GLAECONOMICS

Working Paper 3: Valuing Greenness

Is there a segmented preference for housing attributes in London?

by **Adarsh Varma**

June 2003

Transport for London

MAYOR OF LONDON

LONDON DEVELOPMENT AGENCY

copyright

Greater London Authority
June 2003

Published by
Greater London Authority
City Hall
The Queen's Walk
London SE1 2AA
www.london.gov.uk
enquiries 020 7983 4100
minicom 020 7983 4458

This publication is printed on recycled paper

GLA Economics is funded by the London Development Agency (LDA), Transport for London (TfL) and the Greater London Authority (GLA) and was established by the Mayor of London in May 2002. Its purpose is to provide a firm statistical, factual and forecasting basis for policy decision-making by the GLA group.

Contents

Abstract

This working paper sets out two methodologies used to value green spaces in London. The analysis provides the basis for a new report by GLA Economics, *Valuing Greenness: Green spaces, house prices and Londoners' priorities* (GLA, 2003).

Two forms of hedonic pricing methodology have been used:

- Model 1: A simple semi-log regression model using different housing attributes to explain the variation in house prices.
- Model 2: A pooled semi-log regression model using dummy variables to check for segmented preferences.

We have based the approach used in *Valuing Greenness* on model 1, which uses house values to reflect the spatial variation in 'greenness' and other key indicators in London wards (1998 boundaries). We have tried to infer whether the indicators have a statistically significant impact on house prices. We have also tried to isolate the significance of green spaces in explaining the variation in house prices. Model 2 attempts to show the rent price differential due to open green spaces between two regions of London.

This working paper also accounts for some of the measurement errors and methodological problems associated with this kind of analysis.

Introduction

The importance of green spaces has become a central theme for policy makers. Green spaces not only provide a pleasant and natural environment but also improve the quality of life in urban areas. They ensure sustainable use of housing supplies, jobs and infrastructure in place and planned.

Green spaces play a vital role in the lives of Londoners by promoting healthy living, as a source of education, preserving heritage and culture, providing recreation and tourism, environmental sustainability and community development.

This working paper contains the technical econometric analysis of the results in *Valuing Greenness* (GLA, 2003). The report uses open green spaces along with socio-economic indicators at ward level to explain the variation in house prices.

1. Methods for valuing non-market goods

Environmental goods have a number of different attributes, such as recreation, water and air quality, landscape, biodiversity and amenity values. Moreover, the methods to measure these attributes are also different in the literature on valuing environmental goods and services (Garrod and Willis, 1999). Several of these measurement methods are explained below.

1.1 Expressed preference methods

Contingent valuation methodology
Contingent valuation methodology is a survey-based technique where individuals are asked their maximum willingness to pay for an environmental good or benefit. For example, what is the maximum amount visitors would be willing to pay for accessing a green spaces site for recreational and educational purposes, rather than go without it. The non-use values of public goods such as wilderness, landscape preservation and biodiversity can be measured using this technique. Contingent valuation methodologies value the good in its entirety.

Choice experiment
Economists use choice experiments to determine individuals' preferences for the attributes of a good or service. This is achieved within a questionnaire framework where respondents are asked about various hypothetical alternatives offering different levels of those attributes. If one of these attributes is price, then the respondents' willingness to pay for the other attributes can often be inferred. Most choice experiments involve some form of survey where respondents are asked to make choices based on their preferences for a series of profiles. The defining characteristics of a good are further questioned to obtain their individual values. For example, distinction can be made between the aesthetic and prevention value of a landscape.

Both contingent valuation methodologies and choice experiments, however, fail to capture non-use values of biodiversity. This is mainly because people have widely different preferences for wildlife, so the variance of the mean willingness to pay value is large. Moreover, people may be ignorant of the main benefits and importance of biodiversity. A focus group approach can be used, where groups of participants are given the chance to learn about biodiversity in green spaces before being asked to express their preferences.

1.2 Revealed preference methods

The 'price' of amenities for which markets do not exist such as green spaces can be inferred from observing and analysing the price of goods for which markets do exist such as houses.

Travel cost method
The minimum willingness to pay to enjoy the services of non-market good such as green spaces can be measured by estimating the costs of travelling to them, in addition to any other costs incurred in consuming these services. On the basis of questionnaire survey, data is collected on the number of visits and cost of accessing the site.

Hedonic pricing method
Hedonic pricing methods are based on the consumer theory proposition that the marginal value of a good is based on a wide range of attributes the good possesses. In our analysis, the purchase price of a house is determined by local socio-economic characteristics such as housing densities, accessibility to transport and health services and local amenities such as green spaces. In *Valuing Greenness*, a hedonic pricing method is used due to the availability of data and because it is less time consuming (GLA, 2003). Geographic information system (GIS) data is used to complement available statistical data. The aim of both the models in this paper is to take account of a number of factors, ranging from accessibility to public services, socio-economic and environmental conditions that affect house prices.

2. Variables

The purchase price which a potential buyer is ready to pay for a house is based on a number of housing characteristics, ranging from structural characteristics (e.g. number of rooms, garage space and plot size) to local socio-economic and public sector characteristics (e.g. quality of schools, health services and the local unemployment rate). Other important criteria include local environmental quality, transport links and access to other services. This study used the indicators in Table 1 to explain the variation in house prices.

Table 1 Indicators used in this study to explain variation in house prices

Green spaces†	
1. Green spaces	In this study, green spaces refer to total identifiable 'strategic green spaces' (km^2) for each ward. The identifiable green spaces are the Green Belt, Metropolitan Open Land, Sites of Metropolitan Importance, Sites of Borough Importance and Sites of Local Importance. This is divided by the total area of the ward and expressed as a percentage. Green spaces such as urban parks, private gardens and common green spaces around flats are excluded from this study, except in the Green Belt, as data are not available. *Source: Connecting with London's Nature: The Mayor's Draft Biodiversity Strategy, 2002*
Housing	
2. Overcrowding*	Percentage of households living at densities of 1.00 or more persons per room. *Source: 1991 Census (estimated to 1998 ward boundaries)*
3. Dwelling density	Total dwellings for each ward divided by the ward area, expresses as number of dwellings per km^2. *Source: Valuation Office Agency, 2001*
Deprivation	
4. Income support (IS)*	Income support claimants as a percentage of population over the age of 18 for each ward. *Source: Department for Work and Pensions, 1998*
Education	
5. SATs 2 scores*	Standard Achievement Targets 2 scores. Pupils scoring at less than Level 4 as a proportion of total pupils aged 10. Data are for 1998 and refer to school addresses in the absence of pupil addresses. This means that, in the absence of some heroic form of modelling to attribute pupil addresses to schools, we can only attribute school performance to the ward in which a school is located. Values for schools have therefore been attributed to the wards in which the schools are located (and aggregated across schools where there is more than a single school in a ward). Where there is no school in a ward, the ward has been attributed the average value for all schools in the borough. This is clearly very crude, given the size and complexity of school catchment areas even at primary-school level. It does, nevertheless, reflect something of the areas in which the schools operate (London Index of Deprivation, 2002). *Source: Department for Education and Skills, 1998*

Crime	
6. Domestic burglaries*	Domestic burglaries as a per cent of adult population (18 years+). The dataset was originally compiled with grid references and the number of offences. The grid references and their values were plotted and attributed to wards. The most common reported crime is domestic burglaries. *Source: MPS, (1999/2000).*
Travel accessibility	
7. Travel times†	Travel time zones to central London have been averaged for each ward. Central London is defined as roughly the same as zone 1 of the underground map. Transport for London divides London into 1,019 travel zones. The following modelling periods have been used: morning (07:00-09:59), interpeak (10:00-15:59) and evening peak (16:00-18:59). *Source: Transport for London, 2001*
Health accessibility	
8. Health†	Postcode level data for hospitals, NHS trust sites, dentists and GPs are summed and then mapped to obtain a ward level health indicator. *Source: London Health Observatory, 2002*
Environment	
9. NO_2 average*	Levels of nitrogen dioxide in parts per billion (ppb). The data are derived from mapping of NO_2 concentrations in London. There are a large number of air quality monitoring sites around London, which give valuable information on pollution at specific sites. The continuous surface map is modelled with the use of data on emissions of air pollutants together with weather data and geographical information to calculate the likely pollution concentrations. *Source: South East Institute of Public Health, 1999*
Dummy variable	
10. High affluent	Wards with average house prices greater than £500,000 located within Underground zone 1. This indicator is included to avoid the data being skewed because of large deviation from higher average house prices.

Notes:
† The headline indicator for each domain in the London Index of Deprivation (London Index) is used, as this allows a convenient and useful interpretation of the index. The headline indicator is the one with the greatest conceptual link to the domain and the one with the highest intra-domain correlations. Its use to represent a domain avoids the inclusion of noise from low-scoring indicators included in the multivariate assemblies used by IMD2000. The transformation used in the London Index is readily understandable and transparent, and avoids the impact of negative scores. No weighting is used for domains. (GLA, 2002)
* Computed by using Mapinfo GIS system.

3. Model 1

Model 1 is based on a semi-log regression equation, which is used in *Valuing Greenness* (GLA, 2003).

$$\ln HP = \alpha_0 + \beta_i \sum_{i=1}^{10} X_i + e_i \qquad \ldots 1$$

where HP is average house prices in wards (ONS, 2001)

α_0 is the intercept

β_i are the coefficients of the indicators or a vector of unknown parameters to be estimated

X_i are the main indicators where:
X_1 is Travel time to central London
X_2 is Income support
X_3 is Dwelling density
X_4 is Per cent green
X_5 is SATs (education indicator)
X_6 is Domestic burglary (crime indicator)
X_7 is High affluent dummy variable
X_8 is Over crowded households
X_9 is NO_2 average (environment indicator)
X_{10} is Health facilities indicator
e_i is the error term

We ran the regressions for 760 wards (based on 1998 ward boundaries) where the City of London was grouped as one ward. The spatial patterns of the dependent and independent variables can be found in Chapter 2 of *Valuing Greenness* (GLA, 2003).

3.1 Regression results

Table 2 Model summary

R	R^2	Adjusted R^2	Standard error of the estimate
.811	.658	.653	.262

R (0.811), the multiple correlation coefficient, shows a strong correlation between the observed and predicted values of the dependent variable (home sale price).

R^2 (0.65), the coefficient of determination, is the squared value of the multiple correlation coefficient. It shows that about two-thirds of the variation in home sale prices is explained by model 1.

Table 3 Analysis of variance

Model	Sum of squares	df	Mean square	F	Sig.
Regression	98.736	10	9.874	143.826	.000
Residual	51.419	749	.069		
Total	150.155	759			

The significance value of the F statistic is less than 0.05 and highly significant. The null hypothesis that home sale price is not linearly related to all the explanatory variables can be rejected. The F statistic is the regression mean square (MSR) divided by the residual mean square (MSE). The regression sum of squares (RSS) 98.7 is nearly twice as large as the error sum of squares (ESS) 51.4, indicating model 1 accounts for most of the variation in mean home sale prices (dependent variable).

Table 4 Coefficients

	Unstandardised coefficients		Standardised coefficients	t	Sig.	95% confidence interval for B		Correlations			Collinearity statistics	
Model	**B**	**Standard Error**	**Beta**			**Lower bound**	**Upper bound**	**Zero-order**	**Partial**	**Part**	**Tole-rance**	**VIF**
(Constant)	11.992	.173		69.190	.00	11.651	12.332					
Dwelling density	4.890E-05	.000	.226	6.627	.00	.000	.000	.326	.235	.142	.394	2.540
Income support	-3.671E-02	.003	-.471	-13.094	.00	-.042	-.031	-.324	-.432	-.280	.353	2.830
Education (SAT)	-5.432E-03	.001	-.168	-6.066	.00	-.007	-.004	-.392	-.216	-.130	.600	1.668
Overcrowding	-2.026E-02	.005	-.127	-4.159	.00	-.030	-.011	-.128	-.150	-.089	.488	2.050
Domestic burglaries	9.541E-03	.002	.163	6.070	.00	.006	.013	.188	.217	.130	.638	1.569
Travel time	-1.210E-02	.002	-.338	-7.491	.00	-.015	-.009	-.448	-.264	-.160	.224	4.460
Per cent green	4.140E-03	.001	.191	7.042	.00	.003	.005	-.050	.249	.151	.619	1.616
Health facility	5.079E-03	.002	.058	2.592	.01	.001	.009	.165	.094	.055	.928	1.077
NO_2 avg (ppb)	3.616E-02	.006	.256	6.265	.00	.025	.047	.463	.223	.134	.274	3.652
High affluent dummy	.428	.079	.125	5.403	.00	.272	.583	.390	.194	.116	.857	1.167

The explanatory variables all have the expected signs and significant t-statistics at 95 per cent confidence interval. One can argue that average emission levels should be negatively related with house prices. However, in the case of London, travel time to central London and high demand for houses in central London are key drivers of house prices. The high density of population, higher levels of economic activity and transport emissions are the reasons why we see a positive relationship between emission levels and house prices. The standardised coefficients enable us to compare the relative importance of the explanatory variables. If we converted the data to z-scores before we ran the regression, we would have obtained beta as our unstandardised coefficients. Larger absolute standard coefficient values indicate greater contribution in explaining the model. For example, even though overcrowding has a larger unstandardised coefficient than per cent green, it does not contribute as much as per cent green to the model.

Ranked explanatory power of each explanatory variable on the basis absolute standardised coefficients:

1. Income support
2. Travel time
3. NO2 average
4. Dwelling density
5. Per cent green
6. Standard Achievement Targets (SATs)
7. Domestic burglaries
8. Over crowded households
9. High affluent dummy variable
10. Health facilities indicator

The results above suggest that an absolute change in each of the indicators can explain a relative change in house prices (see Appendix 3). The top nine indicators are significant in explaining the variation in house prices.

The significance of the top five indicators in quantitative terms is explained below. Due to the inherent simple nature of the model the results, though indicative, may result in a downward bias in the regression coefficients.

Our study has found that on average one per cent increase in the amount of green space in a ward can be associated with a 0.3 to 0.5 per cent[1] increase in the average house price in that ward.

A 1 per cent fall in the proportion of income support claimants is associated with a 3.1 to 4.2 per cent higher average house prices. Income support claimants are predominantly located in areas with high proportions of council flats, such as the boroughs of Newham, Tower Hamlets and Hackney, which may depress house prices. This may also affect prices of other residential property in the vicinity as these areas may be seen with a low preference for housing.

---

[1] Using 95 per cent confidence interval for coefficient of percentage of green spaces (given in Table 4).

On average, a one-minute reduction in travel time to central London is associated with a 1 to 1.5 per cent increase in average house price.

On average, dwelling density indicates the level of demand for houses in any region, which is reflected by house prices. Historically the borough of Hackney has dense housing and due to their poor quality is relatively cheap now. From our analysis, a unit increase in dwellings per square kilometre ward area can relate to an increase in average house prices by 0.05 per cent.

Education performance of schools also has some consideration for housing preferences. A 1 per cent decrease in the number of pupils scoring less than Level 4 as a proportion of total pupils (aged 10) can on average be associated with a 0.4 to 0.7 per cent increase in house prices.

Mulitcollinearity

The tolerance[2] (Table 4) is the percentage of the variance in a given predictor that cannot be explained by the other predictors. When the tolerances are close to 0, there is high multicollinearity[3] and the standard error of the regression coefficients will be inflated. The tolerances are significantly far from zero in most cases, the small tolerances for income support, dwelling density and travel time show that 70 per cent of the variance in a given predictor can be explained by the other predictors where as the larger ones such as per cent green, high affluent dummy indicate that their variances cannot be explained by other predictors. A variance inflation factor (VIF)[4] greater than 2 is usually considered problematic but greater than 3 indicates serious multicollinearity problem. As a rule of thumb, one often uses the frontier of R (the multiple correlation coefficient, 0.8). If $R \geq 0.9$ then multicollinearity may be considered 'harmful', if $R \leq 0.9$ then it is not harmful. Further collinearity diagnostics are given in Appendix 2, Table A2.

Model 1 also reveals a plausible correlation between areas with characteristically a higher percentage of green spaces and lower levels of deprivation, as expressed through a negative correlation between percentage of green spaces and income support. In addition, a positive correlation between dwelling density and income support shows that the more crowded areas of London, despite higher housing prices also exhibit high levels of poverty. The correlation coefficients and scatter matrix are shown in Appendix 2, Tables A4 and A5.

[2] A statistic used to determine how much the independent variables are linearly related to one another (multicollinear). The proportion of a variable's variance not accounted for by other independent variables in the equation. A variable with very low tolerance contributes little information to a model, and can cause computational problems. It is calculated as 1 minus R squared for an independent variable when it is predicted by the other independent variables already included in the analysis.

[3] Multicollinearity is the undesirable situation where the correlations among the independent variables are strong, i.e. the effect of explanatory variables is strong and it becomes difficult to isolate the significance of any one indicator.

[4] The reciprocal of the tolerance. As the variance inflation factor increases, so does the variance of the regression coefficient, making it an unstable estimate. Large VIF values are an indicator of multicollinearity.

3.2 Criticism

The regression model suffers from measurement problem related to errors in the observed values of the dependent and explanatory variables (see Chapter 5). For example, dwelling density has been calculated to include green spaces in the area of a ward and is included twice. In model 2, dwelling density has been calculated on the area of the ward less the area of 'green space' in the ward, not on the total areas of each ward. It should be noted that in both models dwelling density is also an inverse measure of the quantity of green space not included within the 'green space' indicator. This could account for those green spaces as mentioned in Table 1 and not included in per cent green, assuming that a small proportion of dwelling density is also inversely related to other land uses, such as commercial and industrial uses. Thus, dwelling density and green space are a measure of green space in both models.

4. Model 2

The results of model 1 led us to believe there may be an inner-outer dichotomy for London with some outliers. There may be a case where the population has segmented preferences for housing. Housing preferences for proximity to central London over shadow the need for a cleaner greener environment for one set of people. The other group values a more suburban and green environment even though commuting times to central London may be long. Hence, the analysis revealed a set of wards with low travel times to central London, high dwelling density, low green spaces and high house prices. On the other hand, there are also wards with high house prices but with high travel times, high green spaces and low dwelling density.

In model 2, we attempt to estimate the house price differential between two segmented parts of London. To test for segmentation and the significance of green spaces in outer London we use the following hypothesis:

> To see whether house prices-green space relationship in ward with travel times to central London greater than 35 minutes is statistically significant, two regression equations are pooled – one containing wards with travel times less than 35 minutes and the other greater than 35 minutes into one equation using a dummy variable[5].

$$\ln HP = \alpha_0 + \alpha_1 D + \beta_i \sum_{i=1}^{7} X_i + \beta_{PG} Pcntgreen + \beta_D (D.Pcntgreen) + e_i \qquad \ldots 2$$

Where:

α_0 is the intercept term for wards with travel time less than 35 minutes.

$\alpha_{`1}$ is the differential intercept coefficient showing us how much the value of the intercept term for wards with travel time greater than 35 minutes differs from the intercept of wards with travel time less than 35 minutes.

HP, β_i and X_i are the same as equation 1. Overcrowding and NO_2 average are omitted to reduce multicollinearity.

β_{PG} is the coefficient for per cent green.

β_D is the differential slope coefficient indicating by how much the slope coefficient of the hedonic function, where travel time is greater than 35 minutes differs from the slope coefficient of the hedonic function where it is less than 35 minutes.

D is dummy variable with value 1 for wards with travel time to central London greater than 35 minutes.

[5] A map showing wards with travel time to central London greater than 35 minutes is given in Appendix 1.

The two equations can be derived as follows

$$\ln HP_n = \alpha_0 + \beta_i \sum_{i=1}^{7} X_i + \beta_{PG} Pcntgreen \qquad \ldots 3$$

where n is the number of wards with travel times to central London less than 35 minutes.

$$\ln HP_m = (\alpha_0 + \alpha_1) + \beta_i \sum_{i=1}^{7} X_i + (\beta_{PG} + \beta_D) Pcntgreen \qquad \ldots 4$$

where m is the number of wards with travel to central London greater than 35 minutes.

Multiplying the dummy variable of D with per cent green in equation 2 helps us to differentiate between the slope coefficients of the two groups (Gujarti, 1978).

4.1 Regression results

Table 5 Model summary

R	R Square	Adjusted R Square	Standard error of the estimate
.797	.636	.631	.270

R and R Square are more or less the same as for Model 1.

Table 6 Analysis of variance

	Sum of Squares	df	Mean Square	F	Sig.
Regression	95.483	10	9.548	130.810	.000
Residual	54.672	749	.073		
Total	150.155	759			

The F statistic is high in explaining the overall significance of the estimated regression at 10 degrees of freedom. Model 2, as it stands, does not really represent an improvement on the original as RSS increases from 51.4 to 54.7.

Table 7 Coefficients

	Unstandardised coefficients		Standardised coefficients	t	Sig.	95% confidence interval for b		Correlations			Collinearity statistics	
Model	**B**	**Std. Error**	**Beta**			**Lower bound**	**Upper bound**	**Zero-order**	**Partial**	**Part**	**Tolerance**	**VIF**
(Constant)	12.985	.081		159.854	.000	12.826	13.145					
Dwelling density	4.632E-05	.000	.214	5.755	.000	.000	.000	.326	.206	.127	.352	2.843
Income support	-4.130E-02	.003	-.530	-15.945	.000	-.046	-.036	-.324	-.503	-.352	.440	2.272
Education (SAT)	-5.800E-03	.001	-.179	-6.323	.000	-.008	-.004	-.392	-.225	-.139	.608	1.646
Domestic burglaries	5.428E-02	.009	.154	5.871	.000	.036	.072	.175	.210	.129	.703	1.422
Travel time	-1.640E-02	.002	-.459	-10.685	.000	-.019	-.013	-.448	-.364	-.236	.264	3.788
High affluent dummy	.504	.081	.147	6.237	.000	.345	.662	.390	.222	.138	.876	1.142
Per cent green (β_{PG})	2.554E-03	.001	.118	2.160	.031	.000	.005	-.050	.079	.048	.163	6.150
D (α_1)	-.123	.041	-.138	-2.994	.003	-.204	-.042	-.354	-.109	-.066	.228	4.383
D.Per cent green (β_D)	2.291E-03	.001	.109	1.846*	.065	.000	.005	-.145	.067	.041	.140	7.158
Health facility	3.944E-03	.002	.045	1.960*	.050	.000	.008	.165	.071	.043	.936	1.069

* Significant at 90 per cent confidence level, 10 degrees of freedom, critical value 1.796

Travel time and income support remain the most significant indicators. Per cent green β_{PG} is not as significant as in model 1 with the inclusion of the dummy variables. However, it is interesting to see that the differential intercept α_1 is statistically significant indicating that the two equations have different intercepts. The differential slope coefficient β_D is also significant, showing that the slopes are also different. The hypothesis $\alpha_1 = \beta_D = 0$ can be rejected as our F statistic is also statistically significant at 10 degrees of freedom. The differential intercept α_1 explains most of the useful split for the inner-outer dichotomy. By accepting our hypothesis we can conclude that the relationship between green spaces and house prices becomes more statistically significant when we look at wards with travel time to central London greater than 35 minutes.

High VIF and tolerances for per cent green, intercept dummy α_1 and slope dummy β_D show that nearly 75-85 per cent of their variance can be explained by other predictors. However, high multicollinearity may arise due to the fact that all of the three variables above are explaining the significance of green spaces affecting house prices and have strong correlations among themselves. This can be verified with the tolerance and VIF of per cent green in model 1, where tolerance was 0.623, implying that per cent green contributes sufficiently to the model. In other words, only 38 per cent of the variance in per cent green can be explained by the other predictors.

The semi-elasticities show that, on average, a 1 per cent increase in green space in wards with travel time less than 35 minutes is associated with a 0.25 per cent increase in house prices. On the other hand, it is associated with a 0.5 ($\beta_{PG} + \beta_D$) per cent increase in house prices in wards with travel time greater than 35 minutes.

Even though green spaces are more significant in explaining the variation in house prices in wards with travel times greater than 35 minutes, travel time is still considerably more significant than green spaces. Hence, it is a plausible conclusion that people may prefer to live in areas with more green spaces but also with good travel links to central London. For example, many boroughs with green spaces such as Richmond upon Thames, Merton, Enfield and Redbridge have low travel times to central London but are located in areas with high open green spaces.

The average house price in wards with less than 35 minutes travel time to central London (347 wards) is £242,043 and for wards with travel time greater than 35 minutes (413 wards) it is £163,526. This tells us that housing preferences in central London or close proximity (by time) to central London coupled with shortages in supply leads to an average figure, which is not a true reflection of average house prices in London. Even though green spaces are more significant in explaining house prices in wards with travel time greater than 35 minutes, the premium attached to residential properties close to central London is much higher. The high significance of travel time also tell us that people may consider the trade off between close proximity to central London and open green spaces. Also, larger plot size with gardens and more privacy is another important criteria. They may choose to spend a considerable amount of time commuting in order to enjoy the high wage and lifestyle of central London.

5. Limitations and considerations

Both models suffer from some methodological and measurement errors.

The maps and ward data in Chapter 2 of *Valuing Greenness* suggest a strong spatial correlation between nearby wards (GLA, 2003). The data, though indicative, is at ward level, which restricts analysis within the statutory boundaries. Services and access to transport infrastructure and green spaces are not restricted within ward boundaries. Where a large open space is included within one ward, but borders another, the first ward is allocated all the space and the second allocated none. Where a facility has a catchment (eg a school, as in Table 1, or a hospital) parts of the catchment often will lie in other wards to those allocated. Postcode level data would show a more local specific area characteristics since there are more detailed geographic boundaries.

The density of dwellings in London is negatively correlated with other land uses, including commercial and institutional uses. Prominent among these other uses are the lesser green spaces for which we had no data for this study: particularly private residential gardens which amount to some 20 per cent of London's land area, but also such places as allotments, cemeteries, school grounds and sports pitches. It is not possible to separate out dwelling density per se from these competing land uses. To this extent dwelling density is an inverse indicator of the green spaces not included within the 'strategic green space' indicator.

Our hedonic pricing models suffer from two main data problems: measurement error and multicollinearity.

- **Measurement error**: There will be errors in the observed values of the dependent and explanatory variables. The statistical model also depends on the choice of and weights attached to significant indicators. Inclusion of some other key indicators such as noise, ethnic mix, specific characteristics of houses (eg number of rooms, detached, terraced, garage), river views and income would add to the robustness of the model.
- **Multicollinearity**: Multicollinearity is a serious problem in hedonic models and arises when the effects of several variables are closely linked. If mulitcollinearity exists, then it becomes hard to determine which of several correlated variables is truly influential. An example of multicollinearity is where the levels of one pollutant may be closely correlated with the levels of another. For example, suspended PM_{10} is closely related to NO_2. We have used only NO_2 concentrations to take account of this problem. It is not believed that there is a significant residual problem with multicollinearity.

Some of the indicators, such as per cent green and overcrowding, are not linearly related to house prices. Other pairs of indicators (as seen from the scatter plot matrix and correlation coefficients in Appendix 2, Tables A4 and A5) also appear to have a curvilinear relationship. Taking squares for these indicators can remove the bias but then it becomes difficult to interpret the data. Moreover, given the large number of degrees of freedom does not cause significant problems with the heterogeneity of the variances.

Consideration of house and family size is also important when determining preferences for housing, which the models do not take into consideration.

This study has primarily looked at London as a place to live, rather than a place to work. Further work would be needed to ascertain the benefits of green spaces to businesses and commuter workers.

The analysis is more limited in looking at the impact on property values and does not really include those parts of the population that do not own their own homes. Tenure mix is a key variable for London, where a high proportion of the population of London live in socially rented accommodation.

Equal weights are attached to health indicators. In other words, the same weight is given to a hospital as to a GP.

The study has also overlooked the relevance of green spaces to visitors especially tourists, which should be quantified as well.

Since the report is a snapshot in time, it proves difficult to get an idea of the relationships between the variables and any implication resulting from changes in them. Due to the unavailability of ward level data we have not been able to undertake a time series analysis, which would prove extremely useful for policy and planning decisions. A time series analysis could realistically show any channels of influence between the various variables.

The property market suffers from market failure because of government intervention in the form of rent ceiling, subsidised housing and planning controls.

Hedonic pricing models do not estimate option, existence or bequest values of green spaces. For example, the non-use values of biodiversity are not measured.

The type, quality and accessibility of green spaces are a major factor determining property prices. For example, Richmond Park has more impact on property prices than an area of green space in the Green Belt.

The study, though indicative, is not complete and comprehensive in its valuation of green spaces. This is mainly because the attributes of green spaces are numerous and each requires a different methodology to measure it. For example, measuring biodiversity values is different from active sports value of green spaces. Survey based techniques can value green space as a whole or an aggregate of individual attributes. Net migration of people to rural areas will give a good picture of the value of green spaces as an indicator of quality of life.

6. Conclusion

This working paper has discussed two models which have been used to explain the significance of green spaces. The analysis, though suffering from some data and methodological problems, provides an indicative measure of open of green spaces. One can say, with a reasonable degree of confidence, that open green spaces have a comparatively greater impact on house prices in outer London. The working paper has also highlighted the key aspects of model 1, which needs to be considered in more detail, along with how to take this work forward. The lack of digitised data on all forms of green spaces is a major hindrance. More data would explain the variation in house prices more effectively.

The working paper has set out a methodology for review and ongoing consultation. It is intended to be a starting point in a series of analysis to be undertaken in this area.

Appendix 1: Travel time map

Map 1 Wards with travel time to central London greater than 35 minutes

Note: Grey areas have travel times greater than 35 minutes.

Appendix 2. Statistical results and analysis

Table A1 Descriptive statistics

	Minimum	Maximum	Mean	Standard deviation	Skewness
Home sale price (£s) 2001	79,121	1,208,150	199,000	115,000	2.94
Area (km^2)	0.24	30.51	2.06	2.26	5.18
Dwellings (no)	230	10,227	4,027.88	1,252.66	0.62
Dwelling density (no./km^2)	53.49	11,686.56	3,142.58	2,054.26	1.28
Income support 18 yrs+ (IS) %	0.62	31.81	10.90	5.71	0.59
Education below threshold (SAT) %	0	73.44	35.53	13.72	0.01
Overcrowding	0.34	29.82	4.15	2.80	2.12
Domestic burglaries 18 yrs+ %	1.58	49.14	14.04	7.58	1.18
Travel time (mins)	2.09	78.10	34.86	12.44	-0.43
Green space (km^2)	0	29.98	0.71	1.81	7.93
Per cent green (%)	0	98.25	19.57	20.57	1.27
Health (no.)	0	40	6.66	5.04	1.43
NO_2 average (ppb)	16.24	33.23	21.78	3.15	1.11

The descriptive statistics for the variables are shown in Table A1. Most of the indicators have a positive skew[6] and their logarithms would have ensured better approximated normal distributions. We have not done that to ensure transparency and avoid other measurement errors like multicollinearity.

[6] Skewness is a measure of distribution which indicates how much a distribution differs significantly from a normal symmetric distribution.

Table A2 Collinearity diagnostics for model 1

Dimension	Eigenvalue	Condition index
1	8.069	1.000
2	1.023	2.808
3	.847	3.086
4	.381	4.604
5	.214	6.147
6	.166	6.971
7	.136	7.704
8	8.504E-02	9.741
9	4.905E-02	12.826
10	2.885E-02	16.722
11	1.937E-03	64.537

The collinearity diagnostics show that model 1 does well to avoid serious multicollinearity. Table A2 checks for overall multicollinearity, and variables causing the greatest problem can be inferred by looking at the tolerances and VIF. Four eigenvalues are close to 0, indicating that the predictors are highly intercorrelated and that small changes in the data values may lead to large changes in the estimates of the coefficients. The condition indices are computed as the square roots of the ratios of the largest eigenvalue to each successive eigenvalue. Values greater than 15 indicate a possible problem with collinearity and greater than 30, means a serious problem. Only two of them are greater than 15, with one of these greater than 30. This indicates some multicollinearity as some variables are correlated. For example, overcrowding has some overlap with dwelling density and the inverse of dwelling density can be interpreted as an indicator of green spaces – including those captured by per cent green and those which are not. The multicollinearity is not severe enough to introduce a heavy bias on the regression coefficients, as nine of the 11 dimensions have condition indices less than 15.

Table A3 Collinearity diagnostics for model 2

Dimension	Eigenvalue	Condition index
1	7.239	1.000
2	1.580	2.140
3	.980	2.717
4	.389	4.313
5	.342	4.603
6	.162	6.689
7	.128	7.515
8	8.532E-02	9.211
9	5.159E-02	11.845
10	3.254E-02	14.915
11	1.053E-02	26.213

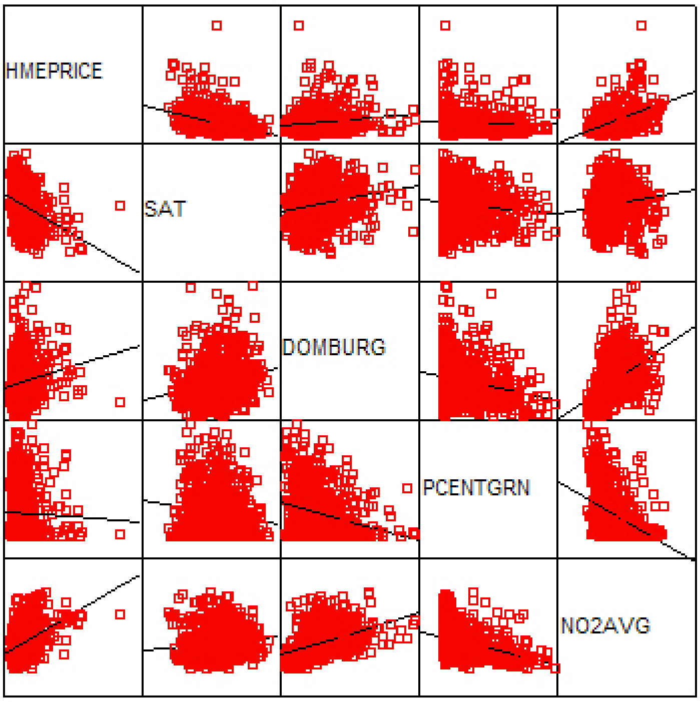
Model 2 avoids serious multicollinearity as only one condition index is greater than 15 but less than 30.

Table A4 Correlations

		Log house prices	Dwelling density	IS	SAT	Over crowding	Domestic burglary	Travel time	Per cent green	Health	NO2 average	High affluent dummy
Pearson Correlation	**Log house prices**	1.000	.326	-.324	-.392	-.128	.188	-.448	-.050	.165	.463	.390
	Dwelling density	.326	1.000	.374	.203	.436	.366	-.696	-.598	.054	.628	.171
	Income support (IS)	-.324	.374	1.000	.603	.663	.493	-.430	-.278	-.024	.376	-.152
	SAT	-.392	.203	.603	1.000	.438	.210	-.143	-.141	-.105	.141	-.133
	Over crowding	-.128	.436	.663	.438	1.000	.381	-.476	-.307	.081	.463	-.009
	Domestic burglary	.188	.366	.493	.210	.381	1.000	-.502	-.238	.108	.460	.084
	Travel time	-.448	-.696	-.430	-.143	-.476	-.502	1.000	.461	-.152	-.841	-.184
	Per cent green	-.050	-.598	-.278	-.141	-.307	-.238	.461	1.000	-.127	-.368	-.031
	Health	.165	.054	-.024	-.105	.081	.108	-.152	-.127	1.000	.114	.017
	NO_2 average	.463	.628	.376	.141	.463	.460	-.841	-.368	.114	1.000	.239
	High affluent dummy	.390	.171	-.152	-.133	-.009	.084	-.184	-.031	.017	.239	1.000
Sig. (1-tailed)	**Log house prices**	.	.000	.000	.000	.000	.000	.000	.085	.000	.000	.000
	Dwelling density	.000	.	.000	.000	.000	.000	.000	.000	.069	.000	.000
	Income support (IS)	.000	.000	.	.000	.000	.000	.000	.000	.250	.000	.000
	SAT	.000	.000	.000	.	.000	.000	.000	.000	.002	.000	.000
	Over crowding	.000	.000	.000	.000	.	.000	.000	.000	.013	.000	.400
	Domestic burglary	.000	.000	.000	.000	.000	.	.000	.000	.001	.000	.011
	Travel time	.000	.000	.000	.000	.000	.000	.	.000	.000	.000	.000
	Per cent green	.085	.000	.000	.000	.000	.000	.000	.	.000	.000	.196
	Health	.000	.069	.250	.002	.013	.001	.000	.000	.	.001	.319
	NO2 average	.000	.000	.000	.000	.000	.000	.000	.000	.001	.	.000
	High affluent dummy	.000	.000	.000	.000	.400	.011	.000	.196	.319	.000	.

Pearson correlation coefficients assume the data are normally distributed and is a measure of linear association between two variables. The values of the correlation coefficient range from -1 to 1. The sign of the correlation coefficient indicates the direction of the relationship (positive or negative). The absolute value of the correlation coefficient indicates the strength, with larger absolute values indicating stronger relationships. The significance of each correlation coefficient is also displayed in the correlation table. The significance level (or p-value) is the probability of obtaining results as extreme as the one observed. If the significance level is very small (less than 0.05) then the correlation is significant and the two variables are linearly related. We can see that dwelling density, income support, SAT, travel time and average emissions are linearly related with house prices. Per cent green, health, overcrowding and crime are weakly related with house prices. It shows most correlations are significant.

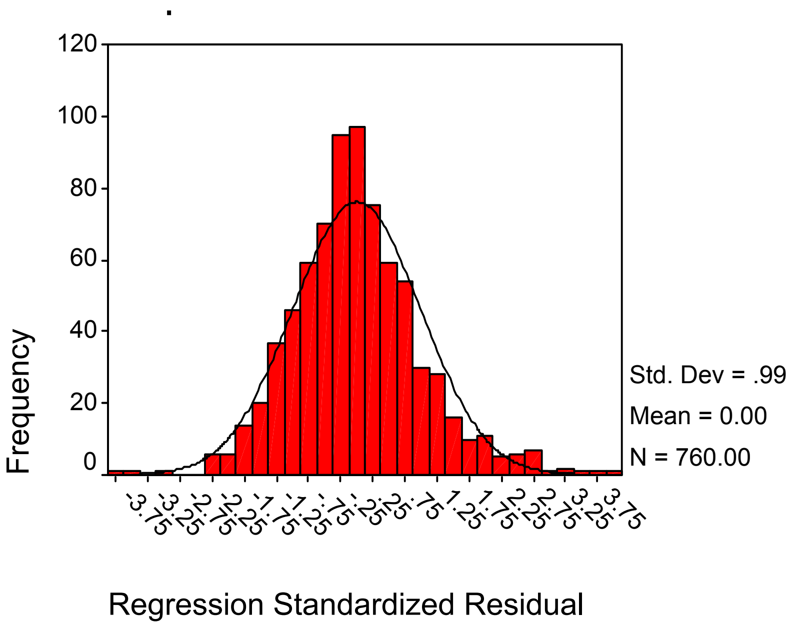
Table A5 Scatter plot matrix



Note:
HMEPRICE – Home sale price (£s) 2001
SAT – Standard Achievement Targets (Educational performance indicator)
DOMBURG – Domestic burglaries as a per cent of adult population
PCENTGRN – Per cent green
NO2AVG – Nitrous dioxide, Environment Indicator

Table A5 shows the fitted lines for each pair of indicators. Poor education performance, the proportion of SAT2 pupils failing to reach level 4 decreases in line with an increase in per cent green in wards. Emission level, overcrowding and crime are also negatively related with per cent green in wards.

Table A6 Regression standardised residual histogram



Note: LNHP – log of home sale price (£s)

The normality of the error term was inferred by plotting the histogram of the regression standardised residual and was found to be acceptably close to the normal curve.

Appendix 3. Semi-log model

The association of variables with house prices as explained by the semi-log model. For example:

$$\ln HP = \alpha_0 + \alpha_{PG} Pcntgreen + e_i$$

$$\alpha_{PG} = \frac{\text{Relative change in house prices}}{\text{Absolute change in per cent green}}$$

The slope coefficient measures the constant relative change in house prices for a given absolute change in the explanatory variables. A relative change multiplied by 100 becomes a percentage change.

Appendix 4. Abbreviations

CPRE	Council for the Protection of Rural England
DfES	Department for Education and Skills
DWP	Department for Work and Pensions
GIS	Geographical Information System
HoLP	Health of Londoners Project
IS	Income support
LFEPA	London Fire and Emergency Planning Authority
MPS	Metropolitan Police Service
NO_2	Nitrous dioxide
NOX	Nitrous Oxides
ONS	Office of National Statistics
PM_{10}	Particulate Matters
SATs	Standard Achievement Targets
SEIPH	South East Institute for Public Health
SO_2	Sulphur dioxide
TfL	Transport for London
VIF	Variance inflation factor

Appendix 5. Bibliography

Brookshire David S, Mark A Thayer, William D Schulze and Ralph C d'Arge 1982, 'Valuing Public Goods: A Comparison of Survey and Hedonic Approaches' *The American Economic Review*, vol. 72, Issue 1, pp 165-177.

Din Allan, Martin Hoesli and André Bender 2001, 'Environmental variables and real estate prices' *Urban Studies*, vol. 38, No. 11, pp 1989- 2000.

DTLR 2002a, *Green Spaces, Better Places: Interim report of The Urban Green Spaces Taskforce*, DTLR, London.

DTLR 2002b, Improving urban parks, play areas and open spaces, DTLR, London.

Freeman III, A Myrick 1979, 'Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues' *Scandinavian Journal of Economics*, vol. 81, pp 154-173.

Garrod Guy and Kenneth G Willis, 1999, Economic Valuation of the Environment. Edward Elgar Cheltenham.

GLA 2002, *The London Index of Deprivation. An alternative to the IMD 2000. Consultation Document*, GLA, London.

GLA 2003, *Valuing Greenness: Green spaces, House prices and Londoners' priorities,* GLA, London.

Gujarati, Damodar 1995, *Basic Econometrics*, 3rd edition, McGraw-Hill, New York.

Rosen Sherwin 1974, 'Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition' *Journal of Political Economy*, vol. 82, pp 34-55.

Wabe, JS 1970, 'A study of house prices as a means of establishing the value of journey time, the rate of time preference and the valuation of some aspects of the environment in the London Metropolitan region', Research Paper 11, Department of Economics, University of Warwick, Coventry.

Other formats and languages

For a large print, Braille, disc, sign language video or audio-tape version of this document, please contact us at the address below:

Public Liaison Unit
Greater London Authority
City Hall
The Queen's Walk
London SE1 2AA

Telephone **020 7983 4100**
Minicom **020 7983 4458**
www.london.gov.uk

You will need to supply your name, your postal address and state the format and title of the publication you require.

If you would like a summary of this document in your language, please phone the number or contact us at the address above.

Chinese

中文
如果需要此文檔的您的母語拷貝，
請致電以下號碼或和下列地址聯係

Hindi

यदि आप इस दस्तावेज़ की प्रति अपनी भाषा में चाहते हैं, तो कृपया निम्नलिखित नम्बर पर फोन करें अथवा दिये गये पता पर सम्पर्क करें।

Vietnamese

Tiếng Việt
Nếu bạn muốn bản sao của tài liệu này bằng ngôn ngữ của bạn, hãy gọi điện theo số hoặc liên lạc với địa chỉ dưới đây.

Bengali

আপনি যদি আপনার ভাষায় এই দলিলের প্রতিলিপি (কপি) চান, তা হলে নীচের ফোন্ নম্বরে বা ঠিকানায় অনুগ্রহ করে যোগাযোগ করুন।

Greek

Αν θα θέλατε ένα αντίγραφο του παρόντος εγγράφου στη γλώσσα σας, παρακαλώ να τηλεφωνήσετε στον αριθμό ή να επικοινωνήσετε στην παρακάτω διεύθυνση.

Urdu

اگر آپ اِس دستاویز کی نقل اپنی زبان میں چاہتے ہیں، تو براہِ کرم نیچے دیئے گئے نمبر پر فون کریں یا دیئے گئے پتہ پر رابطہ قائم کریں.

Turkish

Bize telefon ederek ya da yukarıdaki adrese başvurarak bu belgenin Türkçe'sini isteyebilirsiniz.

Arabic

إذا أردت نسخة من هذه الوثيقة بلغتك، الرجاء الاتصال برقم الهاتف او الكتابة الى العنوان

Punjabi

ਜੇ ਤੁਹਾਨੂੰ ਇਸ ਦਸਤਾਵੇਜ਼ ਦੀ ਕਾਪੀ ਤੁਹਾਡੀ ਆਪਣੀ ਭਾਸ਼ਾ ਵਿਚ ਚਾਹੀਦੀ ਹੈ, ਤਾਂ ਹੇਠ ਲਿਖੇ ਨੰਬਰ 'ਤੇ ਫ਼ੋਨ ਕਰੋ ਜਾਂ ਹੇਠ ਲਿਖੇ ਪਤੇ 'ਤੇ ਰਾਬਤਾ ਕਰੋ:

Gujarati

જો તમને આ દસ્તાવેજની નકલ તમારી ભાષામાં જોઈતી હોય તો, કૃપા કરી આપેલ નંબર ઉપર ફોન કરો અથવા નીચેના સરનામે સંપર્ક સાધો.

GREATER**LONDON**AUTHORITY

City Hall
The Queen's Walk
London SE1 2AA

www.london.gov.uk
Enquiries **020 7983 4100**
Minicom **020 7983 4458**

MoL/May03/CJ D&P/MT