



Area-Based Hedonic Pricing of Urban Green Amenities in Beijing: A Spatial Piecewise Approach

Zhaoyang Liu¹ · Heqing Huang² · Juha Siikamäki³ · Jintao Xu²

Accepted: 1 March 2024
© The Author(s) 2024

Abstract

This study explores a spatial piecewise approach for the hedonic valuation of the area of urban green space at different distances from a property, using a rich census dataset collected from Beijing. We explore three novel empirical strategies that improve the identification of the spatial boundary or threshold distance within which green space is capitalised into housing prices. We first delineated a series of concentric circles surrounding each property and measured the area of green space within each doughnut-shaped ring. We next estimated the hedonic price using three methods. The first is a regression spline model combined with a machine learning type of model selection procedure which objectively selects the exact location of the threshold distance that optimises the model's predictive performance. The second is a novel matching algorithm that minimises covariate imbalance for a continuous treatment variable (i.e., the area of green space) to provide stronger causal evidence on the hedonic prices of green space at different distances. The third is a spatial difference-in-differences approach that further accounts for endogeneity bias associated with unobserved factors. For our dataset, we found that housing prices are more likely to be affected by green space within a 1 km radius.

Keywords Urban green space · Hedonic pricing · Regression splines · Model selection · Covariate balancing matching for continuous treatment · Spatial difference-in-differences

JEL Classification Q51 · Q57 · R30

This paper is an output of the Ecosystem Services Accounting for Development project jointly initiated by the Environment for Development Initiative and the Swedish Environmental Protection Agency (SEPA Reference Number: NV-05230-14), and funded by the Swedish International Development Cooperation Agency (SIDA Reference Number: 13/000230). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The authors sincerely appreciate the very helpful comments provided by Dr. Jessica Alvsilver, Dr. Christopher Campbell-Durufié, Dr. Timothy Hamilton, Prof. Nick Hanley, Prof. Thies Lindenthal, Dr. Richard Mulwa, Dr. Matías Piaggio, Dr. Carolin Hoeltken, Dr. Per Stromberg, Dr. Byela Tibesigwa, Dr. Jane Turpie and Dr. Dawit Woubishet. In addition, the authors are greatly indebted to Mrs Cyndi Berck for editing the paper. Last but not least, the authors gratefully acknowledge two anonymous referees and the editor for their very helpful comments on an earlier version of the paper.

Extended author information available on the last page of the article

1 Introduction

Urban parks and other green amenities provide urban residents with recreational, health and aesthetic benefits, as well as many other desirable ecosystem services (Millennium Ecosystem Assessment 2005). Urban green amenities are one of the primary components of the urban ecosystem accounts of many large-scale ecosystem accounting frameworks, such as the United Nation's System of Environmental-Economic Accounting - Ecosystem Accounting (United Nations 2021), the UK's urban natural capital accounts (Office for National Statistics 2023) and the US pilot urban ecosystem accounting (Heris et al. 2021). These ecosystem accounting frameworks often depend on the monetary value of urban green amenities, which is usually quantified using environmental valuation techniques. The fact that urban green amenities are mostly open to visitors free of charge precludes a market price as a monetary measurement.

Hedonic pricing is one of the tried and trusted methods for the monetary valuation of urban green amenities. It considers green amenities surrounding a property as an attribute of the property, and seeks to derive the implicit price of green amenities as a component of the property's total market value. This paper presents a case study in Beijing. We test three novel empirical strategies to identify the maximum distance from a property within which urban green amenities are capitalised into the property's market value. This facilitates the measurement of the aggregate hedonic price by identifying which green spaces in a city should be included in a hedonic valuation, and thereby contributes to the extensive global efforts in standardising and mainstreaming ecosystem accounting.

There is a vast array of literature on the hedonic pricing of urban green amenities (see Bockarjova et al. 2020; Brander and Koetse 2011; Kovacs et al. 2022; Perino et al. 2014 for systematic reviews). Many previous hedonic studies from Beijing involved green amenities, either as the primary focus or as one of the locational attributes in the hedonic pricing model (e.g., Dong and Wu 2016; Mei et al. 2019, 2021; Wu et al. 2022; Zhang et al. 2020; Zheng et al. 2016; Zheng and Kahn 2008). However, it tends to be more straightforward to value the proximity to the nearest green space. Proximity can be measured, for example, as the length of the shortest route, as described in the studies mentioned above. Moreover, recent developments in applying Geographical Information Systems (GIS) and machine learning techniques to environmental valuation have facilitated hedonic valuations of the views of green amenities that can be seen from a property (e.g., Black 2018; Cavailhès et al. 2009; Daams et al. 2016; Walls et al. 2015; Wu et al. 2022).

In valuing the area (or size) of green space, there is an additional technical consideration as to which green spaces should enter the hedonic price model. The value of a property is less likely to be affected by a sufficiently distant green space, according to Tobler's first law of geography that 'near things are more related than distant things' (Tobler 1970) and distance decay in spatial interactions (Taylor 1983). This implies that the value of a property might be affected only by green spaces within a 'threshold' or maximum distance, and that only these green spaces should enter the hedonic price model. In that regard, a common strategy has been to adopt a predetermined spatial bound, such as a predetermined radius (e.g., Albouy et al. 2020; Czembrowski and Kronenberg 2016; Netusil et al. 2010; Waltert and Schläpfer 2010) or a census block (e.g., Cho et al. 2008; Netusil et al. 2014). Some other studies (e.g., Conway et al. 2010; Nafilyan and Lorenzi 2019; Sander et al. 2010; Stromberg et al. 2021) undertook what we call a spatial piecewise step function analysis. They delineated a series of concentric circles surrounding each property and measured the area of green space within each doughnut-shaped ring. All these green space area variables

entered the hedonic pricing model. It was assumed that the estimated hedonic prices of green space beyond a certain threshold distance would become statistically equal to zero, and this threshold distance would be regarded as the spatial bound of the hedonic valuation.

This study builds upon the standard spatial piecewise step function approach by exploring three novel empirical strategies that improve the identification of the threshold distance. These strategies include a model-selection-based regression spline method, a spatial piecewise matching analysis, and a spatial difference-in-differences (DID) approach. The three strategies offer different strengths and thus reinforce each other in the identification of the threshold distance. To the best of our knowledge, this is the first study that implements and compares the three empirical strategies in the hedonic pricing of urban green amenities.

We start with a model-selection-based regression spline approach, assuming that the hedonic prices of green spaces at different distances within the threshold distance can be estimated as a polynomial function of distance, whereas the hedonic prices of green spaces beyond the threshold distance are statistically equal to zero and no longer vary along distance. We next loop over all possible threshold distances in search of the preferred threshold distance that has the best predictive performance within-sample and/or out-of-sample. This model selection approach echoes that proposed by Fitzpatrick and Parmeter (2021), who focused on the effects of coal mines on housing prices. This approach objectively selects the exact location of the threshold distance that maximises the predictive performance of the hedonic pricing model. This provides a useful twist to the standard step function approach, which often relies on visual assessment and/or on arbitrary cut-off levels of statistical significance.

Next, to enhance the causal strength of the standard step function approach, we use a novel matching algorithm proposed by Fong et al. (2018) for a continuous treatment variable (e.g., the area of green spaces). In comparison, the standard step function analysis of the hedonic price of urban green amenities typically relies on the OLS estimator, which is more prone to endogeneity bias associated with factors that correlate with both housing prices and urban green spaces but are not adequately controlled. This speaks to an increasing emphasis in the hedonic pricing literature on adopting quasi-experimental methods (i.e., matching, difference-in-differences, instrumental variables, and regression discontinuity designs) to mitigate endogeneity bias (Bishop et al. 2020; Kuminoff et al. 2010). Under the spatial piecewise framework, we used the matching algorithm of Fong et al. (2018) to re-estimate the hedonic prices of green spaces at different distances to further assess the robustness of the findings in our regression spline analysis.

Lastly, we adopt a spatial difference-in-differences (DID) approach to further account for endogeneity bias associated with unobserved factors. This is achieved by assessing whether the creation of a new urban green space affects housing prices within a certain radius, taking housing prices outside the radius as the control, as in Haninger et al. (2017), Mei et al. (2021), Muehlenbachs et al. (2015) and Tanaka and Zabel (2018).¹ We repeat the analysis by redefining the 'treated' observations using a series of different radiuses to identify the threshold distance where the treatment effect disappears. This spatial DID approach better accounts for endogeneity bias associated with unobserved factors, but requires

¹ These studies focused on environmental amenities/disamenities other than green space, namely restored brownfields (Haninger et al. 2017), power plants that switched fuels from coal to gas (Mei et al. 2021), shale gas wells (Muehlenbachs et al. 2015), and nuclear power plants (Tanaka and Zabel 2018).

housing price observations before and after the creation of urban green spaces, and thus focuses on only a subset of the green spaces in our dataset to satisfy that requirement.

The remainder of this paper is structured as follows. Section 2 describes the study site, data and variables. Section 3 performs the spatial piecewise analysis of the hedonic prices of green spaces, focusing on the regression spline, matching and spatial DID approaches. Section 4 derives the aggregate hedonic value of green spaces, which can be used for ecosystem services accounting and cost-benefit analysis for urban land use decision-making. Section 5 discusses the results and concludes.

2 Study Area, Data and Variables

The hedonic pricing analysis in this study is applied to data collected from Beijing. The city's real estate sector experienced China's housing reform programme in the late 1990s and became by and large a market-oriented system, where privately-developed new homes generally can be traded freely (Ren and Folmer 2022; Zheng and Kahn 2008). In the first two decades of this century, the city's population increased dramatically, from slightly over 13 million to nearly 22 million (Beijing Municipal Bureau of Statistics 2023), which has led to spiralling housing prices. According to the China Index Academy, the average price of newly built housing in 2023 is around CNY 60,000/m². This implies that a homebuyer has to shell out nearly half a million US dollars for a typical 50-m² apartment.

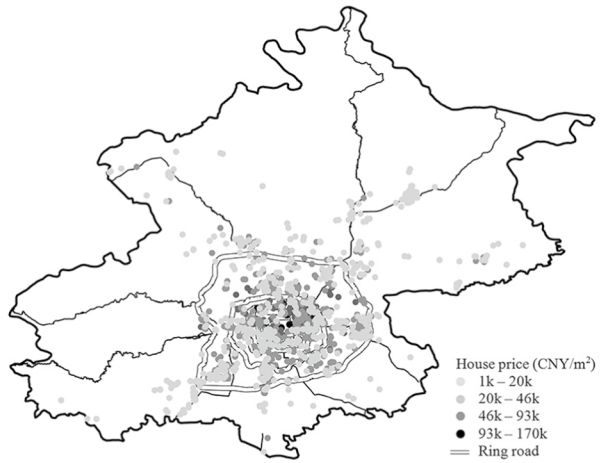
Nevertheless, urban green amenities are extensive in this densely populated city. Beijing's dry climate and inland location have left green spaces as one of the few types of environmental amenities available to its residents. By definition, the scope of urban green amenities is confined to those in Beijing's urban areas, and hence does not include vegetated land in the municipality's outlying rural areas, such as forests and national parks. In 2016, the city's per capita area of urban green space came to 40 m², which was higher than the per capita area of housing (32 m²). The seemingly unbalanced trade-off warrants reassessment on whether it is worthwhile to spare such extensive area of land for green space.

In this context, we undertake a hedonic analysis to investigate the housing price premiums attributable to green space. Our analysis relies on a geographically referenced census dataset that details the first-time transactions of all newly built properties in Beijing from 2006 to 2016. All geographic data were mapped and compiled using ArcGIS. We consulted previous hedonic studies, particularly those from Beijing (i.e., Dong and Wu 2016; Li et al. 2016; Wu and Dong 2014; Zheng et al. 2016; Zheng and Kahn 2008), to guide our data collection and measurement of variables. Table A1 in the appendix defines these variables and presents the descriptive statistics.

Our housing transaction data were obtained from the China Index Academy and were sourced from the housing transaction registration system of the municipal government, which recorded the first-time transactions of all 1270 newly built residential blocks from 2006 to 2016 (see Fig. 1a).² We were only able to obtain the longitude and latitude coordinates of the centroids of these residential blocks. These coordinates were used to map the housing data to green spaces and other urban amenities and infrastructure. Therefore, the green space and other locational variables have the same values for all properties within

² Most new homes in Beijing are units within multi-family residential blocks, which go on the market at about the same time. We excluded data points for publicly subsidised affordable homes.

Fig. 1 Housing prices and urban green space in Beijing



(a) Housing Prices



(b) Green Spaces

each residential block. In light of this, we measured all variables at the residential block level.

Data on urban green amenities were collected by the Beijing Municipal Bureau of Landscape and Forestry through field surveys in 2014 (see Fig. 1b). This dataset is a full inventory of more than 230,000 plots of green space in Beijing’s urban areas. Surveyors digitised the boundaries of these green spaces using GPS trackers, and investigated a number of other attributes, such as the area of a green space and the time when it was created. This enabled us to map property transactions to the green spaces that existed at the time of sale. As illustrated in Figure A1 in the appendix, we calculated the area of green space in each

ring³ around the centroid of a residential block, giving rise to a series of variables indicating the area of green space at different distances.⁴

Further, we constructed a wide range of control variables for other locational characteristics that may affect property values. We calculated the distance from each residential block to Tiananmen Square to indicate a residential block's location relative to the city centre. The dummy variable '*Southern half of Beijing*' takes the value one for those residential blocks located to the south of Tiananmen Square, and zero otherwise. This dummy variable is expected to capture a downward shift in property values, since the southern half of Beijing was historically occupied by economically disadvantaged groups. We digitised the paper-based Beijing Education Map (Beijing Municipal Education Commission 2015) into an ArcGIS data file, which specifies the school district that covers each residential block and hence the schools assigned to it. School quality is proxied by the number of 'demonstration' ('*shifan*') schools/kindergartens in each school district. The title 'demonstration' is usually awarded to the highest ranked and most reputable schools in Beijing. We extracted geographic data on other urban infrastructure and services (including hospitals, railways, highways, roads, subway stations, bus stops, restaurants and shops) from Gaode Maps, a leading web map service in China, and measured their proximity and quantity variables. These services and infrastructure benefit local residents, but may also induce adverse effects such as noise and crowding. Lastly, we generated a set of dummy variables that distinguish the districts, ring roads (representing zones partitioned by the ring roads), and years. The district and ring road dummies control for the price effects of features specific to a district or ring road that do not vary over time. The year fixed effects capture year-specific macro shocks that have a common effect on property values, such as changes in housing and mortgage policies.

3 Spatial Piecewise Analysis of Hedonic Prices for Green Space

As a reference point, we first performed the standard spatial piecewise step function analysis, where housing prices are regressed against a sequence of green space variables that measure the total area of green space in different rings at different distances of each residential block (as described in Section 2), controlling for other property- and location-specific variables listed in Table A1.⁵ The estimated hedonic prices at different distances are graphically reported in Figure A2. However, there appears to be a lack of clear-cut indication of a breakpoint distance where green spaces cease to affect housing prices, and the patterns are sensitive to the choice of the step length (the width of the rings in Figure A1). These difficulties motivate us to explore three novel empirical strategies that facilitate the identification of the breakpoint distance. These strategies are the model-selection-based regression spline, spatial piecewise matching, and spatial difference-in-differences approaches.

³ The ring at the centre is a solid circle.

⁴ We measured 100 variables that indicate the total area of green space in each 100 m-wide ring in a 10 km radius of each residential block. For brevity, Table A1 only describes green space variables for each 1 km-wide ring in a 10 km radius.

⁵ The hedonic pricing model is $\log(\text{HousePrice}_i) = \sum_j p_j \text{Green}_{ij} + \beta \mathbf{x}_i + \epsilon_i$, where i indexes residential blocks and j refers to rings at different distances. p_j is the hedonic price to be estimated. The vector \mathbf{x}_i consists of all other explanatory variables listed in Table A1.

3.1 Regression Spline Analysis

Our regression spline approach was adapted from a standard restricted cubic spline regression (as formally discussed in Orsini and Greenland 2011; Smith 1979; Wegman and Wright 1983). This approach is used in high-quality applied statistical research, notably in medicine (e.g., Austin et al. 2022; Desquilbet and Mariotti 2010; Keogh and Morris 2018). The hedonic pricing model is re-parametrised as Eq. (1). We make two assumptions. First, only green spaces within a threshold distance D_K (the last knot of the spline) affect housing prices, and the hedonic price of these green spaces can be expressed as a polynomial function of distance. Second, outside the threshold distance (or after the last knot), the hedonic price of green spaces no longer varies over distance and is expected to be statistically insignificant.

$$\log(\text{HousePrice}_i) = \alpha \sum_j \text{Green}_{ij} + \sum_{k=1}^{k=K} \gamma_k \sum_j (D_k - d_j)_+^n \text{Green}_{ij} + \beta \mathbf{x}_i + \epsilon_i. \quad (1)$$

In this hedonic pricing model, the subscript i indexes residential blocks, j denotes rings at different distances from each residential block, and k indicates the knots of the regression spline. The vector \mathbf{x}_i consists of all other explanatory variables in Table A1. α , γ_k and β are the parameters to be estimated, d_j represents the radius of the middle of the j th ring, and D_k denotes the location of the k th knot. The positive part function $(D_k - d_j)_+$ truncates $(D_k - d_j)$ at zero. The hedonic price estimates can be recovered through evaluating the function $p_j = \alpha + \sum_k \gamma_k \sum_j (D_k - d_j)_+^n$ at different distances, and their standard errors can be estimated using the delta method described in Greene (2020).⁶

The positions of the knots (and hence the threshold distance) are objectively decided via a model selection procedure. The step function estimates (in the finest 100 m step length setting) suggest that the hedonic price curve along distance is likely to have two conspicuous turning points (around 0.5 km and 1 km, respectively). We therefore allow for a maximum of two knots ($K = 0, 1$ or 2) to accommodate the two turning points. We specified the degree of the spline function to be three ($n = 3$), which is the most commonly adopted choice in the regression spline literature (Orsini and Greenland 2011; Wegman and Wright 1983), to ensure that the estimated hedonic price curve and its first- and second-order derivatives are continuous at the knots (so that the curve is visually smooth). The hedonic

⁶ Suppose we confine the analysis to a 300 m radius using a 100 m step length, and thus have three green space variables for each home i : Green_{i1} for green space in the 0–100 m ring, with $d_1 = 50$ (the radius of the middle of the ring), Green_{i2} for green space at 100–200 m, with $d_2 = 150$, and Green_{i3} for green space at 200–300 m, with $d_3 = 250$. Suppose we have a total of two knots, the first at 100 m and the second at 200 m ($K = 2, D_1 = 100, D_2 = 200$). In that case, the green space variable associated with the parameter α is $(\text{Green}_{i1} + \text{Green}_{i2} + \text{Green}_{i3})$, as in the first term of Eq. (1). The green space variable for γ_1 is 50^3Green_{i1} , and that for γ_2 is $(150^3 \text{Green}_{i1} + 50^3 \text{Green}_{i2})$, as in the second term of Eq. (1). The three parameters α , γ_1 and γ_2 can be identified in the regression because the three green space variables have different variation across homes. The three parameters can be loosely regarded as three components of the hedonic price estimates: for green space at 0–100 m, the hedonic price is $(\alpha + 50^3 \gamma_1 + 150^3 \gamma_2)$; for green space at 100–200 m, the hedonic price is $(\alpha + 50^3 \gamma_2)$; for green space at 200–300 m, the hedonic price is α .

pricing model (Eq. (1)) is then estimated repeatedly using all possible combinations of the knots' locations within a 5 km radius.⁷ to search for the model specification (defined by the knots' locations) that fits the data best.

We adopted two different procedures for model assessment and selection. One is the conventional approach based on within-sample prediction and the Akaike Information Criterion (AIC).⁸ The other is influenced by the rapid proliferation of statistical learning, which suggests that models with better 'out-of-sample' explanatory power (as opposed to the conventional within-sample explanatory power) are increasingly preferred (Varian, 2014). In addition, regression spline models are often ill-adapted to extrapolation beyond the data used to fit these models (Suits et al. 1978). It is thus particularly helpful to assess the out-of-sample predictive performance of our regression spline models. In light of that, we conducted a five-fold cross-validation analysis to select a model with the best out-of-sample prediction accuracy, as per Hastie et al. (2008), James et al. (2017), Jardine and Siikamäki (2014), and Jaya and Folmer (2020).

Our cross-validation analysis consists of the following steps: (1) specify a candidate functional form for Eq. (1) defined by the locations of the two knots D_1 and D_2 . (2) Randomly split our full sample into five equal-sized sub-samples. (3) For the model specified in Step (1), remove a randomly selected sub-sample (test data, indexed by m , with a total number of observations M), and use the remaining four sub-samples together (training data, indexed by $-m$) to estimate the model using OLS. (4) Use the model derived in Step (3) and the explanatory variables of the test data m to predict the dependent variable $\log(\widehat{HousePrice}_m)$. (5) For the test data, use the predicted values $\log(\widehat{HousePrice}_m)$ and the actual values $\log(HousePrice_m)$ to derive three measures of fit, including the Mean Squared Error (MSE):

$$MSE = \frac{1}{M} \sum_{m=1}^{m=M} [\log(HousePrice_m) - \log(\widehat{HousePrice}_m)]^2, \quad (2)$$

the Mean Absolute Error (MAE):

$$MAE = \frac{1}{M} \sum_{m=1}^{m=M} |\log(HousePrice_m) - \log(\widehat{HousePrice}_m)|, \quad (3)$$

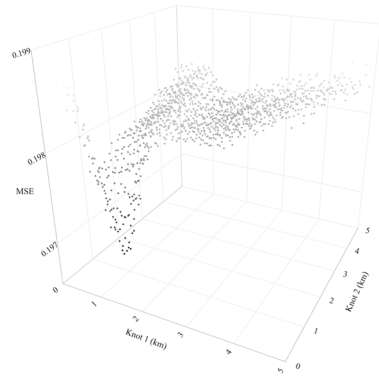
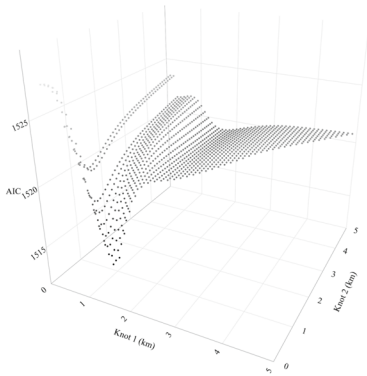
and the the pseudo- R^2 :

⁷ We focus on green space within a 5 km radius because our step function estimates and preliminary regression spline analysis suggest that green space outside 5 km is less likely to affect housing prices. In the 100 m step length setting, Eq. (1) has one possible specification with no knots (where the second term of Eq. (1) is 0), 50 possible specifications with one knot ($D_1 = 100$ m, 200 m, 300 m, ..., or 5 km), and 1,225 possible specifications with two knots (where D_1 and D_2 are combinations of the 50 possible positions ranging from 100 to 5 km, and the total number of such combinations is $C_2^{50} = 1225$). Therefore Eq. (1) has a total of 1276 possible specifications. This total number differs in other step length settings.

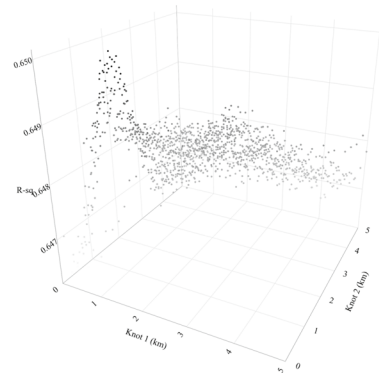
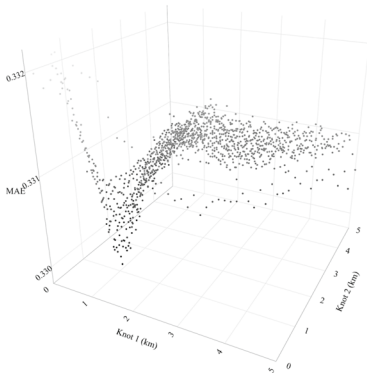
⁸ We have opted for the AIC for within-sample model assessment, instead of the three model performance measures described below for the cross-validation analysis (the mean squared error, the mean absolute error, and the pseudo- R^2). Adding regressors (through, for example, increasing the number of knots) always improves those three measures in within-sample prediction. Therefore, attempting to select a model specification using those three measures in within-sample prediction would lead to overfitting.

Table 1 Knots preferred by different model performance criteria (100 m step length)

Knots	AIC	MSE	MAE	R ²
D ₁ (km)	0.9	0.7	1.0	0.9
D ₂ (km)	1.0	1.2	1.1	1.0



(a) AIC (Within-Sample), 100m Step Length (b) MSE (Out-of-Sample), 100m Step Length



(c) MAE (Out-of-Sample), 100m Step Length (d) R² (Out-of-Sample), 100m Step Length

Fig. 2 Predictive performance of regression spline models (100 m step length)

$$R^2 = \frac{\sum_{m=1}^{m=M} [\log(\widehat{HousePrice}_m) - \log(\overline{HousePrice}_m)]^2}{\sum_{m=1}^{m=M} [\log(HousePrice_m) - \log(\overline{HousePrice}_m)]^2}, \tag{4}$$

where $\log(\overline{HousePrice}_m)$ refers to the mean of the dependent variable for the test data. (6) Repeat Steps (3)–(5) for each of the five sub-samples as the test data. (7) Repeat Steps (2)–(6) 500 times and compute the means of the MSE, the MAE, and the pseudo-R². (8) Repeat Steps (1)–(7) for all possible specifications of Eq. (1).

Fig. 3 Hedonic price estimates at different distances: regression spline estimates (100 m step length). Note: this figure focuses on the segment of the hedonic price curve within 2 km, since the curve no longer changes along distances outside 1 km

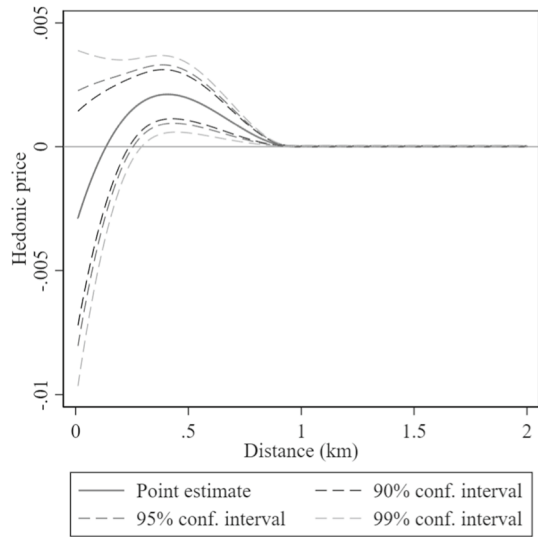


Table 2 Regression spline estimates

DV: log (housing price)	Model 1
<i>Hedonic price</i>	
Green area 0–1 km	$1.03 \times 10^{-3}***$ (2.93×10^{-4})
<i>Regression spline parameters</i>	
$\hat{\gamma}_1$	$-8.51 \times 10^{-11}***$ (2.71×10^{-11})
$\hat{\gamma}_2$	$5.90 \times 10^{-11}***$ (1.79×10^{-11})
$\hat{\alpha}$	1.31×10^{-5} (9.97×10^{-6})
Housing controls (Table A1)	Yes
Obs	1270

Standard errors are in parentheses

Asterisks indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Fig. 2 displays the within- and out-of-sample predictive performance of all possible model specifications in the 100 m step length setting. In each panel of Fig. 2, the horizontal plane represents the full set of all possible combinations of the knots' locations, and the vertical axis denotes the model performance measure (AIC, MSE, MAE, or pseudo- R^2). The preferred model specification has the lowest AIC/MSE/MAE, or the highest pseudo- R^2 . Table 1 summarises the model specifications (or the locations of the knots) preferred by different model performance measures. Reassuringly, it can be seen that all the model performance measures suggest highly comparable (and unambiguous)

breakpoint distances where the hedonic price of green space disappears (D_2 , which ranges from 1.0 to 1.2 km). We have opted for 1.0 km as the breakpoint for the subsequent analyses, for two reasons. First, this is preferred by two model selection measures: one within-sample measure (AIC) and one out-of-sample measure (pseudo- R^2). Second, this is the most conservative (closest) breakpoint across all our selection criteria, which may reduce the risk of overestimating the hedonic value of green space.

As can be seen in Model 1 in Table 2, the estimated parameters ($\hat{\gamma}_1$ and $\hat{\gamma}_2$) that capture the distance-dependent patterns of the hedonic price estimates between the two knots are both strongly significant (p value < 0.01). The hedonic price estimates at different distances (and their confidence intervals) can be recovered from the regression spline, as shown in Fig. 3. The negative sign on $\hat{\gamma}_1$ and the positive sign on $\hat{\gamma}_2$ jointly give the inverted U-shaped curve in Fig. 3 of the hedonic prices inside the second knot (1 km as discussed above). The hedonic price first increases with distance, peaks at 500 m, and then declines with distance until about 1 km.⁹ The three estimates, $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\alpha}$, jointly formulate the hedonic price estimates, as mentioned in Footnote 6; the hedonic prices between the first and the second knots are $[\hat{\alpha} + (D_2 - d_j)^3 \hat{\gamma}_2]$, and the first-order derivative with respect to d_j is $[-3(D_2 - d_j)^2 \hat{\gamma}_2]$. Therefore, the positive sign on $\hat{\gamma}_2$ implies a negative first-order derivative and hence decreasing hedonic prices over d_j between the first and the second knots. Similarly, the hedonic prices inside the first knot are $[\hat{\alpha} + (D_1 - d_j)^3 \hat{\gamma}_1 + (D_2 - d_j)^3 \hat{\gamma}_2]$, and the first-order derivative with respect to d_j is $[-3(D_1 - d_j)^2 \hat{\gamma}_1 - 3(D_2 - d_j)^2 \hat{\gamma}_2]$. An increasing trend in hedonic prices inside the first knot requires $\hat{\gamma}_1$ to be negative if $\hat{\gamma}_2$ is positive. However, in that case, the first-order derivative may not be monotonically positive and may become negative as d_j increases. This would imply an initial increase and then a decrease in hedonic prices inside the first knot.

In monetary terms, the highest estimate appears in the 400–500 m ring (CNY 54.23 or USD 8.17 per m² of housing per ha of green space).¹⁰ The lowest estimate occurs in the 0 m–100 m ring (CNY –43.54 or USD –6.56). The area-weighted average hedonic price within 1 km (1.03×10^{-3})¹¹ translates into CNY 26.98 or USD 4.06, and is highly significant (p value < 0.01; standard error estimated using the delta method). For green space outside the threshold distance, the hedonic price estimate ($\hat{\alpha}$) is statistically insignificant (p value = 0.19) and small (less than 2% of the area-weighted average hedonic price estimate within 1 km). These estimates characterise the hedonic price curve presented in Fig. 3,

⁹ In the context of Beijing, green space provides both amenities (such as views and recreational benefits) and disamenities (such as attracting mosquitoes and green space users, which lead to noise and privacy concerns) (Liu et al. 2020). Both the positive and negative effects are likely to decline over distance. However, the negative effects are likely to decline more quickly over distance than the positive effects, within a short distance to residents' homes (e.g. within 500 m, as we found in Fig. 3). The net effect could be an increase in the hedonic price of green space. Over a longer distance (e.g., beyond 500 m), it is possible that the negative effects become negligible and the spatial pattern of hedonic prices is dominated by the distance decay in the positive effects.

¹⁰ The original log-linear estimate (\hat{p}_j) indicates that the average predicted home price $E(\text{HousePrice}_i)$ would change by $100[\exp(\hat{p}_j \Delta G)]$ per cent if the area of green space changes by ΔG (Wooldridge 2020).

¹¹ We first used the estimated parameters $\hat{\alpha}$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ to derive the hedonic price of each 100 m ring within 1 km: $\hat{p}_j = \hat{\alpha} + \sum_{k=1}^{K=2} \hat{\gamma}_k \sum_j (D_k - d_j)_+^3$, and then computed the area-weighted average of those hedonic price estimates as $\sum_{j=1}^{10} w_j \hat{p}_j$, where w_j refers to the ratio of the mean area of green space in the j th ring to the mean total area of green space inside 1 km.

which has an inverted U-shape within 1 km and then becomes almost indistinguishable from the horizontal axis (though still marginally above zero). This hedonic price curve closely resembles that derived from the step function approach using a 100 m step length (Figure A2a).

Our findings on the preferred model specification and hedonic price estimates are stable if we switch to a 200 m step length.¹² In fact, according to each of the four model selection criteria, all possible model specifications in the 200 m, 500 m and 1 km step length settings are outperformed by the preferred specification in the 100 m step length setting. This has inclined us to focus on the findings from the 100 m step length setting.

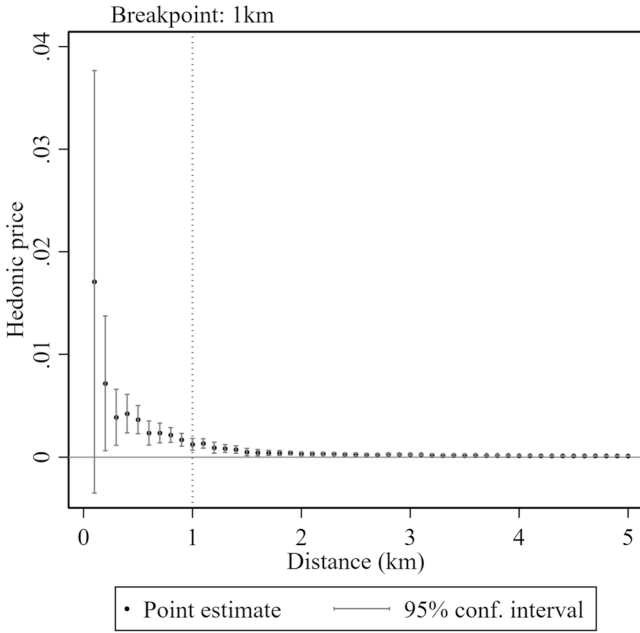
3.2 Matching Analysis

We next switch to a novel matching approach proposed by Fong et al. (2018) to further test the robustness of our findings. Matching has been advocated by the causal econometric literature (e.g., Greenstone and Gayer 2009; Imbens and Rubin 2015, and Imbens and Wooldridge 2009) as a means to better control for endogeneity issues (especially those associated with observed factors), and thereby strengthen an estimate's causal inference. In particular, Fitzpatrick and Parmeter (2021) performed matching estimation to provide stronger causal evidence in their study on the effects of coal mines on housing prices at different distances. However, the matching algorithm of Fitzpatrick and Parmeter (2021) was intended for a binary variable: whether there was a coal mine within a certain distance of a house. By contrast, our study focuses on a continuous variable: the area of green space within a certain distance of a home. Fong et al. (2018) built upon conventional matching methods by accommodating nonbinary treatment variables and by directly optimising sample covariate balance by minimising the correlation between covariates and the treatment. Fong et al. (2018) named this novel matching algorithm the 'covariate balancing generalised propensity score' (CBGPS), where the generalised propensity score refers to the distribution of the treatment conditional on the covariates.

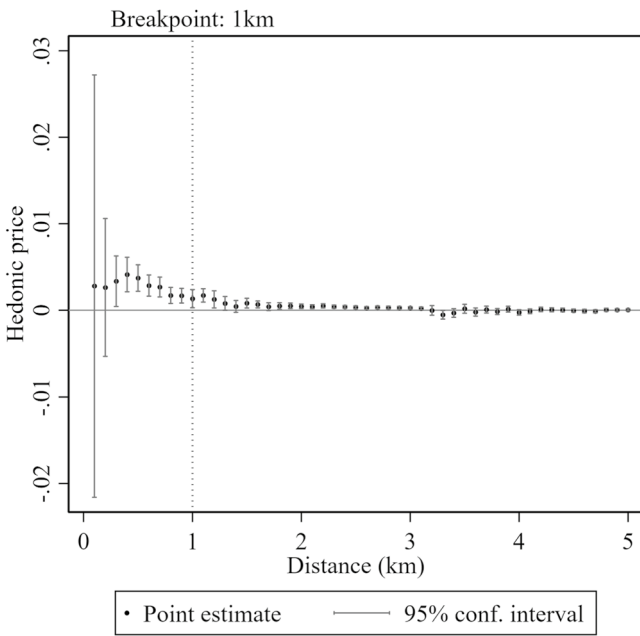
Following the empirical strategy of Fitzpatrick and Parmeter (2021), we used a CBGPS algorithm to re-estimate the hedonic prices of green spaces at different distances, to further assess the robustness of the findings in our regression spline models. We separately used both parametric and non-parametric CBGPS methods. The parametric method assumes that the generalised propensity score is normally distributed. In contrast, the non-parametric method does not depend on any assumptions about the functional form of the generalised propensity score. Our matching analysis was conducted with the following steps: (1) following Fong et al. (2018), we first identified the optimal Box-Cox transformation for the area of green space within a certain distance by searching for the exponent parameter (from the range -2 to 2 with a 0.01 step length) that gives the best approximation of the standard normal distribution.¹³ (2) We then performed the parametric matching algorithm on the covariates listed in Table A2. (3) The weights derived from the matching algorithm were then utilised to estimate a regression

¹² For the 200 m step length, Figure A3 in the appendix presents the predictive performance of all possible specifications of the regression spline model, and Figure A4 in the appendix presents the regression spline estimates of the hedonic prices at different distances.

¹³ We adopted this transformation for both the parametric and non-parametric CBGPS estimation to ensure the estimates' comparability, even though the non-parametric CBGPS does not involve any distributional assumptions for the generalised propensity score.



(a) Parametric CBGPS, 100m Step Length



(b) Non-Parametric CBGPS, 100m Step Length

Fig. 4 Hedonic price estimates at different distances: matching estimates

Table 3 Matching estimates

DV: log (housing price)	Model 2 Parametric CBGPS	Model 3 Non-parametric CBGPS
<i>Hedonic price</i>		
Green area 0–1 km	$1.26 \times 10^{-3***}$ (2.98×10^{-4})	$1.34 \times 10^{-3***}$ (5.37×10^{-4})
Obs	1270	1270

Standard errors are in parentheses

Asterisks indicate statistical significance: * p value < 0.10, ** p value < 0.05, *** p value < 0.01

of housing prices on the transformed green space variable and the three sets of fixed effects listed in Table A1 (to approximate within-cluster matching). (4) Due to the transformation in Step (1), the estimate from Step (3) on the transformed green space variable had to be converted (back) to the semi-elasticity form of the hedonic price estimate, in order to be directly comparable with the estimates from the regression spline models. (5) Steps (1)–(4) were bootstrapped 500 times to derive the standard error and confidence intervals of the hedonic price estimate. (6) Steps (1)–(5) were repeated for 50 green space variables (one at a time), which represent the area of green space within different distances ranging from 100 m to 5 km, with a 100 m step length.

Figure 4 presents the matching estimates of the hedonic prices of green spaces within different distances. Compared to the regression spline estimates (3), the non-parametric CBGPS estimates (Fig. 4b) exhibit a highly similar pattern of hedonic prices over distance. The hedonic prices start with positive but low levels at 100 m and 200 m; slightly increase to the peak level at 400 m; steadily decline along distance; and, after about 1.5 km, converge to very small swings around zero. The parametric CBGPS estimates (Fig. 4a) are notably higher for green spaces within 100 m and 200 m. This is likely because 69% of our observations have no green space within 100 m, and 49% have no green space within 200 m; this might have caused anomalies when the parametric CBGPS algorithm forced the estimated generalised propensity score to be normally distributed. Beyond 200 m, however, the parametric CBGPS estimates closely resemble the pattern of the non-parametric CBGPS estimates. For the 1–1.5 km interval, both the parametric and non-parametric CBGPS estimates are higher than the regression spline estimates. This suggests that the 1 km breakpoint previously identified might be a lower bound of the threshold distance at which the hedonic price starts to disappear.

Finally, we discuss in more detail the matching estimates for green spaces within 1 km. Table A2 in the appendix presents the results of the covariate balance tests. The first column of Pearson correlation coefficients shows considerable pre-matching correlation between each covariate and the (transformed) green space variable. The absolute value of the correlation coefficient is above 0.15 for 20 out of a total of 22 covariates, and above 0.30 for 13 covariates. Such correlation was substantially reduced by the parametric and non-parametric CBGPS matching procedures. None of the covariates has a post-matching correlation coefficient above 0.15 in absolute value, while 19 covariates have a coefficient below 0.10 after the parametric matching, and 21 are below 0.10 after non-parametric matching. This improvement of covariate balance reduces concerns about potential endogeneity issues, since the green space variable has become notably less correlated with the observed covariates in the post-matching sample.

Models 2 and 3 in Table 3 report the hedonic price estimates derived from the parametric and non-parametric matching algorithms. It can be seen that the two matching estimates are qualitatively comparable to the regression spline estimate (in Model 1, Table 2), although 22–30% higher in magnitude.

3.3 Difference-in-Differences Analysis

Our regression spline and matching analyses controlled for a wide range of factors that may correlate with both housing prices and green space. However, it is difficult in those two approaches to explicitly account for all such factors, especially unobserved factors, which might bias or confound the hedonic price estimates. We seek to better address unobserved confounders using a spatial difference-in-differences (DID) approach that better controls for location-specific unobserved confounders at a more detailed level.

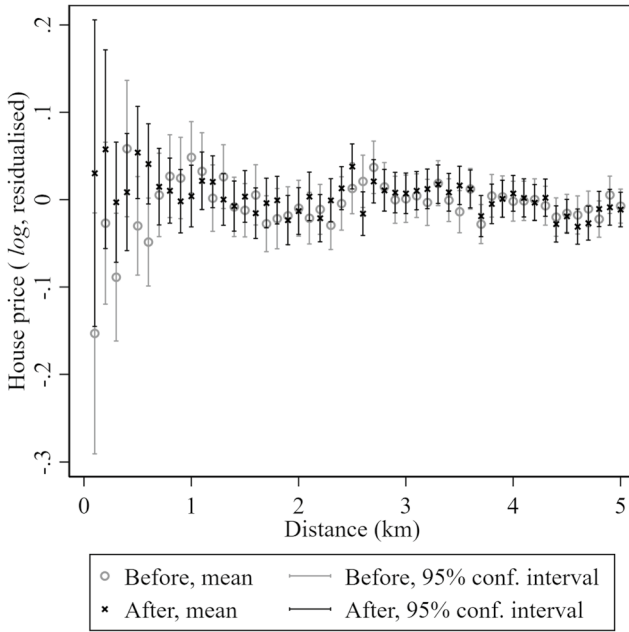
Our DID analysis switches the unit of analysis from residential blocks to locations where new green spaces (larger than 0.5ha each¹⁴) were created in the timespan of our housing dataset (2006–2016). For each of these locations, we searched in a 5 km radius for residential blocks that were sold within two years before or after the new green space was created, and calculated the exact distances. This gave us housing price observations at different distances and different points in time before and after the creation of the green space, as shown in Fig. 5. Housing prices were residualised by regressing out location and time fixed effects and all the other control variables listed in Table A1, which accounted for all location- and time-specific confounders, both observed and unobserved. Figure 5a presents the means of residualised housing prices at different distances, using a 100 m step length. Figure 5b models residualised housing prices as a local linear polynomial function of distance. There is a reasonably discernible pattern that, within a 1 km radius, housing prices became higher after the creation of green spaces. Outside 2 km, housing prices before and after the creation of green spaces converge to similar levels.

We next formally investigate these visual patterns using DID regressions as specified in Eq. (5):

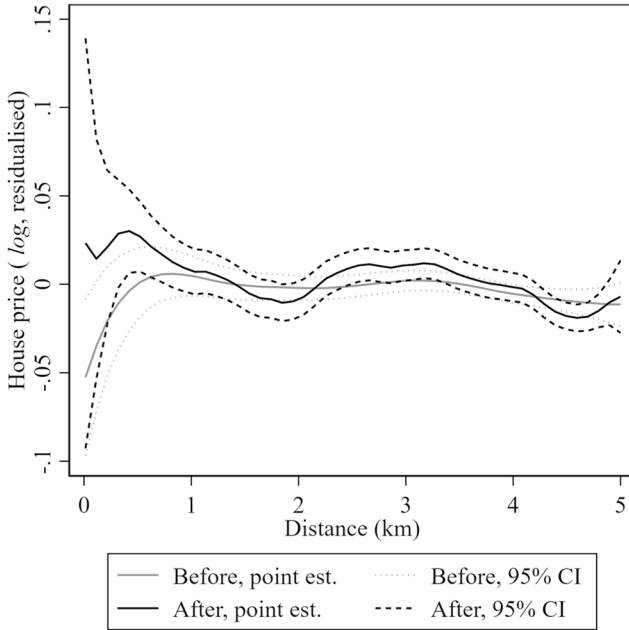
$$\log(\text{HousePrice}_{si}) = \lambda \text{After}_{si} + \tau \text{Radius}_{si} + \delta \text{After}_{si} \times \text{Radius}_{si} + \theta_s + \theta_t + \beta \mathbf{x}_i + \mu_{si}, \quad (5)$$

where the binary variable ‘ After_{si} ’ equals one if housing price i is observed after the creation of a new green space in location s , and zero otherwise, and the binary variable ‘ Radius_{si} ’ indicates whether housing price i is observed within a certain radius (e.g., 1 km) of location s . δ is the DID estimator that captures the ‘treatment effect’ of creating a new green space on housing prices within a certain radius. The vectors θ_s and θ_t represent location and time fixed effects, and \mathbf{x}_i consists of all the other control variables listed in Table A1. This model specification utilises housing prices within a certain radius of location s as treated observations, and those outside as control observations, as in Haninger et al. (2017), Mei et al. (2021), Muehlenbachs et al. (2015) and Tanaka and Zabel (2018). However, those studies exploited housing resale data which allowed for DID analyses at the property level, whereas our analysis is at the location level and relies on one-off housing prices observed in different years to introduce time-wise variation.

¹⁴ We repeated the DID analysis using newly created green spaces that are larger than 1ha, and the findings are qualitatively stable.



(a) Means at Different Distances, 100m Step Length



(b) Local Linear Regression against Distance

Fig. 5 Housing prices before and after the creation of green spaces

Table 4 Difference-in-differences estimates

DV: log (housing price)	Model 4 Radius = 1 km	Model 5 Radius = 2 km	Model 6 Radius = 3 km	Model 7 Radius = 4 km
EV				
After	$-1.04 \times 10^{-2*}$ (5.35×10^{-3})	$-1.01 \times 10^{-2*}$ (5.47×10^{-3})	-9.09×10^{-3} (5.73×10^{-3})	-8.83×10^{-3} (6.52×10^{-3})
Radius	-4.13×10^{-3} (8.04×10^{-3})	-1.01×10^{-3} (4.45×10^{-3})	$1.11 \times 10^{-2***}$ (3.56×10^{-3})	$2.22 \times 10^{-2***}$ (3.64×10^{-3})
After \times Radius	$2.31 \times 10^{-2*}$ (1.23×10^{-2})	4.62×10^{-3} (6.73×10^{-3})	-6.58×10^{-4} (5.38×10^{-3})	-1.11×10^{-3} (5.57×10^{-3})
Green space location fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Housing controls (Table A1)	Yes	Yes	Yes	Yes
Model significance (p value)	< 0.01	< 0.01	< 0.01	< 0.01
R^2 (within)	0.40	0.40	0.40	0.40
Obs	124,137	124,137	124,137	124,137

Asterisks indicate statistical significance: * p value < 0.10, ** p value < 0.05, *** p value < 0.01. Standard errors are in parentheses

We estimated four DID regressions (Models 4–7, Table 4), which have the same specification as in Eq. (5) but differ in whether $Radius_{i,t}$ is defined as within 1, 2, 3 or 4 km. In Model 4 where the radius is 1 km, the positive and statistically significant coefficient on the interaction term ‘After \times Radius’ suggests that a new green space tends to increase housing prices within a 1 km radius. In our DID dataset, the average size of green spaces is 4.14ha. This suggests that, approximately speaking, a new hectare of green space increases housing prices in a 1 km radius by CNY 147.82 or USD 22.26 per m^2 of housing. This is notably higher than the counterpart estimates from the regression spline and matching analyses, although the DID estimate refers to only a subset of the green spaces in the regression spline and matching analyses. In Models 5–7 (where the radius is 2, 3 or 4 km), the coefficient on the interaction term ‘After \times Radius’ is statistically indistinguishable from zero (the lowest p value = 0.49), and the absolute value of the estimated effect is much smaller than that for a 1 km radius. These results provide further evidence that a new green space is more likely to be capitalised into housing prices within a 1 km radius.

Our DID approach has controlled for all location- and time-specific factors, in addition to the control variables listed in Table A1. If there are no other confounders, the mean difference in the (residualised) treated and control housing prices would not vary over time before the creation of the green spaces, known as the parallel trends assumption (Wooldridge 2020). Otherwise, the main causal strength of this DID analysis would be confined to accounting for location-specific time-invariants (both observed and unobserved) at a more detailed level (compared to the regression spline and matching analyses), at the cost of focusing on only a subset of the green spaces in our dataset. Figure A5 in the appendix provides some visual evidence for the applicability of the parallel trends assumption in our analysis. A Wald test further corroborates that the mean differences between the treated and control housing prices are statistically equal across the two pre-treatment years (p value = 0.64). These assessments find no evidence that the parallel trends assumption is violated.

Table 5 Instrumental variable and placebo estimates

DV: log (housing price)	Model 8	Model 9	Model 10	Model 11
	OLS	IV-2SLS	OLS	OLS
Hedonic price				
Green area 0–1 km	$5.90 \times 10^{-4**}$ (2.57×10^{-4})	$1.40 \times 10^{-3**}$ (6.42×10^{-4})		$1.05 \times 10^{-3***}$ (3.73×10^{-4})
Future green area 0–1 km			-3.92×10^{-4} (1.01×10^{-3})	
Housing controls (Table A1)	Yes	Yes	Yes	Yes
Excluded instruments				
Green area 2–3 km	No	Yes	No	No
Green area 3–4 km	No	Yes	No	No
Weak IV test				
Cragg–Donald Wald F stat (H_0 : Weak IV)		110.83		
R^2 (1 st stage)		0.50		
Over-identification test				
Sargan statistic: p value (H_0 : Valid IV)		0.19		
IV redundancy test				
LM test: p value (H_0 : Redundant IV)		< 0.01		
Endogeneity test				
Diff-in-Sargan–Hansen Statistic: p value (H_0 : Exogenous ‘Green area 0–1 km’)		0.17		
Hedonic price model sig.: p value	< 0.01	< 0.01	< 0.01	< 0.01
Hedonic price model R^2	0.68	0.68	0.65	0.66
Obs	1270	1270	889	889

Standard errors are in parentheses

Asterisks indicate statistical significance: * p value < 0.10, ** p value < 0.05, *** p value < 0.01

3.4 Further Robustness Tests

The DID analysis, despite focusing on a subset of the green spaces in our dataset, lends further support to our finding that housing prices are less likely to be affected by green spaces outside a 2 km radius. On the other hand, the spatial distribution of green space in neighbouring locations in Beijing tends to be positively correlated (Wang et al. 2023; Zhang et al. 2021). A location is likely to have more green space if its neighbouring locations have more green space. In fact, for the residential blocks in our dataset, the area of green space within 1 km has a correlation coefficient above 0.60 with the area of green space at 1–2 km, 2–3 km, and 3–4 km. In other words, green space outside a 2 km radius of a property is less likely to directly influence housing price, but tends to be highly correlated with green space within a 1 km radius. This suggests the possibility of instrumenting green space within a 1 km radius using green space outside a 2 km radius, as a means to further explore the implications of potential endogeneity issues. Bayer et al. (2009) adopted a similar ‘spatial lag’ type of instrument for air quality.

To facilitate the instrumental variable (IV) estimation, we first estimated an OLS hedonic pricing regression (Model 8, Table 5). The right-hand side has only one green

space variable, which represents the total area of all green space within a 1 km radius, controlling for other housing and location-specific variables listed in Table A1. The OLS estimate of the hedonic price (the estimate for ‘*Green area 0–1 km*’ in Model 8 in Table 5) is smaller than that derived from the regression spline approach (Model 1 in Table 2), but remains statistically significant at the 5% level. In Model 9, ‘*Green area 0–1 km*’ is instrumented using the area of green space at 2–3 km and 3–4 km. The *F* statistic from the weak IV test (110.83) is markedly greater than the conventional rule of thumb (10) and exceeds a more recently recommended threshold (104.7) (Lee et al. 2022), which provides evidence against the null hypothesis of weak identification. The *p* value from the over-identification test is well above the critical value (0.10), which further justifies the validity of the instruments. Although the *p* value from the endogeneity test does not reject the null hypothesis of no endogeneity bias at the conventional critical level (0.10), the IV estimate for ‘*Green area 0–1 km*’ is notably higher than the OLS and regression spline estimates. This suggests that the OLS and regression spline estimates are likely to be conservative estimates of the true hedonic price of green space.

Lastly, we undertook placebo tests to further assess whether the foregoing analyses have picked up some distance-dependent patterns other than the effects of green space on housing prices. For each residential block, we measured the total area of green space at different distances that was created after the residential block was sold. In theory, such ‘future’ green space is less likely to affect housing prices, compared to preexisting green space. We next re-estimated Model 8 after replacing the variable ‘*Green area 0–1 km*’ (green space inside 1 km that existed *before* housing transactions) with ‘*Future green area 0–1 km*’ (green space inside 1 km that was created *after* housing transactions), to test whether the model would falsely attribute any unexpected effect to future green space. As in Model 10 in Table 5, the estimate on ‘*Future green area 0–1 km*’ is statistically insignificant (*p* value = 0.70), which is in line with intuition. Note that this placebo analysis is confined to housing transactions that happened before 2014, because our green space data were collected in 2014, and thus do not include any future green space for housing transactions from 2014 onward. We thus re-estimated Model 8 using housing transactions before 2014 to test whether the estimates could be affected by dropping subsequent housing transactions. This gave rise to Model 11, where the estimate on ‘*Green area 0–1 km*’ remains positive and statistically significant, although larger in size compared to the estimate for the full sample. Moreover, we performed another placebo test by adding to Model 10 the area of future green space at 1–2 km, 2–3 km, 3–4 km and 4–5 km. As can be seen in Figure A6 in the appendix, all these placebo estimates are statistically insignificant; the lowest *p* value is above 0.35. In comparison, the non-placebo estimates in Figure A6 were derived from an expanded version of Model 11, which contains preexisting green space at different distances; here, the estimate on ‘*Green area 0–1 km*’ is almost identical to that in Model 11.

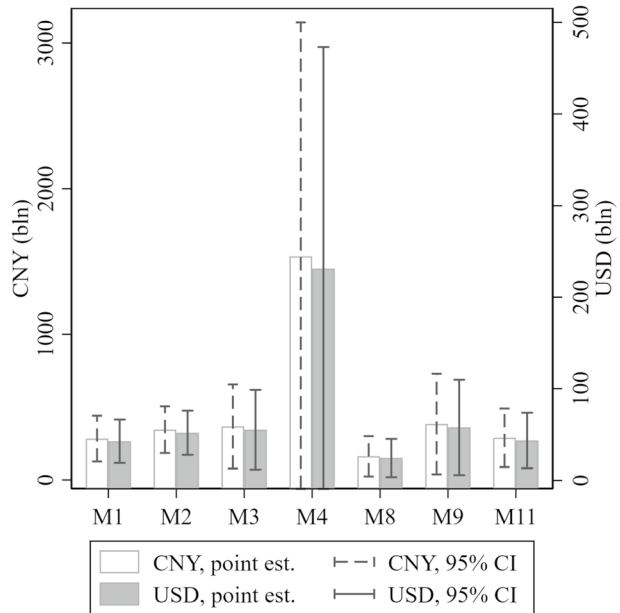
4 Aggregate Hedonic Prices of Green Space in Central Beijing

This section derives the aggregate hedonic prices of green space using the unit hedonic price estimates from Sect. III. To reduce computational workloads, we confined the analysis to green spaces that are larger than 0.5ha and located in Beijing’s six central districts. For each of these green spaces, we first searched in our dataset for all residential blocks located in the green space’s 1 km radius. The aggregate hedonic price was then calculated by multiplying the unit hedonic price (the coefficient on the variable ‘*Green area 0–1 km*’



Fig. 6 Hedonic prices (billion CNY in 2016 prices) of green spaces in Central Beijing

Fig. 7 Aggregate hedonic prices (in 2016 prices) of green spaces in Central Beijing. Note: M1: Model 1, regression spline; M2: Model 2, parametric covariate balancing generalised propensity score (CBGPS) matching; M3: Model 3, non-parametric CBGPS matching; M4: Model 4, difference-in-differences (sub-sample); M8: Model 8, ordinary least squares (OLS); M9: Model 9, instrumental variable; M11: Model 11, OLS (sub-sample)



after being converted from a semi-elasticity estimate to a monetary marginal effect estimate) by the total area of the green space and the total floor area of the residential blocks within the green space’s 1 km radius. As an example, Fig. 6 maps the aggregate hedonic prices of these green spaces individually (based on the regression spline estimate in Model (1)). Admittedly, owing to the nature of the hedonic approach, the spatial distribution of the

aggregate hedonic prices largely depends on the area of green spaces and the density of housing (in this study, newly built housing). Still, these results provide instrumental information that can be directly fed into cost-benefit analysis for removing or creating a green space. Figure 7 presents the total hedonic price estimates of all these green spaces. These estimates, although concerning only a subset of Beijing's green spaces, are sizeable: the annual average of the most conservative aggregate hedonic price (based on the OLS estimate in Model 8) is comparable to 0.5% of Beijing's GDP in 2022 (CNY 4,161.09 billion or USD 618.65 billion) (Beijing Municipal Bureau of Statistics 2023).

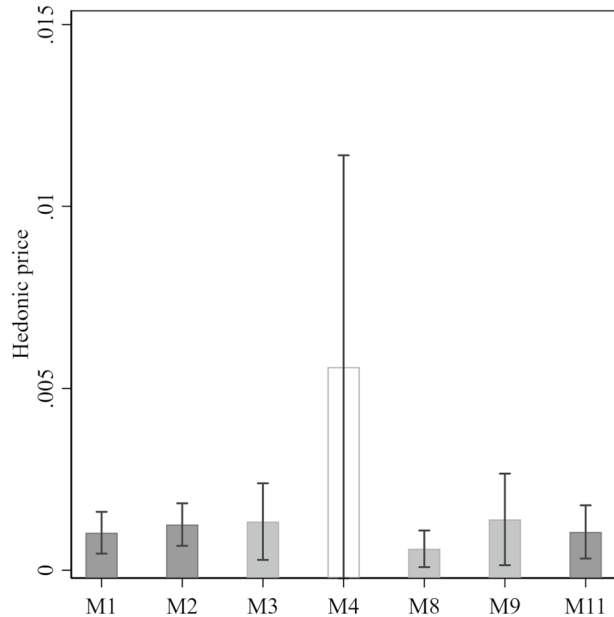
5 Discussion and Conclusion

There have been extensive studies that valued urban green amenities by measuring the ensuing housing price premiums (or hedonic prices). Nevertheless, in valuing the size (or area) of green spaces, there exist difficulties in defining the spatial scope as to which green spaces should be included in the valuation. Some previous studies resorted to a spatial piecewise step function approach, which regresses housing prices on a series of green space variables representing the area of green spaces at different distances from a home, in an attempt to discern the spatial limit where the hedonic value of green spaces disappears. This study builds upon that approach by exploring three novel empirical strategies that facilitate the identification of the threshold distance under a spatial piecewise framework. Among the three strategies, the model-selection-based regression spline method provides objective indications of the threshold distance that optimises the predictive performance of the hedonic pricing model. By contrast, the spatial piecewise matching and spatial difference-in-differences (DID) analyses provide stronger causal evidence for the hedonic price estimates.

Using a rich dataset from Beijing, we found that green space is more likely to be capitalised into housing prices within a 1 km radius. This conclusion is based on the highly comparable findings from the three empirical strategies, which offer different strengths and thus reinforce each other. In particular, the joint implementation of the three strategies has allowed us to assess how the exact threshold distance identified by the regression spline analysis compares to the findings from the matching and DID analyses, which better address endogeneity bias. The results of the three analyses taken together suggest that 1 km is likely to be a lower bound of the threshold distance for our dataset collected from Beijing. In comparison, previous spatial piecewise analyses by Conway et al. (2010) and Sander et al. (2010) found much shorter threshold distances (300ft in California and 250 m in Minnesota, respectively). Nafilyan and Lorenzi (2019) opted to focus on a 500 m radius in England and Wales. Such discrepancies regarding the threshold distance may relate to certain contextual differences. For example, homes in the US and the UK typically have a private green space in the backyard, whereas urban residents in Beijing mostly rely on communal and public green spaces at some distance. Some other studies have focused on the distance-dependent pattern of the hedonic price of the proximity to the closest green space (e.g., Daams et al. 2016, 2019; Łaszkiwicz et al. 2022; Melichar and Kaprová 2013; Wu et al. 2022), and found that the hedonic price does not disappear until several kilometres away.¹⁵ However,

¹⁵ These studies all focus on green spaces in urban areas. Regarding other types of vegetated areas, such as national parks in rural areas, housing prices can be affected by the proximity to national parks much farther away, such as 46.7 km on average, as found by Gibbons et al. (2014).

Fig. 8 Hedonic price estimates (semi-elasticities) from different estimation methods. *Note:* Bars: point estimates (white: $p < 0.10$; lighter grey: $p < 0.05$; darker grey: $p < 0.01$). Capped spikes: 95% confidence intervals. M1: Model 1, regression spline; M2: Model 2, parametric covariate balancing generalised propensity score (CBGPS) matching; M3: Model 3, non-parametric CBGPS matching; M4: Model 4, difference-in-differences (sub-sample, original estimate divided by the average size of green spaces so as to be comparable with other models' estimates); M8: Model 8, ordinary least squares (OLS); M9: Model 9, instrumental variable; M11: Model 11, OLS (sub-sample)



this could be a reflection of the inherent difference in the hedonic prices of the proximity to a green space and the size of a green space. Moreover, those studies did not adopt multiple quasi-experimental strategies to further account for potential endogeneity issues stemming from various sources.

Focusing on green spaces within a 1 km radius, we estimated the hedonic price using several estimation methods, including OLS, regression spline, CBGPS matching, spatial DID, and instrumental variable estimation. These methods gave a range of hedonic price estimates and their confidence intervals, which are summarised in Fig. 8. The lowest (CNY 15.41 or USD 2.32 per m² of housing per ha of green space) was from the OLS estimation (Model 8, Table 5). The highest (CNY 147.82 or USD 22.26, same unit as above) was from the DID analysis (Model 4, Table 4). All other estimates are within a narrower range, from CNY 26.98 or USD 4.06 (the regression spline estimate, Model 1, Table 2) to CNY 36.56 or USD 5.51 (the instrumental variable estimate, Model 9, Table 5).

As mentioned previously, these estimation methods offer different strengths and some are based on only a subset of our data. Policymakers should take these methodological considerations into account when choosing estimates. The DID estimate might offer a stronger causal interpretation and might be more applicable to a subset of green spaces created in 2006–2016 (as described in Sect. 3.3). The OLS estimate is more prone to endogeneity bias and represents the most conservative estimate. The middle-range estimates avoid the higher and lower ends.

The magnitudes of these unit hedonic price estimates seem limited, but the aggregate hedonic price estimates are considerable. Even focusing on the most conservative estimate and a subset of green spaces in central Beijing, the annual average of the aggregate hedonic price is comparable to half a per cent of Beijing's GDP in 2022. In comparison, the aggregate hedonic price estimated by Nafilyan and Lorenzi (2019) for green space in England and Wales in 2016 is about 4% of the UK's GDP in that year, which is also sizeable.

In communicating our findings to policymakers, several other considerations deserve comment. First and foremost, green space provides many valuable ecosystem services that

are not capitalised into housing prices and hence cannot be captured by hedonic prices. For instance, in 2022, despite Covid, Beijing accommodated 182 million tourists from China and abroad (equivalent to more than half of the US population), who spent CNY 252 billion (~USD 37 billion) in the city (Beijing Municipal Bureau of Statistics 2023). A large proportion of them may have visited the city's historical parks (such as the Temple of Heaven Park) and other green amenities (such as the Olympic Park). In that case, the recreational value of green space materialised in the form of attracting tourists and contributing to tourism revenues. Therefore, the hedonic value of homes may represent only a fraction of green spaces' total value. Second, our dataset only concerns housing newly built in 2006–2016. The aggregate hedonic price of green space would be much more sizeable had we taken into account all housing in the city. In addition, some benefits of green space may be better captured by non-monetary measures, such as subjective residential satisfaction, as in Ren and Folmer (2017). That said, on account of the abundance of green space currently in Beijing, adding more green space may eventually reduce their per unit value, due to diminishing marginal utility. In fact, in Model 8, adding the quadratic form of the variable 'Green area 0–1 km' yields a negative coefficient with a notably low p value (0.009), implying the possibility of a diminishing hedonic price if more green space is added. Lastly, we caution against literally extrapolating the specific results of this study, such as the threshold distance we identified (1 km), to other cases. For example, in a less populated city with a more open layout, housing prices may be affected by green space outside a 1 km radius.

Funding This study was funded by the Swedish International Development Cooperation Agency (SIDA Reference Number: 13/000230).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albouy D, Christensen P, Sarmiento-Barbieri I (2020) Unlocking amenities: estimating public good complementarity. *J Public Econ* 182:104110
- Austin PC, Fang J, Lee DS (2022) Using fractional polynomials and restricted cubic splines to model non-proportional hazards or time-varying covariate effects in the Cox regression model. *Stat Med* 41:612–624
- Bayer P, Keohane N, Timmins C (2009) Migration and hedonic valuation: the case of air quality. *J Environ Econ Manag* 58:1–14
- Beijing Municipal Bureau of Statistics (2023) Beijing statistical yearbook 2023. China Statistics Press, Beijing
- Beijing Municipal Education Commission (2015) Beijing education map. Beijing Evening News

- Bishop KC, Kuminoff NV, Banzhaf HS, Boyle KJ, von Gravenitz K, Pope JC, Smith VK, Timmins CD (2020) Best practices for using hedonic property value models to measure willingness to pay for environmental quality. *Rev Environ Econ Policy* 14:260–281
- Black KJ (2018) Wide open spaces: estimating the willingness to pay for adjacent preserved open space. *Reg Sci Urban Econ* 71:110–121
- Bockarjova M, Botzen WJW, van Schieff MH, Koetse MJ (2020) Property price effects of green interventions in cities: a meta-analysis and implications for gentrification. *Environ Sci Policy* 112:293–304
- Brander LM, Koetse MJ (2011) The value of urban open space: meta-analyses of contingent valuation and hedonic pricing results. *J Environ Manag* 92:2763–2773
- Cavaillès J, Brossard T, Foltête J-C, Hilal M, Joly D, Tourneux F-P, Tritz C, Wavresky P (2009) GIS-based hedonic pricing of landscape. *Environ Resour Econ* 44:571–590
- Cho S-H, Poudyal NC, Roberts RK (2008) Spatial analysis of the amenity value of green open space. *Ecol Econ* 66:403–416
- Conway D, Li CQ, Wolch J, Kahle C, Jerrett M (2010) A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *J Real Estate Finance Econ* 41:150–169
- Czembrowski P, Kronenberg J (2016) Hedonic pricing and different urban green space types and sizes: insights into the discussion on valuing ecosystem services. *Landsc Urban Plan* 146:11–19
- Daams MN, Sijtsma FJ, van der Vlist AJ (2016) The effect of natural space on nearby property prices: accounting for perceived attractiveness. *Land Econ* 92:389–410
- Daams MN, Sijtsma FJ, Veneri P (2019) Mixed monetary and non-monetary valuation of attractive urban green space: a case study using Amsterdam house prices. *Ecol Econ* 166:106430
- Desquilbet L, Mariotti F (2010) Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med* 29:1037–1057
- Dong G, Wu W (2016) Schools, land markets and spatial effects. *Land Use Policy* 59:366–374
- Fitzpatrick LG, Parmeter CF (2021) Data-driven estimation of treatment buffers in hedonic analysis: an examination of surface coal mines. *Land Econ* 97:528–547
- Fong C, Hazlett C, Imai K (2018) Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *Ann Appl Stat* 12:156–177
- Gibbons S, Mourato S, Resende GM (2014) The amenity value of English nature: a hedonic price approach. *Environ Resour Econ* 57:175–196
- Greene WH (2020) *Econometric analysis*, 8th edn. Pearson Education Limited, New York
- Greenstone M, Gayer T (2009) Quasi-experimental and experimental approaches to environmental economics. *J Environ Econ Manag* 57:21–44
- Haninger K, Ma L, Timmins C (2017) The value of brownfield remediation. *J Assoc Environ Resour Econ* 4:197–241
- Hastie T, Tibshirani R, Friedman JH (2008) *The elements of statistical learning: data mining, inference, and prediction*. Springer
- Heris M, Bagstad KJ, Rhodes C, Troy A, Middel A, Hopkins KG, Matuszak J (2021) Piloting urban ecosystem accounting for the United States. *Ecosyst Serv* 48:101226
- Imbens GW, Rubin DB (2015) *Causal inference for statistics, social, and biomedical sciences*. Cambridge University Press
- Imbens GW, Wooldridge JM (2009) Recent developments in the econometrics of program evaluation. *J Econ Lit* 47:5–86
- James G, Witten D, Hastie T, Tibshirani R (2017) *An introduction to statistical learning: with applications in R*. Springer
- Jardine SL, Siikamäki JV (2014) A global predictive model of carbon in mangrove soils. *Environ Res Lett* 9:104013
- Jaya IGNM, Folmer H (2020) Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia. *J Geogr Syst* 22:105–142
- Keogh RH, Morris TP (2018) Multiple imputation in Cox regression when there are time-varying effects of covariates. *Stat Med* 37:3661–3678
- Kovacs K, West G, Nowak DJ, Haight RG (2022) Tree cover and property values in the United States: a national meta-analysis. *Ecol Econ* 197:107424
- Kuminoff NV, Parmeter CF, Pope JC (2010) Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *J Environ Econ Manag* 60:145–160
- Łaszkiwicz E, Heyman A, Chen X, Cimburova Z, Nowell M, Barton DN (2022) Valuing access to urban greenspace using non-linear distance decay in hedonic property pricing. *Ecosyst Serv* 53:101394
- Lee DS, McCrary J, Moreira MJ, Porter J (2022) Valid t-ratio inference for IV. *Am Econ Rev* 112:3260–3290
- Li S, Yang J, Qin P, Chonabayashi S (2016) Wheels of fortune: subway expansion and property values in Beijing. *J Reg Sci* 56:792–813

- Liu Z, Hanley N, Campbell D (2020) Linking urban air pollution with residents' willingness to pay for greenspace: a choice experiment study in Beijing. *J Environ Econ Manag* 104:102383
- Mei Y, Gao L, Zhang P (2019) Residential property price differentials of waste plants: evidence from Beijing, China. *Appl Econ* 51:5952–5960
- Mei Y, Gao L, Zhang W, Yang F-A (2021) Do homeowners benefit when coal-fired power plants switch to natural gas? Evidence from Beijing, China. *J Environ Econ Manag* 110:102566
- Melichar J, Kaprová K (2013) Revealing preferences of Prague's homebuyers toward greenery amenities: the empirical evidence of distance-size effect. *Landsc Urban Plan* 109:56–66
- Millennium Ecosystem Assessment (2005) *Ecosystems and human well-being: synthesis*. Island Press
- Muehlenbachs L, Spiller E, Timmins C (2015) The housing market impacts of shale gas development. *Am Econ Rev* 105:3633–3659
- Nafilyan V, Lorenzi L (2019) Valuing green spaces in urban areas: a hedonic price approach using machine learning techniques
- Netusil NR, Chattopadhyay S, Kovacs KF (2010) Estimating the demand for tree canopy: a second-stage hedonic price analysis in Portland, Oregon. *Land Econ* 86:281–293
- Netusil NR, Levin Z, Shandas V, Hart T (2014) Valuing green infrastructure in Portland, Oregon. *Landsc Urban Plan* 124:14–21
- Office for National Statistics (2023) *Urban natural capital accounts methodology guide*, UK
- Orsini N, Greenland S (2011) A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata J* 11:1–29
- Perino G, Andrews B, Kontoleon A, Bateman I (2014) The value of urban green space in Britain: a methodological framework for spatially referenced benefit transfer. *Environ Resour Econ* 57:251–272
- Ren H, Folmer H (2017) Determinants of residential satisfaction in urban China: a multi-group structural equation analysis. *Urban Stud* 54:1407–1425
- Ren H, Folmer H (2022) New housing construction and market signals in urban China: a tale of 35 metropolitan areas. *J Hous Built Environ* 37:2115–2137
- Sander H, Polasky S, Haight RG (2010) The value of urban tree cover: a hedonic property price model in Ramsey and Dakota counties, Minnesota, USA. *Ecol Econ* 69:1646–1656
- Smith PL (1979) Splines as a useful and convenient statistical tool. *Am Stat* 33:57
- Stromberg PM, Öhrner E, Brockwell E, Liu Z (2021) Valuing urban green amenities with an inequality lens. *Ecol Econ* 186:107067
- Suits DB, Mason A, Chan L (1978) Spline functions fitted by standard regression methods. *Rev Econ Stat* 60:132–139
- Tanaka S, Zabel J (2018) Valuing nuclear energy risk: Evidence from the impact of the Fukushima crisis on U.S. house prices. *J Environ Econ Manag* 88:411–426
- Taylor PJ (1983) Distance decay in spatial interactions. *Geo Books*
- Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46:234–240
- United Nations (2021) *System of environmental-economic accounting—ecosystem accounting (SEEA EA)*
- Walls M, Kousky C, Chu Z (2015) Is what you see what you get? The value of natural landscape views. *Land Econ* 91:1–19
- Waltert F, Schläpfer F (2010) Landscape amenities and local development: a review of migration, regional economic and hedonic pricing studies. *Ecol Econ* 70:141–152
- Wang C, Wang S, Cao Y, Yan H, Li Y (2023) The social equity of urban parks in high-density urban areas: a case study in the core area of Beijing. *Sustainability* 15:13849
- Wegman EJ, Wright IW (1983) Splines in statistics. *J Am Stat Assoc* 78:351–365
- Wooldridge JM (2020) *Introductory econometrics: a modern approach*, 7th edn. Cengage Learning, Boston
- Wu W, Dong G (2014) Valuing the “green” amenities in a spatial context. *J Reg Sci* 54:569–585
- Wu C, Du Y, Li S, Liu P, Ye X (2022) Does visual contact with green space impact housing prices? An integrated approach of machine learning and hedonic modeling based on the perception of green space. *Land Use Policy* 115:106048
- Zhang L, Wu J, Liu H, Zhang X (2020) The value of going green in the hotel industry: evidence from Beijing. *Real Estate Econ* 48:174–199
- Zhang S, Liu J, Song C, Chan C-S, Pei T, Wenting Y, Xin Z (2021) Spatial-temporal distribution characteristics and evolution mechanism of urban parks in Beijing, China. *Urban For Urban Green* 64:127265
- Zheng S, Kahn ME (2008) Land and residential property markets in a booming economy: new evidence from Beijing. *J Urban Econ* 63:743–757
- Zheng S, Hu W, Wang R (2016) How much is a good school worth in Beijing? Identifying price premium with paired resale and rental data. *J Real Estate Finance Econ* 53:184–199

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Zhaoyang Liu¹  · Heqing Huang² · Juha Siikamäki³ · Jintao Xu²

✉ Zhaoyang Liu
zl290@cam.ac.uk

Heqing Huang
hqhuang2017@nsd.pku.edu.cn

Juha Siikamäki
Juha.Siikamaki@iucn.org

Jintao Xu
xujt@pku.edu.cn

¹ Department of Land Economy, University of Cambridge, Cambridge, UK

² National School of Development, Peking University, Beijing, China

³ International Union for Conservation of Nature, Washington, DC, United States