hypothesis tests were performed for spatial stationarity under different bandwidth strategies [22,23]. The $p$-values of the hypothesis tests were all less than 1%, indicating statistical significance. The spatial non-stationary characteristic test was carried out for the regression coefficient [34]. As shown in Table 2, the results demonstrated that property fees, greening rate, FAR, subway station, primary school, and junior high schools all presented spatial non-stationarity.

Table 2. The spatial non-stationary characteristic test of the RGWR and GWR model.

<table>
<thead>
<tr>
<th>Model</th>
<th>p-Value</th>
<th>LNPF</th>
<th>LNFAR</th>
<th>LNGR</th>
<th>NA</th>
<th>LNS</th>
<th>LNH</th>
<th>LNP</th>
<th>LNM</th>
<th>LNPS</th>
<th>LNMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR (Fixed)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>RGWR(Fixed)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>E-RGWR(Fixed)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.086 *</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>GWR(Adaptive)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>RGWR(Adaptive)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.052 *</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>E-RGWR(Adaptive)</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.062 *</td>
<td>0.000 ***</td>
<td>0.006 **</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
<td>0.000 ***</td>
</tr>
</tbody>
</table>

Note: *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

3. Results

3.1. Model Performance Comparison

3.1.1. Comparison of Model Results

The models were evaluated according to the $R^2$ value, RMSE, mean square error (MSE), adjusted $R^2$ ($R^2_{adj}$), and AICc; the experimental results are shown in Table 3. The results indicate the superiority of the E-RGWR model in the considered study area and scene. GWR achieved the worst prediction result, both under the fixed and adaptive bandwidth strategies, with minimum $R^2$ and $R^2_{adj}$ values and highest prediction error according to RMSE and MSE. In view of the obvious spatial discrete heterogeneity of the housing price distribution in the study area, GWR cannot accurately detect the internal relationship and spatial fluctuation between discrete areas. The RGWR model was capable of detecting intrinsic relationships and spatial fluctuations between discrete regions to a certain extent. Under the fixed bandwidth strategy, the $R^2$ of RGWR was increased by
7.47% compared with GWR, $R^2_{adj}$ was increased by 7.49%, and RMSE and MSE were about 26.33% and 14.18% lower than those of the GWR model, respectively. Under the adaptive bandwidth strategy, the $R^2$ and $R^2_{adj}$ of RGWR were also improved by 6.08% and 6.09% compared with GWR, and the RMSE and MSE were about 22.16% and 13.92% lower than the GWR model, respectively. The E-RGWR model showed the best prediction performance. Under the fixed bandwidth strategy, the $R^2$ and $R^2_{adj}$ of E-RGWR were increased by 11.23% and 11.01%, compared with the GWR model, and the RMSE and MSE were about 38.71% and 21.69% lower than the GWR model, respectively. Under the adaptive bandwidth strategy, the $R^2$ and $R^2_{adj}$ of E-RGWR were improved by 7.11% and 7.13% compared with the GWR model, and the RMSE and MSE were about 26.00% and 13.92% lower than the GWR model, respectively. Overall, regardless of whether we considered a fixed or adaptive bandwidth strategy, E-RGWR presented significantly improved results compared with the other models and showed better predictive ability. It can be seen that E-RGWR had the most significant improvement effect under the fixed bandwidth strategy, and the model estimated the highest efficiency, which can accurately explain the spatial evolution of housing prices in the main urban areas within the coverage of the Beijing subway. Therefore, a fixed bandwidth strategy was used for comparison and analysis of the estimation results in the following analyses.

**Table 3.** The values of the E-RGWR, RGWR and GWR models.

<table>
<thead>
<tr>
<th>Models</th>
<th>Bandwidth/Nth</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>MSE</th>
<th>RMSE</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWR (fixed)</td>
<td>10,012.4</td>
<td>0.7789</td>
<td>0.7785</td>
<td>0.0695</td>
<td>0.2637</td>
<td>1775</td>
</tr>
<tr>
<td>RGWR (fixed)</td>
<td>10,012.4</td>
<td>0.8371</td>
<td>0.8368</td>
<td>0.0512</td>
<td>0.2263</td>
<td>−212.5276</td>
</tr>
<tr>
<td>E-RGWR (fixed)</td>
<td>10,012.4</td>
<td>0.8644</td>
<td>0.8642</td>
<td>0.0426</td>
<td>0.2065</td>
<td>−1322.3</td>
</tr>
<tr>
<td>GWR (adaptive)</td>
<td>900</td>
<td>0.7848</td>
<td>0.7844</td>
<td>0.0677</td>
<td>0.2601</td>
<td>966.0492</td>
</tr>
<tr>
<td>RGWR (adaptive)</td>
<td>900</td>
<td>0.8325</td>
<td>0.8322</td>
<td>0.0527</td>
<td>0.2295</td>
<td>−895.3239</td>
</tr>
<tr>
<td>E-RGWR (adaptive)</td>
<td>900</td>
<td>0.8406</td>
<td>0.8403</td>
<td>0.0501</td>
<td>0.2239</td>
<td>−1315.7</td>
</tr>
<tr>
<td>E-RGWR (fixed)/GWR (fixed) Improvement</td>
<td>–</td>
<td>10.98%</td>
<td>11.01%</td>
<td>38.71%</td>
<td>21.69%</td>
<td></td>
</tr>
<tr>
<td>RGWR (fixed)/GWR (fixed) Improvement</td>
<td>–</td>
<td>7.47%</td>
<td>7.49%</td>
<td>26.33%</td>
<td>14.18%</td>
<td></td>
</tr>
<tr>
<td>E-RGWR (adaptive)/GWR (adaptive) Improvement</td>
<td>–</td>
<td>7.11%</td>
<td>7.13%</td>
<td>26.00%</td>
<td>13.92%</td>
<td></td>
</tr>
<tr>
<td>RGWR (adaptive)/GWR (adaptive) Improvement</td>
<td>–</td>
<td>6.08%</td>
<td>6.09%</td>
<td>22.16%</td>
<td>13.92%</td>
<td></td>
</tr>
</tbody>
</table>

3.1.2. Comparison of Model Fitting Effects

By comparing the predicted values and residuals of housing prices obtained by the E-RGWR, RGWR, and GWR models, it was possible to intuitively explore the fitting effect of the models. Figure 4 shows the residual distributions of E-RGWR, RGWR, and GWR under fixed and adaptive bandwidth policies. The X-axis represents the predicted values of different models under different bandwidth strategies, while the Y-axis represents the residuals. In Figure 4, the closer the residual point is to zero, the better the fitting effect of the model.

Under the same bandwidth strategy, the points of E-RGWR were concentrated closer to zero than those of RGWR and GWR, indicating that the fitting effect of the E-RGWR model is significantly improved, compared to those of RGWR and GWR. Similarly, between the fixed bandwidth strategy and the adaptive bandwidth strategy, the point distribution of E-RGWR under the fixed bandwidth was closer to zero than that with the adaptive bandwidth. At the same time, the $R^2$ value of the E-RGWR model under the fixed bandwidth strategy was 0.8644, 2.83% higher than that under the adaptive bandwidth. This indicates that, in the data environment of this article, the fitting effect of the E-RGWR model under the fixed bandwidth strategy was better than that of the E-RGWR model under the adaptive bandwidth strategy.
3.2. Analysis of Variables Related to House Prices

As the E-RGWR model was more effective than the GWR model and could achieve more accurate detection of the spatially discrete heterogeneity of housing prices in the research area compared to RGWR, we only focused on the local estimates of E-RGWR. E-RGWR generates local parameter estimates that reflect possible spatial continuous and discrete heterogeneity in the processes affecting residential land prices. The E-RGWR model performance and its spatial discrete heterogeneity were explored visually by mapping the local coefficient estimates of the variables. The distributions of variables derived from the E-RGWR were examined to demonstrate the significance of the proposed model in socioeconomic and public policy research.

First, we analyzed the structure covariates and temporal covariates—specifically, the log of the green ratio (LNGR), log of the FAR (LNFAR), log of the property fees (LNPF), and normalized building age (NA)—as shown in Figure 5.

The housing prices were mainly positively correlated with the LNPF, and the representative areas were the southern and northern outer fifth ring areas of Beijing. In the core area of Beijing, Haidian, Xicheng, Dongcheng, and Chaoyang, the housing prices were mainly weakly negatively correlated with the LNPF. We speculate that communities where second-hand houses are located have been established for a long time, and the property fees are generally not high. At the same time, the property fees do not represent the quality

![Figure 4](image-url) Fitting effect distribution of GWR, RGWR, and E-RGWR. (a) Predicted values and residuals of the housing prices and residual by the GWR model under the fixed bandwidth strategy; (b) predicted values of the housing prices and residuals by the RGWR model under the fixed bandwidth strategy; (c) predicted values of the housing prices and residuals by the E-RGWR model under the fixed bandwidth strategy; (d) predicted values of the housing prices and residuals by the GWR model under the adaptive bandwidth strategy; (e) predicted values of the housing prices and residuals in the RGWR model under the adaptive bandwidth strategies; (f) predicted values of the housing prices and residuals by the E-RGWR model under the adaptive bandwidth strategy.
of residential communities. Therefore, the LNPF in these areas presented no significant impact on house prices.

First, we analyze the structure covariates and temporal covariates—specifically, the log of the green ratio (LNGR), log of the FAR (LNFAR), log of the property fees (LNPF), and normalized building age (NA)—as shown in Figure 5.

The housing prices were mainly positively correlated with the LNPF, and the representative areas were the southern and northern outer fifth ring areas of Beijing. In the core areas, the impact of the NA on house prices was generally not significant. However, in some parts, (i.e., six regions) of Beijing, there was a weakly negative correlation; that is, the longer since a house has been built, the lower the house price. Outside these six regions, there was a weakly positive correlation; that is, the longer since the building has been
completed, the higher the house price. This is inconsistent with intuition and previous research [48]. We speculate, as the second-hand housing market in Beijing has a “special full five” regional policy—that is, no personal income tax will be charged for resale of second-hand housing after five years, while a large amount of personal income tax is required to be paid for resale within five years, and these taxes are often borne by the buyers—the listing prices of second-hand houses less than five years old are often much lower than those of second-hand houses in the same area that are five years old or older. The number of sub-new houses in the six districts of Beijing in recent years is very low, and most houses have been built more than five years ago. The weak negative correlation between the construction time and housing price in the six districts of the old city is in line with intuition. Outside the six regions of the city, a large number of secondary new houses cannot be sold at higher prices due to public policies, such that the NA was slightly positively correlated to the house price.

The impact of LNGR on housing prices presented obvious regional discrete differences. Dongcheng and Xicheng are located in the core area of Beijing, with a small area, complete public supporting facilities, high average education quality, and obvious location advantages. Some courtyard houses and school district houses in the area have shown greatly increased housing prices, and so the LNGR has no significant impact on housing prices, showing a feeble negative correlation. Chaoyang is the core economic area of Beijing, and there are a large number of CBDs in the area. Fengtai is interspersed with a large number of railways, military bases, and astronomic areas, and the land for residential areas is limited. As one-sided pursuit of green space may occupy road width and parking spaces, the impact of greening rate in these areas was negatively correlated with housing prices. The mountains around Mentougou have added green resources, such that the requirements for the greening rate by the community are not high. In other regions, the LNGR has a significant positive effect on housing prices, with the strongest effects in the west of Haidian, Changping, Shunyi, Daxing, and other regions. In these regions, the supply of second-hand housing is sufficient, and residential areas with low building density and large green areas are more marketable.

The relationship between LNFAR and housing prices changes from a relatively large negative correlation to a weak positive correlation from the urban core to the periphery. There was a negative correlation between the east and south of the city, and a weak positive correlation between the west and the north. In the core area of the city, the land for the residential area is tight, and the lower the floor area ratio is, the more marketable the residential area. Whereas, in areas in the northwest, which are slightly away from the core area, high FARs correspond to new high-rise residences, which is the underlying reason for the weak positive correlation in these areas.

Second, we examined the influence of the neighborhood covariates—specifically, log of the distance to the nearest subway station (LNS), log of the distance to the nearest park (LNP), log of the distance to the nearest market (LNM), log of the distance to the nearest hospital (LNH), log of the distance to the nearest primary school (LNPS), and log of the distance to the nearest middle school (LNMS)—as shown in Figure 6.

The impact of LNS on housing prices was negatively correlated globally. Ordinarily, the farther away from the subway, the lower the corresponding housing price. In the core area of the city, there is a general degree of negative correlation. This is because the core area corresponds to a high concentration of employment. People who buy houses in these areas are not far away from their workplaces. Therefore, the demand for commuting through the subway channel is not as high. In the west and north of Haidian and Changping, and west of Fengtai, these areas have a very high demand for subway commuting, thus showing a strong negative correlation. On the contrary, in Mentougou, Shunyi and the northeast of Daxing, due to the single subway line and the short opening time of the only subway line, the impact of the distance to the subway station on housing prices is relatively weak.
Figure 6. Cont.
This potentially indicates that the medical resources of the Economic Development Zone Third Ring Road are full of businesses and residents can shop conveniently. The LNM has places where office workers live, with a relatively high degree of convenience around them, phenomenon. Dongcheng, Xicheng, Haidian, Chaoyang, and other core areas of Beijing's parks in the area; meanwhile, the south of Changping is an emerging residential area with
ences. Due to the concentration of high-quality educational resources and the large number
negative correlation was concentrated at the junction of Daxing and Tongzhou. This area is
indicating that Beijing's medical resources are relatively balanced. An area showing a strong
a strong negative correlation between LNM and housing prices in the Mentougou area,
and the correlation between the LNM and housing prices is relatively weak. There was
a feeble correlation with housing prices. Fangshan and Changping are newly-emerging
hing inconvenience posed by traffic.
neighbhorhood of Xiangshan Forest Park in this area greatly supplements the green resources, and the
distance to the park had a significant positive correlation with housing prices. We speculate
the closer to Xiangshan Forest Park in this area greatly supplements the green resources, and the
closer to Xiangshan Forest Park, the more inconvenience posed by traffic.

The impact of LNM on housing prices shows a relatively obvious regional discrete phenomenon. Dongcheng, Xicheng, Haidian, Chaoyang, and other core areas of Beijing’s Third Ring Road are full of businesses and residents can shop conveniently. The LNM has a feeble correlation with housing prices. Fangshan and Changping are newly-emerging places where office workers live, with a relatively high degree of convenience around them, and the correlation between the LNM and housing prices is relatively weak. There was a strong negative correlation between LNM and housing prices in the Mentougou area, indicating that business facilitation in this area needs to be strengthened urgently.

The impact of LNH on housing prices was relatively weak in the overall correlation, indicating that Beijing’s medical resources are relatively balanced. An area showing a strong negative correlation was concentrated at the junction of Daxing and Tongzhou. This area is
the Economic Development Zone designated by Beijing, and there are many enterprises. This potentially indicates that the medical resources of the Economic Development Zone need to be strengthened.

The impact of LNPS on housing prices also showed obvious regional discrete differences. Due to the concentration of high-quality educational resources and the large number

Figure 6. Neighborhood covariate maps of the study area. (a) Log of the distance to the nearest subway station (LNS) map of the study area; (b) log of the distance to the nearest park (LNP) map of the study area; (c) log of the distance to the nearest market (LNM) map of the study area; (d) log of the distance to the nearest hospital (LNH) map of the study area; (e) log of the distance to the nearest primary school (LNPS) map of the study area; and (f) log of the distance to the nearest middle school (LNMS) map of the study area.
of primary schools in Xicheng, Dongcheng, and the northern part of Haidian, LNPS has little impact in these areas. In economic development zones, there are a large number of enterprises in the area, many factories spanning a large area, and the educational resources of primary schools are relatively unbalanced; as such, LNPS shows a strong negative correlation in these areas—that is, the closer to primary schools, the higher the housing price. Other areas showed a trend from a large negative correlation to a weak positive correlation from the urban core to the periphery, which is also in line with previous research [49]. There are not many primary schools in the suburbs of the city, the educational resources are relatively balanced, and there is a lack of high-quality primary schools. Therefore, LNPS has little impact on housing prices in these areas.

The regional LNMS in Xicheng close to the central area of the capital had little impact on housing prices. Other regions generally showed a trend from a strong positive correlation to a weak positive correlation from the urban core to the periphery, which is consistent with intuition and previous research [50,51].

The regression functions of the E-RGWR and GWR models were similar, thus validating the proposed model. However, the GWR model was associated with higher errors. These analyses highlight the high potential value of the proposed valuation model in socioeconomic and social governance policy research.

4. Discussion

4.1. Model Performance in Exploring Spatial Heterogeneity

Our main aim of exploring the spatial heterogeneity of residential land prices was to provide appropriate guidance to strengthen classified regulation and control of the real estate market. In this study, we developed the E-RGWR method, which expands the definition of regional factors by incorporating those factors affected by education into the GWR weight matrix. Then, we conducted an empirical analysis on housing sales prices in Beijing as an example, in order to prove the effectiveness of introducing regional factors affected by education. Furthermore, the proposed method focuses on the construction of a regional weight matrix for a discrete dataset as opposed to a spatiotemporal matrix. In other words, the E-RGWR gives the maximum weight to points in the same area or similar areas by adjusting the spatial weight matrix, while assigning different weights to points in different areas according to the regional impact factors.

We found that, in the Beijing housing sale price model, under the fixed and adaptive bandwidth strategies, the regional factors affected by education added into the E-RGWR are reasonable, and its performance was significantly better than those of the GWR and RGWR models. The introduction of regional factors affected by education in specific research areas can bring substantial benefits. The results in Table 3 demonstrate that, compared with the normal GWR model, the mean square error of the E-RGWR model was reduced by 38.71%, and, compared with the extended RGWR of GWR, the mean square error of the E-RGWR model was reduced by 16.80%. More importantly, E-RGWR had a significantly improved $R^2$ value, which was 10.98% higher than GWR and 3.26% higher than RGWR. These results demonstrate that the E-RGWR model is superior to the GWR and RGWR models, and also show that the administrative region affected by education has a significant impact on housing prices in Beijing. The E-RGWR model, using sample data, effectively addresses the spatially discrete heterogeneity caused by educational factors, and creatively extends the feasibility of introducing regional educational factors into the GWR model.

4.2. Spatial Discrete Heterogeneity of Commercial Housing Prices in Beijing

According to the estimation results of the E-RGWR, RGWR, and GWR models, we drew house price spatial distribution maps, Figure 7a shows the actual house prices, while Figure 7b–d shows the house prices predicted by GWR, RGWR, and ERGWR, respectively. The spatial trend of housing price distribution in the four figures is basically the same, with prices showing a gradual downward trend when moving outward from the central city area. The housing price in the core urban area of the Third Ring Road were very high,
the housing prices in the Fifth Ring Road to the Third Ring Road were moderate, and the housing price in the peripheral areas of the city were low, consistent with previous research [52]. Housing prices higher than 60,000 yuan were concentrated in the six urban districts, while those less than 60,000 yuan were basically concentrated outside the six urban districts. In this result, RGWR and ERGWR are consistent with reality. The GWR prediction results for the Changping District in the north of the study area were not consistent with the actual results. At the same time, the spatial discrete heterogeneity of housing prices in Beijing was significant. Among the six regions of the city, most of the super-high housing prices in Xicheng and Dongcheng were concentrated, and Haidian and Chaoyang were second only to Dongxicheng. Shijingshan and Fengtai had the lowest housing prices of the six districts of the city. In this regard, RGWR and E-RGWR performed better. Although the house prices estimated by GWR conformed to the trend of gradual decline from the center to the outside overall, they failed to reflect the spatial discrete heterogeneity of house prices in different regions and circles. Except for the super-high housing prices in the Xicheng and Dongcheng districts, there was no obvious dispersion in other areas. In contrast, the estimation results of E-RGWR and RGWR well-reflected the spatial dispersion of housing prices. Xicheng and Dongcheng were the most expensive, followed by Haidian and Chaoyang, while Shijingshan and Fengtai were in the third grade, consistent with the actual situation. These results indicate that E-RGWR and RGWR can better explain the spatial discrete heterogeneity of housing prices in Beijing.

We chose the confluence area of Shijingshan, Haidian, and Fengtai for close-up comparison between different models, as shown in Figure 8. It can be clearly seen that, in the actual housing price distribution, there is an obvious spatial discrete phenomenon in the intersection of the three regions. The housing prices in Haidian are concentrated above 90,000 yuan, while those in Shijingshan and Fengtai are mostly below 90,000 yuan. The different models showed obvious differences in estimating this spatially discrete heterogeneity. GWR had the worst performance, and there was no obvious gap between the three regions. RGWR performed slightly better than GWR and could explain the spatial discrete heterogeneity of housing prices between Haidian and Fengtai, to some extent, but the gap between Haidian and Shijingshan was not completely differentiated. E-RGWR presented the best performance. There were obvious differences between Haidian, Shijingshan, and Fengtai in housing prices, which is consistent with reality. Due to the effect of high-quality education in Haidian, the overall house prices in Haidian are significantly higher than those in Shijingshan and Fengtai.

In summary, we found that E-RGWR is more suitable for use in Beijing, a city with spatial discrete heterogeneity caused by educational factors, and the overall estimation result of the model was better than that of GWR and RGWR. Furthermore, E-RGWR can more accurately estimate the house prices in different regions at the local intersection of regions, more accurately reflect the phenomenon of spatial dispersion, and improve the accuracy of the model in terms of global and local estimation.
Figure 7. Actual housing prices and the housing prices predicted by different models. (a) Commercial housing prices in Beijing; (b) housing prices predicted by the GWR model; (c) housing prices predicted by the RGWR model; and (d) housing prices predicted by the E-RGWR model.
In summary, we found that E-RGWR is more suitable for use in Beijing, a city with spatial discrete heterogeneity caused by educational factors, and the overall estimation result of the model was better than that of GWR and RGWR. Furthermore, E-RGWR can more accurately estimate the house prices in different regions at the local intersection of regions, more accurately reflect the phenomenon of spatial dispersion, and improve the accuracy of the model in terms of global and local estimation.

4.3. Policy Impact Analysis

Abnormal fluctuations in housing prices affect the risk and uncertainty faced by central and local governments when making decisions related to land auctions, mortgage policy, land supply, monetary policy, and fiscal policy. A high house price is detrimental to consumer welfare, but low house prices are harmful to government revenue. Reforms in China’s system of urban housing, an important part of the “reform and opening up” policy initiated in 1978, have led to a general and significant improvement in accommodations for most of the urban population in the country. Over the past 40 years, housing prices in Beijing have been determined by typical market factors. However, in this study, it can still be seen that some public policies have a great impact on housing prices.

First, we considered the impact of education policy on housing prices. Primary schools in Beijing implement a strict “nearby enrollment” admission policy. This has led to higher housing prices in administrative regions where high-quality educational resources are concentrated. On the contrary, the distance from the primary school presents a discrete and stratified phenomenon in which the influence of the core area is small, the impact of the middle area is large, and the impact of the peripheral area is small. This is because the two administrative regions of Dongcheng and Xicheng in the core area are small in size and have many high-quality primary schools, and so the distance is not a significant influencing factor. In the Haidian and Chaoyang Districts, although there are many high-quality educational resources, the area is large and the distance from primary schools is a significant factor affecting housing prices. Therefore, education policy not only affects the balance of education, but also affects the level of regional housing prices, thus having a huge impact on Beijing’s economy.
The second factor was the impact of second-hand real estate sales policies on housing prices. In the research of many scholars, the construction time has been taken as an important factor affecting housing prices [53–55]. In this study, a significant difference from previous studies was that the construction time in Beijing was not considered as the main factor affecting housing prices. This has been greatly affected by the “full five” policy in the second-hand housing market in Beijing. High-value new houses cannot be sold at a high price within five years, as owners have to pay personal income tax. If they are forced to sell, they will even be cheaper than the surrounding second-hand houses that have been built for a longer time.

The third factor was that the establishment of economic development zones leads to the fact that the impact of public facilities in a given area on housing prices may differ from that in other areas. In this study, the negative correlation between LNH and LNPS on housing prices in the Beijing Economic Development Zone was significant, potentially indicating that the large area of factory buildings and industrial parks in the Economic Development Zone have affected the balanced distribution of hospital and primary education resources. Policy makers should focus on optimizing the distribution of medical resources and primary school resources in economic development zones, balancing medical and educational resources, and improving the living satisfaction of residents.

4.4. Extended Significance of the Model

Since 2016, real estate in different cities in China has begun to implement “one city, one policy”. Different policies and different regional environmental factors have led to different trends in China’s real estate in its various cities and regions. Therefore, it was meaningful to build the E-RGWR model based on the traditional GWR and the expanded RGWR, thus improving the quality of the model evaluation and allowing us to monitor the spatial dispersion heterogeneity of Beijing housing prices according to the current situation of housing prices in Beijing.

We can see that, in modern society, more and more emphasis is being placed on refined social governance, and higher requirements are being put forward for the estimation accuracy of relevant models. Basic models are more suitable for macro-estimation at a large spatial scale, while research refined to the county or city level requires model expansion and construction according to local conditions for different regions. Therefore, the exploration of model expansion in this paper is extremely meaningful.

5. Conclusions

In this study, the regionally geographically weighted regression affected by education (E-RGWR) model was outlined and applied to a case study of housing prices in Beijing, China. The analysis revealed that the GWR model is suitable under the situation of spatial non-stationarity, RGWR is helpful to analyze the influence of discrete heterogeneity, and E-RGWR is more suitable for the study of spatial discrete heterogeneity caused by education in the selected study area.

We used the experimental results of the Beijing housing price case study to demonstrate that the modeling accuracy of the proposed E-RGWR model is better than that of the GWR and RGWR models. On the basis of solving the discrete heterogeneity of RGWR, E-RGWR further considers the regionalization factors affected by education, according to the characteristics of the study area, thus improving the accuracy of the model. Compared with the GWR model, E-RGWR increased the $R^2$ and $R^2_{adj}$ from 0.7789 and 0.7785 to 0.8644 and 0.8642, respectively, and reduced the $MSE$ and $RMSE$ from 0.0635 and 0.2637 to 0.0426 and 0.2065, respectively. The $AICc$ standard was also 3039.3 lower than that of GWR. Compared with the RGWR model, E-RGWR increased the $R^2$ and $R^2_{adj}$ from 0.8371 and 0.368 to 0.8644 and 0.8642, respectively, and reduced the $MSE$ and $RMSE$ from 0.0512 and 0.2263 to 0.0426 and 0.2065, respectively. The $AICc$ standard was also 1109.7724 lower. Statistical tests demonstrated significant differences between E-RGWR,
RGWR, and GWR, thus we concluded that it is meaningful to incorporate regional impact factors affected by education into the GWR model.

Meanwhile, our empirical research proved that the property fee, greening rate, floor area ratio, distance to subway station, distance to market, distance to primary school, and distance to middle school variables all have significant effects on residential land prices in Beijing. Notably, the impacts of greening rate, distance to market, distance to hospitals, and construction time on housing prices in Beijing presented significant spatial discrete heterogeneity. At the same time, the relationships between the enrollment policy, real estate sales policy, and house prices were analyzed. Too high housing prices can lead to social problems such as unfair education, social stratification, and reduced consumer welfare. However, if housing prices are low, it will jeopardize government revenue and affect the overall economy of Chinese society, especially in the context of the current COVID-19 pandemic. Therefore, in addition to macro-control, the government should implement precise policies targeting different regions to promote education equity and boost the economy.

Furthermore, there were also some limitations to our case study. For example, this research was mainly aimed at the research area of Beijing, and it is debatable whether it is generally applicable to other areas. In this study, administrative regions were still used as the basic analysis unit. When faced with a more refined division method, whether the model still has better accuracy than the GWR is worth exploring. Therefore, future research should focus on the specific improvement of applying the method to different research areas and different research objects, further exploring the zoning method, and exploring the applicability of the model in a more refined zoning method to further improve the associated prediction accuracy. In general, E-RGWR provides an effective complementary method for large-scale housing price estimation and provides an idea of tailoring the model to local conditions for refined social governance, providing evidence for policy makers to understand the nature of varying relationships within a housing price data set in Beijing.

Author Contributions: Conceptualization, Z.W. and Y.Z.; methodology, Z.W., F.Z. and Y.Z.; software, Z.W.; validation, Z.W. and Y.Z.; resources, F.Z. and Y.Z.; data curation, Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, Z.W., F.Z. and Y.Z.; visualization, Z.W.; funding acquisition, F.Z. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC, Grant No.42001343).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Chai, Z.; Yang, Y.; Zhao, Y.; Fu, Y.; Hao, L. Exploring the Effects of Contextual Factors on Residential Land Prices Using an Extended Geographically and Temporally Weighted Regression Model. Land 2021, 10, 1148. [CrossRef]
2. Wang, Z.; Wang, Y.; Wu, S.; Du, Z. House Price Valuation Model Based on Geographically Neural Network Weighted Regression: The Case Study of Shenzhen, China. ISPRS Int. J. Geo-Inf. 2022, 11, 450. [CrossRef]

11. Qu, Y.; Jiang, G.; Li, Z.; Shang, R.; Zhou, D. Understanding the multidimensional morphological characteristics of urban idle land: Stage, subject, and spatial heterogeneity. Cities 2020, 97, 102492. [CrossRef]


17. Mok, H.M.; Chan, P.P.; Cho, Y. A hedonic price model for private properties in Hong Kong. J. Real Estate Financ. Econ. 1995, 10, 37–48. [CrossRef]


19. Duan, J.; Tian, G.; Yang, L.; Zhou, T. Addressing the macroeconomic and hedonic determinants of housing prices in Beijing Metropolitan Area, China. Habitat Int. 2021, 113, 102374. [CrossRef]


25. Liu, J.; Zhao, Y.; Xu, S.; Zhao, Y.; Wang, Y.; Zhang, F. A Geographically Temporal Weighted Regression Approach with Travel Distance for House Price Estimation. Entropy 2016, 18, 303. [CrossRef]


31. Li, H.; Wei, Y.D. Spatial inequality of housing value changes since the financial crisis. J. Geogr. Syst. 2020, 23, 217–232. [CrossRef]


34. Wang, Z.; Zhao, Y.; Zhang, F. Simulating the Spatial Heterogeneity of Housing Prices in Wuhan, China, by Regionally Geographically Weighted Regression. ISPRS Int. J. Geo-Inf. 2022, 11, 129. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.