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

Zhaoyang **Liu**, Heqing **Huang**, Juha **Siikamäki**, Jintao **Xu**

C-EENRG *Working Papers, 2023-2*

C-EENRG Research Series - February 2023

UNIVERSITY OF CAMBRIDGE

CEENRG
Cambridge Centre for Environment, Energy and Natural Resource Governance

Please cite this paper as:

Z. Liu, H. Huang, J. Siikamäki, J. Xu, 2023, 'Area-Based Hedonic Pricing of Urban Green Amenities in Beijing: A Spatial Piecewise Approach'. C-EENRG Working Papers, 2023-2. pp.1-43. Cambridge Centre for Environment, Energy and Natural Resource Governance, University of Cambridge.

Dr Zhaoyang Liu
Department of Land Economy, University of Cambridge, UK

Heqing Huang
National School of Development, Peking University

Dr Juha Siikamäki
International Union for Conservation of Nature

Prof Jintao Xu
National School of Development, Peking University

Contact:
Dr Zhaoyang Liu
Department of Land Economy, University of Cambridge
19 Silver Street, Cambridge CB3 9EP
zl290@cam.ac.uk

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

Zhaoyang Liu, Heqing Huang, Juha Siikamäki, Jintao Xu

ABSTRACT

This study explores a spatial piecewise approach as a means to facilitate the hedonic valuation of per unit area of urban green spaces at different distances from a property. In particular, we adapted the step function and regression spline estimation methods to statistically and visually identify the spatial boundary where green spaces cease to be capitalised into house prices. In comparison, existing literature on the hedonic prices of green spaces, despite being extensive, has mostly focused on the proximity to and views of urban green spaces, instead of the area of green spaces. Yet arguably, the hedonic price of the area of green spaces can be more relevant to cost-benefit analysis for urban land use decision making, which typically concerns a particular area of land. The empirical approach proposed in this study was applied to a rich census dataset collected from Beijing. Our hedonic price estimates are robust to an instrumental variable estimator and a novel matching algorithm that minimises covariate imbalance for a continuous treatment variable (the area of green spaces).

Key words: Urban green space; Hedonic pricing; Spatial piecewise regression; Regression splines; Model selection; Instrumental variable estimation; Covariate balancing matching for continuous treatment

ACKNOWLEDGEMENTS

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.

1. Introduction

Urban parks and other green amenities provide metropolitans with spiritual, aesthetic, educational and recreational benefits, as well as many other desirable ecosystem services (Millennium Ecosystem Assessment, 2005). Nevertheless, creating green amenities in compact urban areas typically incurs prohibitively expensive costs, in light of fiercely competing demands for land. Economically optimal decision making in this regard often rests upon a quantitative understanding of the benefits and costs attached to urban green amenities. The opportunity costs tend to be more tangible, such as investment in green infrastructures (e.g. trees, paddocks, exercise equipment, playgrounds and other facilities) and foregone profits of alternative land uses (e.g. residential and commercial developments). In comparison, monetary value of the benefits is often largely obscure and much debated. Urban green amenities are usually open to visitors free of charge, which precludes a market price as a monetary measurement of the benefits accruing to users. Take Beijing as an example. In 2017, the entire municipality's fixed investment (sourced from both public and private sectors) in green infrastructures ran to CNY 11.8 billion or USD 1.8 billion (Beijing Municipal Bureau of Statistics, 2018). This sum represents 28% of the municipality's public spending on healthcare in the same year (Beijing Municipal Bureau of Statistics, 2018). Yet, it remains unclear whether such heavy investments in urban green spaces are economically worthwhile, owing to the absence of a solid monetary valuation of the ensuing benefits.

A variety of non-market valuation techniques have been developed to overcome this hurdle. Hedonic pricing is one of the tried and trusted methods. It describes green amenities surrounding a property as one of its attributes, and seeks to discern the implicit value of green amenities via their quantitative relation to property prices. There has been a vast array of existing literature dedicated to this subject (see Bockarjova et al. 2020, Brander et al. 2011, Kovacs et al. 2022 and Perino et al. 2014 for systematic reviews). Hedonic studies from Beijing that involve green amenities are not unprecedented either (e.g. Dong et al. 2016, Mei et al. 2019, 2021, Wu et al. 2014, Zhang et al. 2020, Zheng et al. 2016 and Zheng and Kahn 2008). Despite that, they have mostly valued the *proximity* to the nearest green space, namely the value added to a house if it is located in closer proximity to a green space. On pragmatic grounds, it might be difficult to utilise this measurement to derive an aggregate value for a particular green space,[1] yet understanding the total benefits of a green space is essential for the cost-benefit analysis of urban land use. Recent developments in applying Geographical Information Systems (GIS) and machine learning to environmental valuation have facilitated hedonic valuations of the *views* of green amenities (e.g. Black 2018, Cavailhès et al. 2009, Daams et al. 2016, Walls et al. 2015 and Wu et al. 2022). This value is more amenable to aggregation, particularly if views are measured as a continuous variable (such as the area of green amenities in a property's view shed). Even so, this is likely to be an incomplete estimate of the total benefits of green amenities, since such benefits arise from both passive aesthetic values and other active onsite use values. The latter values may be better captured by the proximity to green amenities. Yet, attempting to aggregate both valuations undoubtedly adds an additional layer of complexity.

In comparison, valuing the *area* (or *size*) of green space would produce a more comprehensive measurement of the capitalisation of green amenities in property prices. In particular, it would directly inform decision making in urban planning, by providing a straight answer on the hedonic price of a green space that occupies a certain land area. However, this practice is notably less common in existing literature. A pivotal technical hurdle appears to be associated with a model specification problem as to which green spaces should enter the hedonic price model. Intuition suggests that the value of a property is less likely to be affected by a sufficiently distant green space. This implies that perhaps the value of a property is only affected by green spaces within a 'threshold' or maximum distance, and it is these green spaces that should enter the hedonic price model. In a few rare examples of hedonic studies valuing the size of green amenities, the most common strategy was 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 and Waltert and Schläpfer 2010) or a census

[1] One could hypothetically remove or create a green space and recalculate the distance between each property and its nearest green space, which would give rise to the aggregate impact on property values attributable to this particular green space. But this might be an incomplete estimate of the hedonic price of the green space, which relates to our subsequent discussion of the hedonic pricing of the views of green amenities.

block (e.g. Cho et al. 2008 and Netusil et al. 2014). Two other cases (i.e. Conway et al. 2010 and Sander et al. 2010) undertook what this study refers to as a spatial piecewise analysis. They delineated a series of concentric rings surrounding each house and measured the green space coverage within each doughnut-shaped area. All these green space coverage variables entered the hedonic pricing model. It was assumed that the estimated hedonic prices of green spaces 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 can be formalised into a space-wise variant of the 'step function' approach. In regression settings, the conventional step function approach typically involves breaking an explanatory variable into bins and fitting a constant in each bin (James, Witten, Hastie, & Tibshirani, 2017). In contrast, the spatial step function approach assumes that the hedonic price per unit area of green spaces (or the coefficient on the area of green spaces) can be expressed as a step function of their distance from the property, and therefore differs from the standard practice where the dependent variable (house prices in this case) is assumed to be a step function of an explanatory variable (e.g. area of green spaces). A similar variant of the step function approach was adopted by Schlenker and Roberts (2009), who investigated the nonlinear effects of temperature on agricultural production via regressing crop yields against accumulated growth time at different temperatures. Henceforth, we will borrow the terminology 'step function' to refer to this model specification.

However, this approach can hardly avoid adopting arbitrary cut-off levels of the statistical significance of the estimated hedonic prices to decide the threshold radius where green spaces cease to contribute to house prices, which may become increasingly debatable amid a recent flood of cautions against statistical studies basing conclusions primarily on arbitrary cut-off levels of statistical significance (e.g. Ferraro and Shukla 2020 and Wasserstein et al. 2019). Moreover, the spatial thresholds derived using the spatial piecewise approach by Conway et al. and Sander et al. are counter-intuitively small (300ft and 250m respectively). In contrast, other studies valuing the proximity to green space found that the association between house prices and proximity to green space would not disappear up until 2, 3.2 or 7km away (Daams et al., 2016; Mansfield, Pattanayak, McDow, McDonald, & Halpin, 2005; Melichar & Kaprová, 2013). Admittedly, urban residents' perceptions of the size of and proximity to green space could be inherently different. But the striking difference still provokes a suspicion that that the estimated hedonic prices of the size of green amenities beyond the purported bounds (300ft and 250m) may be under-identified, owing to high correlation among green space coverage variables for neighbouring locations (Irwin, 2002). Perhaps this is what underlies the paucity of hedonic valuations of the size of green amenities.

In addition to the step function approach, this study proposes another empirical strategy that circumvents these pitfalls using regression splines and model selection procedures. We assume that hedonic prices of green spaces at different distances within the threshold distance is 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. This represents a fitted curve consisting of polynomial splines before the last knot and a right tail restricted to be flat, which can be derived using a regression model (Orsini & Greenland, 2011; Wegman & Wright, 1983). We next loop over all possible threshold distances up to 10km with a step length of 100m in search of the preferred threshold distance, using both within-sample and out-of-sample predictions for model assessment and selection. This allows us to reparametrise the hedonic price model by approximating hedonic prices at different distances as a piecewise polynomial function of distance, and thereby helps avoid directly including a large number of green space area variables for all spatial segments as regressors. This model specification resembles a hybrid of the polynomial function and piecewise approaches of Schlenker and Roberts (2009), which were adopted in their study as alternatives to the step function approach. There have been hedonic studies that undertook polynomial regressions to explore the nonlinear patterns of house prices with regard to the proximity to restored brownfields (Haninger, Ma, & Timmins, 2017), power plants that switched fuels from coal to gas (Mei et al., 2021), shale gas wells (Muehlenbachs, Spiller, & Timmins, 2015), and nuclear power plants (Tanaka & Zabel, 2018). These studies identified the spatial scope of the hedonic valuation through assessing whether the price of a house at a certain distance of an environmental disamenity (a brownfield, power plant or shale gas well) was affected by an external shock (a brownfield restoration programme, fuel switching, shale gas drilling or a nuclear accident) that changed (perceptions towards) the disamenity. The regression spline approach proposed in this study provides a useful addition to their approach through directly estimating the *hedonic price* of an environmental amenity as a function of distance, whereas the three studies mentioned above estimated *house prices* as a function of distance and compared the estimated functions with and without an external shock. Our approach is therefore particularly applicable in the absence of such external shocks to environmental amenities/disamenities. Moreover, those three studies resorted to informal visual assessment to decide whether a shock caused a visually discernible difference in house prices in a certain radius. In contrast, our approach identifies the threshold distance using a model selection procedure which provides formal statistical justification.

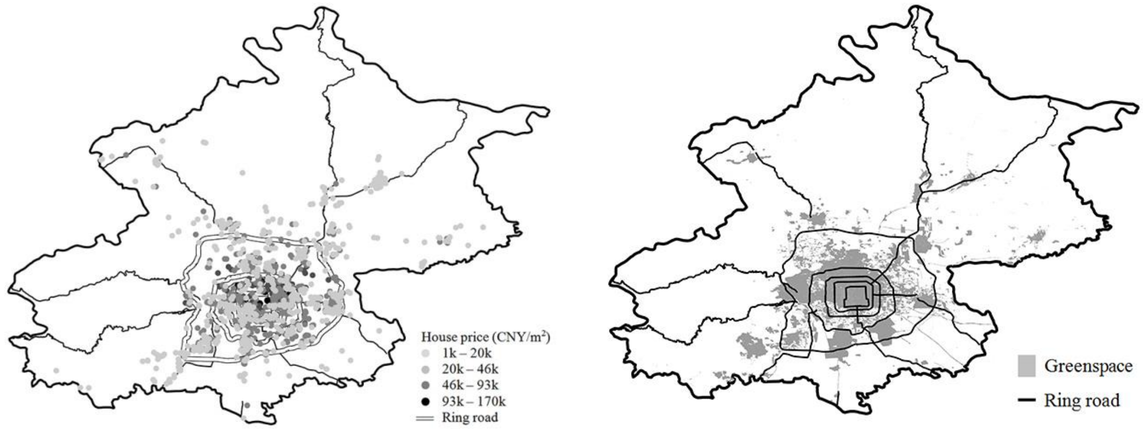
Moreover, the identification of the threshold distance justifies the validity of utilising outer green spaces 'just' beyond the threshold distance to instrument inner green spaces that are being valued, since such outer green spaces are likely to be highly correlated with the inner green spaces but are less likely to directly affect property values. In addition, we further tested the robustness of our hedonic price estimates using a novel matching approach proposed by Fong et al. (2018), which minimises covariate imbalance for a continuous treatment variable (the area of green spaces). These empirical strategies speak to an increasing emphasis on adopting quasi-experimental methods (i.e. matching, difference-in-differences and fixed effects, instrumental variables, and regression discontinuity designs) to mitigate endogeneity bias in hedonic price estimates (Bishop et al., 2020; Kuminoff, Parmeter, & Pope, 2010).

The remainder of this paper is structured as follows. Section 2 describes the study site, data and variables. Section 3 performs the spatial piecewise estimation of hedonic prices for green spaces, respectively using the step function and the regression spline approaches. In addition, this section reports the instrumental variable and matching estimates which better account for potential endogeneity issues. Section 4 provides a demonstration of applying our results to the aggregation of the 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 is similar in size to Greater London (or New York City), and accommodates nearly 20 million people. In respect of the administrative hierarchy, Beijing consists of 16 districts, which are further organised into 147 sub-districts ('*jiedao*'). Below the sub-district level, there exist 2,932 communities ('*shequ*'), each of which encompasses a number of residential blocks ('*xiaoqu*'). In 2016, the size of Beijing's economy was on a par with Norway. This megacity has attracted floods of migrants from all over China, and its urban population has almost doubled during the past 15 years, which has led to spiralling house prices. According to the China Index Academy, our housing data provider, the new build average price within the 6th ring road[2] has exceeded CNY 60,000/m^2 in the first half of 2017, which implies that a homebuyer has to shell out nearly half a million US dollars for a typical 50-m^2 apartment. The government attempts to cool off the overheating housing market by regulating strict eligibility criteria for home buying, tightening mortgage rules and creating more publicly subsidised affordable housing. But at bottom, Beijing's real estate sector is by and large a market-oriented system (Zheng & Kahn, 2008).

[2] The city centre of Beijing is surrounded by five ring roads (numbered from the innermost 2nd ring road to the most remote 6th ring road).



(a) House prices **(b) Green spaces**

Figure 1 Visualisation of residential blocks and urban green spaces in Beijing

Nevertheless, urban green amenities are well developed in this densely populated city. Beijing's dry climate and inland location have left green spaces one of the few types of environmental amenities available to its residents. By definition, the scope of urban green amenities is confined to those situated in Beijing's urban areas, and hence does not include vegetated land (mostly forests) in the outlying rural areas of the municipality. Beijing's urban green spaces comprise five broad categories, including public green spaces (e.g. parks), productive green spaces (e.g. tree nurseries), protective green spaces (e.g. noise buffers and windbreaks), affiliated green spaces (those attached to residential blocks, public bodies and businesses, etc.) and other green spaces. Chief among them are public and affiliated green spaces, which respectively represent 37% and 41% of the total area of green spaces. The municipal government is responsible for the development and maintenance of public green spaces, whereas affiliated green spaces are created and/or maintained by the entities they are attached to. In 2016, the per capita area of green spaces came to 40m^2, which went above the average dwelling size (32m^2). The seemingly unbalanced trade-off arouses curiosity about the value of these green assets.

In this context, this study undertakes a hedonic analysis to investigate the house price premiums attributable to green spaces. Our analysis relies on a geographically referenced census dataset that details the first-time transactions of all new-build properties in Beijing between 2006 and 2016. All geographic data were mapped and compiled in ArcGIS. We consulted previous hedonic studies, particularly those from Beijing (i.e. Dong et al. 2016, Li et al. 2016, Wu et al. 2014 and Zheng et al. 2016; 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 data were obtained from the China Index Academy and were allegedly sourced from the housing transaction registration system of the municipal government, which recorded the first-time transactions of all 1,270 new-build residential blocks between 2006 and 2016 (as in Figure 1a).[3] We were only able to obtain the longitude and latitude coordinates of the centroid of each residential block. These coordinates were used to map the housing data to the urban infrastructure data. Therefore, the urban infrastructure variables would have the same values for all properties within 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 (as in Figure 1b). This dataset is a full inventory of more than 230,000 plots of green spaces in Beijing's urban areas. Surveyors delineated digital boundaries of these green spaces using GPS trackers, and investigated a number of other attributes such as, in particular, the time when a green space was created. This enabled us to map property transactions to the green spaces that existed at the selling time. As illustrated in Figure A1 in the appendix, we calculated the area of green spaces in each ring[4] around the centroid of a residential block, giving rise to a series of variables indicating the area of green spaces at different distances.[5]

Further, we constructed a wide range of control variables for 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. This dummy

[3] We excluded data points for publicly subsidised affordable homes.

[4] The one at the centre is a solid circle.

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

variable is expected to capture a downward shift in property values, since the southern half of Beijing was historically occupied by lower income groups and disadvantaged ethnic 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' schools/kindergartens in each school district. The title 'demonstration' ('*shifan*') is usually awarded to the highest ranked and most reputable schools in Beijing. We extracted geographic data on other urban infrastructures and services (including hospitals, railways, highways, regular roads, subway stations, bus stops, restaurants and shops) from Gaode Maps, a leading web map service in China, and measured the proximity and quantity variables for them. These infrastructures and services facilitate a more comfortable and convenient life, but may also induce adverse effects such as noise and crowdedness. Lastly, we generated a set of district, ring road (representing zones partitioned by the ring roads) and year dummies. The district (ring road) dummies control for the price effect of district (ring road) specific features that do not vary over time. The year fixed effects capture macro shocks that have a uniform effect on property values, such as changes in housing and mortgage policies.

3. Spatial Piecewise Estimation of Hedonic Prices For Green Spaces

This section reports the estimation results of the step function and the regression spline approaches. Moreover, we used instrumental variable estimation and a novel matching approach to formally explore the robustness of our results to potential endogeneity bias.

3.1. Step Function Estimate

The departure point of our data analysis is the step function approach, where the hedonic pricing model contains a sequence of variables that measure the total area of green spaces within each ring illustrated in Figure A1 in the appendix:

$$log(HousePrice_i) = \sum_j p_j Green_{ij} + \boldsymbol{\beta x_i} + \varepsilon_i. \qquad \text{Eq.1}$$

In this equation, the subscript i indexes observations (at the residential block level), whilst j denotes the ordinal numbers of rings (where a larger number indicates an outer ring). This is the most flexible specification which does not presume any functional form for the dependency of the hedonic price of green spaces on distance. The vector $\boldsymbol{x}$ consists of all other explanatory variables specified in Section 2. All estimation in this study was undertaken in Stata. Due to the large number of green space variables, their coefficient estimates (which represent hedonic prices at different distances) are graphically reported in Figure 2, and other regression output is omitted for brevity (available upon request).

As can be seen in Figure 2a, when using green space variables for 100m wide rings (or using a 100m step length), the estimated hedonic price over distance roughly follows an inverted U-shaped pattern until about 1km away from a residential block. The hedonic price curve first increases with distance, peaks at 500m, and then declines with distance until about 1km. Beyond that point, the estimated hedonic price over distance curve exhibits a largely flat trend up to 10km, despite small swings around zero. There appears to be a breakpoint near 1km, but the estimated hedonic prices at different distances are mostly statistically insignificant inside 1km, except in the 400m–500m and 600m–700m rings, where the estimates are statistically significant at the 10% level. Since house prices are measured in logarithms, the estimated coefficient on the area of green spaces in the 400m–500m ring translates into a CNY 57.05 (~USD 8.59) increase in the average house price per m^2 in response to a 1ha increase in green spaces in this ring, ceteris paribus.[6] In comparison, the estimated hedonic price in the 600m–700m ring drops to CNY 40.08 (~USD 6.04) per m[2] of housing per ha of green spaces. To provide a flavour of the magnitude of the aggregate hedonic value, we measured the difference in the total predicted property value[7] between 1) the baseline scenario where all green space coverage variables took the actual values, and 2) a hypothetical scenario where we removed green spaces that were created after 2006 in the 400m–500m and 600m–700m rings of all residential blocks in our dataset. Despite the seemingly limited magnitude of the hedonic price estimates, the aggregate hedonic value of these green spaces amounts to a sizeable CNY 58 billion (~USD 9 billion) in 2016 prices, owing to the vast total floor area of the 1,270 residential blocks. This figure is likely to be a lower bound of these green spaces' hedonic value, since our housing transaction dataset only concerns houses built between 2006 and 2016. Even so, this lower bound still outweighs Beijing's fixed investment in all green infrastructures between 2006 and 2016 (CNY 26 billion or USD 4 billion in 2016 prices) (Beijing Municipal Bureau of Statistics, 2018).[8] However, we have also observed statistically significant hedonic price estimates in some further away locations (e.g. at 3km). The swings in the magnitude and statistical significance of the hedonic price estimates add to the difficulty of identifying the spatial scope for the hedonic valuation.

Moreover, the statistical significance of hedonic price estimates at different distances is found to be sensitive to the step length of the analysis. When using 200m rings, as shown in Figure 2b, we find a similar (albeit 'coarsened') hedonic price over distance curve: it still has an inverted U shape within 1km and becomes virtually flat afterwards. Despite that, the estimated hedonic prices are statistically significant at the 5% level within the 400m–800m belt, which is much wider than the 400m–500m and 600m–700m rings that have a statistically significant hedonic price estimate derived

---

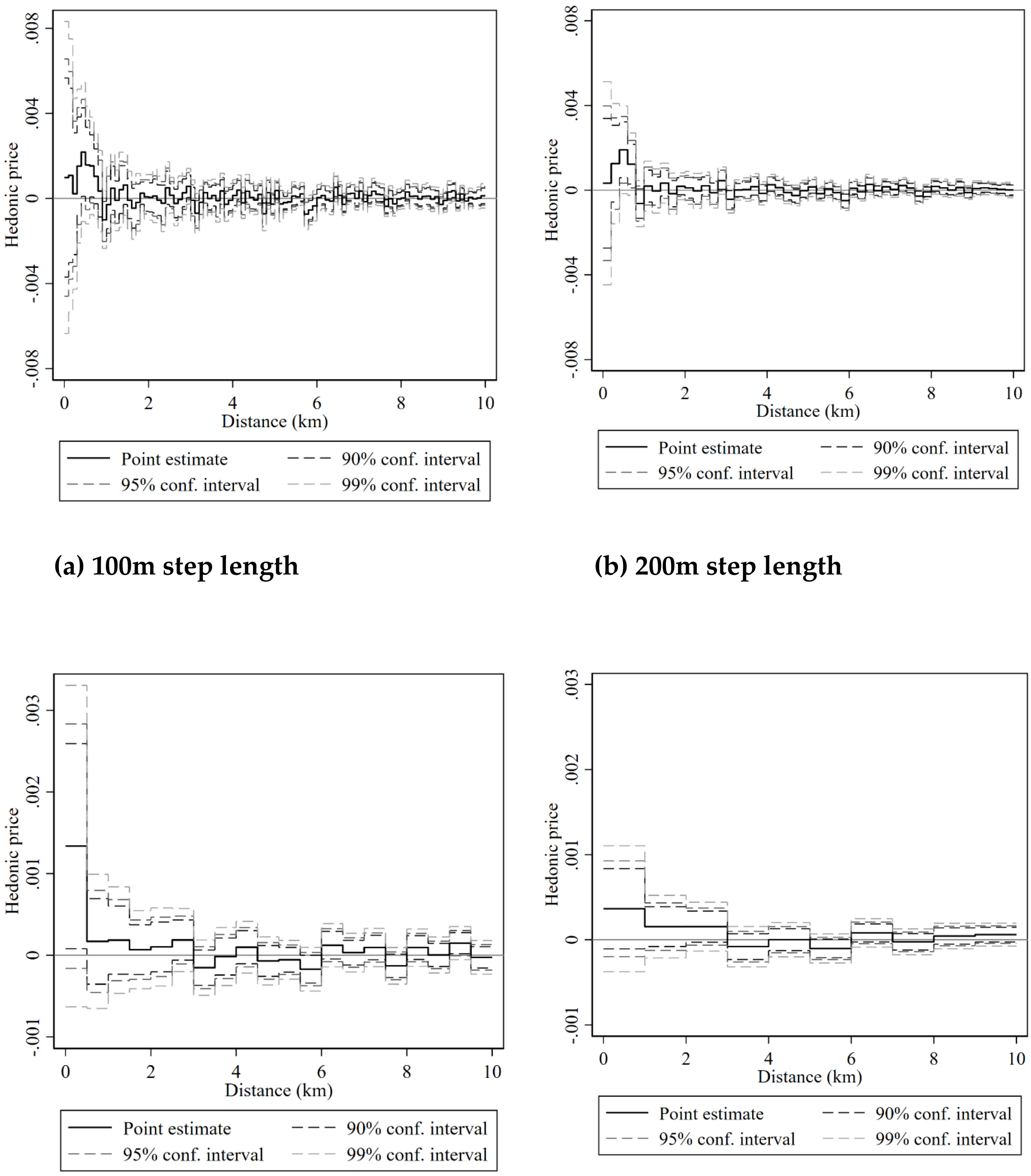
[6] The original log-linear estimate ($\hat{p}_j$) indicates that the average predicted house price $E(\widehat{House\ Price_i})$ would be changed by $100[exp(\hat{p}_j \Delta G)]$ per cent if the area of green spaces is changed by ΔG on average (Wooldridge, 2013).

[7] Following Cameron and Trivedi (2009), house prices were predicted as $\widehat{HousePrice_i} = exp\ (\sum_j \hat{p}_j Green_{ij} + \widehat{\boldsymbol{\beta}} \boldsymbol{x_i}) exp\ (0.5\hat{\sigma}^2)$, where $\hat{\sigma}$ is the root mean squared error of Eq.1 and all other notations have the same meaning as in Eq.1.

[8] The monetary values of green spaces and municipal fixed-asset investment in previous years (2006–2015) were converted to 2016 values using a discount rate of 8%, as recommended by the National Development and Reform Committee of China and the Ministry of Housing and Urban-Rural Development of China (2006). The undiscounted raw sums are respectively CNY 44 billion (~USD 7 billion) and CNY 20 billion (~USD 3 billion).

using a 100m step length as we mentioned above. The magnitudes of the estimated hedonic prices (CNY 49.88 or USD 7.51 per m^2 of housing per ha of green spaces for the 400m–600m ring, and CNY 32.12 or USD 4.84 for the 600m–800m ring) resemble those for the 400m–500m and 600m–700m rings in the 100m step length setting. Yet, the aggregate hedonic value of green spaces created after 2006 in the 400m–800m belt comes to CNY 99 billion (~USD 15 billion) in 2016 prices, which nearly doubles the counterpart estimate in the 100m step length setting. There still exist statistically significant hedonic price estimates for outer rings (e.g. the 2.8km–3km ring). Turning next to Figure 2c, with a 500m step length, we only find a statistically significant hedonic price estimate within the 500m circle (CNY 34.97 or USD 5.27 per m^2 of housing per ha of green spaces, p-value = 0.08). Beyond the 500m circle, the statistical significance of the hedonic price estimates disappears and the magnitude declines rapidly to practically zero, which somewhat resembles the results of Conway et al. (Conway et al., 2010) and Sander et al. (Sander et al., 2010) who adopted a similar approach. The aggregate hedonic value of green spaces created after 2006 inside 500m shrinks notably to CNY 47 billion (~USD 7 billion) in 2016 prices, which is smaller than that given by a 100m step length. Moreover, the hedonic price estimate becomes statistically significant in the 9km–9.5km ring, which is largely counterintuitive. Lastly, it can be seen from Figure 2d that hedonic price estimates for 1km wide rings are statistically insignificant everywhere (even at the 10% level), which substantially deviate from our previous findings.

Overall, there appears to be a lack of clear-cut indication of a breakpoint distance where green spaces cease to affect house prices. In terms of statistical significance, it is difficult to find a breakpoint distance where the p-value structurally changes from being below a conventional threshold level (at all distances inside the breakpoint distance) to being above that level (at all distances outside the breakpoint distance). In addition, the practice of basing research conclusions primarily on arbitrary cut-offs of the p-value regardless of the magnitude of the estimates has been increasingly cautioned against (e.g. Ferraro and Shukla 2020 and Wasserstein et al. 2019). For instance, the 100m step length models found a statistically significant hedonic price estimate for green spaces at 3km. However, this estimate has a much smaller magnitude than the estimates for green spaces within 1km, and the general pattern of the hedonic price estimates at other distances around 3km suggest that these are more likely to be small swings around zero. In light of both the magnitude and statistical significance, the 100m and 200m step length models seem to suggest a breakpoint around 1km, yet this conjectured breakpoint is mostly based on informal visual assessment rather than on formal statistical criteria, which resembles the visual assessment approach of Haninger et al. (2017), Mei et al. (2021), Muehlenbachs et al. (2015) and Tanaka and Zabel (2018). Moreover, these patterns are largely sensitive to the step length, which calls for justifiable and replicable means to assess which step length is preferable. A natural solution might be to prioritise the step length that gives rise to regression models with higher levels of explanatory power or goodness of fit. This also implies the possibility of identifying the breakpoint distance according to measures of goodness of fit, if the choice of the breakpoint's position can be converted to a model selection problem. We next formally explore this possibility through a regression spline approach which endogenously chooses the position of the breakpoint using model selection procedures.



(a) 100m step length

(b) 200m step length

(c) 500m step length

(d) 1km step length

Figure 2 Hedonic price estimates at different distances: Step function estimates

3.2. Regression Spline Estimates

We propose a regression spline approach adapted from Schlenker and Roberts (2009) and Orsini and Greenland (2011). The hedonic pricing model is reparametrised as a restricted regression spline (Eq.2), assuming that 1) only green spaces within a threshold distance D_K (which corresponds to the last knot of the spline) affect house prices, and the hedonic price of these green spaces can be expressed as a polynomial function of distance, and 2) 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.

$$log(House\ Price_i) = \alpha \sum_j Green_{ij} + \sum_{k=1}^{k=K} \gamma_k \sum_j (D_k - d_j)_+^n Green_{ij} + \boldsymbol{\beta x_i} + \varepsilon_i. \quad \text{Eq.2}$$

In this hedonic pricing model, α, γ_k and $\boldsymbol{\beta}$ are the parameters to be estimated, d_j represents the radius of the middle of the *j*-th ring, k indicates the order of knots, 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 (2012).

The positions of the knots (and hence the threshold distance) are endogenously decided via a model selection procedure. The step function estimates (in the finest 100m step length setting) suggest that the hedonic price over distance curve is likely to have two conspicuous turning points (around 0.5km and 1km, respectively). We have therefore allowed 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 & Greenland, 2011; Wegman & 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 pricing model (Eq.2) is then estimated repeatedly using 5,151 possible combinations of the locations of the knots within a 10km radius (the possible locations of the knots differ in other step length settings), to search for the model specification (defined by the locations of the knots) that fits the data best.

We propose two different procedures for model assessment and selection. One is the conventional approach based on within-sample prediction and the Akaike Information Criterion (AIC).[9] On the other hand, the rapid proliferation of statistical learning suggests that models with better 'out-of-sample' explanatory power (as opposed to the conventional within-sample explanatory power) are increasingly preferred (Varian, 2014). However, regression spline models are in general considered particularly ill-adapted to extrapolation beyond the data used to fit these models (Suits, Mason, & Chan, 1978). In light of that, we conducted a 5-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), and Jardine and Siikamäki (2014). We first randomly split our data into five equal-sized sub-samples. For each sub-sample, we fit a hedonic pricing model using the other four sub-samples (training data), and calculate the mean squared error (MSE) of the fitted model when predicting the unused sub-sample (test data). This is repeated over all five sub-samples (for the same model specification) and the resulting average MSE measures the model's out-of-sample prediction accuracy. We next repeat this procedure 500 times to obtain stable results.

Figure 3a displays the AIC measures from within-sample predictions of all 5,151 possible model specifications considered in the 100m step length setting, where the horizontal plane represents the full set of all possible combinations of the knots' locations, and the vertical axis denotes the corresponding AIC values. The preferred model specification (that gives the lowest AIC) has the first knot at 0.9km and the second knot at 1km. In comparison, Figure 3b shows that the out-of-sample cross validation procedure favours a slightly different model specification, which has the first knot at 0.7km and the second knot at 1.2km. These two findings, albeit not identical, both provide much more evident and unambiguous evidence as to the spatial scope of the hedonic valuation (as indicated by the location of the second knot), compared to the results from the step function approach. The preference between the two model selection procedures is largely a case specific decision, depending on whether the priority is to best explain a particular dataset or to obtain higher out-of-

[9] We have opted for the AIC instead of the mean squared error (MSE) for within-sample model assessment and selection. Adding regressors (through, for example, increasing the number of knots) always reduces the MSE of within-sample prediction. Therefore attempting to select a model specification that minimises the MSE of within-sample prediction would lead to overfitting.

sample predictive power. In this study, we prioritise out-of-sample performance and therefore focus on the model specification recommended by the out-of-sample cross validation procedure. That said, it is reassuring to see that the two procedures suggest considerably similar breakpoint distances where the hedonic price of green spaces disappears (1km and 1.2km, respectively).

As can be seen in Model 1 in Table 1, the estimated parameters (γ_1 and γ_2) that capture the distance-dependent patterns of the hedonic price estimates between the two knots are both strongly significant (p-value < 0.001). The hedonic price estimates (and their confidence intervals) at different distances can be recovered from the regression spline, as shown in Figure 4a. In monetary terms, the highest estimate appears in the 300m–400m ring (CNY 54.19 or USD 8.16 per m^2 of housing per ha of green spaces) and the lowest occurs in the 0m–100m ring (CNY –42.24 or USD –6.36). The area weighted average hedonic price within 1.2km is estimated to be CNY 18.39 or USD 2.77, and is highly significant (p-value < 0.01, standard error estimated using the delta method). For green spaces outside the threshold distance, the hedonic price estimate (α) is statistically insignificant and considerably small in size (less than 2% of the weighted average hedonic price estimate within 1.2km). These estimates characterise the hedonic price curve presented in Figure 4a, which has an inverted U-shape within 1.2km 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 100m step length (Figure 2a).

The aggregate hedonic value of green spaces created after 2006 inside 1.2km amounts to CNY 125 billion (~USD 19 billion) in 2016 prices, which is obtained using the same prediction-based method described in the Section 3.1. This aggregate value, which refers to green spaces in the entire circular area within a 1.2km radius, is unsurprisingly much higher than the highest estimate given by the step function approach, since the latter only concerns green spaces in a few segments of the 1.2km circle that have a statistically significant hedonic price estimate in the step function model.

Our findings on the preferred model specification, hedonic price estimates and the aggregate hedonic value are reasonably stable if we switch to a 200m step length.

More specifically, Figure 3c shows that the within-sample model selection procedure using a 200m step length prefers a threshold distance at 1km, which is identical to that derived using a 100m step length. In addition, as suggested in Figure 3d, the out-of-sample cross validation procedure also favours the same threshold distance (1.2km) as in the 100m step length setting. If we still focus on the threshold distance with the best out-of-sample prediction accuracy (1.2km), the monetary hedonic price estimates inside 1.2km range from CNY –10.28 or USD –1.55 (for the 0m–200m ring) to CNY 53.13 or USD 8.00 (for the 200m–400m ring) per m^2 of housing per ha of green spaces, and average out to CNY 14.96 or USD 2.25. The hedonic price estimate beyond 1.2km remains small and statistically insignificant. Overall, the hedonic price curve (Figure 4b) hardly differs from that derived using a 100m step length (Figure 4a). The aggregate hedonic value of green spaces created after 2006 within 1km (CNY 112 billion or USD 17 billion in 2016 prices) is somewhat lower than that in the 100m step length setting (CNY 125 billion or USD 19 billion), but the difference is much smaller than that from the step function approach, in both absolute and relative terms.

More importantly, a palatable property of the regression spline approach is that the same model selection criteria can be applied to model assessment across different step length settings. Intuitively, the regression spline approach should be less amenable to coarsened step lengths that cannot capture detailed variation in green spaces' hedonic price along distance. In an extreme case, if we use a 5km step length, the only identifiable threshold distance would be 5km and we would not have the scope to explore any other possible thresholds. In our case, a 500m or 1km step length would already considerably limit the possible thresholds that can be explored. As shown in Figure A2 in the appendix, the preferred model specification in a 500m or 1km step length setting has a threshold distance in an extremely far away location (e.g. 8km–9km), which is less explicable in the contexts of Chinese cities, where residents primarily rely on neighbourhood gardens in close proximity for leisure activities (Chen & Jim, 2011). That said, all possible model specifications in the 200m, 500m and 1km step length settings are outperformed by the preferred specification in the 100m setting step length, in terms of both within- and out-of-sample prediction accuracy. This has inclined us to lean towards the findings derived from a 100m step length and regard a 1.2km radius as the spatial scope of the hedonic valuation.

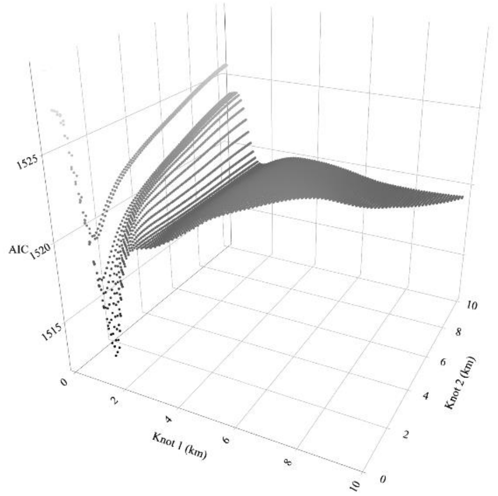
3.3. Instrumental Variable Estimates

We next formally explore the implications of potential endogeneity bias for our hedonic price estimates. Admittedly, despite the extensive number of control variables in our hedonic price model, it is difficult to explicitly account for all determinants of house prices. If omitted factors are associated with both house prices and green spaces, our estimates might be biased. For instance, a greener residential block may intentionally target the high-end housing market and is thus likely to have other desirable features (such as stylish interior design) that may come with a price premium yet are not controlled for in our hedonic price model, which would impart an upward bias to the estimated hedonic prices of green spaces.

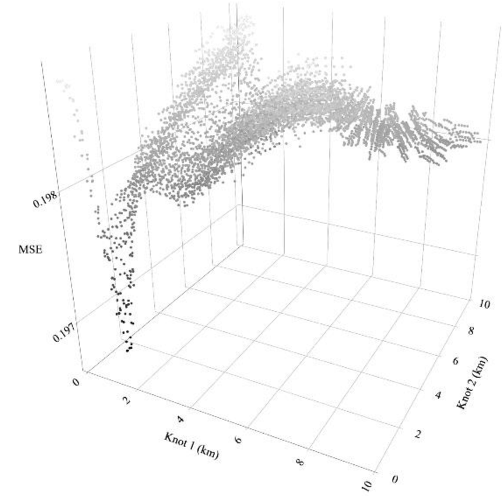
We attempt to assess potential endogeneity bias via instrumenting the area of green spaces. Now we are more inclined to believe that green spaces further than 1.2km away from a property would not directly influence its price, but tend to be highly correlated with green spaces closer to the property. In other words, green spaces beyond 1.2km only affect the dependent variable via its correlation with the suspected endogenous regressor (green spaces within 1.2km), and thus qualify as instruments. Bayer et al. (2009) adopted a similar 'spatial lag' type of instrument for air quality. Unfortunately, we cannot directly instrument green spaces in this manner in the step function or regression spline estimation, since all green spaces within a 10km radius have already entered the hedonic pricing model as explanatory variables to search for the threshold distance.

Therefore, to facilitate the instrumental variable (IV) estimation, we reestimated the hedonic pricing model (Eq.1) using only one green space variable that represents the total area of all green spaces within a 1.2km radius. The OLS estimate of the hedonic price of these green spaces (the estimate for the 'Green area 0m–1.2km' variable in Model 2 in Table 1) is smaller than that given by the regression spline approach (Model 1 in Table 1), but remains statistically significant at the 5% level. In Model 3, the 'Green area 0m–1.2km' variable is instrumented using the area of green spaces in the first three 1km wide rings outside 1.2km (‘Green area 1.2km–2.2km’, ‘Green area 2.2km–3.2km’ and ‘Green area 3.2km–4.2km’).[10] The *F* statistic from the weak IV test, which is markedly greater than the frequently invoked rule of thumb (10), provides substantive 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 cannot reject the null hypothesis of no endogeneity bias at the conventional critical level (0.10), the IV estimate for ‘Green area 0m–1.2km’ is notably higher than the OLS estimates in Models 1 and 2, which suggests that the OLS estimates have likely underestimated the true hedonic value of green spaces due to endogeneity bias. We reestimated Model 3 using various subsets of the three excluded instruments, and the results are practically identical.

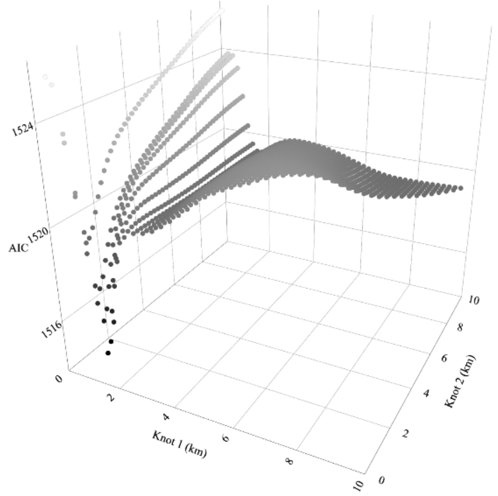
[10] We have opted for these instruments because green space variables for further rings do not have sufficient explanatory power for the area of green spaces inside 1.2km.



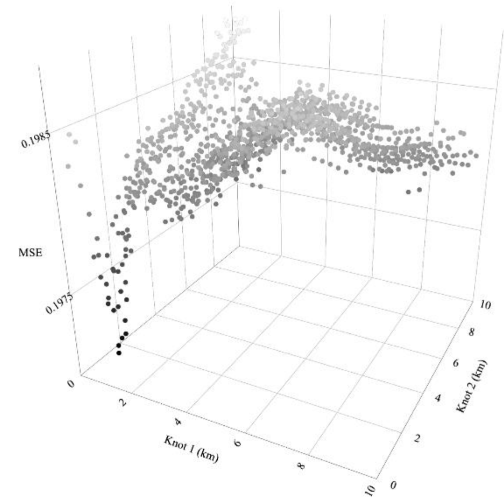
(a) Within-sample, 100m step length



(b) Out-of-sample, 100m step length

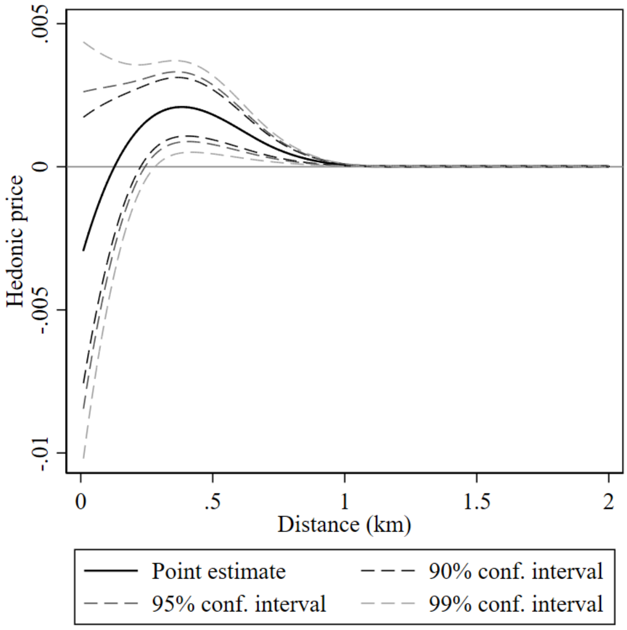


(c) Within-sample, 200m step length

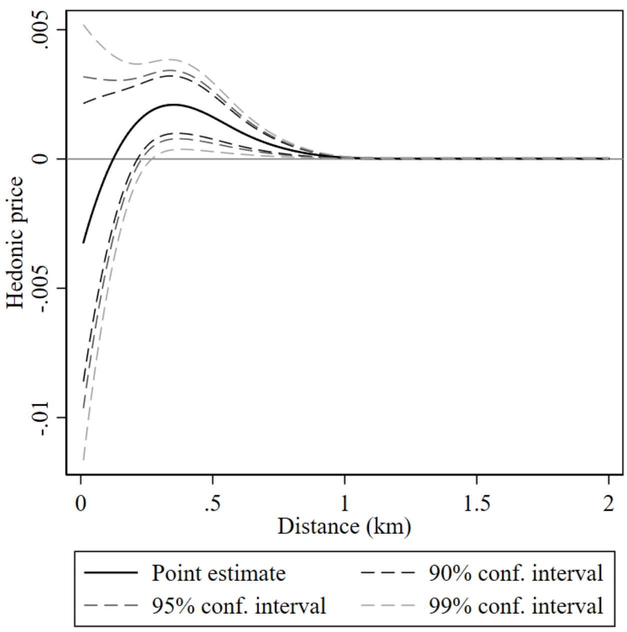


(d) Out-of-sample, 200m step length

Figure 3 Predictive performance of regression spline models (100m and 200m step lengths)



(a) 100m step length



(b) 200m step length

Figure 4 Hedonic price estimates at different distances: Regression spline estimates

Note: Figures 4a and 4b focus on the segment of the hedonic price curve within 2km, since the hedonic price estimates no longer change based on distance outside 1.2km.

Table 1 Regression spline estimates (100m step length)

DV: *Log (House price)*	**Model 1**	**Model 2**	**Model 3**
	OLS	OLS	IV–2SLS
Hedonic price:			
Green area 0m–1.2km	***7.03×10^{-4}******	***4.82×10^{-4}*****	***9.28×10^{-4}*****
	(2.07×10^{-4})	***(2.02×10^{-4})***	***(3.93×10^{-4})***
Regression spline parameters:			
γ_1	***-3.99×10^{-11}******		
	(1.53×10^{-11})		
γ_2	***6.14×10^{-12}******		
	(1.84×10^{-12})		
α	1.32×10^{-5}		
	(9.99×10^{-6})		
Control variables (Table A1)	Yes	Yes	Yes
Excluded instruments:			
Green area 1.2km–2.2km	No	No	Yes
Green area 2.2km–3.2km	No	No	Yes
Green area 3.2km–4.2km	No	No	Yes
Weak IV test			
Cragg-Donald Wald *F* statistic			137.16
(H_0: Weak IV)			
R^2 (1st stage)			0.60
Over-identification test			
Sargan statistic: *p*-value			0.20
(H_0: Valid IV)			
IV redundancy test			
LM test: *p*-value			0.00
(H_0: Redundant IV)			
Endogeneity test			
Diff-in-Sargan-Hansen Statistic: *p*-value			0.19
(H_0: Exogenous 'green area 0m–1.2km')			
Hedonic price model sig.: *p*-value	0.00	0.00	0.00
Hedonic price model R^2	0.68	0.68	0.68
Obs.	1,270	1,270	1,270

Note: * *p*-value < 0.10, ** *p*-value < 0.05, *** *p*-value < 0.01. Standard errors are in parentheses.

3.4. Matching Estimates

Lastly, we further tested the robustness of the aforementioned hedonic price estimates using a novel matching approach proposed by Fong et al. (2018). Matching has been advocated by the causal econometric literature (e.g. Greenstone & Gayer 2009, Imbens & Rubin 2015 and Imbens & Wooldridge 2009) as a means to better control for endogeneity issues and thereby strengthen an estimate's causal inference. Fong et al. (2018) built upon conventional matching methods by, on the one hand, accommodating nonbinary treatment variables, and on the other, directly optimising sample covariate balance through 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. The CBGPS matching approach is well suited for this study, because we have a nonbinary 'treatment' variable, the area of green spaces.

We reestimated the hedonic price of the area of green spaces within a 1.2km radius using both the parametric and non-parametric CBGPS methods. The parametric method assumes the generalised propensity score to be normally distributed. In contrast, the non-parametric method does not depend on any assumptions about the functional form of the generalised propensity score. Following Fong et al. (2018), we first identified the optimal Box-Cox transformation of the variable 'Green area 0m–1.2km' 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.[11] We then performed the matching algorithm on the covariates listed in Table A2. The weights derived from the matching algorithm were then utilised to estimate a regression of house prices on the transformed green space variable and the three sets of fixed effects listed in Table A1 (to approximate within-cluster matching). Finally, the estimate on the transformed green space variable was utilised to compute the hedonic price estimate in the semi-elasticity form as in Table 1. The entire procedure was bootstrapped 500 times to derive the standard error and confidence interval.

[11] We adopted this transformation for both the parametric and non-parametric CBGPS estimation to ensure the estimates' comparability, despite that the non-parametric CBGPS does not involve any distributional assumptions for the generalised propensity score.

Table A2 in the Appendix present 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, and 19 covariates has a coefficient below 0.10. 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.

Table 2 reports the hedonic price estimates derived from the two matching procedures. It can be seen in the first column of results that the parametric CBGPS matching gave a hedonic price estimate which highly resembles the IV estimate, in terms of both the magnitude and statistical significance. The second column of results shows that the hedonic price estimate from the non-parametric matching procedure becomes slightly higher than that the IV estimate, but remains qualitatively comparable.[12] These findings lend further support to the robustness of our results against potential endogeneity bias.

[12] We assessed the implications of controlling for the covariates in the post-matching regressions. The hedonic price estimates become smaller (parametric CBGPS matching: 5.34×10^{-4}; non-parametric CBGPS matching: 9.26×10^{-4}), but remain statistically significant and qualitatively comparable to the estimates in Tables 1 and 2.

Table 2 Matching estimates (100m step length)

DV: *Log (House price)*	**Model 4** Parametric CBGPS	**Model 5** Non-parametric CBGPS
Hedonic price:		
Green area 0m–1.2km	***9.26×10^{-4}******	***11.16×10^{-4}*****
	(2.68×10^{-4})	***(5.07×10^{-4})***
	[4×10^{-4}, 14×10^{-4}]	***[2×10^{-4}, 22×10^{-4}]***

Note: * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01. Standard errors are in parentheses. 95% confidence intervals are in brackets.

4. Aggregate Hedonic Prices of Green Spaces in Central Beijing

This section provides a demonstration of deriving aggregate hedonic prices using our estimation results. To reduce calculation 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 1.2km radius. The aggregate hedonic price was then calculated through multiplying the unit hedonic price (the coefficient on the variable 'Green area 0m–1.2km' in Models 1–5 after being converted from a semi-elasticity estimate to a 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 1.2km radius. Figure 5 maps the aggregate hedonic prices of these green spaces individually (based on the unit hedonic price estimate in Model 3). 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 (new builds in this study). Still, these results provide instrumental information that can be directly fed into cost-benefit analysis for removing or creating a green space. Table 3 reports the total hedonic price estimates of all these green spaces. These estimates, although only concerning a subset of Beijing's green spaces in our dataset, are already considerably sizeable: the annual average of the aggregate hedonic price is comparable to 1% of Beijing's GDP in 2016 (CNY 2,566.91 billion or USD 386.58 billion).

Figure 5 Hedonic prices (billion CNY in 2016 prices) of green spaces in central Beijing

Table 3 Unit and total hedonic prices of green spaces

		Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Unit hedonic price	CNY	18.37	12.60	24.25	24.20	29.17
(per ha of green space per m² of floor area)	USD	2.77	1.90	3.65	3.64	4.39
Aggregate hedonic price	CNY bln	192.56	132.02	254.18	253.64	305.68
(all green spaces > 0.5ha in central Beijing)	USD bln	29.00	19.88	38.28	38.20	46.04

5. Discussion and Conclusion

There have been extensive studies that valued urban green amenities via measuring the ensuing house value premiums (or hedonic prices). Nevertheless, previous studies have mostly focused on valuing the *proximity* to and *views* of green spaces. We argue in this paper that an aggregable valuation of the *size* (or *area*) of green spaces would better capture the overall benefits of a particular green space, and therefore would be more applicable to urban land use decision making and ecosystem services accounting. Yet this practice is subject to difficulties in identifying the spatial limits of the capitalisation of green spaces in house prices. This study explored a spatial piecewise approach as a means to overcome such difficulties. We regress house prices on a series of green space variables representing the area of green spaces at different distances from a property, in an attempt to discern the spatial limit where the house value premium associated with green spaces is just about to disappear. Moreover, we propose a novel regression spline approach, which reparametrises the hedonic price model by approximating hedonic prices at different distances as a piecewise polynomial function of distance. We next loop over all possible threshold distances in search of the preferred threshold distance, using both within-sample and out-of-sample predictions for model assessment and selection. These techniques allow us to derive the spatial scope of the hedonic valuation based on formal statistical justification. Our hedonic price estimates are robust to an instrumental variable estimator and a novel matching algorithm which better account for potential endogeneity issues.

In communicating our findings to policymakers, however, we draw attention to a number of caveats. First and foremost, green spaces provide many valuable ecosystem services that are not capitalised in housing values and hence cannot be captured by hedonic prices. For instance, in 2016 Beijing accommodated 285 million tourists (close to the US population), who spent CNY 502 billion (~USD 76 billion) in the city (Beijing Municipal Bureau of Statistics, 2018). A large proportion of them may have visited the city's world-renowned 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 spaces has materialised in the form of attracting tourists and contributing to tourism revenues, instead of residential housing value premiums. Therefore, the hedonic value may only partly represent green spaces' total value. Second, our dataset only concerns housing newly built during 2006–2016. The aggregate value of green spaces would be much more sizeable had we taken into account the associated price premiums of all housing in the city. Lastly, we would caution against literally extrapolating the results to other cases. This study seeks to undertake a methodological exploration and demonstration. The specific results (such as the threshold distance) may not be directly transferrable to other contexts. For example, for a less populated city with a more open layout, the spatial scope of hedonic valuation may well exceed the 1.2km radius. That said, the internal validity of our results is considerably robust to a number of alternative estimation methods. This can be taken as an indication of the robustness and reliability of the estimation procedure we proposed, which constitutes the primary contribution of this study.

References

Albouy, D., Christensen, P., & Sarmiento-Barbieri, I. (2020). Unlocking amenities: Estimating public good complementarity. *Journal of Public Economics*, 182, 104110. https://doi.org/10.1016/j.jpubeco.2019.104110

Bayer, P., Keohane, N., & Timmins, C. (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, *58*(1), 1–14. https://doi.org/10.1016/j.jeem.2008.08.004

Beijing Municipal Bureau of Statistics. (2018). *Beijing statistical yearbook 2018*. Beijing: China Statistics Press.

Beijing Municipal Education Commission. (2015). *Beijing Education Map*. Beijing: Beijing Evening News.

Bishop, K. C., Kuminoff, N. V, Banzhaf, H. S., Boyle, K. J., von Gravenitz, K., Pope, J. C., … Timmins, C. D. (2020). Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality. *Review of Environmental Economics and Policy*, *14*(2), 260–281. https://doi.org/10.1093/reep/reaa001

Black, K. J. (2018). Wide open spaces: Estimating the willingness to pay for adjacent preserved open space. *Regional Science and Urban Economics*, 71:110–121. https://doi.org/10.1016/j.regsciurbeco.2018.06.001

Bockarjova, M., Botzen, W. J. W., van Schiec, M. H., & Koetseb, M. J. (2020). Property price effects of green interventions in cities: A meta-analysis and implications for gentrification. *Environmental Science & Policy*, 112:293–304. https://doi.org/10.1016/j.envsci.2020.06.024

Brander, L. M., & Koetse, M. J. (2011). The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. *Journal of Environmental Management*, *92*(10), 2763–2773. https://doi.org/10.1016/j.jenvman.2011.06.019

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. College Station, Tex.: Stata Press.

Cavailhès, J., Brossard, T., Foltête, J.-C., Hilal, M., Joly, D., Tourneux, F.-P., … Wavresky, P. (2009). GIS-Based Hedonic Pricing of Landscape. *Environmental and Resource Economics*, *44*(4), 571–590. https://doi.org/10.1007/s10640-009-9302-8

Chen, W. Y., & Jim, C. Y. (2011). Resident valuation and expectation of the urban greening project in Zhuhai, China. *Journal of Environmental Planning and Management*, *54*(7), 851–869. https://doi.org/10.1080/09640568.2010.537552

Cho, S.-H., Poudyal, N. C., & Roberts, R. K. (2008). Spatial analysis of the amenity value of green open space. *Ecological Economics*, *66*(2), 403–416. https://doi.org/10.1016/j.ecolecon.2007.10.012

Conway, D., Li, C. Q., Wolch, J., Kahle, C., & Jerrett, M. (2010). A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values. *The Journal of Real Estate Finance and Economics*, *41*(2), 150–169. https://doi.org/10.1007/s11146-008-9159-6

Czembrowski, P., & Kronenberg, J. (2016). Hedonic pricing and different urban green space types and sizes: Insights into the discussion on valuing ecosystem services. *Landscape and Urban Planning*, *146*, 11–19.

https://doi.org/10.1016/j.landurbplan.2015.10.005
Daams, M. N., Sijtsma, F. J., & van der Vlist, A. J. (2016). The Effect of Natural Space on Nearby Property Prices: Accounting for Perceived Attractiveness. *Land Economics*, *92*(3), 389–410. https://doi.org/10.3368/le.92.3.389
Dong, G., & Wu, W. (2016). Schools, land markets and spatial effects. *Land Use Policy*, *59*, 366–374. https://doi.org/10.1016/j.landusepol.2016.09.015
Ferraro, P. J., & Shukla, P. (2020). Feature—Is there a Replicability Crisis on the Horizon for Environmental and Resource Economics? *Review of Environmental Economics and Policy*, 14:339–351. https://doi.org/10.1093/reep/reaa011
Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, 12(1), 156–177. https://doi.org/10.1214/17-AOAS1101
Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Upper Saddle River: Prentice Hall.
Greenstone, M., & Gayer, T. (2009). Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management*, 57(1), 21–44. https://doi.org/10.1016/j.jeem.2008.02.004
Haninger, K., Ma, L., & Timmins, C. (2017). The Value of Brownfield Remediation. *Journal of the Association of Environmental and Resource Economists*, *4*(1), 197–241. https://doi.org/10.1086/689743
Hastie, T., Tibshirani, R., & Friedman, J. H. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751
Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1), 5–86. https://doi.org/10.1257/jel.47.1.5
Irwin, E. G. (2002). The Effects of Open Space on Residential Property Values. *Land Economics* , *78*(4), 465–480. https://doi.org/10.2307/3146847
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R*. London: Springer.
Jardine, S. L., & Siikamäki, J. V. (2014). A global predictive model of carbon in mangrove soils. *Environmental Research Letters*, *9*(10), 104013. https://doi.org/10.1088/1748-9326/9/10/104013
Kovacs, K., West, G., Nowak, D. J., & Haight, R. G. (2022). Tree cover and property values in the united states: A national meta-analysis. *Ecological Economics*, 197:107424. https://doi.org/10.1016/j.ecolecon.2022.107424
Kuminoff, N. V, Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, *60*(3), 145–160. https://doi.org/10.1016/j.jeem.2010.06.001
Li, S., Yang, J., Qin, P., & Chonabayashi, S. (2016). Wheels of fortune: Subway expansion and property values in Beijing. *Journal of Regional Science*, *56*(5), 792–813. https://doi.org/10.1111/jors.12284
Mansfield, C., Pattanayak, S. K., McDow, W., McDonald, R., & Halpin, P. (2005). Shades of Green: Measuring the value of urban forests in the housing market. *Journal of Forest Economics*, *11*(3), 177–199. https://doi.org/10.1016/j.jfe.2005.08.002
Mei, Y., Gao, L., & Zhang, P. (2019). Residential property price differentials of waste

plants: evidence from beijing, china. *Applied Economics*, 51:5952–5960. https://doi.org/10.1080/00036846.2019.1640858

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. *Journal of Environmental Economics and Management*, 110:102566. https://doi.org/10.1016/j.jeem.2021.102566

Melichar, J., & Kaprová, K. (2013). Revealing preferences of Prague's homebuyers toward greenery amenities: The empirical evidence of distance–size effect. *Landscape and Urban Planning*, *109*(1), 56–66. https://doi.org/10.1016/j.landurbplan.2012.09.003

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Washington, DC: Island Press.

Muehlenbachs, L., Spiller, E., & Timmins, C. (2015). The Housing Market Impacts of Shale Gas Development. *American Economic Review*, 105(12), 3633–3659. https://doi.org/10.1257/aer.20140079

National Development and Reform Committee of China, & Ministry of Housing and Urban-Rural Development of China. (2006). *Methods and parameters for the economic evaluation for construction projects (in Chinese)*. Beijing: China Planning Press.

Netusil, N. R., Chattopadhyay, S., & Kovacs, K. F. (2010). Estimating the Demand for Tree Canopy: A Second-Stage Hedonic Price Analysis in Portland, Oregon. *Land Economics*, 86(2), 281–293. Retrieved from http://www.jstor.org/stable/27821424

Netusil, N. R., Levin, Z., Shandas, V., & Hart, T. (2014). Valuing green infrastructure in Portland, Oregon. *Landscape and Urban Planning*, *124*, 14–21. https://doi.org/10.1016/j.landurbplan.2014.01.002

Orsini, N., & Greenland, S. (2011). A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal*, *11*(1), 1–29. https://doi.org/10.1177/1536867X1101100101

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. *Environmental and Resource Economics*, *57*(2), 251–272. https://doi.org/10.1007/s10640-013-9665-8

Sander, H., Polasky, S., & Haight, R. G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics*, *69*(8), 1646–1656. https://doi.org/10.1016/j.ecolecon.2010.03.011

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Suits, D. B., Mason, A., & Chan, L. (1978). Spline Functions Fitted by Standard Regression Methods. *The Review of Economics and Statistics*, *60*(1), 132–139. https://doi.org/10.2307/1924341

Tanaka, S., & Zabel, J. (2018). Valuing nuclear energy risk: Evidence from the impact of the Fukushima crisis on U.S. house prices. *Journal of Environmental Economics and Management*, *88*, 411–426. https://doi.org/10.1016/j.jeem.2017.12.005

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Walls, M., Kousky, C., & Chu, Z. (2015). Is What You See What You Get? The Value of Natural Landscape Views. *Land Economics*, *91*(1), 1–19. https://doi.org/10.3368/le.91.1.1

Waltert, F., & Schläpfer, F. (2010). Landscape amenities and local development: A review of migration, regional economic and hedonic pricing studies. *Ecological Economics*, *70*(2), 141–152. https://doi.org/10.1016/j.ecolecon.2010.09.031

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < 0.05." *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wegman, E. J., & Wright, I. W. (1983). Splines in Statistics. *Journal of the American Statistical Association*, *78*(382), 351–365. https://doi.org/10.1080/01621459.1983.10477977

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Mason, Ohio: South-Western Cengage Learning.

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. https://doi.org/10.1016/j.landusepol.2022.106048

Wu, W., & Dong, G. (2014). Valuing the "green" amenities in a spatial context. *Journal of Regional Science*, *54*(4), 569–585. https://doi.org/10.1111/jors.12099

Zhang, L., Wu, J., Liu, H., & Zhang, X. (2020). The value of going green in the hotel industry: Evidence from beijing. *Real Estate Economics*, 48:174–199. https://doi.org/10.1111/1540-6229.12225

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. *The Journal of Real Estate Finance and Economics*, *53*(2), 184–199. https://doi.org/10.1007/s11146-015-9513-4

Zheng, S., & Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics*, *63*(2), 743–757. https://doi.org/10.1016/j.jue.2007.04.010

APPENDICES

Table A1 Description of variables

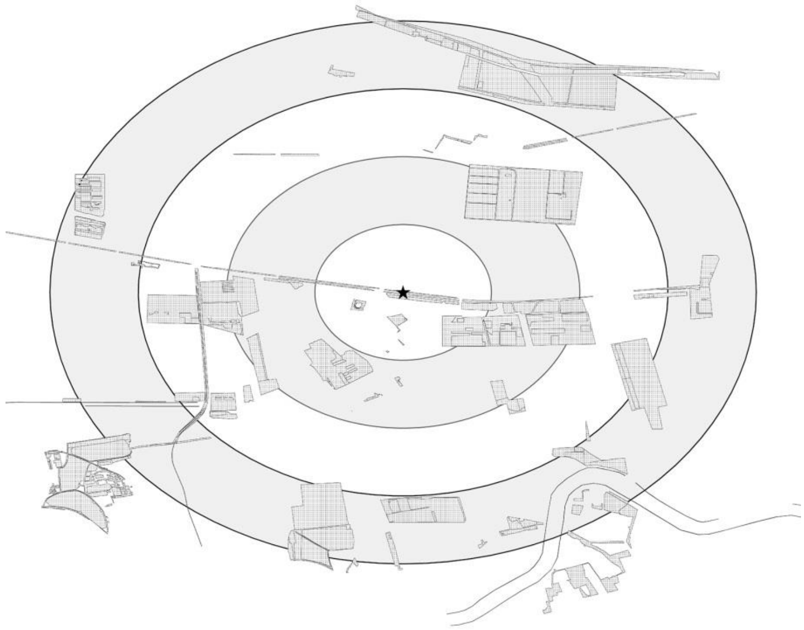
Variable	Definition	Mean	SD	Min	Max
Dependent variable:					
House price	Average house prices at the residential block level (CNY/m^2).	26,136.07	21,859.21	1,000	170,000
Explanatory variable:					
Green area 0–1km	Total area (ha) of green spaces 0–1km.	86.46	61.46	0.00	735.81
Green area 1–2km	Total area (ha) of green spaces 1–2km.	265.56	154.48	0.00	1,408.76
Green area 2–3km	Total area (ha) of green spaces 2–3km.	418.19	231.90	0.00	1,307.77
Green area 3–4km	Total area (ha) of green spaces 3–4km.	563.95	305.91	0.00	1,338.65
Green area 4–5km	Total area (ha) of green spaces 4–5km.	689.36	395.52	0.00	1,793.35
Green area 5–6km	Total area (ha) of green spaces 5–6km.	812.97	462.34	0.00	2,480.88
Green area 6–7km	Total area (ha) of green spaces 6–7km.	912.21	539.15	0.00	2,874.51
Green area 7–8km	Total area (ha) of green spaces 7–8km.	1,021.27	609.64	0.00	3,132.88
Green area 8–9km	Total area (ha) of green spaces 8–9km.	1,120.00	678.24	0.00	3,436.32
Green area 9–10km	Total area (ha) of green spaces 9–10km.	1,228.31	745.53	0.00	3,330.72
House	Binary: 1 = Detached or semi-detached houses; 0 = Flats.	0.11	0.32	0	1
High-rise	Binary: 1 = High-rise; 0 = Otherwise.	0.15	0.36	0	1
Decoration status	Binary: 1 = Fully decorated; 0 = Otherwise.	0.31	0.46	0	1
School	Number of 'demonstration' kindergartens, primary schools and middle schools.	5.26	4.96	0	17
Distance to city centre	Distance (m) to Tiananmen Square.	38,598.73	31,062.49	621.53	149,208.90
Southern half of Beijing	Binary: 1 = To the south of Tiananmen Square; 0 = Otherwise.	0.43	0.50	0	1
Distance to hospital	Distance (m) to the nearest 3A hospital.	12,157.54	18,106.05	44.88	93,955.29
Distance to railway	Distance (m) to the nearest railway.	3,108.72	3,394.63	0.31	32,167.34
Distance to	Distance (m) to the nearest highway.	697.43	952.00	0.33	19,781.63

highway					
Distance to road	Distance (m) to the nearest road (aside from highways).	286.29	877.43	0.03	19,790.42
Distance to tube	Distance (m) to the nearest tube station.	11,262.11	18,689.13	25.42	99,998.54
Distance to bus stop	Distance (m) to the nearest bus stop.	485.32	719.47	2.16	18,176.29
Restaurants 0–1km	Number of restaurants within 0–1km.	31.89	50.47	0	403
Restaurants 1–3km	Number of restaurants within 1–3km.	236.00	310.14	0	1,638
Restaurants 3–5km	Number of restaurants within 3–5km.	417.92	536.08	0	2,817
Restaurants 5–7.5km	Number of restaurants within 5–7km.	741.72	898.93	0	3,951
Restaurants 7.5–10km	Number of restaurants within 7.5–10km.	959.84	1,095.65	0	4,448
Shops 0–1km	Number of shops within 0–1km.	47.52	71.88	0	539
Shops 1–3km	Number of shops within 1–3km.	354.63	404.14	0	1,922
Shops 3–5km	Number of shops within 3–5km.	623.70	711.40	0	3,280
Shops 5–7.5km	Number of shops within 5–7.5km.	1,098.08	1,213.23	0	4,954
Shops 7.5–10km	Number of shops within 7–10km.	1,418.90	1,536.63	0	6,020
District fixed effects	15 dummies indicating 16 districts.			0	1
Ring road fixed effects	5 dummies indicating 6 zones partitioned by 5 ring roads.			0	1
Year fixed effects	10 dummies indicating 11 years (2006–2016).			0	1

Note: CNY 6.64 = USD 1 in 2016. 3A hospitals are the highest-quality hospitals in China according to the classification system of the country's Ministry of Health.

Table A2 Pearson correlation coefficients between each covariate and the green space variable (0m–1.2km) before and after matching

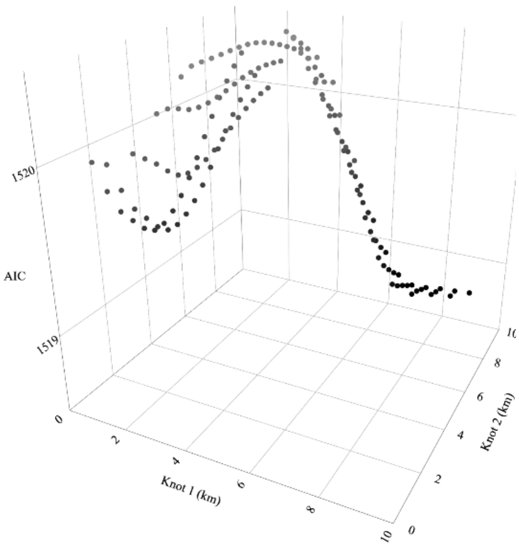
Covariate	Unmatched	Parametric CBGPS		Non-parametric CBGPS	
	Correlation coefficient	Correlation coefficient	Reduction (%) in absolute value	Correlation coefficient	Reduction (%) in absolute value
House	–0.22	–0.04	80.37	–0.04	83.84
High-rise	0.14	0.03	82.41	–0.02	86.80
Decoration status	0.19	0.04	79.35	–0.03	85.06
School	0.29	0.03	88.83	0.04	84.85
Distance to city centre	–0.37	–0.08	78.97	0.03	90.76
Southern half of Beijing	0.03	–0.01	53.60	–0.06	–96.72
Distance to hospital	–0.31	–0.10	67.39	0.06	80.67
Distance to railway	–0.25	–0.06	74.71	0.00	98.55
Distance to highway	–0.15	–0.07	55.50	–0.02	89.24
Distance to road	–0.19	–0.12	38.76	–0.06	68.10
Distance to tube	–0.32	–0.09	71.91	–0.01	97.23
Distance to bus stop	–0.29	–0.11	61.29	–0.07	76.61
Restaurants 0–1km	0.37	0.01	96.90	0.06	84.53
Restaurants 1–3km	0.42	0.04	90.20	0.02	94.21
Restaurants 3–5km	0.40	0.03	91.90	0.01	97.00
Restaurants 5–7.5km	0.37	0.03	93.06	0.00	99.35
Restaurants 7.5–10km	0.36	0.03	90.98	–0.02	93.89
Shops 0–1km	0.37	0.02	95.79	0.05	87.41
Shops 1–3km	0.43	0.04	90.14	0.02	96.09
Shops 3–5km	0.40	0.02	93.77	0.01	96.81
Shops 5–7.5km	0.35	0.02	94.49	0.00	99.20
Shops 7.5–10km	0.35	0.03	92.87	–0.01	97.05

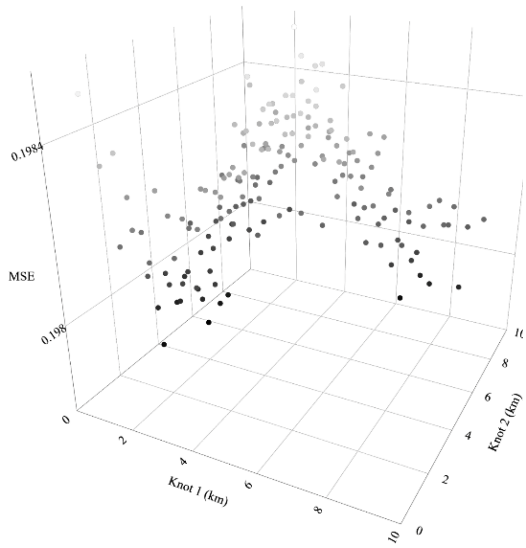


★ Centroid of residential block

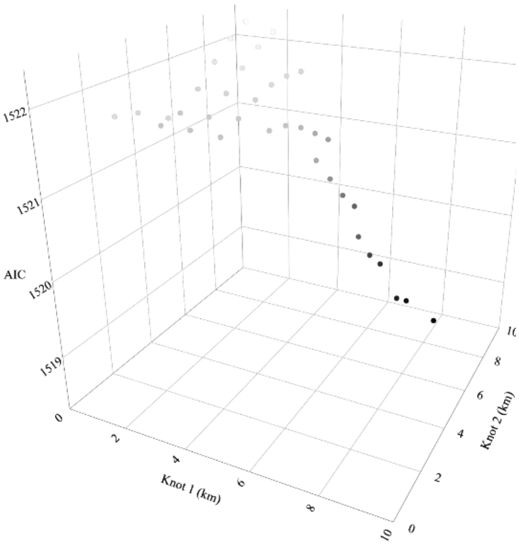
Greenspace Ring

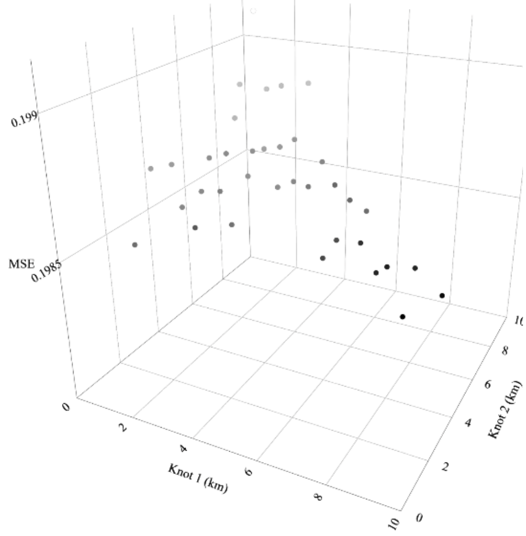
Figure A1 Schematic diagram of a residential block and its surrounding green spaces





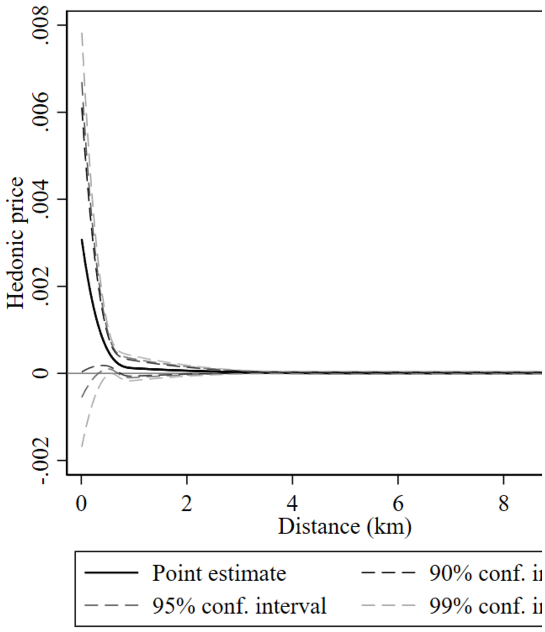
(a) Within-sample, 500m step length **(b) Out-of-sample, 500m step length**



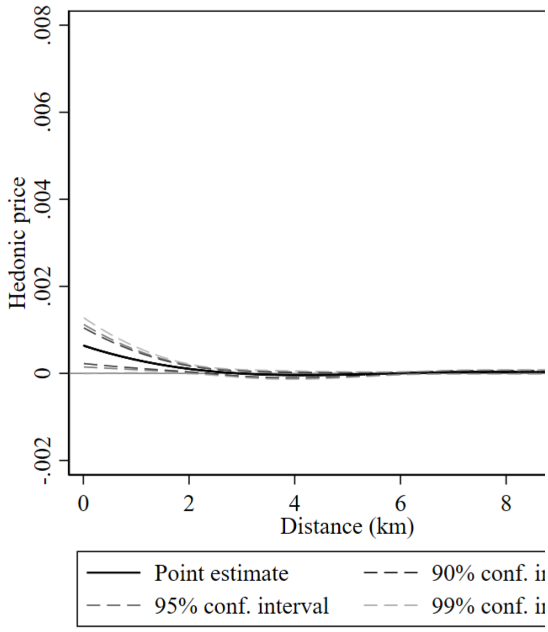


(c) Within-sample, 1km step length **(d) Out-of-sample, 1km step length**

Figure A2 Predictive performance of regression spline models (500m and 1km step lengths)



(a) 500m step length



(b) 1km step length

Figure A3 Regression spline estimates (500m and 1km step lengths)