Article

Exploring the Spillover Effects of Urban Renewal on Local House Prices Using Multi-Source Data and Machine Learning: The Case of Shenzhen, China

Xiaojun Li 1,†, Jieyu Wang 2,†, Ke Luo 1, Yuanling Liang 1 and Shaojian Wang 2,*

1 Guangzhou Urban Planning & Design Survey Research Institute, Guangzhou 510060, China
2 Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China
* Correspondence: wangshj8@mail.sysu.edu.cn
† These authors contributed equally to this work.

Abstract: Urban renewal is a current hotspot for research in the field of urban geography and urban planning. However, few studies have been able to quantify the impact of urban renewal on local house prices. Taking Shenzhen as an example, this paper measures the added premium effect of urban renewal on local house prices through econometric models and multi-source data and explores the spillover effect of urban renewal on house prices using an integrated model based on machine learning and Geo-detector analysis. The main findings are: (1) Shenzhen’s urban renewal had a significant positive premium effect on the unit transaction price of local housing. (2) The population characteristics and the accessibility to transport in the context of urban renewal are the main drivers for premiums on house prices. (3) There is spatial heterogeneity with respect to the housing premium effect due to urban renewal, among which optimization of the density of the road network is most closely associated with the premium effects. The interaction of the road density network and the population density of particular streets drives the medium premium effect. Our findings have important implications for refinements in management practices for urban renewal in the context of the housing market.

Keywords: urban renewal; house prices; premium; machine learning; Shenzhen

1. Introduction

Urban renewal, an important means to revitalize the infrastructure of towns and cities, is one of the elements that have received much attention in China in recent times [1]. Since the reform and opening up of the country, China’s urban economy has undergone phenomenal growth, and the speed of urban construction has accelerated accordingly. However, against the background of “the paid use of state-owned land” and reforms of the socialist market economy system and commercialization of housing, a series of problems have emerged in urban areas, such as an increasing population density, a decrease in the availability of suitable land, inefficient land allocation, crowded and dilapidated housing, insufficient municipal and public infrastructure, poor protection of historical relics and monuments, and an unappealing landscape [2]. At present, China’s economy is transitioning from a high-speed growth stage to one of high-quality development, and urban development has been transformed from crude extensiveness to intensive connotation; accordingly, a people-centered approach to urbanization has gained widespread attention. Urban renewal was highlighted in a government report for the first time in 2021, and this topic has become a focus for high-quality urban development. The topic covers the modernization of cities, the transformation of villages, the renovation of neighborhoods and the urban landscape, and the protection of our cultural heritage, etc. Along with physical aging and the structural and functional decline of built-up areas, as represented...
by some of the old urban areas, urban renewal has become an important way to tap the potential of the stock of the land and promote quality in urban construction [2]. A series of changes brought about by urban renewal (changes in building attributes, development of transportation and road networks, changes in population density, changes in commercial services, and changes in land use) can reflect the influences of government forces on urban land rent in the supply and demand process of the market [3]. In this context, it is of practical significance to study the impact of urban renewal on local house prices and, at the same time, recognize that the mechanism of the premium effect of urban renewal on house prices can provide a policy reference for urban planning and construction development in the future.

The term “urban renewal” was used first to refer to the urban regeneration strategies that emerged in Western countries after World War II in response to the phenomenon of counter-urbanization and suburbanization. Urban renewal aimed to solve the problems of residential division and social conflict, as well as urban decay caused by the shift of the urban population and industry to the suburbs [4]. Research on urban renewal first focused on the transformation of the physical environment, and then studies extended gradually to urban design, urban management, and economic and social analysis. A series of empirical studies emerged on topics such as cooperation between government and enterprises, ecological balance, development patterns, neighborhood relations, and the creation of rent gaps and distribution in urban renewal [5–7]. With accelerated industrialization and the changes in economic structure which occurred in China in the middle and late 20th century, research on urban renewal in China gradually began to gain the attention of academics. This research mainly addressed three aspects. The first was the analysis of experience and reflection of urban renewal projects, such as demolition and resettlement issues and the mechanism for the distribution of benefits in the process of renovation [8,9], the development history and experience of urban renewal [10,11], and a comparative analysis of urban renewal and renovation methods in China and the United States [12]. The second aspect was theoretical and applied research, whereby the application of differential rent and location theories on old city renewal was examined [13,14]. The third aspect was an empirical study of “three old urban transformations” [15].

Existing geographic studies on house prices have mainly analyzed spatial variation, growth patterns, and factors that influence house prices at the medium and macro scales [16–18], but no unified conclusion has been reached regarding the impact of urban renewal on house prices. Most researchers believe that urban renewal can improve the environment and enhance the lifestyle by changing the structure of the housing supply, but at the same time, this will have a crowding-out effect on low-income people and push up local house prices; moreover, this externality has spatial decay characteristics [18–22]. For example, Liang et al. [23] estimated the impact of urban renewal on neighborhood housing prices through difference-in-difference methods and spatial econometrics and found that the impact of urban renewal induces a sustained response in neighborhood housing prices even before reconstruction is completed. Using small- and medium-sized cities in central and northern Italy as case studies, Albanese et al. [24] found that urban renewal projects led to higher house prices by improving public sector interventions. Using Ulsan, Korea, as the study area, Uho et al. [25] found that house price increases were most pronounced in project locations where residents showed a high willingness to participate. In these areas, house price increases were found even before the final plans were released. Some studies also point out that urban renewal has insignificant or even negative effects on local house prices [26,27]. Other studies have focused mainly on analyzing the average impact of urban renewal on residential house prices and the spatio-temporal heterogeneity characteristics at the macroscopic scale [26,28]. However, few studies have analyzed the mechanisms underpinning the impact of urban renewal on the housing premium effect from the perspective of the microphysical built environment and the demographic composition of the community.

In terms of research methods, economists have mostly adopted econometric models to empirically study the effect of the response of the housing market on urban renewal
policies [29–31]. Alternatively, geographers have mostly used the characteristic price method to study the impact of urban renewal on local house prices [32,33]. Additionally, the double difference method (DID), which effectively compensates for the omitted variables and endogeneity problems, has become a powerful tool that has been adopted in related studies in recent years. For example, through DID, Dehring et al. [34] found no significant effect of new sports facilities on the local house prices in Dallas, USA. Kavetsos [35] applied DID to explore the changes in the prices of local housing before and after the completion of the Olympic Games in London and found that the prices increased by 2.1–3.3%. Lee et al. [36] used DID to explore the impact of urban renewal projects on house prices of surrounding residential areas within a perimeter of 800 m in Taipei City and found that urban renewal generated a premium of TWD 14,880 per m$^2$ for the affected residences. Wang et al. [20] used DID to analyze the spillover effect on the prices of local housing for the renovation of shanty town areas and found that there was a positive spillover effect and a negative phase effect on house prices within 1 km of the renovation area.

Despite such progress in existing studies, there is a lack of discussion and attention given to whether DID satisfies the parallel hypothesis, given that we cannot exclude the possibility of interference brought by non-urban renewal factors. In addition, GIS and remote sensing technologies have been widely used in urban renewal studies [37–42], and these tools provide strong support for assessing the built urban environment and formulating strategies for urban renewal and renovation implementation.

In summary, many studies have focused on urban renewal, and little attention has been paid to the mechanism of urban renewal’s influence on house prices, which means that the existing literature evidences a relatively poor understanding of the link between urban renewal and house prices. Meanwhile, previous studies have mainly conducted regression analysis based on traditional econometric models, with less consideration of variable omission and endogeneity. Therefore, the purpose of this paper is to attend to this gap in the study field, and the contribution of this paper mainly includes two aspects. To be specific, first of all, we quantify the premium effect of urban renewal on house prices since existing studies have paid insufficient attention to the spillover effects and mechanisms of housing prices in the context of urban renewal. Quantifying the impact of urban renewal renovation is a relatively new research topic. Second, our study may provide policy tools for urban renewal through impacts on house prices. Explicitly, the empirical results of our study can provide a reference for refinements in management practice for urban development and assist in the delivery of sound city planning and construction. Third, compared with the regression analysis based on traditional econometric models in previous studies, this paper adopts multi-source data, DID, and machine learning to provide a multi-faceted and dynamic analytical approach to study the impact of urban renewal on house prices, which can more accurately analyze the impact mechanism and the role relationship between urban renewal and housing price spillover. Therefore, this paper will contribute to studies related to the impact of urban renewal.

2. Study Area

Shenzhen, located in the south of Guangdong Province in southern China, forms the research area addressed in this study (Figure 1). As part of China’s reform and opening, Shenzhen was established as a special economic zone in the 1980s and transformed from a small fishing village to one of the four first-tier cities in China after Beijing, Shanghai, and Guangzhou. Shenzhen has an area of 1997 km$^2$ and a population of about 17.68 million in 2021. With a gross domestic product (GDP) of RMB 3070 billion, it ranks third in China. Shenzhen is the pioneer area for urban renewal in China. The long-standing space problem of insufficient land being available for urban renewal has made transformation and development of the stock infrastructure an urgent issue to be addressed. Shenzhen became the first city in China to fully transform its land supply for stock renovation and thus became an exemplar for good urban renewal research and practice in China. The Shenzhen Urban Renewal Measures promulgated by the Shenzhen Municipal Government
defines urban renewal as "an activity of comprehensive improvement, functional change or demolition and reconstruction of a specific urban built-up area by a corresponding body". Specifically, the definition includes the following three categories: (1) Comprehensive improvement: without modifying the main structure and function of the building, improving fire protection facilities, improving infrastructure and public service facilities, improving street facades, environmental improvement, and energy-saving renovation of existing buildings; (2) Functional change: changing part or all of the functions of the building, but without changing the main body of land use rights and period of use, retaining the original structure of the building; (3) Demolition and reconstruction: strictly in accordance with the provisions of unit planning for urban renewal of housing units, and implementation of an annual plan for urban renewal.

![Location of the study area](image)

**Figure 1. Location of the study area.**

### 3. Materials and Methods

#### 3.1. Methods

**3.1.1. The Difference-in-Difference Model Based on Propensity Score Matching (PSM-DID)**

Confirmation that urban renewal generates a significant premium on local house prices is the premise for this research. Urban renewal may be regarded as a policy experiment in some areas of Shenzhen, and an evaluation of the effect of the policy can usually be analyzed by DID [43]. In this study, residences within 500 m of the site of urban renewal are considered to be the experimental group and affected by urban renewal, while residences outside the 500 m zone are considered as the control group. Given that the parallel trend assumption for DID may not be satisfied for the experimental and control groups, this study also adopts the difference-in-difference and propensity score matching (PSM-DID) method. The PSM-DID method can strip out the effects of other factors, thus alleviating the problem of selection bias of treatment effects as much as possible and obtaining a better estimate of the house price spillover effect due to urban renewal.

Using the PSM-DID method allows the experimental and control groups to satisfy the parallel trend assumption. The district where the urban renewal project is located is different from other districts in terms of location conditions, surrounding facilities, and other factors. Traditional methods such as the DID cannot exclude the influence of these factors or endogenous factors on house prices, thus compromising the assessment of the spillover effect on house prices due to urban renewal. The PSM can match, as far as possible, similar samples in the experimental and the control groups based on the covariate X. The PSM model uses Logit regression to regress the binary explanatory variables (whether
urban renewal was conducted or not) and then fits a propensity score for each sample, which reflects the probability of a sample being in the experimental group, based on using the regression coefficients of each matching variable as weights. Then, the experimental group and the non-experimental group are matched with the nearest neighbors based on put-back and one-to-many matching of the propensity scores, and finally, matching of the control group with the experimental group is obtained [44,45]; in effect, this helps to solve the problem that the experimental group and the control group may not satisfy the parallel trend assumption.

\[ \text{Price}_{it} = \alpha_0 + \gamma_1 \text{Treat}_{it} + \gamma_2 \text{Time}_{it} + \gamma_3 \text{Time}_{it} \times \text{Treat}_{it} + \gamma_4 \text{X}_{it} + \epsilon_{it} \]  

(1)

The DID model, as shown in Equation (1), is used to regress the successfully matched samples in the PSM model. The dependent variable (Price) indicates the unit price per m\(^2\) of a housing transaction, and the unit is 10,000 yuan. The explanatory variable (Treat) indicates whether to accept the treatment, i.e., whether to carry out urban renewal, and if yes, then assumes the value is 1; otherwise, it is 0. Time indicates the time effect, and the sample transaction time is 0 before the urban renewal is carried out, and is 1 on completion of renewal. The coefficient of the cross-product term reflects the net effect of urban renewal on the house price spillover. In addition, if the urban renewal has a positive spillover effect on house prices, the term will be significantly greater than 0. X represents a set of covariates. According to previous studies [36,46], the explanatory variables selected in this paper are the number of points of interest (POI) of six types of factors around each housing sample, including transportation, medical, education, food and beverages, leisure, and business. \(\epsilon\) is a random interference term.

From regression Equation (1), it can be seen that for the local areas with urban renewal (Treat = 1), the house prices before and after renewal are \(\alpha_0 + \gamma_1\) and \(\alpha_0 + \gamma_1 + \gamma_2 + \gamma_3\), respectively, and the change in house prices before and after urban renewal is \(\Delta \text{Price}_1 = \gamma_2 + \gamma_3\). Similarly, for the local areas without urban renewal (Treat = 0), the change in house prices before and after urban renewal is \(\Delta \text{Price}_0 = \gamma_2\). Therefore, by subtracting \(\Delta \text{Price}_0\) from \(\Delta \text{Price}_1\), the net effect of urban renewal on the change in house prices can be obtained as \(\Delta \text{Price} = \gamma_3\). The coefficient of \(\gamma_3\) obtained from the regression is significantly positive if urban renewal generates a positive premium on the surrounding house prices.

3.1.2. Random Forest Model

The random forest (RF) model is a machine-learning algorithm that belongs to the integrated learning category. The RF model trains, predicts, and generates classifiers for samples through multiple decision trees, which in turn aggregates a large number of classification trees. The RF model is more robust with respect to the treatment of missing data and unbalanced data and can provide relatively accurate prediction results [47]. The model can be used not only for classification but also for regression analysis and further exploration of the relative importance of the factors which influence the results [48]. In this study, a classifier to identify the degree of the housing premium is trained by the RF model using the CART tree algorithm. Based on this, the Gini impurity is calculated, and each feature (i.e., factor of influence) that affects the classification is ranked in relative importance to filter the factors of influence whereby there is a high contribution to the housing premium. The formula for calculating the Gini impurity is as follows:

\[ GI_{im} = \sum_{k=1}^{[k]} \sum_{k' \neq k} p_{imk} p_{imk'} = 1 - \sum_{k=1}^{[k]} p_{imk}^2 \]  

(2)

where \(GI_{im}\) denotes the Gini impurity at the m node on the i decision tree; and \(p_{imk}\) denotes the probability that the k class is drawn from the overall sample at the m node on the i decision tree, i.e., the percentage of samples of the k class at this node.
If the decision tree is branched at node \( m \) into node \( x \) and node \( y \) by feature \( j \), then the amount of change in the Gini impurity before and after the branch, or in other words, the importance of feature \( j \) at node \( m \), is calculated as follows:

\[
VIM_{imj}^{(\text{Gini})} = GI_{im} - GI_{ix} - GI_{iy}
\]  

(3)

where \( VIM_{imj}^{(\text{Gini})} \) represents the importance score (based on the Gini impurity) of feature \( j \) at the \( m \) node on the \( i \) decision tree; and \( GI_{ix} \) and \( GI_{iy} \) represent the Gini impurity of the \( x \) node and the \( y \) node on the \( i \) decision tree, respectively.

If all nodes on a decision tree that branch through feature \( j \) are the point set \( M \), and there are a total of \( n \) decision trees in the RF classifier, then the importance score for feature \( j \) is as follows:

\[
VIM_j^{(\text{Gini})} = \sum_{i=1}^{n} \sum_{m \in M} VIM_{imj}^{(\text{Gini})}
\]

(4)

where \( VIM_j^{(\text{Gini})} \) is the importance score of feature \( j \) in the classifier. By normalizing the importance scores for all the features, the more important features (i.e., influencers with higher contributions) can be identified.

3.1.3. Geo-Detector Analysis

Based on the premise that the housing premium is influenced by various factors, it is crucial to further analyze the role of those factors. The Geo-detector approach can be used to explore the factors influencing the spatial variation of house prices and their interactions. The Geo-detector approach consists of four main components [49]: risk detector, factor detector, ecological detector, and interaction detector. The risk detector is used to identify significant differences in attributes between two regions. The factor detector is used to analyze the degree of influence of factors on the explanation of the study object. The ecological detector can investigate whether there are significant differences in the spatial distribution of different factors on the study object. Finally, the interaction detector identifies the explanatory power of two factors on the study object under interaction [44]. This study focuses on an analysis of the factor detector and the interaction detector.

The Geo-detector approach helps diagnose the drivers of spatial heterogeneity in house price premiums, and if the strength of the effect of a factor is spatially significant and consistent with the price movement of second-home transactions, that variable has a significant influence on the housing premiums [45]:

\[
q = 1 - \frac{SSW}{SST} = 1 - \frac{1}{N\sigma^2} \sum_{h=1}^{L} N_h \sigma_h^2
\]

(5)

where \( q \) denotes the determinant of the spatial variation of the housing premium by the factor of influence. \( N \) and \( N_h \) are the price fluctuations of all second-hand traded housing in the study area and the price fluctuation of second-hand traded housing affected by urban renewal projects, respectively. \( \sigma^2 \) and \( \sigma_h^2 \) represent the discrete variance of the price fluctuation of all second-hand traded housing in the study area and the discrete variance of second-hand traded housing affected by urban renewal projects, respectively. \( L \) denotes the number of types of each factor of influence. The value of \( q \) is in the range of \([0, 1]\), and the closer the value is to 1, the greater the influence of the factor on the spatial variation of the house price premium; \( q = 0 \), indicates that the spatial variation of the house price premium is not driven by the factor. On this basis, the interaction detector is applied to determine the effect of the interaction of different factors of influence on the house price premiums.

The Geo-detector approach can also identify the strength, direction, linearity, or non-linearity of the interaction factor. The interaction detector can classify specifically the relationships between the factors into five categories (Table 1) and is able to form a new \( q \)-value for the two-by-two superposition relationship of the geo-probing factors to obtain
the magnitude of the influence on the spatial differentiation of housing premiums as a result of the interaction of the factors.

Table 1. The relationships of factors in the interaction detector.

<table>
<thead>
<tr>
<th>Factor Relationship</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>q(X1 \∩ \ X2) &lt; Min(q(X1), q(X2))</td>
<td>Non-linear weakening</td>
</tr>
<tr>
<td>Min(q(X1), q(X2)) &lt; q(X1 \∩ \ X2) &lt; Max(q(X1), q(X2))</td>
<td>Single non-linear weakening</td>
</tr>
<tr>
<td>q(X1 \∩ \ X2) &gt; Max(q(X1), q(X2))</td>
<td>Two-factor enhancement</td>
</tr>
<tr>
<td>q(X1 \∩ \ X2) = q(X1) + q(X2)</td>
<td>Independent</td>
</tr>
<tr>
<td>q(X1 \∩ \ X2) &gt; q(X1) + q(X2)</td>
<td>Non-linear enhancement</td>
</tr>
</tbody>
</table>

3.2. Data Sources

The data presented in this paper mainly relate to data for urban renewal projects in Shenzhen. Considering the time lag between the implementation process for such projects and the occurrence of the housing premium effect, the data for 2010–2018 were selected for analysis. The ancillary data include basic urban mapping data and census data for buildings (for estimating the contribution from construction before and after urban renewal), which were obtained from the website of the Shenzhen Municipal Bureau of Planning and Natural Resources (http://pnr.sz.gov.cn, accessed on 18 March 2022). According to the statistics, there were plans for 717 urban renewal projects in Shenzhen from 2010 to 2018. The housing transaction data (including the transaction prices and the geographical coordinates of the properties) in Shenzhen during the study period were obtained from the website of Lianjia (https://gz.lianjia.com, accessed on 20 March 2022). The built environment data of Shenzhen adopted satellite remote sensing data, which were obtained through the natural resources satellite remote sensing cloud service platform (http://sasclouds.com/chinese/home, accessed on 20 March 2022) and the China Remote Sensing Data Sharing Network (http://eds.ceode.ac.cn/sjglb/dataservice.htm, accessed on 20 March 2022). The Shenzhen city map data, road traffic network data, and related POI data were obtained from the Guihuayun website (http://guihuayun.com, accessed on 26 March 2022). The population density for each district in Shenzhen was obtained from the Shenzhen Statistical Yearbook of the appropriate year, and the population data for each street were obtained from the relevant statistical bulletin of each district.

In the PSM-DID method, houses are used as the sample N. In RF regression, the number of housing premiums around each urban renewal project is the dependent variable, and the number of changes in commercial facilities, public service facilities, transportation facilities, and population characteristics due to urban renewal are the independent variables. In the analysis of geo-detector analysis, the study sample is urban renewal projects.

The software used for the PSM-DID model, random forest model, and Geo-detector are Stata, Matlab, and Geodetector, respectively.

4. Results

4.1. Effects of Urban Renewal on the Housing Premium

Using the PSM-DID method, a Logit regression was performed on the control variables with Time × Treat as the dependent variable, and the propensity scores were obtained. The regression results showed that the number of POI in each of the six categories reflecting the local environmental factors had a significant effect on the dependent variable (Table 2).
Table 2. Results of Logit regression.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Renew</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.00441***</td>
</tr>
<tr>
<td></td>
<td>(19.41)</td>
</tr>
<tr>
<td>Medical</td>
<td>0.00561***</td>
</tr>
<tr>
<td></td>
<td>(13.88)</td>
</tr>
<tr>
<td>Education</td>
<td>−0.00699***</td>
</tr>
<tr>
<td></td>
<td>(−24.98)</td>
</tr>
<tr>
<td>Food and Beverages</td>
<td>0.00030***</td>
</tr>
<tr>
<td></td>
<td>(4.45)</td>
</tr>
<tr>
<td>Leisure</td>
<td>0.00252***</td>
</tr>
<tr>
<td></td>
<td>(5.62)</td>
</tr>
<tr>
<td>Business</td>
<td>0.00084***</td>
</tr>
<tr>
<td></td>
<td>(5.36)</td>
</tr>
<tr>
<td>Cons</td>
<td>−0.54159***</td>
</tr>
<tr>
<td></td>
<td>(−43.51)</td>
</tr>
</tbody>
</table>

N 101,914

Note: Renew stands for Time × Treat. t-values are in parentheses. *** indicates significant at the 1% statistical level.

According to the results of the balance test of PSM (Figure 2), the absolute values of %bias for all covariates were less than 10%, and all of the values were significantly smaller than the bias before matching; the absolute values for the biases were also significantly lower than those before matching. These results indicate that for the covariates, the values between the experimental and control groups after matching do not differ, and the original hypothesis that “there is no systematic bias in the values of the covariates between the two groups” is not rejected, indicating that the samples are matched well, and the screening of “whether to carry out urban renewal” is effective.

![Figure 2. Balance test results for PSM.](image)

The results for the common support hypothesis test of the PSM (Figure 3) show that the vast majority of samples in the treatment and control groups are within the common range of values, while very few samples are not within the common range and the propensity score values are more extreme, thus confirming the robustness of the PSM results.
Figure 2. Balance test results for PSM.

The results for the common support hypothesis test of the PSM (Figure 3) show that the vast majority of samples in the treatment and control groups are within the common range of values, while very few samples are not within the common range and the propensity score values are more extreme, thus confirming the robustness of the PSM results.

Figure 3. Common support hypothesis test results for PSM.

According to the estimated results for PSM-DID (Table 3), the core explanatory variable is significantly positive in all regressions, indicating that urban renewal imposes a significant positive premium on the transaction prices of houses (ranging from 8800 yuan to 13,817 yuan per m²). In addition, based on the actual number of samples participating in each regression, the number of samples used in column (4) which satisfy the common support hypothesis is only 16 less than the full-sample benchmark regression in column (1), indicating that the vast majority of samples are involved in matching. This finding is consistent with the results of the common support hypothesis test, as indicated in Figure 3.

Table 3. Results of PSM-DID.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>FE</td>
<td>Weight</td>
<td>On_Support</td>
</tr>
<tr>
<td>Renew</td>
<td>0.87772 ***</td>
<td>0.94245 ***</td>
<td>1.38708 ***</td>
<td>0.94255 ***</td>
</tr>
<tr>
<td></td>
<td>(8.66)</td>
<td>(85.37)</td>
<td>(101.49)</td>
<td>(85.37)</td>
</tr>
<tr>
<td>Transportation</td>
<td>−0.00092</td>
<td>−0.00624 ***</td>
<td>−0.00759 ***</td>
<td>−0.00624 ***</td>
</tr>
<tr>
<td></td>
<td>(−0.19)</td>
<td>(−27.32)</td>
<td>(−26.78)</td>
<td>(−27.33)</td>
</tr>
<tr>
<td>Medical</td>
<td>−0.00719 *</td>
<td>−0.00752 ***</td>
<td>−0.00384 ***</td>
<td>−0.00753 ***</td>
</tr>
<tr>
<td></td>
<td>(−2.33)</td>
<td>(−21.02)</td>
<td>(−9.44)</td>
<td>(−21.02)</td>
</tr>
<tr>
<td>Education</td>
<td>0.00313</td>
<td>0.00251 ***</td>
<td>0.00160 ***</td>
<td>0.00252 ***</td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
<td>(10.65)</td>
<td>(5.80)</td>
<td>(10.68)</td>
</tr>
<tr>
<td>Food and Beverages</td>
<td>0.00035</td>
<td>0.00031 ***</td>
<td>0.00069 ***</td>
<td>0.00031 ***</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(5.04)</td>
<td>(9.17)</td>
<td>(5.03)</td>
</tr>
<tr>
<td>Leisure</td>
<td>−0.00151</td>
<td>0.00416 ***</td>
<td>0.00237 ***</td>
<td>0.00416 ***</td>
</tr>
<tr>
<td></td>
<td>(−0.57)</td>
<td>(11.07)</td>
<td>(5.72)</td>
<td>(11.05)</td>
</tr>
<tr>
<td>Business</td>
<td>0.00272</td>
<td>0.00473 ***</td>
<td>0.00413 ***</td>
<td>0.00473 ***</td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
<td>(31.43)</td>
<td>(22.43)</td>
<td>(31.42)</td>
</tr>
</tbody>
</table>
Table 3. Cont.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>3.96364 **</td>
<td>4.19067 ***</td>
<td>3.82565 ***</td>
<td>4.19069 ***</td>
</tr>
<tr>
<td>(12.88)</td>
<td>(349.88)</td>
<td>(221.65)</td>
<td>(349.84)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>101,914 (Total sample size)</td>
<td>101,914 (Total sample size)</td>
<td>67,069 (Number of matched samples)</td>
<td>101,898 (Number of matched samples)</td>
</tr>
</tbody>
</table>

Note: The renew term is Time × Treat, and its coefficient represents the effect of urban renewal on house price premiums. The t-values are in parentheses. *** denotes significant at the 1% statistical level.

The results for the above model show that the PSM matches the sample to a high degree and is able to make the experimental group (urban renewal occurred) and the control group (urban renewal did not occur) as similar as possible in all aspects of the characteristics, reducing the selection bias. The final results show that urban renewal produces a significant positive premium on the surrounding house prices, and a premium effect of 0.88 million yuan to 13.9 million yuan on the surrounding house prices in Shenzhen (from 2010 to 2018) is calculated with the other housing attribute factors being stripped out of the calculation.

4.2. Drivers for the Impact of Urban Renewal on Housing Premiums

4.2.1. Random Forest Results

Based on the results obtained by the PSM-DID, a distance of 500 m around the site of the renewal project was deemed to be the area of influence. In this way, the level of the housing premium of the average secondary house price transaction for each urban renewal project was measured. Based on the established attributes and characteristics of urban renewal and with reference to relevant studies, a system of indicators for the impact of the urban renewal factors on the housing premium was developed (Table 4). The system of indicators was changed with reference to the independent variable to conduct RF regressions on the premiums of the renewal projects, and the important factors that influence the housing premiums were identified according to the values for the Gini impurity.

Table 4. Index system for the impact of urban renewal on the housing premium.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Dimension</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing premium for urban renewal projects</td>
<td>Business location</td>
<td>Change in the number of medium and high-end hotels (3-star and above) within the area of influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in the number of business office buildings within the area of influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in the number of restaurants in the area impacted</td>
</tr>
<tr>
<td></td>
<td>Public Services</td>
<td>Change in leisure facilities within the area impacted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in educational facilities within the area of influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in medical facilities within the area of influence</td>
</tr>
<tr>
<td></td>
<td>Transportation</td>
<td>Change in traffic facilities within the area impacted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change in the density of the road network within the area impacted</td>
</tr>
<tr>
<td></td>
<td>Demographic characteristics</td>
<td>Average years of schooling (15+) for streets within the area of influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average age for streets within the area of influence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population density for streets within the area of influence</td>
</tr>
</tbody>
</table>

In the study, 1000 trees were selected for RF regression. The correlation coefficient reached 0.85, which means that RF regression was fit for purpose (Figure 4). The out-of-bag data were stable below 0.5 after the number of trees exceeded 100, indicating that the error for the model was small.
Based on the operational results of the PSM-DID model on 300 samples, it was possible to input to the Geo-detector for analysis of the factors that influence the results. Therefore, the RF algorithm was input to the Geo-detector for analysis of the factors that influence the results.

### 4.2.2. Geo-Detector Analysis

Figure 5. Results for the Gini impurity (importance of indicator).

According to the results for the Gini impurity (Figure 5), the scores for importance for each indicator were obtained. The average years of education (students of 15 years and above) living in the streets, the average age of people living in the streets, the population density for the streets, the number of changes made to the traffic facilities, and the density of the road networks ranked in the top five factors (in the order given), with the Gini impurity exceeding 0.7, indicating the highest importance for the indicators. In particular, the indicator for the average number of years of education has a Gini impurity of 1.85, which is the most important indicator. These results indicate that the demographic characteristics and the degree of improvement of road traffic facilities around urban renewal projects contribute the most to the housing premium.

![Image](a.png)

**Figure 4.** Number of trees for random forest model and the results. (a) number of trees, (b) the results of random forest model.

The dominant factors revealed by the RF model based on the machine-learning algorithm were input to the Geo-detector for analysis of the factors that influence the results. Based on the operational results of the PSM-DID model on 300 samples, it was possible to effectively exclude the interference of non-urban renewal factors around the housing for 300 samples, which were distributed in the Nanshan, Futian, and Luohu districts of
Shenzhen. The average value of the urban renewal, which drove the premium generated by local housing from 2010 to 2018, was 12,100 yuan/m². According to the natural breakpoint method, the housing premiums may be divided into three classes: high premium (24,000–49,400 yuan), medium premium (2400–10,700 yuan), and low premium (−0.62–10,700 yuan). The spatial distribution pattern for the premium effect on housing for urban renewal in Shenzhen (from 2010 to 2018) is shown in Figure 6, and this was used as a basis for the Geo-detector analysis of the impact factors.

![Spatial distribution of housing premium as a result of urban renewal](image)

**Figure 6.** Spatial distribution of housing premium as a result of urban renewal.

The urban renewal projects that exhibited a high premium effect on house prices were mainly located in Nanshan District and Yantian District. Urban renewal in Nanshan District focused on urban villages, old industrial areas, and old urban areas, with demolition and reconstruction serving as examples of renewal such that comprehensive improvement and functional change accounted for the main part of the renewal of the landscape in the district. Urban renewal in the Futian District focused on improving the core functions of the central business district and improving the level of coverage of transport and other critical services, which would have a clear effect on the house price premium. From the results of the Geo-detector analysis (Table 5), it can be seen that the largest factor influencing the high housing premium was the extent of improvement in the road network density, where the q value was 0.9930.

<table>
<thead>
<tr>
<th>Detection Factors</th>
<th>Low Premium</th>
<th>Medium Premium</th>
<th>High Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in traffic facilities (X₁)</td>
<td>0.7277</td>
<td>0.6758</td>
<td>0.7049</td>
</tr>
<tr>
<td>Change in medical provision (X₂)</td>
<td>0.5569</td>
<td>0.5366</td>
<td>0.7144</td>
</tr>
<tr>
<td>Average years of education for students in the street (&gt;15 years old) (X₃)</td>
<td>0.3981</td>
<td>0.3240</td>
<td>0.4303</td>
</tr>
<tr>
<td>Average age of population in the street (X₄)</td>
<td>0.3984</td>
<td>0.3240</td>
<td>0.4303</td>
</tr>
<tr>
<td>Population density in the street (X₅)</td>
<td>0.3984</td>
<td>0.3240</td>
<td>0.4303</td>
</tr>
<tr>
<td>Road network density (X₆)</td>
<td>0.9836</td>
<td>0.8816</td>
<td>0.9930</td>
</tr>
<tr>
<td>Dominant interaction factor</td>
<td>X₆ ∩ X₄ or X₅</td>
<td>X₆ ∩ X₃ or X₄ or X₅</td>
<td>X₆ ∩ X₁ or X₂ or X₃ or X₄ or X₅</td>
</tr>
<tr>
<td>Dominant interaction factor: q value</td>
<td>0.9838</td>
<td>0.9832</td>
<td>0.9932</td>
</tr>
</tbody>
</table>
Based on the above analysis, a framework for the driving mechanism for the variation of housing premiums in the context of urban renewal was constructed (Figure 7). The variation of the housing premiums in urban renewal was mainly influenced by a combination of factors, i.e., population characteristics, public services, commercial location, and transport. The dominant factors for the “population characteristics” dimension are the average age of the population, the years of education, and the population density in the street(s) where the urban renewal project is located. The leading factors for the “location of public services” dimension include the growth of leisure, education, and medical facilities within a distance of 0.5 km from the site of the urban renewal project. The leading factor in the “commercial location” dimension includes the growth in the number of medium and high-end hotels, business offices, and restaurants in the vicinity. The leading factors concerning the “location of transport” dimension are the growth in the number of transport facilities and the improvement in the road network density. The drivers may be classified into three levels: a strong driver, a medium driver, and a weak driver, according to the differences in the degree of their impact on the housing premiums. Among them, the strongest driver for achieving a high premium on house prices is the location of transport, while the remaining factors constitute weak drivers. The strong and medium drivers for the low premium level are the location of transport and overcrowding, while public services and the location of commercial organizations are weak drivers at this level. The basic drivers of urban renewal on housing at negative premium levels are all weak drivers.

![Figure 7](image_url)

**Figure 7.** The mechanism whereby key factors influence housing premiums in urban renewal.

### 5. Conclusions and Policy Recommendations

Based on using multi-source data, innovative integrated econometric models, machine learning, and geographic detector analysis, this study revealed the impact and mechanism of urban renewal on the housing premiums for local housing. First, PSM-DID was used to identify the premium caused by urban renewal by conducting Logit regression on the binary explanatory variable (whether urban renewal is conducted or not). The PSM obtains the value of the propensity score for each sample and matches the urban renewal areas (experimental group) with the non-renewed areas (control group). The DID conducted on this basis allows the experimental and control groups to satisfy the parallel hypothesis and mitigate the problem of selection bias of treatment effects as much as possible. Subsequently, a multivariate model was constructed by an integrated algorithm to simulate and reflect the affected housing prices associated with urban renewal projects; empirically validate the different factors of influence, including their strengths; and identify the main factors of influence. Then, the Geo-detector method was used to analyze the main factors of influence and the interaction relationships in urban renewal projects at different levels of premium. The following conclusions were made:

Firstly, during the period of 2008–2018, there was a significant positive premium effect of urban renewal on the overall unit price of local housing transactions, with the premium concentrated in the range of 9000–14,000 yuan per m². This finding is similar to the results of empirical studies for cities in Hong Kong, Taiwan, Berlin, and South Yorkshire,
which show that urban renewal can lead to significant increases in house prices, including commercial property rent \([36,50–52]\). This is due to the fact that urban renewal can eliminate neighborhood and housing negative externalities and will also have a gentrification effect, raising local housing prices and squeezing out low-income earners, which in turn will continue to drive up house prices \([21]\).

Secondly, demographic characteristics such as the population density and years of education of the student population (age > 15 years), as well as the optimized level of transport facilities and the density of the road networks, were the main factors that determined the housing premium effect generated by urban renewal projects. The projects located in high-density built-up areas and in aging communities had a more pronounced effect on the local housing premium. The improvement in road traffic conditions was an important factor driving the effect of urban renewal on local house prices.

Thirdly, different drivers, such as the location of transport, the location of public services, the location of commercial activities, and the demographic characteristics and their interactions, had different effects on housing premiums for the various urban renewal projects. Among them, the housing premiums generated by urban renewal projects with high premium effects were driven mainly by the interaction of the road network density and the increase in the number of transport facilities. The main drivers for the low premium effect of urban renewal projects on local housing were the increase in the population density in the street(s) and the density of the road network. The small number of negative premium samples was mainly influenced by the short-term unmeasured negative externalities connected with the implementation of the urban renewal projects. Based on the above findings, it may be concluded that this research has revealed the divergent driving factors for the spillover effect of urban renewal on house prices.

The above empirical results of our study provide practical evidence for the link between urban renewal and the real estate market. Results reveal more precisely the influence of urban renewal on local house price premiums and can provide useful background information and act as a reference for urban renewal and refinement of urban developments in Shenzhen and other fast-growing economies well into the future. To be specific, the government should promote a sustainable urban renewal model, improve local and microregulation of the real estate market while promoting urban renewal projects, establish and improve the long-term mechanism of the real estate market, and promote the healthy development of urban renewal and the real estate market.

**Author Contributions:** Conceptualization, S.W.; methodology, J.W.; software, J.W.; validation, J.W., K.L. and Y.L.; formal analysis, J.W.; investigation, J.W.; resources, S.W.; data curation, S.W.; writing—original draft preparation, X.L.; writing—review and editing, X.L., J.W., K.L., Y.L.; visualization, S.W.; supervision, S.W.; project administration, S.W.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Humanities and Social Science Foundation of the Ministry of Education of China (no. 21YJAZH087).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

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

**References**

43. Liu, R.M.; Zhao, R.J. Western development: Growth drive or policy trap—An analysis based on PSM-DID method. *China Ind. Econ.* 2015, 6, 32–43.
52. Liu, X. Housing renewal policies, house prices and urban competitiveness. *Appl. Geogr.* 2010, 30, 221–228. [CrossRef]