Value of nature implicit in property prices – Hedonic Pricing Method (HPM) methodology note

This note explains the methodology used to estimate the value of green and blue space implicit in property prices in urban areas in Great Britain.

Table of contents

1 . Introduction

This note explains the methodology used to estimate the value of green and blue space implicit in property prices in urban areas in Great Britain in the initial urban ecosystem accounts published on 12 July 2018.

2 . Strengths and limitations of using Hedonic Pricing Method

The Hedonic Pricing Method (HPM) relies on the assumption that a class of differentiated products can be broken down in to a number of characteristics. A combination of these characteristics and the external factors that affect the product determines the price. The most common example of this is property values, where the market price of a property is determined by a combination of structural characteristics (floor area, number of bedrooms, garden, garage and so on) and the socio-economic and environmental characteristics of the surrounding area (quality of schools, access to retail, transport, levels of water/air pollution, proximity to green space and so on)

HPM can be used to estimate the extent to which characteristics affect price by modelling property prices as a set of explanatory variables, including the structural, socio-economic and environmental characteristics. Assuming that nature is implicit in property prices, this methodology can then be used to extract values for environmental goods or services from market-based transactions by including measures of access to natural areas in the regression model.

HPM returns market-based marginal prices of ecosystem services comparable to those used for standard market goods (Day, 2013), which is important for consistency with the national accounts.

Further, data on property transactions and characteristics are readily available and it potentially offers a means of estimating the value of everyday recreational visits to local green spaces that do not require admission fees or the use of public or private transport. Such trips are currently excluded by the expenditure-based method currently used in the UK Natural Capital accounts.

A limitation of HPM is that to obtain accurate and robust estimates large datasets with very detailed information about property characteristics and the surrounding environment are needed. Omitting key determinants of property prices that are correlated with availability of natural amenities may result in biased estimates and the results are sensitive to model specification, therefore care is needed when building the models.

We cannot be certain that increasing the amount of green and blue space would result in an increase in property prices. To interpret the coefficients causally, we need to assume that our models capture all variables that influence both property prices and areas of blue and green spaces. Whilst this assumption cannot be directly tested, the high R-squared (our models explain over 80% of the variance in property prices) suggests that we include many factors relevant to property prices in our models, leaving less room for unobserved heterogeneity.

The results demonstrate it is possible to apply HPM for national statistics. However, further work is needed, in particular interpreting regression results into annual flow and stock values.

3 . Which ecosystem services could have been captured?

It can be difficult to determine which ecosystem services are captured through the Hedonic Pricing Method (HPM) . It could be posited that the buyers of properties must be aware of the services provided by natural capital for those services to be reflected in property prices. We work on the assumption that the majority of the value captured is that from cultural services, such as recreation and attractive views, rather than regulating services such as carbon sequestration and temperature regulation which people are less likely to be aware of.

Table 1 summarizes the types of services which could be captured and the potential overlap with other ecosystem services already included in the accounts.

Table 1: Ecosystem services potentially captured by the Hedonic Pricing Method and likelihood of overlap existing service estimates

Ecosystem service	Likelihood overlap with existing service estimates
Physical health benefits	Dependent upon how health benefits and the associated service are defined and measured. Needs additional investigation
Recreation	HPM can potentially capture the additional value of local recreational trips not incurring cash expenses, which are not currently included in monetary accounts. There would need to be consistency between the average distance of local walked trips and the radius used in HPM.
Urban cooling	No overlap. The methodology for temperature regulation involves averaging across urban areas and focuses on savings in output and energy costs.
Carbon sequestration	No overlap
Provisioning services, such as food and material production	No overlap. Property prices will be sensitive to the size of a domestic garden but not necessarily specific uses of it such as garden produce. Gardens are excluded from the hedonic results.
Air quality regulation	Potential for overlap, although air filtration accounting estimates are not modelled at the same localised scale as HPM. Whilst air pollution levels do affect property prices, we are interested in the role of air filtration by vegetation, which is more marginal. Needs additional investigation.
Noise regulation	Potential for overlap, needs additional investigation. Whilst noise pollution clearly affects property prices, we are interested in the role of vegetation in reducing noise, which is more marginal.

Source: Office for National Statistics

4 . Data sources

A large amount of data is needed to conduct hedonic regression analysis, and for the purposes of the UK natural capital accounts it also needs to be available on a national scale. The following datasets are used:

ACORN classification: A data set which provides a well-established geographic segmentation of the UK. Produced and licensed by CACI Ltd it segments UK neighbourhoods and postcodes in to six categories, eighteen groups and sixty-two types by analysing significant behavioural and social factors. For more information see the ACORN user guide on "The consumer classification".

ACORN is currently a key determinant in the hedonic regression and is also used in the production of the ONS Land and Property Services House Price Index and the current ONS House Price Index.

Zoopla: A UK based property website. The dataset was provided by Zoopla Limited, © 2018 to the Urban Big Data Centre (UBDC) collection and includes information on over a million properties sold in Great Britain between 2009 and 2016. Information includes location, number of bedrooms, number of receptions rooms, property type, for sale or rent, asking price, sale price and so on. The Zoopla data also provide a description of the property which we use to extract additional characteristics of the property for example, whether it has a garage or not, has been recently renovated. The description is also used to fill in missing information about property type.

Ordnance Survey: We have been fortunate to be able to collaborate with the Ordnance Survey (OS) who have created a wide range of variables for the purposes of the HPM. These variables have been derived through the geospatial analysis of multiple OS datasets, both open data and premium data available through the Public-Sector Mapping Agreement (PSMA), as well as other third-party datasets.

Datasets used include the following.

OS AddressBase Plus

A vector address point dataset for Great Britain that contains current properties and addresses sourced from local authorities identified using a Unique Property Reference Number (UPRN), matched to a series of third-party datasets including:

- Royal Mail's Postcode Address File (PAF)

- Valuation Office Agency's council tax and non-domestic rates

- OS large scale vector mapping

OS MasterMap Topography Layer

A vector dataset providing individual real-world topographic features in Great Britain represented by point, linestring, and polygon geometries, captured at 1:1250, 1:2500 and 1:10 000 scales in urban, rural and mountain /moorland areas respectively. For buildings represented using polygon geometries in the Topographic Area feature type, a series of building height properties are available through a complimentary OS MasterMap Building Height Attribute dataset.

OS MasterMap Highways Network:

A vector dataset providing a topologically structured road and path network which has been heighted and is attributed with associated information including classification, road name(s) and numbering. It brings together geographic information from OS and third-party sources such as the National Street Gazetteer (NSG) to create a single authoritative view of the road and path network.

OS Open Greenspace

A vector polygon dataset providing the location and extent of green spaces such as parks and gardens, playing fields and sports facilities that are likely to be accessible to the public. Where appropriate, it also includes access point geometries to show how people get into these sites.

We link Zoopla data to OS data using the Unique Property Reference Number (UPRN). We include in our sample residential properties listed for sale between 2009 and 2016. After removing outliers (defined as properties with price lying in the top or bottom 0.1% of the price distribution), we have a sample of 2,634,013 properties.

5 . Model specification

The specific form of the regression we estimate is described along with a more detailed analysis of the structural, neighbourhood, environmental and socio-economic variables that we use.

The hedonic pricing method doesn't have a pre-defined functional form. Rosen (1974) however, suggests that there are many reasons to believe that the relationship between property price and the environmental variable to be non-linear in nature. Non-linearity is expected as "purchasers cannot treat individual housing attributes as discrete items from which they can pick and mix until the desired combination of characteristics is found" (Kong page 242 (2007)). This has led to many previous studies either using semi-logarithmic or double-logarithmic models as the log transformation generates the desired linearity in parameters.

Taking this in to account in addition to the mitigation of high variation in property prices via taking the log of property prices we have chosen to employ a semi-logarithmic model.

The semi-log model takes the following form:

$$\ln\left(\text{price}_{i,t}\right) = \text{env}_i \beta + hc_{i,t} Y + n_i \delta + year_{i,t} \lambda + \varepsilon_i$$

Where the dependent variable ln (price $_{i,t}$) is the log of sale price for each property transaction "i" at time "t". env$_i$ is a vector of environmental characteristics, and in our main specification includes area of natural cover, blue spaces and functional green spaces within a 200 metre radius of the property [1].

The vector hc $_{i,t}$ contains structural characteristic variables, which have been shown in previous studies to have the greatest influence on property price. It includes usual housing attributes such as number of bedrooms, property and garden area (square feet) and property type. It also includes a set of attributes retrieved from the description, such as period of the house (for example, Georgian, Victorian, Edwardian), and features that are expected to influence property prices (for example, garage, presence of original features, whether the property has been renovated recently).

The neighbourhood and geographical variables in vector n$_i$ include distance to amenities other than green and blue spaces such as transport infrastructures (for example, bus station, train station and so on), retail area, and workplace centroid. The socio-economic characteristics of the local area are captured by including dummies for each local area (MSOAs[2]) and for the socio-economic "type" of neighbourhood based on a well-established socio-economic segmentation tool of the UK derived by ACORN. This segments postcodes by analysing significant social factors and population characteristics, such as country of birth, family structure, health and so on.

Our primary focus is on the estimation of , a vector of coefficients indicating how areas of green and blue spaces are associated with the log of house prices. The coefficients can be interpreted as semi-elasticities and represent the percentage change in property prices associated with the presence of green space, natural land cover or bluespace within a certain radius of the property and its size.

Whilst we include a range of neighbourhood characteristics in addition to local areas fixed effects, there are likely unobserved variables that affect property prices. For instance, the presence of an outstanding school may increase prices of all properties in a neighbourhood. Omitting important determinants that affect all properties in a given neighbourhood would generate spatial autocorrelation, that is the residuals for properties in the same area would be correlated. The spatial autocorrelation would threaten traditional inference on the parameters , as the p-value assume independence. In addition, if these unobserved factors affecting price are also correlated with the area of green and blue space, the estimated coefficients would be biased. The direction of the bias would depend on the direction of the correlation between unobserved determinants of prices and area of green and blue space.

To mitigate spatial dependence, we include local area (MSOAs) fixed-effects in our models, and cluster standard errors at the MSOA level. Therefore, we rely on the variation within MSOA to estimate the relationship between area of green and blue spaces and property prices.

Several models are estimated where additional explanatory variables are introduced progressively to assess the robustness of our results. For example, the first model we estimate will only include environmental variables. The independent variables included in the regression analysis are summarised in Table 2.

Table 2: Variables for hedonic regression

Characteristic Vector	Component Variables	Sources
Structural Characteristics	No. of bedrooms	Zoopla
	Property area (sq. feet)	Ordnance Survey
	Property type: house, bungalow, flat	
	Property attributes based on description	
Neighbourhood Characteristics	Distance to train station	Ordnance Survey Office for National Statistics
	Distance to local labour market	
	Distance to nearest transport infrastructure	
	Distance to nearest retail cluster	
	TTWA	
	MSOA	
Socio-economic characteristics	ACORN	CACI
Environmental Amenities	Distance to green space	
	Distance to blue space	
	Area of Natural Features in radius of property (100m, 200m, 500m)	Ordnance Survey
	Area of functional green space in radius of property (100m,200m,500m)	
	Function of green space	
	Area of Residential Garden	

Source: Office for National Statistics

6 . Regression results

Table 3 shows the results for the regression analysis using four different model specifications, each one adding more control variables to the model. We only display the coefficients and standard errors for the explanatory variables of interest, the "area of publicly accessible green space within 200 metres", the area of blue space within 200 metres" and the "area of natural land cover within 200 metres". 200 metres has been chosen as the appropriate distance for the radius around the property in which to measure the area as the average distance to publicly accessible green space is 257 metres.

Table 3: Property prices and areas of green and blue spaces within 200 meters

	Log(Price)			
	(1)	**(2)**	**(3)**	**(4)**
Small functional green space within 200m	-0.0409***	0.0059***	0.0066***	0.0053***
	(0.0024)	(0.0012)	(0.0012)	(0.0010)
Medium functional green space within 200m	-0.0440***	0.0059***	0.0069***	0.0060***
	(0.0026)	(0.0013)	(0.0013)	(0.0011)
Large functinoal green space within 200m	-0.0335***	0.0116***	0.0127***	0.0107***
	(0.0028)	(0.0014)	(0.0014)	(0.0011)
Very large functional green space within 200m	-0.0102***	0.0161***	0.0170***	0.0145***
	(0.0034)	(0.0017)	(0.0017)	(0.0014)
Small blue space within 200m	0.0269***	0.0097***	0.0091***	0.0093***
	(0.0029)	(0.0015)	(0.0015)	(0.0012)
Medium blue space within 200m	0.0262***	0.0104***	0.0099***	0.0102***
	(0.003)	(0.0016)	(0.0016)	(0.0013)
Large blue space within 200m	0.0113***	0.0107***	0.0106***	0.0126***
	(0.0034)	(0.0017)	(0.0017)	(0.0014)
Very large blue space within 200m	-0.0069	0.0269***	0.0269***	0.0358***
	(0.0045)	(0.0025)	(0.0025)	(0.0023)
Small area of natural land cover within 200m	-0.1127***	-0.0373***	-0.0365***	-0.0335***
	(0.0176)	(0.0139)	(0.0137)	(0.0093)
Medium area of natural land cover within 200m	-0.1171***	-0.0464***	-0.0471***	-0.0411***
	(0.0177)	(0.0139)	(0.0138)	(0.0094)
Large area of natural land cover within 200m	-0.1157***	-0.0489***	-0.0498***	-0.0418***
	(0.0178)	(0.0139)	(0.0138)	(0.0094)
Very large area of natural land cover within 200m	-0.1077***	-0.0494***	-0.0506***	-0.0439***
	(0.0179)	(0.014)	(0.0138)	(0.0094)
MSOA	Yes	Yes	Yes	Yes
Acorn	No	Yes	Yes	Yes
Distance to other amenities	No	No	Yes	Yes
House characteristics	No	No	No	Yes
Observations	2,628,660	2,628,660	2,628,660	2,626,103
R2	0.3667	0.7199	0.7231	0.8048
Adjusted R2	0.3666	0.7199	0.723	0.8047

Residual Std. Error	0.5010 (df = 2628423)	0.3332 (df = 2628361)	0.3313 (df = 2628344)	0.2781 (df = 2625781)

Notes:

* p less than 0.1.

** p less than 0.05.

*** p less than 0.01.

The coefficients can be interpreted as the elasticity of price with respect to the presence and area of green and blue spaces. The price elasticity indicates by how much property prices changes (in percent) when one of the categorial variables, for example, a small functional green space within 200 metres, is present in comparison to if it wasn't. For example, a coefficient of 0.01 for "small FGS within 200 metres" would indicate that its presence is associated with a 1% increase in property price compared to if there was no functional green space within 200 metres.

Model 4 further adds house characteristics (number of bedrooms, garage, building area, area of residential garden and so on) and is our preferred specification as it includes the wider set of covariates. All coefficients for the different size categories for the FGS within 200 metres of property are positive and significant. The table shows that the presence of a small FGS within 200 metres is associated with a rise in property price of 0.5%. The presence of a very large FGS within 200 metres is associated with a rise in property price of 1.5%.

7 . Valuation of monetary stock

Based on the analysis presented we build a model to predict house prices based on structural and environmental variables. We split our dataset into a training and a test dataset, and use the training dataset to estimate the model and the test dataset to assess the predictive power of the model using the Root Mean Square Error (RMSE) as a metric of predictive performance. We compute the RMSE of three different model specifications. The first specification includes all the variables used in our preferred specification (see earlier "Model specification" section) except information on green and blue spaces. The second model adds areas of blue and green spaces within a 200 metre radius. The third model further includes areas of blue and green spaces within 100, 200 and 500 metres of the properties. We find that the third model has the lowest RMSE.

Using this model, the predicted average property prices in our test data is £246,010. The predicted average property price in the absence of green and blue spaces is £241,197 which is £4,813 lower that the predicted average price using the real data. We conclude that in the absence of green and blue spaces property prices in Great Britain would be £4813 lower and this reflects the value of services provided by green and blue spaces.

By multiplying this figure by the number of residential properties (27.2 million), we obtain an estimate of £130.9 billion for the stock value of blue and green spaces. It must be noted however that this estimate relies on the assumption that the sample used is representative of the property stock, which may not be the case. For example, small starter homes sell much more frequently than other types of property and therefore will be over-represented in the data. It also assumes that the value implicit in property price is the same in Northern Ireland.

Notes for: Valuation of monetary stock

1. We randomly allocate 20% of our data to the training dataset and 80% to the test dataset.