A Methodology for Estimating Spatial Non-Stationarity in Ecosystem Service Values

By Austin Troy,

Assistant Professor, University of Vermont School of Natural Resources ● austin.troy@uvm.edu

Abstract

This research, which is part of the Baltimore Ecosystem Study, analyzes housing transactions in West Baltimore to determine whether there is spatial variability in willingness to pay for proximity to environmental amenities, including trees and parks, using an expanded form of hedonic analysis called geographically weighted regression (GWR). GWR is a regression method that allows parameter estimates and test statistics to vary continuously over space. By enabling such spatial variability, we can visualize spatial patterns in complex socio-economic relationships and assess how valuations of environmental amenities change over space. Moreover, we can use this information to help visualize and delineate socio-economic patches, or areas where these relationships are homogeneous. The study finds that valuations of tree cover and proximity to parks is spatially non-stationary within the study area.

Introduction

Hedonic analysis is an econometric method that is used to disaggregate housing prices into a schedule of marginal, unobserved, implicit attribute prices (e.g. the marginal value of a fifth bathroom) . Because housing prices capitalize the value of surrounding amenities, hedonic analysis has frequently been used to value marginal willingness to pay for environmental goods and services, such as trees, views, open space, protected areas, clean water, and clean air. Because of its usefulness in deriving such values, hedonic studies are a common source of data in ecosystem service valuation.

One common assumption of the hedonic model—and with regression modeling in general—is that the relationship between price and attributes is globally constant across the extent of the modeled population. In some cases this is appropriate, but in others spatially static model parameters serve to obscure local heterogeneity. When that variation is great, this can lead to misleading results.

This is a particularly salient point in modeling housing markets because the relationship between price and many housing attributes is clearly non-stationary. For example, the marginal implicit price of an additional bedroom is likely to vary from one neighborhood to the next. In some cases this spatial variation in a marginal attribute price can be controlled for without taking a non-stationary approach, simply by adding a variable that proxies space, such as distance to employment centers. In others cases, however, not enough variables can be operationalized within a stationary model to control for that spatial heterogeneity because of the elusive factors that define "place," in which case a spatially non-stationary modeling approach may be preferable. In the latter case, a geographically non-stationary modeling method can help to elucidate the local variation that would otherwise be obscured by a global model. Moreover, it is an excellent method for diagnosing patterns or patchiness that may be otherwise undetectable and for isolating and defining the elusive boundaries of housing submarket—that is, geographically contiguous areas where price-attribute relationships remain relatively constant.

Background on Geographically Weighted Regression

In this study, Geographically Weighted Regression (GWR) (designed by Fotheringham et al. 2000, Fotheringham et al 2002, who built upon the works of Hastie and Tibshirani 1990 and Loader 1999,) was used to generate non-stationary parameter estimates of the relationships between housing price and attributes in West Baltimore. In particular, it was hoped to determine whether marginal willingness to pay for proximity to environmental amenities—and the ecosystem services they deliver—varies across space in a way that cannot readily be explained through inclusion of control variables in a global model. The amenities that were looked at were trees and parks.

GWR is a non-stationary regression method that uses the formula

$$Y(x) = \alpha\,(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_k$$

where α and β vary continuously as a function of location (u,v) at each point *i*, unlike in a global regression model where parameters are constant. In this case Y(x) is housing price function, x_k is a vector of attributes and Beta is vector of parameters whose value is a function of location.

In this method, separate regressions are run centered on each observation, with a spatial kernel determining which observations are included in the population of each individual regression and how they are weighted, based on a Gaussian distance decay function. An adaptive or fixed size kernel can be used to determine the number of local points that will be included, depending on the spacing of the data. For this project, adaptive kernels were used since the data were not evenly distributed. The method used to determine optimal bandwidth was minimization of the Akaike Information Criterion (AIC)

$$\mathrm{AIC}_c = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n + \mathrm{tr}(\mathbf{S})}{n - 2 - \mathrm{tr}(\mathbf{S})}\right\}$$

where tr(**S**) is the trace of the *hat matrix* and n is the number of observations.

Methods

Property data for West Baltimore were obtained from Maryland Property View, ranging from 1992-2000. Data were assigned numerous spatial attributes, such as distance to highways, subway stops, parks and trees (tree canopy data were derived from IKONOS satellite imagery as part of the Strategic Urban Forests Assessment Project [Irani and Galvin 2002]) and elevation. Additionally block-group level socio-economic attribute from the 2000 Census were assigned, such as median household income and percent vacancy. Finally, structural attributes were assigned from the Property View data set, including assessed improvement value, number of bathrooms, building material, single family or multi-family, and age of structure.

Once attributes were coded, log-transformed price was regressed against these 14 variables using GWR software (Fotheringham et al. 2003). The model output allowed for testing of the global model for improvement against the local model. It also allowed for testing of each parameter for global significance and for local spatial variability. Model output included a GIS point file of the observations with parameter values and t statistics for each point (representing the results of the individual kernel –based regressions centered on each point). These were plotted to look for spatial patterns in the coefficients and t statistics on predictor variables, particularly distance to trees, distance to parks and the dummy variable for trees within ten meters.

GLOBAL MODEL RESULTS				LOCAL RESULTS				
Var Name	Coeff	Std Err	T Stat	Lwr Quartile	Median	Upr Quartile	P (H_0= stationary)	
Intercept	10.787979	0.039974	269.8756	10.457601	10.7445	11.525506	0.000	**
NFMIMPVL	0.000007	0.000000	35.4648	0.000005	6E-06	0.000006	0.210	
D2SUBWAY	0.000003	0.000006	0.4362	-0.000172	-2.8E-05	0.000032	0.000	**
D2HIWAY	0.000045	0.000021	2.1824	-0.000019	4.1E-05	0.000211	0.000	**
D2PARK	0.000047	0.000017	2.8358	-0.000163	-4.1E-05	0.000055	0.000	**
ELV	-0.002603	0.000229	-11.3732	0.000699	0.0015	0.004242	0.000	**
BATH	0.046810	0.008367	5.5943	0.010501	0.02899	0.049988	0.390	
ASBEST	-0.026298	0.041685	-0.6309	-0.083831	0	0	0.390	
SFH	0.130338	0.020262	6.4325	-0.398552	0.11919	0.213064	0.000	**
YRSOLD	0.000197	0.000273	0.7211	-0.000798	-0.0002	0.000903	0.020	*
MHHI	0.000004	0.000000	12.0501	0.000001	2E-06	0.000003	0.010	**
PVAC	-0.531325	0.086270	-6.1589	-0.567788	-0.19903	-0.028381	0.110	*
TREES5	0.058003	0.017272	3.3582	-0.022553	0.01742	0.047081	0.000	**
D2TREES	0.000164	0.000204	0.8056	-0.001819	-0.00032	0.000067	0.000	*
* Test statistic on non-stationarity significant at 95% confidence level								
**Test statistic on non-stationarity signficiant at 99% confidence level								

Table 1. Global and Local (GWR) Regression Results

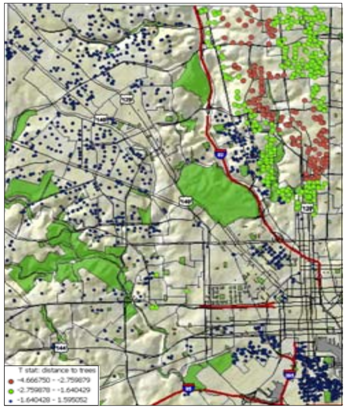

Figure 1. Plot of T statistics on distance to nearest trees. Large circles represent parameters significant at the 95% level

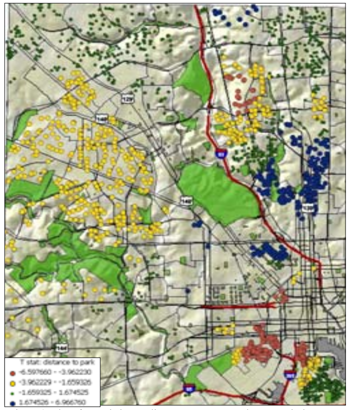

Figure 2. Plot of T statistics on distance to nearest park. Large circles represent parameters significant at the 95% level

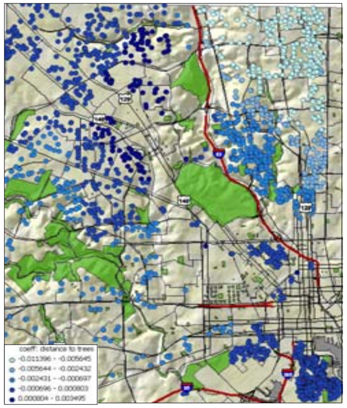

Figure 3. Plot of coefficient on distance to nearest trees. As figure 1 shows, not all of those observations are significant

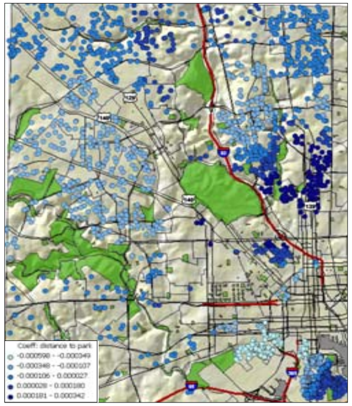

Figure 4. Plot of coefficient on distance to nearest park. As figure 2 shows, not all of those observations are significant

Results

Several statistics indicated that the local model offered a significant improvement in fit over the global model, including the change in the Akaike Information Criterion (AIC[global]=1234, AIC[local]= 869) and an ANOVA (F=6.0167) testing the null hypothesis that the local model offers no improvement over the global. R-squared values also were increased with the local model (R^2=.78) over the global model (R^2=.72) . Fourteen of the parameters were found to be spatially non-stationary at the 95% confidence level, according to a Monte Carlo significance test, where the variance of parameter estimates is compared against an experimental distribution (Hope 1968). The p-values from this test are shown in Table 1 below, along with global parameter coefficients and mean, lower and upper quartile GWR parameter estimates. The differences between upper and lower quartiles on most parameters show how variable they are across space.

The other output of this analysis was an ARC/INFO point file, showing parameters, t statistics and standard errors for each local regression. To look for patterns, the t statistics (figures 1-2) and coefficients (figure 3-4) were plotted out for the variables on distance to nearest tree (D2TREES) and distance to nearest park (D2PARK).

Plots of the t statistics showed that there was only a significant relationship between price and attributes for certain areas (large dots are significant at the 95% confidence level, small dots are insignificant). For instance, Figure 1 shows there is a significant negative relationship between price and distance to nearest trees (indicating the positive value of trees) in the north eastern part of the study area (Roland Park, Hampden, Medfield, Homeland), but elsewhere this is not the case. The magnitudes of this relationship can be visualized by looking at the coefficients plot in Figure 3, which shows a clear pattern of increasingly negative coefficients towards the northeast. The t statistic on distance to nearest park (Figure 2) displays clear spatial patchiness, showing that for the most part parks are capitalized positively (coefficient on distance is negative), while in one area in the east (Hampden, Remington) parks are negatively capitalized, and in several other areas coefficients are insignificant. This plot, along with the coefficient plot (Figure 4), shows that houses strongly and positively capitalize parks in the south (Washington Village), positively capitalize them to a lesser extent on the west side and in patches in the east, and negatively capitalize them in one patch in the east. The plot for the t statistic and coefficient on the dummy variable for trees within 5 meters was not given here in the interests of space and because only one area in the south of the city positively capitalized trees within 5 meters.

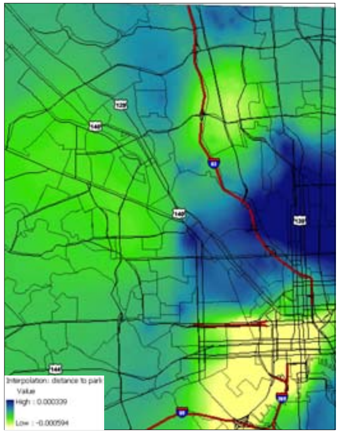

Figure 5. Inverse Distance Weighting Interpolation of coefficient on distance to park

References

Fotheringham, A.S., Charlton, M.E and Brunsdon, C. 2003. *Geographically Weighted Regression*. Software.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 2002, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Chichester: Wiley

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 2000, *Quantitative Geography*, London: Sage

Hastie, T.J. and Tibishrani, R.J. 1990 *Generalized Additive Models*. London, Chapman & Hill

Hope, A.C. 1968 A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society Series B* **30:**582-98.

Irani, F.M. and Galvin, M.F. 2002. *Strategic Urban Forests Assessment: Baltimore Maryland*, Maryland Department of Natural Resources

Loader, C. 1999. *Local Regression and Likelihood* . New York, Springer Verlag.

Acknowledgements: This research was made possible through funding from the Baltimore Ecosystem Study (National Science Foundation Long-Term Ecological Research program, grant number DEB 9714835) and the US Forest Service