

Article

Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data

Yixiong Xiao ¹ , Xiang Chen ² , Qiang Li ^{1,*}, Xi Yu ¹, Jin Chen ³ and Jing Guo ¹

¹ Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; yxx@mail.bnu.edu.cn (Y.X.); yx@mail.bnu.edu.cn (X.Y.); guojing_92@163.com (J.G.)

² Department of Emergency Management, Arkansas Tech University, Russellville, AR 72801, USA; chenxiangpeter@gmail.com

³ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China; chenjin@bnu.edu.cn

* Correspondence: liqiang@bnu.edu.cn; Tel.: +86-010-5880-2876

Received: 9 October 2017; Accepted: 13 November 2017; Published: 15 November 2017

Abstract: The housing market in Chinese metropolises have become inflated significantly over the last decade. In addition to an economic upturn and housing policies that have potentially fueled the real estate bubble, factors that have contributed to the spatial heterogeneity of housing prices can be dictated by the amenity value in the proximity of communities, such as accessibility to business centers and transportation hubs. In the past, scholars have employed the hedonic pricing model to quantify the amenity value in relation to structural, locational, and environmental variables. These studies, however, are limited by two methodological obstacles that are relatively difficult to overcome. The first pertains to difficulty of data collection in regions where geospatial datasets are strictly controlled and limited. The second refers to the spatial autocorrelation effect inherent in the hedonic analysis. Using Beijing, China as a case study, we addressed these two issues by (1) collecting residential housing and urban amenity data in terms of Points of Interest (POIs) through web-crawling on open access platforms; and (2) eliminating the spatial autocorrelation effect using the Eigenvector Spatial Filtering (ESF) method. The results showed that the effects of nearby amenities on housing prices are mixed. In other words, while proximity to certain amenities, such as convenient parking, was positively correlated with housing prices, other amenity variables, such as supermarkets, showed negative correlations. This mixed finding is further discussed in relation to community planning strategies in Beijing. This paper provides an example of employing open access datasets to analyze the determinants of housing prices. Results derived from the model can offer insights into the reasons for housing segmentation in Chinese cities, eventually helping to formulate effective urban planning strategies and equitable housing policies.

Keywords: Points of Interest (POIs); hedonic pricing model; accessibility; amenity value; spatial autocorrelation

1. Introduction

Over the last decade, the housing markets in Chinese metropolises have become inflated to an unprecedented extent. Due to rapid urbanization and domestic migration, cities are not able to accommodate medical, educational, transportation, and cultural needs [1]. These services or amenities, such as hospitals, schools, and transportation hubs, are distributed unevenly across urban landscapes. As a result, the housing markets tend to manifest a clear pattern of spatial heterogeneity in terms of property values.

Although housing prices vary according to proximity to conveniences, the values of these amenities cannot be easily quantified. Exploring the correlation between the amenity value and housing prices is the first step to understand the complexities of real estate valuation and to formulate more effective housing policy initiatives [2]. In the past, the hedonic pricing model has been pervasively explored in an attempt to establish this correlation. The amenity variables are typically categorized into structural (e.g., size, age of buildings), locational (e.g., availability of public transit, accessibility to the Central Business District [CBD]), and environmental factors (e.g., availability of green spaces, scenic views) near the property [3–7].

Unlike most US cities that underwent urban decentralization or exurban settlement, which required greater reliance on personal cars [8], city dwellers in Asian countries are more heavily reliant on various forms of mass transit. According to a recent survey, only a limited percentage (10.9%) of commuters in Beijing utilize private cars, while the majority (48.6%) are bus and subway riders [9]. In such a public transit-oriented environment, living close to urban amenities and transit stations becomes a necessity. This dependency on urban amenities and the pursuit of convenient transportation is largely the impetus for the popularity of centralized community development, where location advantage outperforms other non-spatial factors in the formulation of housing prices [10]. To demystify the intricacies of real estate valuation in Asian cities, recent studies have focused on exploring the correlations between property values and geographical accessibility to services and facilities, including retailers [11,12], metro stations [13–15], hospitals [16], and schools [17,18].

These studies, however, are limited by two methodological obstacles that make it a challenge to analyze housing prices in China. The first is problems with data collection, due to the fact that information on housing prices and amenities is not available to the public. Traditional methods such as onsite investigations or manual georeferencing are relatively labor-and-time intensive and cannot be performed on a large scale. The second obstacle pertains to the spatial autocorrelation effect inherent in property values, or the correlation of spatial attributes in nearby properties. As a result, failing to account for this effect in the hedonic regression obfuscates the estimation results [19].

Using Beijing as a case study, we addressed these two issues through an enhanced hedonic pricing model. First, we gathered a total of 6959 residential housing prices and 76,786 urban amenities in terms of points of interest (POIs) through web-crawling on open access platforms. Second, we employed an improved model to eliminate spatial autocorrelation in hedonic regression using the Eigenvector Spatial Filtering (ESF) method. The results of the study offer insights into urban space planning strategies that will stabilize the housing market.

This paper is organized as follows: Following the introduction, Section 2 briefly reviews the history of housing development and issues of data collection in China. It also includes discussions of past studies on the hedonic regression model and attempts to address the spatial autocorrelation effect; Section 3 includes a formulation of the enhanced hedonic pricing model; Section 4 describes the procedure of data collection on open access platforms and the implementation of the model to the study area; Section 5 discusses the results, followed by Section 6, which specifies the strengths, limitations, and implications for urban studies.

2. Literature Review

2.1. Housing Development and Data Issues in China

Housing markets in cities differ significantly in terms of locations and prices. For example, research of Western cities has shown a clear segmentation pattern due to racial segregation and economic exclusion [20]. In addition to the political and ethical impact of these trends, the value of a single property is always subject to structural, locational, and environmental variables in its neighborhood [21]. While the mantra “location, location, location” is critical for determining investment potential in the housing market [22], locational (e.g., distance to CBD; [23]) and environmental variables (e.g., scenic views; [4,24]) are often relatively difficult to quantify.

In China, housing prices are less determined by neighborhood segmentation than their proximity to urban amenities. Since the economic reform of the 1990s that restructured urban land use, the housing market has gradually transitioned from government-assigned units into a free-trade system [25]. Before the reform, housing was characterized by work-unit compounds (the Chinese *danwei*), where residents lived and worked within the same community and were provided with a wide range of amenities, such as retail stores, schools, and hospitals [26,27]. The commodification of residential housing has substantially transformed the urban landscape and, therefore, residents' commuting patterns. Due to heavy reliance on public transit systems, resident mobility has become the primary momentum for determining property value in the housing market [28]. In other words, housing units with relative proximity to urban centers, amenities, and transportation hubs tend to have considerably higher market value. For example, in a recent study on the effect of the proximity to Beijing railway stations on the values of surrounding properties, it was found that while both non-transfer and transfer stations positively affect the price, the impact of transfer stations nearby was more significant [13]. The vital role that public transit plays can also be extended to suburban regions, where the housing units located within the transit-oriented development zones are priced much higher than the units further away [29]. In addition, property value increases due to other retail amenities, such as convenience stores, supermarkets, and drugstores that cater to nearby residents [11,12,16]. In this regard, exploring the effect of price determinants, especially accessibility to public transit and retail amenities, is a fundamental step towards understanding the Chinese housing market.

The spatial granularity of datasets is another essential aspect of urban studies in China. With the constant emergence and shifting of Chinese cities and townships, there has been an inconsistency in the definition of administrative units for the census and economic purposes [30]. This lack of clear demarcation is further exacerbated when the hierarchy of administrative units is directly involved in the identification of zonal patterns. However, the fine-scale, urban geospatial datasets (e.g., transport infrastructure, housing units) are relatively scant due to flaws in the data sharing policies and confidentiality concerns. Although there has been a recent resurgence of volunteered geographic information (VGI) collected from social media platforms [31], there remains a pressing need for alternative sources of urban datasets with more clearly defined spatiotemporal granularity.

Recent explorations of POI data on open access platforms have provided new opportunities to address these data issues. Particular amenities in terms of POIs can be extracted and categorized, as POIs represent the finest spatial resolution of the built environment of urban landscapes. Many scholars have employed POIs to evaluate potential city growth [32,33], as well as to analyze land use and classification strategies [34]. There is an enormous need to utilize POIs for housing price studies in China, as the pace at which new communities sprout and the sheer number of geospatial variables generated in the process cannot be efficiently captured in a traditional manner.

2.2. Hedonic Pricing Model and Spatial Autocorrelation

Studies on property values have often resorted to the hedonic pricing model [21], the theoretical foundation of which stems from the consumer theory [35]. In the later theory, it is assumed that merchandise is composed of a collection of quantifiable characteristics, and consumers make purchases by logically ranking the characteristics that best suit their needs. Residential properties, as a particular form of merchandise, can thus be appraised as a function of structural, locational, and environmental determinants [3,6].

The hedonic pricing model is plagued by the critical issue of the statistical effect of spatial autocorrelation [19], which violates the assumption of independence in regression analysis. Neglecting the spatial autocorrelation in the hedonic pricing model may lead to biased estimation and insufficient confidence levels [36]. Thus, two spatial autoregressive approaches, the spatial error model and the spatial lag model, have been proposed to ameliorate this effect [37,38]. If the regression is expressed as $y = \alpha + \beta X + e$, the spatial error model considers that the autocorrelation effect exists in the component of the error (e) [6]; the spatial lag model, on the other hand, considers the autocorrelation effect in the

dependent variable (y) [38,39]. In spite of the fact that these two approaches have been employed in many cases of hedonic modeling, two challenges remain. First, it is uncertain if the autocorrelation effect exists in the random error e or in the dependent variable y [40]. This lack of specification raises fundamental questions about the usefulness of this model. Second, incorporating the autoregressive method poses technical barriers to model estimation with regard to the ordinary least square regression (OLS) [41].

Based on Moran's autocorrelation coefficient (also known as Moran's I), the ESF method has been proposed to tackle these two methodological challenges [41]. Compared to the nonlinear approaches, the ESF method provides a linear formulation and solution. This method has been implemented in various cases of spatial analysis, such as crime analyses [42], migration [43], and species distribution [44]. However, the model has rarely been applied to real estate valuation [45].

3. Method

The fundamental premise of the hedonic pricing model is that a product is marketed as a set of characteristics, which can be formulated in a matrix $\mathbf{X} = (x_1, x_2, x_3, \dots, x_n)$ [46]. The market price of the product can thus be illustrated as a function of these characteristics or $y = f(x_1, x_2, x_3, \dots, x_n)$. The general form of the hedonic pricing model is shown in Equation (1):

$$y = \alpha + \beta\mathbf{X} + e \quad (1)$$

where y is the price vector, \mathbf{X} is the matrix of independent variables, α , β are the vectors of coefficients, and e is the vector of random error. Researchers have noted that the distribution of both y and x_n could be skewed, leading to the problem of heteroscedasticity [21,47,48]. Thus, the Box-Cox transformation has been applied to Equation (1) during implementation of the model.

The hedonic price, as discussed in the last section, is influenced by the spatial autocorrelation effect. The theoretical foundation to address the issue is rooted in the findings that spatial correlation can be interpreted using a set of eigenvectors (EVs) from Moran's I [49,50]. The ESF method was proposed to justify using EVs to represent the spatial autocorrelation effect in regression analysis [51]. Here, a set of selected eigenvectors $\mathbf{E}_k = (EV_1, EV_2, EV_3, \dots, EV_k)$ are presented [52], as shown in Equation (2). By introducing $\gamma\mathbf{E}_k$ (where γ is the coefficient of the eigenvectors), the method is able to isolate the spatial autocorrelation effect in the hedonic regression:

$$y = \alpha + \beta\mathbf{X} + \gamma\mathbf{E}_k + e \quad (2)$$

The eigenvectors \mathbf{E}_k in Equation (2) are selected from all the eigenvectors $\mathbf{E}_n = (EV_1, EV_2, EV_3, \dots, EV_n)$ that correspond to the matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)$, according to the Moran's I, as shown in Equation (3):

$$\text{Moran's } I = \frac{n}{\mathbf{1}^t\mathbf{C}\mathbf{1}} \frac{\mathbf{X}^t(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)\mathbf{X}}{\mathbf{X}^t(\mathbf{I} - \mathbf{1}\mathbf{1}^t/n)} \quad (3)$$

where \mathbf{C} is an n -by- n binary spatial weight matrix, \mathbf{I} is an n -by- n identity matrix, $\mathbf{1}$ is an n -by-1 vector of 1, and n is the number of observations.

Employing all n EVs is not practical, as it may cause over-specification when the number of variables is equal to or greater than the number of observations [53]. Therefore, we selected k EVs as a subset of n EVs ($k \ll n$) using a two-step procedure. First, we determined a threshold to preselect m EV candidates ($k \leq m < n$), as proposed in a previous study [54]. The threshold was set to Moran's $I_m/\text{Moran's } I_{\max} > 0.25$. The m EV candidates and the 20 independent variables (i.e., the variables in the case study) were employed to construct the original model. Second, we applied a stepwise regression to the original model to select the set with the best fit, including the k EVs and l ($l \leq 20$) independent variables in the final model.

4. Case Study

4.1. Data Collection

Beijing, the capital of China, has long been regarded as the cultural, financial, and political center of the nation. With an estimated population of 22 million, Beijing had a gross domestic product (GDP) value of 2.49 trillion RMB in 2016. This capital city is demarcated by the Fifth Ring Road (Figure 1), circumscribing an urban area of approximately 750 square kilometers. The properties beyond this ring road were excluded from the study, because these suburban areas support very few urban functions.

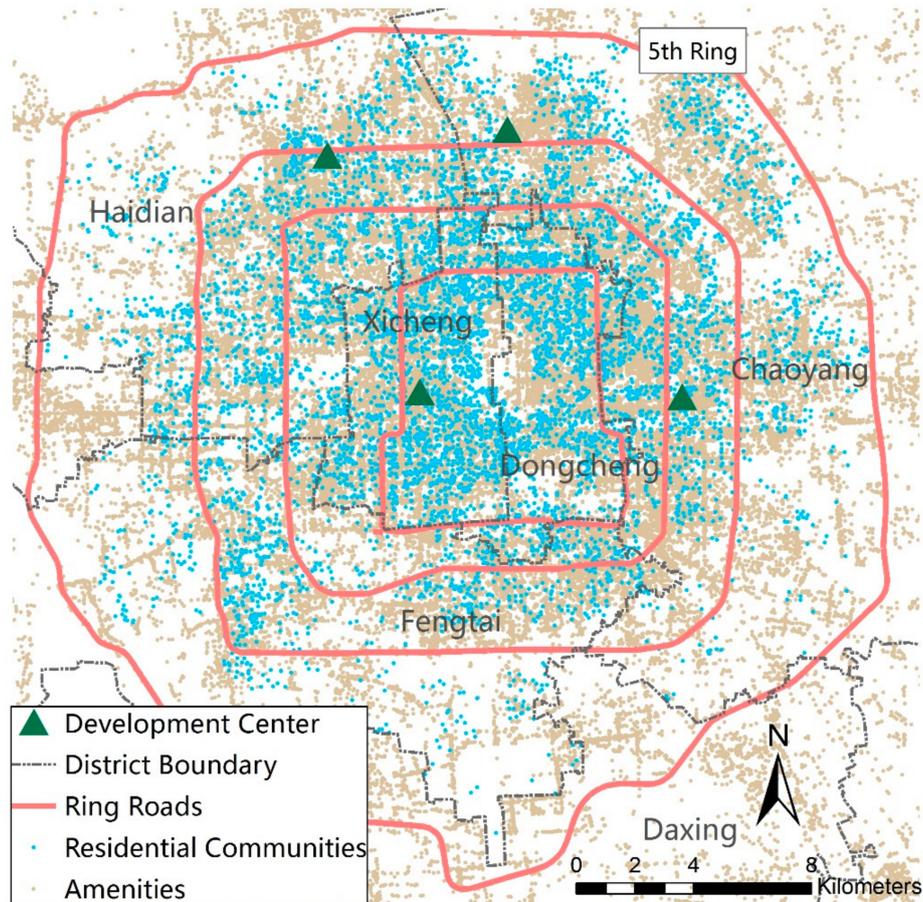


Figure 1. The distribution of housing and amenity POIs in Beijing.

The datasets for this study included both POIs of the properties and those of amenities nearby. First, the amenity POI data were gathered from Gaode Maps (<http://ditu.amap.com/>), the primary product of the AutoNavi Software. The amenity POI data were gathered using a POI extractor developed based on the Gaode Map JavaScript application programming interface (API). The distribution of the POIs included 76,786 amenities, as shown in Figure 1.

Due to the lack of real estate information in the Gaode Maps, we augmented the POI database with data sourced from *Lianjia* (<http://www.lianjia.com/>), the largest real estate brokerage firm in Beijing. We collected 6959 purchase transactions in the third and fourth quarters of 2015 using a web crawling Python program and georeferenced them by property addresses. Most of these properties were typical housing units, apartments within a gated community (the Chinese *xiaoqu*). We further classified the property characteristics into structural, locational, and environmental variables to explore their impact on housing prices, as shown in Table 1. In addition, we determined the local Air Quality Index (AQI) at each location from an interpolated surface of the mean AQI in 2015 [55], as shown in

Figure 2. This index serves as an essential environmental indicator of the city that has been shrouded in smog pollution [56]. The amenities variables included eight subcategories (Table 2). In reference to other research on housing prices [57,58], we also added four location dummy variables to examine the differences among administrative regions (i.e., *Xicheng*, *Dongcheng*, *Chaoyang*, and *Haidian*) and a time dummy variable to account for price fluctuations over the two quarters. To compare the intensity of the influences, the variables in Tables 1 and 2 were all rescaled.

Table 1. Classification of property characteristics.

| Category | Variable | Definition | Min | Max | Mean | Std. |
|-------------|--------------------|---|-----------|---------|---------|--------|
| Structure | Size | Floor space of the property in square meters | 15.00 | 310.00 | 75.58 | 33.70 |
| | Bedrooms | Number of bedrooms | 1.00 | 6.00 | 1.92 | 0.74 |
| | Living rooms | Number of living rooms | 0.00 | 4.00 | 1.07 | 0.49 |
| | Orientation | Number of windows facing north (1), west (2), east (3), and south (4). The value is the total number of windows in the living room and the master bedroom | 2.00 | 8.00 | 5.83 | 1.82 |
| | Age | 2015 minus the year built | 1.00 | 61.00 | 18.75 | 9.16 |
| Location | Development center | Time cost (seconds) to the nearest development center ¹ (Figure 1) | 1.00 | 3746.00 | 1821.47 | 462.81 |
| | <i>Xicheng</i> | Property located in the Xicheng District (1) or otherwise (0) | 0.00 | 1.00 | 0.16 | 0.37 |
| | <i>Dongcheng</i> | Property located in the Dongcheng District (1) or otherwise (0) | 0.00 | 1.00 | 0.12 | 0.33 |
| | <i>Chaoyang</i> | Property located in the Chaoyang District (1) or otherwise (0) | 0.00 | 1.00 | 0.42 | 0.49 |
| | <i>Haidian</i> | Property located in the Haidian District (1) or otherwise (0) | 0.00 | 1.00 | 0.16 | 0.37 |
| | Amenities | Geographical accessibility to type-N amenity | (Table 2) | | | |
| Environment | Air quality | Local Air Quality Index ² (Figure 2) | 109.58 | 121.80 | 115.76 | 3.03 |
| Time | Time dummy | Time of the transaction made in the third (0) or fourth (1) quarter of 2015 | 0.00 | 1.00 | 0.56 | 0.50 |

¹ The development centers were derived from the Beijing Municipal Commission of Urban Planning (<http://www.bjghw.gov.cn/>). ² The local AQI at each property location was extracted from an interpolated surface based on the mean AQI of 35 monitoring stations in 2015. This data was collected from the Beijing Municipal Environment Center (<http://www.bjmecm.com.cn/>).

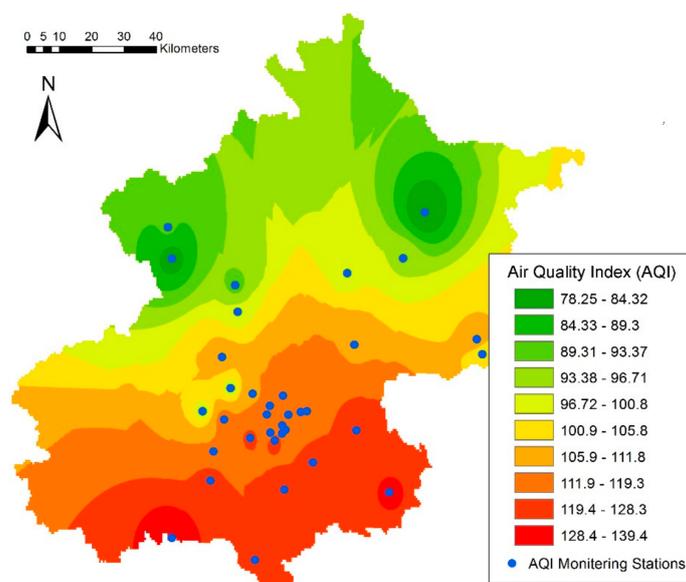


Figure 2. The distribution of AQI and monitoring stations in Beijing (2015).

Table 2. Accessibility measures of amenities by subcategory.

| Variable | Min | Max | Mean | Std. |
|--------------------|------|--------|-------|-------|
| Bus stops | 1.67 | 11.99 | 6.23 | 1.82 |
| Convenience stores | 3.32 | 48.73 | 17.71 | 6.85 |
| Metro stations | 0.00 | 2.90 | 0.63 | 0.34 |
| Parking lots | 6.15 | 205.01 | 70.61 | 24.61 |
| Primary schools | 0.03 | 7.43 | 2.17 | 1.14 |
| Supermarkets | 2.65 | 31.80 | 12.59 | 4.10 |
| Shopping malls | 0.00 | 22.07 | 2.69 | 3.08 |
| Gyms | 1.67 | 11.99 | 6.23 | 1.82 |

4.2. Accessibility Measure

One widely-utilized method to quantify the amenity value is the accessibility measure. In transportation research, the geographical accessibility to amenities can be evaluated as the function of distance or other cost variables, such as time, money, discomfort, and risk [59]. Another facet of the accessibility measure is the emphasis on the spatiotemporal constraints on human activities with regard to time geography [60]. As the exploration of various accessibility metrics was not the focus of this study, we applied a classic accessibility measurement, the cumulative opportunity method, to estimate the effect of spatial separation, as shown in Equation (4). This accessibility index (A_{ij}) summarizes the total number of amenities near the property, where d_{ij} is the Euclidean distance from property i to amenity j and D is the threshold distance. We included a negative-linear function [59,61] to determine the distance decay. As a walkable distance to urban centers has been identified as in the range of 0.5–2 miles [62], we chose 1.5 km as the threshold D . The accessibility indices by amenity subcategory are shown in Table 2.

$$A_{ij} = \sum f(d_{ij})$$

$$f(d_{ij}) = \begin{cases} (1 - d_{ij}/D) & \text{for } d_{ij} < D \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

5. Results and Discussion

In this study, we utilized both OLS and ESF models for the hedonic analysis. The results are shown in Table 3. The relatively low variation inflation indicator (VIF) suggested that no severe collinearity existed within the estimated variables. The Breusch-Pagan test indicated no significant heteroscedasticity in the regression residuals of the two models. Although the Jarque-Berar test results show that the residuals in both models are not normally distributed, they do not affect the asymptotic normality of the OLS estimators when the sample size is large [63]. While there was a positive spatial correlation in the residuals of the OLS model (Moran's $I = 0.2324$, $p < 0.001$), no significant autocorrelation was detected in the ESF residuals (Moran's $I = -0.03$, $p > 0.1$). This result indicated that the ESF model is capable of eliminating the spatial autocorrelation effect in the regression analysis. The adjusted R^2 values (OLS = 0.8384, ESF = 0.8768) suggested that the data had a better fit for the ESF model. The OLS model failed to identify the correlation with the accessibility to bus stops (ESF $\beta = -0.0059$), metro stations (ESF $\beta = 0.0045$), and parking lots (ESF $\beta = 0.0119$).

We will now discuss the results in more detail with respect to the structural, locational, and environmental characteristics. As both models revealed, the structural variables had a significantly positive influence on housing prices, as the property value is dictated mainly by the floor space (size, ESF $\beta = 0.1736$) and floor plans (bedrooms, ESF $\beta = 0.0221$; living rooms, ESF $\beta = 0.0287$). The correlation with the orientation variable (ESF = 0.0098) can be explained by consumers' preferences for properties with plentiful natural light. Due to the centralized location of Beijing, proximity in terms of travel time to the nearest development center was negatively correlated (ESF $\beta = -0.006$). The positive β values for the four district dummy variables (*Xicheng*, *Dongcheng*, *Chaoyang*, *Haidian*)

suggested that property prices in the inner cities were significantly higher than those in suburban regions; and the prices in the *Xicheng* district outnumbered other districts. The accessibility to most amenities was positively correlated because these amenities fulfill people's needs for transportation (metro stations, ESF $\beta = 0.0045$), shopping (convenience stores, ESF $\beta = 0.0126$; shopping malls, $\beta = 0.0058$), and exercise (gyms, ESF $\beta = 0.0117$). The positive correlation with parking lots (ESF $\beta = 0.0119$) can be evidenced by the rapid increase in popularity of private cars in Beijing over the past decades; thus, parking spaces have become a valued resource and an indispensable element of new community development [64,65]. While the accessibility to primary schools was identified as positively correlated with housing prices, we feel that the strength of this correlation was underestimated in this study (ESF $\beta = 0.0083$). The influence of schools on the neighborhood housing market depended not only on proximity but, more importantly, on the good reputation of nearby schools [66]. The case of Beijing is more complex because school zoning is not a direct reflection of proximity but is affiliated with individual communities or *xiaoqu*. The demarcation by community in lieu of physical distance, has been largely criticized for creating inequality of educational opportunities and causing social segregation throughout the city [67].

Table 3. Regression analysis results of housing price determinants utilizing OLS and ESF.

| | | OLS | | | ESF | | |
|-------------------------|--------------------|---------|--------|-----------------|---------|--------|-----------------|
| | | β | VIF | <i>p</i> -Value | β | VIF | <i>p</i> -Value |
| Structure | Size | 0.1765 | 3.614 | *** | 0.1736 | 3.9289 | *** |
| | Bedrooms | 0.0213 | 2.7127 | *** | 0.0221 | 2.8425 | *** |
| | Living rooms | 0.0291 | 1.6303 | *** | 0.0287 | 1.7267 | *** |
| | Orientation | 0.0114 | 1.2112 | *** | 0.0098 | 1.2598 | *** |
| | Age | −0.0251 | 2.0304 | *** | −0.0318 | 2.4164 | *** |
| Location | Development center | −0.0162 | 1.4022 | *** | −0.006 | 2.9418 | ** |
| | <i>Xicheng</i> | 0.2258 | 2.3692 | *** | 0.2237 | 4.6954 | *** |
| | <i>Dongcheng</i> | 0.184 | 2.0835 | *** | 0.1676 | 3.8533 | *** |
| | <i>Chaoyang</i> | 0.0282 | 2.7626 | *** | 0.0198 | 5.9633 | *** |
| | <i>Haidian</i> | 0.1179 | 3.4012 | *** | 0.1261 | 5.8366 | *** |
| | Bus stops | NA | NA | NA | −0.0059 | 2.4996 | *** |
| | Convenience stores | 0.0044 | 2.3815 | * | 0.0126 | 4.3491 | *** |
| | Metro stations | NA | NA | NA | 0.0045 | 2.1238 | ** |
| | Parking lots | NA | NA | NA | 0.0119 | 3.282 | *** |
| | Primary schools | 0.0037 | 1.5073 | * | 0.0083 | 2.468 | *** |
| | Supermarkets | −0.0065 | 1.8605 | *** | −0.0126 | 3.4714 | *** |
| | Shopping malls | 0.0096 | 1.3314 | *** | 0.0058 | 2.7566 | ** |
| | Gyms | 0.0061 | 1.7607 | *** | 0.0117 | 3.6287 | *** |
| Environment | Air quality | −0.066 | 2.3913 | *** | −0.0601 | 5.7194 | *** |
| Time | Time dummy | 0.0162 | 1.0043 | *** | 0.0182 | 1.0146 | *** |
| Adjusted R ² | | 0.8384 | | *** | 0.8768 | | *** |
| Residuals | Breusch-Pagon | 0.5071 | | 0.4764 | 2.2747 | | 0.1315 |
| | Jarque-Bera | 3351.4 | | *** | 17021 | | *** |
| | Moran's I | 0.2324 | | *** | −0.03 | | 1 |

* Result is significant at *p*-value = 0.05; ** Result is significant at *p*-value = 0.01; *** Result is significant at *p*-value = 0.001.

We were surprised by our finding that housing prices were negatively correlated with accessibility to bus stops (ESF $\beta = -0.0059$) and supermarkets (ESF $\beta = -0.0126$). We feel this result can be explained by the formation of large-scale housing estates in suburban Beijing, such as *Huilongguan* in the northwest and *Wangjing* in the northeast. These estates were developed in the early 2000s to ease conflicts caused by limited land use in the inner city. Although housing prices have skyrocketed in these communities over the last decade, certain types of amenities are still limited. This phenomenon is very likely the result of the traditional residential-oriented communities, which were literally surrounded by walls that prevented easy access to nearby amenities [27]. These communities have also been criticized for generating excessive traffic due to the spatial mismatch of job opportunities and housing units [68,69]. Thus, recent urban planning guidelines in China have focused on the removal of walls

and fences in the new community designs in an attempt to ease traffic congestion and improve access to amenities [70].

Not surprisingly, we found that the AQI was negatively correlated with housing prices (ESF $\beta = -0.0601$). This finding is in line with previous hedonic analyses of housing prices based on air pollutants such as PM-10 and sulfur dioxide (SO₂) in Chicago [71] and Seoul [39]. This result revealed that the property value can be affected by the natural environment near the housing unit. Thus, alleviating air pollution and promoting sustainable urban design should be a priority for both community developers and urban planners.

6. Conclusions

The use of the hedonic pricing model has a long tradition in urban studies. In this study, we have developed a complete housing and amenity POI database via web-crawling and employing an open access API. We also incorporated the ESF method to eliminate the spatial autocorrelation effect inherent in hedonic analysis. It is our hope that the results, in addition to revealing the roles that structural, locational, and environmental variables play in the housing market, will help urban planners and stakeholders to better understand the local real estate economy [72].

The contribution of the paper is two-fold. We utilized the ESF method to eliminate the spatial autocorrelation effect in hedonic analyses of housing prices. While most studies on price determinants fail to consider spatial autocorrelation, our study revealed that the level of correlation may fluctuate when a set of eigenvectors are incorporated into the hedonic price. On the other hand, this study provides a solid example of acquiring fine-scale urban datasets from open sources. Unlike most Western countries where geospatial data is public record on various geographical scales (e.g., the US TIGER products), in China datasets containing georeferenced urban features are not available in the public sector. Proprietary data are often limited in size, granularity, and timeliness of information delivery. The increasing transparency of data sharing policies and the emergence of the map service providers such as Google, Bing, Baidu, and Gaode have made this geographical information more accessible to public users. However, full caution is needed when employing such open access platforms. For example, housing prices and structural variables in this study were sourced from a private company and might not adequately represent their actual market value. Therefore, triangulation of the dataset with other sources is necessary to ensure data quality.

Methodological limitations of this study should be recognized by other scholars as opportunities for further research. First, we did not consider the service capacity of amenities and variations of distance decay in the accessibility assessment; and the spatial separation to amenities can be network-based [62,73] or follows a power law [12], thus mediating the correlation [13]. Second, in the future, researchers should focus on communities located in a smaller urban sector rather than an entire city to minimize the impact of geographical heterogeneity. Being aware of these methodological limitations is a prerequisite for model improvement.

In conclusion, in addition to the three tiers of price determinants explored in the study, the Chinese housing markets are largely dictated by local policies that restrict property transactions. In Beijing, the desire to buy a new property can be thwarted by economic externalities, such as lack of a residency permit (the Chinese *hukou*), educational resources, and investment potential. For example, in Beijing, school zoning is a dominant factor fueling the housing market and creating residential economic segregation [74]. The enforcement of limits on home purchases and residency requirements is another set of factors that affect the health of the housing market [75]. To fully understand the facets of the housing market in Chinese cities, these policy initiatives and regulations must be thoroughly assessed.

Author Contributions: Yixiong Xiao designed the experiment, collected the data, performed the analysis, and drafted the manuscript; Xiang Chen performed the accessibility measurements and reorganized the manuscript; Qiang Li supervised the data analysis and created a research agenda for project implementation; and Xi Yu, Jin Chen, and Jing Guo assisted with adjusting the model and analyzing the data.

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

References

1. Cohen, B. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technol. Soc.* **2006**, *28*, 63–80. [[CrossRef](#)]
2. D'Acci, L. Monetary, subjective and quantitative approaches to assess urban quality of life and pleasantness in cities (hedonic price, willingness-to-pay, positional value, life satisfaction, isobenefit lines). *Soc. Indic. Res.* **2014**, *115*, 531–559. [[CrossRef](#)]
3. Jim, C.Y.; Chen, W.Y. External effects of neighbourhood parks and landscape elements on high-rise residential value. *Land Use Policy* **2010**, *27*, 662–670. [[CrossRef](#)]
4. Kaltenborn, B.P.; Andersen, O.; Nellemann, C. Amenity development in the Norwegian mountains. *Landsc. Urban Plan.* **2009**, *91*, 195–201. [[CrossRef](#)]
5. Yamagata, Y.; Murakami, D.; Yoshida, T.; Seya, H.; Kuroda, S. Value of urban views in a bay city: Hedonic analysis with the spatial multilevel additive regression (SMAR) model. *Landsc. Urban Plan.* **2016**, *151*, 89–102. [[CrossRef](#)]
6. Hui, E.C.M.; Chau, C.K.; Pun, L.; Law, M.Y. Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Build. Environ.* **2007**, *42*, 2333–2343. [[CrossRef](#)]
7. Kong, F.; Yin, H.; Nakagoshi, N. Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landsc. Urban Plan.* **2007**, *79*, 240–252. [[CrossRef](#)]
8. Clark, J.K.; McChesney, R.; Munroe, D.K.; Irwin, E.G. Spatial characteristics of exurban settlement pattern in the United States. *Landsc. Urban Plan.* **2009**, *90*, 178–188. [[CrossRef](#)]
9. Lin, D.; Allan, A.; Cui, J. Exploring differences in commuting behaviour among various income groups during polycentric urban development in China: New evidence and its implications. *Sustainability* **2016**, *8*, 1188. [[CrossRef](#)]
10. Wen, H.; Tao, Y. Polycentric urban structure and housing price in the transitional China: Evidence from Hangzhou. *Habitat Int.* **2015**, *46*, 138–146. [[CrossRef](#)]
11. Chiang, Y.H.; Peng, T.C.; Chang, C.O. The nonlinear effect of convenience stores on residential property prices: A case study of Taipei, Taiwan. *Habitat Int.* **2015**, *46*, 82–90. [[CrossRef](#)]
12. Jang, M.; Kang, C.-D. Retail accessibility and proximity effects on housing prices in Seoul, Korea: A retail type and housing submarket approach. *Habitat Int.* **2015**, *49*, 516–528. [[CrossRef](#)]
13. Dai, X.; Bai, X.; Xu, M. The influence of Beijing rail transfer stations on surrounding housing prices. *Habitat Int.* **2016**, *55*, 79–88. [[CrossRef](#)]
14. Zhang, M.; Meng, X.; Wang, L.; Xu, T. Transit development shaping urbanization: Evidence from the housing market in Beijing. *Habitat Int.* **2014**, *44*, 545–554. [[CrossRef](#)]
15. Xu, Y.; Zhang, Q.; Zheng, S. The rising demand for subway after private driving restriction: Evidence from Beijing's housing market. *Reg. Sci. Urban Econ.* **2015**, *54*, 28–37. [[CrossRef](#)]
16. Peng, T.; Chiang, Y. The non-linearity of hospitals' proximity on property prices: Experiences from Taipei, Taiwan. *J. Prop. Res.* **2016**, *9916*, 1–21. [[CrossRef](#)]
17. Wen, H.; Zhang, Y.; Zhang, L. Do educational facilities affect housing price? An empirical study in Hangzhou, China. *Habitat Int.* **2014**, *42*, 155–163. [[CrossRef](#)]
18. Nguyen-Hoang, P.; Yinger, J. The capitalization of school quality into house values: A review. *J. Hous. Econ.* **2011**, *20*, 30–48. [[CrossRef](#)]
19. Basu, S.; Thibodeau, T. Analysis of spatial autocorrelation in house prices. *J. Real Estate Financ. Econ.* **1998**, *17*, 61–85. [[CrossRef](#)]
20. Palm, R. Spatial Segmentation of the urban housing market. *Econ. Geogr.* **1978**, *54*, 210–221. [[CrossRef](#)]
21. Chau, K.W.; Chin, T.L. A critical review of literature on the hedonic price model. *Int. J. Hous. Sci. Its Appl.* **2003**, *2*, 145–165.
22. Kiel, K.A.; Zabel, J.E. Location, location, location: The 3L approach to house price determination. *J. Hous. Econ.* **2008**, *17*, 175–190. [[CrossRef](#)]
23. McMillen, D.P. The return of centralization to Chicago: Using repeat sales to identify changes in house price distance gradients. *Reg. Sci. Urban Econ.* **2003**, *33*, 287–304. [[CrossRef](#)]
24. Jim, C.Y.; Chen, W.Y. Value of scenic views: Hedonic assessment of private housing in Hong Kong. *Landsc. Urban Plan.* **2009**, *91*, 226–234. [[CrossRef](#)]

25. Wu, F. Changes in the structure of public housing provision in urban China. *Urban Stud.* **1996**, *33*, 1601–1628. [[CrossRef](#)]
26. Ta, N.; Chai, Y.; Zhang, Y.; Sun, D. Understanding job-housing relationship and commuting pattern in Chinese cities: Past, present and future. *Transp. Res. Part D Transp. Environ.* **2017**, *52*, 562–573. [[CrossRef](#)]
27. Huang, Y.; Jiang, L. Housing inequality in transitional Beijing. *Int. J. Urban Reg. Res.* **2009**, *33*, 936–956. [[CrossRef](#)]
28. Li, S.M.; Siu, Y.M. Residential mobility and urban restructuring under market transition: A study of Guangzhou, China. *Prof. Geogr.* **2001**, *53*, 219–229. [[CrossRef](#)]
29. Mathur, S.; Ferrell, C. Measuring the impact of sub-urban transit-oriented developments on single-family home values. *Transp. Res. Part A Policy Pract.* **2013**, *47*, 42–55. [[CrossRef](#)]
30. Ma Laurence, J.C.; MaCui, G. Administrative changes and urban population in China. *Ann. Assoc. Am. Geogr.* **1987**, *77*, 373–395. [[CrossRef](#)]
31. Chen, X.; Yang, X. Does food environment influence food choices? A geographical analysis through “tweets”. *Appl. Geogr.* **2014**, *51*, 82–89. [[CrossRef](#)]
32. Yue, Y.; Zhuang, Y.; Yeh, A.G.O.; Xie, J.-Y.; Ma, C.-L.; Li, Q.-Q. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *Int. J. Geogr. Inf. Sci.* **2016**, 1–18. [[CrossRef](#)]
33. Liu, X.; Long, Y. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environ. Plan. B Plan. Des.* **2016**, *43*, 341–360. [[CrossRef](#)]
34. Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira, J.; Pereira, F.C. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36–46. [[CrossRef](#)]
35. Lancaster, K.J. A new approach to consumer theory. *J. Polit. Econ.* **1966**, *74*, 132. [[CrossRef](#)]
36. Malpezzi, S. Hedonic pricing models: A selective and applied review. In *Housing Economics and Public Policy*; Blackwell Science Ltd.: Oxford, UK; pp. 67–89, ISBN 9780470690680.
37. Conway, D.; Li, C.Q.; Wolch, J.; Kahle, C.; Jerrett, M. A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *J. Real Estate Financ. Econ.* **2010**, *41*, 150–169. [[CrossRef](#)]
38. Anselin, L.; Le Gallo, J. Interpolation of air quality measures in hedonic house price models: Spatial aspects. *Spat. Econ. Anal.* **2006**, *1*, 31–52. [[CrossRef](#)]
39. Kim, C.W.; Phipps, T.T.; Anselin, L. Measuring the benefits of air quality improvement: A spatial hedonic approach. *J. Environ. Econ. Manag.* **2003**, *45*, 24–39. [[CrossRef](#)]
40. Anselin, L.; Rey, S. Properties of tests for spatial dependence in linear regression models. *Geogr. Anal.* **1991**, *23*, 112–131. [[CrossRef](#)]
41. Getis, A.; Griffith, D. A comparative spatial filtering in regression analysis. *Geogr. Anal.* **2002**, *34*, 130–140. [[CrossRef](#)]
42. Chun, Y. Analyzing space-time crime incidents using eigenvector spatial filtering: An application to vehicle burglary. *Geogr. Anal.* **2014**, *46*, 165–184. [[CrossRef](#)]
43. Chun, Y.; Griffith, D.A. Modeling network autocorrelation in space–time migration flow data: An eigenvector spatial filtering approach. *Ann. Assoc. Am. Geogr.* **2011**, *101*, 523–536. [[CrossRef](#)]
44. Diniz-Filho, J.A.F.; Bini, L.M. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Glob. Ecol. Biogeogr.* **2005**, *14*, 177–185. [[CrossRef](#)]
45. Helbich, M.; Griffith, D.A. Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Comput. Environ. Urban Syst.* **2016**, *57*, 1–11. [[CrossRef](#)]
46. Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Polit. Econ.* **1974**, *82*, 34. [[CrossRef](#)]
47. Ottensmann, J.R.; Payton, S.; Man, J. Urban location and housing prices within a hedonic model. *J. Reg. Anal. Policy* **2008**, *38*, 19–35.
48. Cropper, M.L.; Leland, B.D.; McConnell, K.E. On the choice of functional form for hedonic price functions. *Rev. Econ. Stat.* **1988**, *70*, 668–675. [[CrossRef](#)]
49. De Jong, P.; Sprenger, C.; van Veen, F. On extreme values of Moran’s I and Geary’s c. *Geogr. Anal.* **1984**, *16*, 17–24. [[CrossRef](#)]
50. Tiefelsdorf, M.; Boots, B. The exact distribution of Moran’s I. *Environ. Plan. A* **1995**, *27*, 985–999. [[CrossRef](#)]
51. Griffith, D.A. A linear regression solution to the spatial autocorrelation problem. *J. Geogr. Syst.* **2000**, *2*, 141–156. [[CrossRef](#)]

52. Haining, R.P. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2003; ISBN 0521774373.
53. Patuelli, R.; Griffith, D.A.; Tiefelsdorf, M.; Nijkamp, P. Spatial filtering and eigenvector stability: Space-time models for German unemployment data. *Int. Reg. Sci. Rev.* **2011**, *34*, 253–280. [CrossRef]
54. Tiefelsdorf, M.; Griffith, D.A. Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environ. Plan. A* **2007**, *39*, 1193–1221. [CrossRef]
55. Beijing Municipal Environment Monitoring Center. Available online: <http://www.bjmemc.com.cn/> (accessed on 13 November 2017).
56. Kay, S.; Zhao, B.; Sui, D. Can social media clear the air? A case study of the air pollution problem in Chinese cities. *Prof. Geogr.* **2015**, *67*, 351–363. [CrossRef]
57. Helbich, M.; Brunauer, W.; Vaz, E.; Nijkamp, P. Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Stud.* **2014**, *51*, 390–411. [CrossRef]
58. Helbich, M.; Brunauer, W.; Hagenauer, J.; Leitner, M. Data-driven regionalization of housing markets. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 871–889. [CrossRef]
59. Kwan, M.-P. Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geogr. Anal.* **1998**, *30*, 191–216. [CrossRef]
60. Chen, X.; Kwan, M.-P. Choice set formation with multiple flexible activities under space–time constraints. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 941–961. [CrossRef]
61. Black, J.; Conroy, M. Accessibility measures and the social evaluation of urban structure. *Environ. Plan. A* **1977**, *9*, 1013–1031. [CrossRef]
62. Chen, X.; Clark, J. Interactive three-dimensional geovisualization of space-time access to food. *Appl. Geogr.* **2013**, *43*, 81–86. [CrossRef]
63. Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, 5th ed.; South Western Cengage Learning: Mason, USA, 2015; ISBN 1111531048.
64. Wang, R.; Yuan, Q. Parking practices and policies under rapid motorization: The case of China. *Transp. Policy* **2013**, *30*, 109–116. [CrossRef]
65. Wang, Y.L.; Wang, X.; Zhang, M.C. Current situation and analysis of parking problem in Beijing. *Procedia Eng.* **2016**, *137*, 777–785. [CrossRef]
66. Jayantha, W.M.; Lam, S.O. Capitalization of secondary school education into property values: A case study in Hong Kong. *Habitat Int.* **2015**, *50*, 12–22. [CrossRef]
67. Wu, Q.; Zhang, X.; Waley, P. Jiaoyufication: When gentrification goes to school in the Chinese inner city. *Urban Stud.* **2016**, *53*, 3510–3526. [CrossRef]
68. Zhao, P. Sustainable urban expansion and transportation in a growing megacity: Consequences of urban sprawl for mobility on the urban fringe of Beijing. *Habitat Int.* **2010**, *34*, 236–243. [CrossRef]
69. Miao, P. Deserted streets in a jammed town: The gated community in Chinese cities and its solution. *J. Urban Des.* **2003**, *8*, 45–66. [CrossRef]
70. Zhao, W.; Zou, Y. Un-gating the gated community: The spatial restructuring of a resettlement neighborhood in Nanjing. *Cities* **2017**, *62*, 78–87. [CrossRef]
71. Chattopadhyay, S. Estimating the demand for air quality: New evidence based on the Chicago housing market. *Land Econ.* **1999**, *75*, 22–38. [CrossRef]
72. Li, H.; Wei, Y.D.; Yu, Z.; Tian, G. Amenity, accessibility and housing values in metropolitan USA: A study of Salt Lake County, Utah. *Cities* **2016**, *59*, 113–125. [CrossRef]
73. Chen, X. Take the edge off: A hybrid geographic food access measure. *Appl. Geogr.* **2017**, *87*, 149–159. [CrossRef]
74. Zheng, S.; Hu, W.; Wang, R. How much is a good school worth in Beijing? Identifying price premium with paired resale and rental data. *J. Real Estate Financ. Econ.* **2016**, *53*, 184–199. [CrossRef]
75. Du, Z.; Zhang, L. Home-purchase restriction, property tax and housing price in China: A counterfactual analysis. *J. Econom.* **2015**, *188*, 558–568. [CrossRef]

