Housing Studies,
Vol. 20, No. 4, 549–569, July 2005

Routledge
Taylor & Francis Group

Community Choices and Housing Demands: A Spatial Analysis of the Southern Appalachian Highlands

SEONG-HOON CHO*, DAVID H. NEWMAN** & DAVID N. WEAR[†]

*Department of Agricultural Economics, University of Tennessee, Knoxville, TN, USA, **Warnell School of Forest Resources, University of Georgia, USA, [†]Southern Research Station, US Department of Agriculture, Forest Service, USA

(Received September 2003; revised October 2004)

ABSTRACT *This paper examines housing demand using an integrated approach that combines residential decisions about choices of community in the Southern Appalachian region with the application of a Geographical Information System (GIS). The empirical model infers a distinctive heterogeneity in the characteristics of community choices. The results also indicate that socio-economic motives strongly affect urban housing demands while environmental amenities affect those of rural housing demand.*

KEY WORDS: Community choices, housing decisions, spatial econometrics

Introduction

Rapid development of rural areas can change the biophysical structure of landscapes and the complement of services that flow from them. This is especially the case in the Southern Appalachians region of the USA where the environmental amenities that provide a strong draw for immigration are impacted by the resulting development. To date, research on development in the Southern Appalachians has focused on understanding the propensity of land to be developed as it relates to topographic, locational and ownership variables of individual sites (see Wear & Bolstad, 1998 for a review). The research by Wear & Bolstad (1998) evaluated and developed a forecasting model for land-use change in the region. They indicate that modelling land-cover change needs to be extended by modelling the human drivers of landscape change such as housing demand. This is because residential development is the dominant driving force of land-use change in the Southern Appalachian Highlands.

Correspondence Address: Seong-Hoon Cho, Department of Agricultural Economics, University of Tennessee, Knoxville, TN 37996-4518, USA. Email: scho@mail.ag.utk.edu

ISSN 0267-3037 Print/1466-1810 Online/05/040549-21 © 2005 Taylor & Francis Group Ltd
DOI: 10.1080/02673030500114433

Housing demand has been estimated by modelling an individual household's consumption of houses as a function of environmental, structural, and neighbourhood characteristics. Most households choose a community before they choose a specific neighbourhood. By contrast, the existing housing demand literature has seldom modelled the choice of a specific community as part of the household's optimisation problem. An exception is Rapaport (1997), who developed an econometric model that incorporated the housing demand of an individual unit within the choice of a specific community. The model specified housing demand with the incorporation of community choice as a function of household characteristics but did not address residential housing patterns for a group of neighbours.

This research, organised mainly by classic locational rent theory, leaves unaddressed important elements of the underlying choice analysis. Ultimately it is the expression of the individual choice analysis upon the complex biophysical/social landscapes of the Southern Appalachians that will define future landscape conditions in the region. The study posits that decisions regarding housing location are not conducted in isolation but start with a selection of the community which provides the suite of infrastructure, services, and other characteristics desired by the decision maker. Once the broader community is selected, site-specific considerations enter into the selection of a specific neighbourhood.

Many of the factors that affect housing demand are spatially heterogeneous. It is said of real estate that what matters most is location. A significant advantage of a spatially explicit model is that it can readily incorporate substantial spatial detail, allowing analysis of how various locational factors influence housing demand. The role of locational factors in housing demand can be examined in two interrelated ways. One form of geographic influence involves externalities associated with the location of the house. These types of externalities are called adjacency effects because they capture the spatial spillover on a given structure by a neighbouring structure. In addition to spillovers from neighbouring structures, overall neighbourhood characteristics such as accessibility also enter into housing demand. These kinds of influences may be called neighbourhood effects.

The combination of spatial analysis and a Geographic Information System (GIS) provides the optimal framework for investigating both types of locational factors in housing demand. GIS serves as the research platform both to manage spatial data and to implement spatial analysis methods (Can, 1998). There are multiple benefits to housing research in terms of collection and integration of a very large database. Two of the most significant benefits are the ability to layer data from multiple sources and to look at data at different scales or geographies. The model here is estimated using spatially explicit data through the application of GIS and spatial statistics. This feature also allows more accurate analysis by providing flexibility in specifying models and measuring variables (e.g. Ding, 2001; Geoghegan *et al.*, 1997; Lake *et al.*, 2000).

Economic models of land use have been applied to both broad units and fine units, based on the spatial scale of land use. Models of broad units examine patterns of land use from a macro viewpoint. These models generally use counties or county groupings as units to highlight how socio-economic factors and physical landscape features influence land-use allocations (Alig, 1986; Hardie & Parks, 1997; Hardie *et al.*, 2000; Miller & Plantinga, 1999; Plantinga, 1996). On the other hand, models of fine units provide analyses of spatially explicit land-use decisions. These models estimate the direct influence of site-specific factors measured at a fine resolution. For example, the road construction and access influences on land development (e.g. Chomitz & Gray, 1996; Dale *et al.*, 1993; Nelson & Hellerstein, 1997) and the influences of location, topography, and ownership (Spies *et al.*, 1994; Turner *et al.*, 1996) are analysed in this framework.

While each type of model independently serves a valuable function, both have limitations. Macro-scale analyses do not capture information in a spatially explicit framework, while micro-scale analyses may miss out on broader physical and social phenomena. Wear & Bolstad (1998) explain the limits of land-use models for different units. They point out that land-use models of spatially broad units may not provide direct insights into the fine-scale socio-economic and physical consequences of land-use changes. They also discuss the limitations of fine-scale units, including the resolution of the definition of land use. For example, residential presence in the satellite images of forest cover (e.g. Wear & Flamm, 1993; Turner *et al.*, 1996) may not capture site-specific land uses.

This paper attempts to bridge the broad and fine scales of analysis by examining choices of community (broad units) in conjunction with site-specific housing demand (fine-scale units). Census blocks are used in this study because the characteristics of block data fit with the model of housing demand for a group of neighbours with the incorporation of the community choice. The census blocks are small enough to be used as fine units for the site-specific choice model for a group of neighbours while they can also be grouped and classified as broad units for the community choice model.

The remainder of this paper is organised as follows. The next section presents community choice and housing demand models that can be estimated with the 1990 US Census block data of the Southern Appalachian region in two stages. The subsequent section presents a description of the study area and data. The estimation results of both community choice and housing demand models follow, and the paper closes with a discussion of the results and their possible interpretation.

The Empirical Model

Communities are characterized by residents' income and preferences and administrative circumstances. This paper models households' choices of urban and rural communities and housing demand. Housing is a continuous variable, but the community is a discrete choice. In particular, a household is constrained in its choice of a specific neighbourhood by the household's choice of community.

Because households choose the community before the specific neighbourhood, the housing demand needs to be modelled in the context of community choice. The model of housing demand for a group of neighbours (instead of an individual housing unit) is treated as a system of equations to incorporate socio-economic effects that originate in one's residential neighbourhood. This approach helps clarify the identifying conditions for neighbourhood effects (Ioannides & Zabel, 2003). The model of housing demand is derived and estimated for a group of neighbours that incorporates the choices of community.

Following Rapaport's (1997) estimation technique, the model is estimated in two stages. First, the probability of a household's choice of different types of urban or rural community is estimated as a function of community and household characteristics. For better specification of the community types, urban and rural communities are sub-categorized into urban-dominant, urban-moderate, rural-moderate and rural-dominant communities. The types of communities are represented by the types of blocks in this study. The classifications of the types of blocks for the types of communities are explained in the next section. The households' choices of the four types of communities are modelled in a multinomial logit framework. The estimates of the framework are examined to check whether the effects of community characteristics are heterogeneous for the households' community decisions. Second, the housing demand is estimated conditional upon the choice of type of urban or rural community. Following the approach by Ioannides & Zabel (2003), the model of aggregate housing demand is treated for a group of neighbours.

Treating the aggregate housing demand conditional upon the type of community makes it possible to test the hypothesis that the types of urban and rural communities are relevant partitioning criteria for the aggregate housing demand estimation. The number of housing units within a census block reflects aggregate housing demand of the location of the census block. Because the size of each census block is different, the housing count within a given area is used to represent aggregate housing demand. The aggregate housing demand is then modelled separately for the four types of communities to test the hypothesis that housing choices are not different under the four different communities.

The model is estimated using 1990 US Census block data of the Southern Appalachian region. The Southern Appalachian region is chosen for this study because residential development plays an increasingly important role in the region. The Southern Appalachian region provides a less complicated study site for testing our methodology because institutional factors such as land-use regulations have only a minor influence on the area's development and the region contains distinctive urban and rural communities.

The estimates from the aggregate housing demand model using housing count conditional upon the type of community may be useful to policy makers who issue housing permits in different communities. If population density for a type of community is projected under a ceteris paribus assumption, it is possible to predict the number of housing developments for each census block. Policy makers can then use these forecasts to make management plans (e.g. plan for expanding public water and sewer services for additional housing development). In effect, this would allow the projection of differential rates of growth for communities in the region.

In order to check whether a self-selection bias arises in the formation of the community-type choice, a self-selection variable is added in each of the four aggregate housing demand equations. The self-selection variables are formed by incorporating the estimates of the community choice models into the housing demand equations. The description of the self-selection variable is presented in the 'The Aggregate Housing Demand Model' section. The self-selection variables detect whether or not households' choices of community have different effects on the households' site-specific housing demands.

The Community Choice Model

Suppose a household tries to choose a community from among four possible types of communities. The types of communities are based on degree of urbanisation. Let u_j^* be the household's expected utility from choosing a type of community j. The community j is indexed as 1, 2, 3 and 4 for urban-dominant, urban-moderate, rural-moderate and rural-dominant communities, respectively:

$$u_j^* = Z'\gamma_j + e_j \tag{1}$$

where Z is a vector of household characteristics and community attributes influencing the choice of the community and e_j is a residual capturing errors in perception and optimisation by the household. The household's utility in choosing an alternative community is not observable, but their choice of a community is observed. Let J be a polychotomous index denoting the household's type of community.

$$J = j \text{ if and only if } u_j^* = \max(u_1^*, u_2^*, u_3^*, u_4^*). \tag{2}$$

Maddala (1983) shows that if the residuals e_j are independently distributed with an extreme value distribution, then the choice of the type of community can be represented by a multinomial logit model (Maddala, 1983, p. 60). Following McFadden (1973), disturbances are assumed to be independent and identically distributed with a Weibull distribution. This implies that the probability of choosing a type of community j by the household can be expressed as

$$P_j \equiv \Pr(J = j) = \frac{\exp(Z' \cdot \gamma_j)}{\sum_{i=1}^{J} \exp(Z' \cdot \gamma_i)}, \quad j = 1, 2, 3, 4. \tag{3}$$

The estimated equations provide the set of probabilities for J community choices. To avoid indeterminacy, the parameter vector of urban-dominant community γ_1 is normalised to zero. This normalisation renders the estimated γ_1 parameters as un-interpretable. However, inferences can be drawn from the computed 'marginal effects' of elements of Z relative to sample averages. The marginal effects in the model are partial derivatives of the probability with respect to the determinants:

$$m_j = \frac{\partial P_j}{\partial Z} = P_j\left(\gamma_j - \sum_{j=1}^{J} P_j \gamma_j\right) = P_j(\gamma_j - \bar{\gamma}). \tag{4}$$

The statistical significances of these effects are estimated by the asymptotic covariance matrix of m_j (Greene, 1997, pp. 916–917). While the parameter vector γ_1 is normalised to zero, the vector of marginal effects δ_j is constrained to sum to zero. This normalisation means that δ_j can be interpreted as the net effects of an increase in the value of determinants Z on the decision to live in community j.

There are several difficulties in dealing with community characteristics in the multinomial logit model. First, the attributes of the community are chosen attributes and are thus subject to self-selection bias. Second, some community characteristics are by definition directly related to the dependent variable since the sub-categorisation of the community is based on demographic features that are correlated with some community characteristic variables. These problems suggest a potential for simultaneity, endogeneity and misspecification if raw characteristics of the actual observed choice are included in the multinomial logit model. The strategy here to deal with these problems is based on the approach by Feridhanusetyawan & Kilkenny (1996). The study normalised local relative to maximum levels of each characteristic and identified the extent to which the normalised measure deviated from the expected, by community. These residuals were used as the explanatory variables in the multinomial logit model.

The normalisation procedure is as follows. The community characteristics of each census block are each expressed relative to the maximum value among all the census blocks. This converts the measure of each community characteristic to a number ranging from 0 to 1. Allowing Z_b^* to denote the normalised local attribute:

$$Z_b^* = Z_b/Max(Z_b) \tag{5}$$

where b is 3687 census blocks.

The normalised local attributes were then regressed on community types and housing density that were used for the grouping of the community to control potential bias caused by simultaneity, endogeneity and misspecification. The regression model for community characteristics is:

$$Z_b^* = \alpha_1 URBANM_b + \alpha_2 RURALD_b + \alpha_3 RURALM_b + \alpha_4 H_b + \varepsilon_b \tag{6}$$

where $URBANM_b$ is the dummy variable, equal to 1 if the census block b is urban-moderate community and equal to 0 otherwise; $RURALD_b$ is the dummy variable, equal to 1 if the census block b is rural-dominant community and equal to 0 otherwise; $RURALM_b$ is the dummy variable, equal to 1 if the census block b is rural-moderate community and equal to 0 otherwise; H_b is housing units within 1 km^2 of a census block b; ε_b is random disturbance. The predicted residual is $\Delta Z_b = Z_b^* - \hat{Z}_b^*$, where $\hat{Z}_b^*$ is the predicted relative community characteristic level estimated using equation (6). By construction, the predicted residual is not correlated with the systemic classification of the community.

Previous studies (e.g. Nechyba & Strauss, 1998; Rapaport, 1997) suggest that individual community choices are specified as a function of household characteristics and community attributes. Here, consideration is given to the influence of individual-specific characteristics (the household characteristics of education level and political view) and choice-specific attributes (the community attributes of population density, crime level, stability, and level of air pollution).

The Aggregate Housing Demand Model

Under the assumption that the housing market is in equilibrium and that households find optimal community types in the first stage, the households make optimal site-specific housing demand decisions within their choices of community types in the second stage. Conventional demand analysis postulates a relationship between the quantity of a good demanded and its relative price, given the income of the household and other household characteristics. In this perspective, the following can be identified

$$h_i = \alpha_i + \beta_i y_i + \delta_i(p_i) + \eta_i x_{ik} + \mu_i \tag{7}$$

where h_i is housing demand by individual $i, i = 1, \ldots, N$; y_i is individual income; p_i is housing price; x_{ik} $(k = 1, \ldots, K)$ is a vector of other socio-economic and environmental variables affecting housing demand. Equation (7) is typically applied to individual housing datasets. The study is interested in the aggregate housing demand for a group of neighbours because the model helps clarify the identifying conditions for neighbourhood effects (Ioannides & Zabel, 2003). The census block is used as units for a group of neighbours. The linear aggregation of the individual housing demand is possible if (a) income and other variables are growing at the same rate in each location—or exhibit a common stochastic trend—and (b) the structures of the housing markets are the same over space (Meen & Andrew, 1998). The census block data have been shown to meet these criteria in previous studies. These studies have found that specifying neighbourhood variables and aggregating housing data at the census block-group level led to robust hedonic price estimations (Cao & Cory, 1981; Geoghegan *et al.*, 1997; Goodman, 1977).

Aggregate housing demand at the block level is estimated as a function of the socio-economic and environmental characteristics of the block, in addition to a self-selection variable in the formation of community choice. Since the aggregate data reflect market equilibrium of both demand and supply, price and quantity of house data of the census block level are endogenous variables. Under the assumption that the rest of the variables are exogenous, we estimate the following equation systems in a two-stage least square estimation (2sls):

$$\hat{H}_{bj} = \alpha_j + \beta_j Y_j + \delta_j(\hat{P}_j) + \eta_j X_{jk} - \theta_j \hat{\lambda}_j + e_j \tag{8}$$

$$\hat{P}_{bj} = \alpha_j + \beta_j Y_j + \delta_j(\hat{H}_j) + \eta_j X_{jk} - \theta_j \hat{\lambda}_j + e_j \tag{9}$$

where $\hat{H}_{bj}$ is the aggregate housing units within 1 km^2 of a census block b at community j; Y_j is per capita income at community j; $\hat{P}_j$ is housing price at community j; X_{jk} is a vector of other socio-economic and environmental variables affecting the housing demand at community j; $\hat{\lambda}_j$ is a self-selection variable at community j. (A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied or, if vacant, is intended for occupancy, as separate living quarters.) The self-selection variable is estimated using the following equation (Lee, 1983):

$$\hat{\lambda}_j = \phi[\Phi^{-1}(\hat{P}_j)]/\hat{P}_j, \tag{10}$$

where $\hat{P}_j = \frac{\exp(Z' \cdot \gamma_j)}{\sum_{i=1}^{J} \exp(Z' \cdot \gamma_i)}$ from the estimates of the first stage. The form of the self-selection variable incorporates community choice into the residential block decision. Explanatory variables X_k are considered to include socio-economic variables describing income, population density, crime rate, length of residency, education, political view, travel time to work, distance to any city, distance to a major city, distance to major roads and a road index. The environmental variables of distance to major open spaces, distance to lakes, air pollution level, elevation, stream index, and an open space index are considered (see the Tables 1 and 2 for definition of all the variables and their mean values by each community).

Following Greene (1997, pp. 740–742), in the first stage of the estimation, the reduced forms of the equations are estimated using OLS. Predicted values from the reduced form equations are then estimated. In the second stage, the equations are re-estimated after replacing the predicted values of $\hat{P}$ and $\hat{H}$ from the first-stage estimation.

The application of GIS at the census block level provides unique spatial variables including spatial indices along with commonly used distance variables. For example, the distance variables can measure the effect of distance to the nearest open space but not the effect of open space in the neighbourhood. The effect of open space in the neighbourhood has typically been measured by a dummy variable in spatial econometric analyses (e.g. Mahan *et al.*, 2000). However, the dummy variable only measures the effect of existence of open space in the neighbourhood, while it does not reflect relative abundance of open space in the neighbourhood. The open space index, ratio of total area of open space to total

Table 1. Definition of variables

Variable	Initials	Definition
Dependent Variables		
Community index	CI	Index for a type of community of urban-dominant, urban-moderate, rural-moderate, rural-dominant
Housing count (per km^2)	HD	Number of houses within 1 km^2 of area
Socio-economic Variables		
Housing value ($1000)	HV	Median value of owner-occupied houses in $1000
Income ($1000)	IC	Per capita income in $1000
Population density (per km^2)	PD	Population within 1 km^2 of area
Crime rate	CR	Number of reported crimes, from vehicle theft to murder
Stability (%)	ST	Ratio of occupancies with 5 years or more to total occupancies
Education (year)	ED	Median school years
Political view (%)	PV	Ratio of population with political outlook very conservative and somewhat conservative to total population
Travel time to work (min)	TW	Travel time to work per employee in minutes
Distance to any city (km)	DA	Distance from a centre of each block to the nearest city, town or village in km
Distance to major city (km)	DM	Distance from a centre of each block to the nearest city with more than 50 000 population in km
Distance to major road (km)	DR	Distance from a centre of each block to the nearest primary highway with limited access, interstate highways and toll highways, in km
Road index (km)	RI	Total distance of all roads in km within 1 km^2 of area
Environmental Variables		
Distance to major open space (km)	DO	Distance from a centre of each block to the nearest major open space including national park service land, national forest or other federal land, state or local parks or forests in km
Distance to lake (km)	DL	Distance from a centre of each block to the nearest major lake or reservoir in km
Pollution	PL	NO_2 level
Elevation (km)	EL	Mean elevation of each block in km
Stream index (km)	SI	Total distance of streams and rivers of each block in km within 1 km^2 of area
Open space index (%)	OS	Ratio of total area of major open space to total area of each block

area of the block, is created to measure the effects of both relative abundance and existence of open space in the neighbouring blocks. Similarly, the stream index and road index are created to measure the effects of both relative abundance and existence of streams and roads respectively in the neighbouring blocks.

Table 2. Mean values of variables for different communities

	Urban-dominant	Urban-moderate	Rural-moderate	Rural-dominant
Dependent Variables				
Housing count (per km^2)	0.0004	0.0003	0.0001	0.00002
Socio-economic Variables				
Housing value ($1000)	59.990	63.337	63.745	56.245
Income ($1000)	12.51	13.31	12.69	11.09
Population density (per km^2)	0.093	0.063	0.014	0.006
Crime rate	137.64	110.87	49.04	45.97
Stability (%)	0.51	0.55	0.59	0.64
Education (year)	11.77	11.84	11.51	11.27
Political view (%)	0.41	0.41	0.42	0.43
Travel time to work (min)	16.45	16.58	19.24	20.56
Distance to any city (km)	2.68	3.01	4.57	7.43
Distance to major city (km)	32.73	40.65	52.36	64.11
Distance to major road (km)	4.46	6.78	10.01	18.52
Road index (km)	0.033	0.028	0.014	0.010
Environmental Variables				
Distance to major open space (km)	15.12	16.69	16.74	17.07
Distance to lake (km)	6.04	6.43	6.05	7.39
Pollution	97.99	94.33	91.97	92.54
Elevation (km)	0.38	0.39	0.42	0.56
Stream index (km)	0.004	0.004	0.004	0.004
Open space index (%)	0.004	0.002	0.001	0.001

The aggregate housing demand equations are estimated using cross-sectional data. Because the block size and characteristics of residential decisions differ across blocks, heteroscedasticity is likely to be present. The null hypothesis of no heteroscedasticity was tested using the Lagrange Multiplier (LM) test suggested by Greene (1997, pp. 653–658). The null hypothesis is rejected at the 1 per cent significance level for each equation. Heteroscedasticity was corrected using the technique suggested by Kmenta (1986, pp. 270–276). The transformed equation system was then estimated using the SUR estimator.

It is a challenge to incorporate all the independent variables for the housing demand equations because they may be collinear. Although there have been many suggestions about how to detect multicollinearity, there are no certain guidelines. A commonly used rule of thumb is that if the correlation coefficient between the values of two regressors is greater than 0.8 or 0.9, then multicollinearity is a serious problem (Judge *et al.*, 1982, p. 620). Few of the correlation coefficients are shown to be close to 0.8 (e.g. correlation between housing values and education level, income and education level, housing values and income, and road index and population density). Testing for the seriousness of the multicollinearity was carried out by deletion of the regressors involved with high correlation coefficients. No serious fluctuations were detected in the coefficients, nor serious changes of statistical significance resulting from the deletion of the regressors with high correlation coefficients. Thus, the suspected multicollinearity does not appear to be a serious problem in the aggregate housing demand equation.

Study Area and Data

The study area is the Blue Ridge region of the Southern Appalachian Highlands which includes all of the mountainous portions of western North Carolina, northern Georgia, southeastern South Carolina, eastern Tennessee, southwestern Virginia and southeastern West Virginia. This region makes up 3687 blocks of the 1990 US Census (see Figure 1). The eastern portion of the region is dominated by the Blue Ridge Mountains, which rise abruptly from the Piedmont province, forming a rugged and diverse landscape. Regionwide, the area of developed land has increased considerably over the past 20 years. Much of this development has been at the expense of cropland and pasture. Although the region has the greatest concentration of federally-owned land in the eastern USA, the vast majority of the region's land is privately owned. The population of the region increased by 27.8 per cent between 1970 and 1990. Despite this growth, the population density in the study area remains below the average for the six states that contain the study area (US Forest Service, 1996).

Two principal data sources were used in this study: Applied Geographic Solutions, Thousands Oaks, California, which collects demographic, housing, crime risk and pollution data from the US Census, the FBI and the EPA; and Geography Network, a web service which provides geographic data from the Environmental System Research Institute (ESRI), Redlands, California. ArcView GIS software was employed to generate the database, using the data from the two principal sources. Distance calculations were made using a raster system where all data were arranged in grid cells. Distances were measured as the Euclidean distance from the centroid of a census block to the nearest edge of a feature. The sum of length and the sum of area were calculated using ArcScripts, downloaded from ESRI.

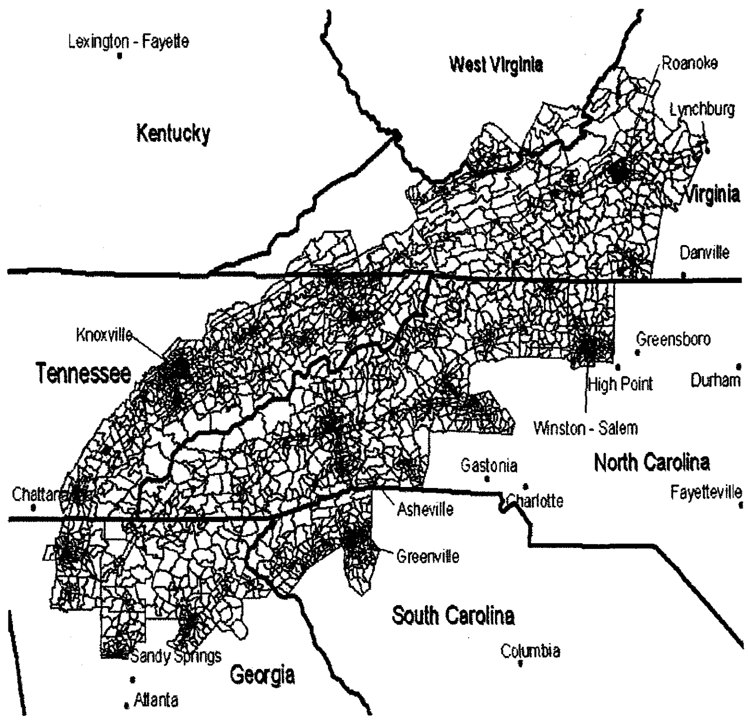

Figure 1. Study area

Although most people intuitively think of census blocks as being rectangular or square, of about the same size, and occurring at regular intervals, as in many large cities of the USA, census block configurations actually are quite different. The pattern, size, and shape of census blocks vary within and between areas. Factors that influence the overall configuration of census blocks include topography, the size and spacing of water features, the land survey system and the extent, age, type, and density of urban and rural development. The census blocks in remote areas may be large and irregular and may contain many square miles (US Census Bureau, 1990).

The dependent variable of the community choice model is a community index. The index to classify each block was constructed into different communities. The classification of the four communities is based on information about housing types from the US Census. The US Census divides housing types into urban core, urban non-core, rural farm and rural non-farm, based on the population of each block. Specifically, the ratios of housing types of urban core and urban non-core to all housing types were calculated for each block. A block is identified as an urban-dominant, urban-moderate, rural-moderate and rural-dominant community based on the ratio of the housing types. Definition of variables and mean values of variables for different communities are shown in Tables 1 and 2 respectively. Number of blocks and percentage of total study land area for the four communities are given in Table 3.

Table 3. Number of blocks and percentage of total study land area for different communities

	Urban-dominant	Urban-moderate	Rural-moderate	Rural-dominant
% of urban core and urban non-core housing types	100%	Greater than or equal to 50% and less than 100%	Greater than 0% and less than 50%	0%
Number of blocks (% of total study land area)	554 (1%)	1027 (6%)	495 (10%)	1611 (83%)

Estimation Results

Estimates of the Community Choice Model

Descriptive statistics of predicted-normalised value of community characteristics, $\Delta Z_b = Z_b^* - \hat{Z}_b^*$, used in the multinomial-logit model are shown in Table A1 in the Appendix. Parameter estimates and marginal effects of independent variables on the choice of urban-moderate, rural-moderate and rural-dominant communities for the multinomial logit model are presented in Table 4. Coefficients for the urban-dominant community are not presented because they are constrained to be 0 as explained in the 'The Community Choice Model' section.

The results show that community choice is significantly affected by educational level, a household characteristic. Educational level is positively correlated with a choice of urban-moderate community, but it is negatively correlated with choices of rural communities. Households are more likely to choose to locate in the urban-moderate communities, but they are less likely to choose to locate in the rural communities if education level of the households in the communities is higher. The marginal effect of education is 0.06 per cent in the rural-moderate model and it increases to 0.24 per cent in the rural-dominant model. This means higher education level pulls households away from rural-moderate communities, and even more so from rural-dominant communities. Political view is also correlated with choices of the rural-moderate and rural-dominant communities. Conservative households are more likely to choose to locate in rural communities. These results indicate that more educated households tend to choose to locate in urban-moderate communities, while a greater number of households are attracted to rural communities with more conservative political views.

The results also show that community choice is significantly affected by population density, a community attribute. Population density is negatively correlated with the choices across all three communities. Households are less likely to choose to locate in all three types of communities consistently if the population density of the communities increases. The marginal effect of population density is 0.95 per cent in the rural-dominant community model, which is greater than the 0.35 per cent of rural-moderate and 0.57 per cent of urban-moderate model. The higher marginal effect of population density in the rural-dominant model may be explained by households' greater degree of discomfort with higher population densities in rural-dominant communities. The community choice is significantly affected by crime rate in the urban-moderate model. An increase in the crime rate decreases the choice of urban-moderate communities. The insignificant crime rate in the rural-community model reflects the minor role of a lower crime rate in the rural area.

Households are more likely to choose to locate in all three types of communities if the communities hold more stable neighbours. Households are more likely to choose to locate in the rural-dominant communities if the rural-dominant communities have a lower level

Table 4. Parameter estimates and marginal effects for the multinomial logit model of community choices

	Urban-moderate		Rural-moderate		Rural-dominant	
	Parameter estimates	Marginal effects	Parameter estimates	Marginal effects	Parameter estimates	Marginal effects
Constant	6.1529**	−0.0150**	−1.1680*	−0.9912**	10.6391**	1.9369**
Education	0.0112**	0.0013**	−0.0115*	−0.0006*	−0.0252**	−0.0024**
Political view	0.0049	0.0002	0.0081*	0.0003*	0.0073*	0.0008*
Population density	−0.1452**	−0.0057**	−0.1505**	−0.0035**	−0.1463**	−0.0095**
Crime rate	−0.0028**	−0.0011**	0.0071	−0.0008	0.0021	0.0004
Stability	0.0003**	0.0010**	0.0024**	0.0001**	0.0067*	0.0015*
Pollution	0.0037	0.0003	0.0061	0.0005	0.0016*	−0.0004*

Log likelihood, −4689.60; **indicates statistical significance at the 1% level; *indicates statistical significance at the 5% level. The unit of community characteristics in the model is deviation from the predicted-normalized values, multiplied by 1000 (see equations (5) and (6)).

of air pollution. However, the households' choices of the urban-moderate and rural-moderate communities are not significantly affected by air quality. These findings suggest that living with cleaner air quality is a more significant concern to households who choose to locate in rural-dominant communities than in urban-moderate or rural-moderate communities.

Estimates of the Housing Demand Model

Estimates of the equation systems of housing demand and housing price for the housing demand model are listed in Table 5 and Table A2. It should be noted that the signs of the statistically significant coefficients in the housing demand equation are consistent with the signs of the statistically significant coefficients in the housing price equation. The focus is on a discussion of the results of the housing demand models for the four different types of communities in Table 5. Of the 76 housing demand coefficients (19 variables in each of the 4 equations), 38 are significant at the 5 per cent level. The system weighted R^2 is between 0.86 and 0.93.

The self-selection variables are taken from the multinomial logit model. There is substantial evidence that self-selection occurred in the households' community choices. The coefficients of the self-selection variable λ are statistically significant at the 5 per cent level in all the communities. This suggests community choices would not have the same effects on housing demand. This implies a distinctive heterogeneity in the characteristics found in the community types observed in the region. The coefficient of the self-selection variable is negative in the rural-moderate community while it is positive in the rest of the communities. The interpretation of negative self-selection can be attributed to the least probability of households' choices of the rural-moderate community.

To determine if the estimated coefficients are statistically different by the four types of communities, a Chow test is used to test the hypothesis that the regression coefficients are the same across the communities. This tests the null hypothesis of equal disturbance variances from the different regressions. If they are not the same, heteroscedasticity exists in the estimation of the pooled data. The F-values for pair-wise tests that the regression coefficients are the same between the four communities are 5.23, 6.02 and 4.55 respectively. All the F-values are greater than the critical value, 1.57, so the hypothesis that all the regression coefficients are the same in the four types of communities at the 5 per cent level is rejected. Based on these tests, it is concluded that housing demand values differ under various socio-economic and environmental variables in the four types of communities.

The parameter estimates of the housing demand equations show that variables affecting housing demand vary across the communities. Housing demands are affected more by socio-economic variables in urban communities, while they are affected more by environmental variables in rural communities. Of the 24 socio-economic coefficients (12 variables in each of the dominant and moderate equations), 15 in the urban communities and 11 in the rural communities are statistically significant at the 5 per cent level. Of the 12 environmental coefficients (6 variables in each of the dominant and moderate equations), no variables in the urban communities and 8 in the rural communities are statistically significant at the 5 per cent level.

Table 5. Parameter estimates for the housing demand equations for alternative community choices

	Urban-dominant	Urban-moderate	Rural-moderate	Rural-dominant
Constant	−0.3534*	0.1332**	−0.0399*	−0.06421**
Socio-economic Variables				
Housing value	−0.0016**	−0.0003**	−0.0001**	−0.00006**
Income	0.0064**	0.0036**	0.0003	−0.00004
Population density	3.4255**	3.7256**	3.2177**	3.1957**
Crime rate	−0.0001*	0.00002	−0.00003	−0.00002*
Stability	−0.1125**	−0.0130	0.0046	−0.0019
Education	0.0221**	−0.0173**	0.0019	0.0039**
Political view	−0.0484	−0.0040	−0.0067	−0.0020*
Travel time to work	0.0028*	0.0013*	0.0001	0.0001**
Distance to any city	0.0028	0.0011	−0.0004	0.0001
Distance to major city	−0.00001	−0.00002	−0.00002	0.00002**
Distance to major road	−0.0025**	0.0001	0.0001	0.00002
Road index	5.3893**	2.4607**	1.8707**	1.2484**
Environmental Variables				
Distance to major open spaces	0.0002	−0.00004	0.00005	0.00003*
Distance to lakes	−0.0002	−0.0003	−0.0002**	−0.00003**
Air pollution level	−0.0004	−0.0004	−0.0068	0.00006
Elevation	−0.0331	0.0181	0.0117**	0.0053**
Stream index	−1.4739	−0.6132	1.8363**	1.1552**
Open space index	0.0957	0.0495	0.0635	0.0115**
Self-selection Variable				
λ	0.0337*	0.0036*	−0.0091**	0.0007**
System weighted R^2	0.91	0.88	0.86	0.93
Sample size	554	1027	495	1611

**indicates statistical significance at the 1% level; *indicates statistical significance at the 5% level.

The effects of socio-economic variables on housing demands across urban and rural communities also vary, although differences in socio-economic effects are not as drastic as differences in environmental effects. Population density in both urban and rural-dominated communities commonly affects housing demands in both urban and rural-dominated communities. A higher population density requires more housing. The marginal effects of population density on the urban communities are higher than those of the rural communities. This suggests that an equal increase in population density increases housing demand more in the urban communities than it does in the rural communities. This finding may provide evidence that housing developments in urban communities are more responsive to increased population than housing developments in rural communities. Alternatively, it may simply reflect the fact that rural households tend to have more people in them than in urban households.

A lower crime rate and higher levels of education attract more housing, both in urban-dominated communities and rural-dominated communities. The marginal effects of these two variables in urban communities are higher than those in rural communities. They indicate that safety and the education level of the community are common concerns of urban and rural households, but the degree of the concern is greater in urban communities.

A less conservative political viewpoint is correlated with more housing in rural-dominant communities. This reflects the fact that more conservative neighbours are inclined toward land-use policy that may restrict residential development in a rural-dominant community. The coefficient for travel time to work is positive and statistically significant at the 1 per cent level in rural-dominant community. This indicates that an increase in travel time to work increases housing demand in rural-dominant community. This surprising result may be explained by the fact that there are many retirees and second homeowners who enjoy better environmental amenities at greater distance to work in rural-dominant community. The coefficient for the road index is positive and statistically significant at the 1 per cent level in all the communities. This suggests that road accessibility is important to houses in any type of community.

Housing value, income, stability, and the distance to major roads have significant effects on housing demand in urban communities. Housing demand is negatively associated with housing value in all types of communities. This is evidence that supports the notion that the law of demand, the inverse relationship between price and quantity, is at work in the housing market. Housing demand is positively associated with income in urban communities. It is negatively associated with the stability of households in urban-dominant communities. This may imply that households in urban communities are younger and closer to the beginning of career and family path, thus more mobile. Housing demand is higher in urban-dominated communities, where the houses are closer to a major road.

Four of six environmental variables are statistically significant at the 1 per cent level in the rural-dominated communities. Households are more likely to locate in the blocks that are closer to lakes, at higher elevations, and with greater access to streams and open space within the rural-dominated communities. Environmental variables did not have a substantial impact on the housing demands of urban communities. Clear differences in the effects of environment factors on housing demands between urban and rural communities imply heterogeneity in the characteristics found in the community choices observed in the region; this confirms significant self-selection.

All coefficients for the distance to a lake are negative across the urban and rural communities, although the coefficients of only the rural communities are significant at the

1 per cent level. This shows that both urban and rural households enjoy the environmental amenities of lakes but the attractions are only substantial to rural households. Elevation and access to streams are statistically significant at the 1 per cent level in both rural-moderate and rural-dominated communities. This indicates that the environmental amenities of higher elevation and a greater access to streams draw a substantial number of households to rural communities. The coefficient for the open space index is positive and statistically significant at the 1 per cent level only in rural-dominated communities. This suggests that access to open space is significantly important only to rural-dominated households.

Distance to the closest city is not a significant factor across the communities, and distance to the closest major city is not a significant factor in urban communities. This result may be explained by the relatively smaller and fewer cities observed in the region. The impact of distance to the closest major city is positive and significant at the 1 per cent level only in rural-dominant communities. This implies that rural-dominated households enjoy remoteness more than the positive utilities of being close to major cities. Air pollution is not a significant factor in housing decisions across the communities, perhaps reflecting that air quality under each community choice of the region is relatively homogeneous. Thus, the air quality is not a significant factor of housing choice within each community, even though it is a significant factor of alternative community choices, as shown in the estimates of the community model.

Conclusion

This paper develops an econometric model that incorporates housing demand within a selected community through the application of spatial statistics with GIS. Census block data are used to present a coherent multi-scale model of housing demand in the Southern Appalachian region.

The first-stage analysis yields estimates of the marginal effects of household characteristics and community attributes in community choices. It was found that more conservative, less crowded, safer and more stable communities attract more households regardless of types of communities while more educated communities attract more households in the urban-moderate communities and cleaner air quality attracts more households in the rural communities. The second-stage analysis yields the marginal effects of the socio-economic and environmental characteristics in the residential choices for different communities. There is a distinctive heterogeneity of the characteristics found in the community choices observed in the region. The socio-economic motives of urban communities and the environmental motives of rural communities are more weighted in their housing decisions. Specifically, housing development in urban communities is more responsive to increased population density than housing development in rural communities. Safety and the education level of the community are a greater concern to urban households. The law of demand, the inverse relationship between price and quantity people want to buy, is at work in the housing market. The higher income in urban communities attracts more housing. Households in urban communities are younger and closer to the beginning of career and family path, and are thus more mobile. Houses are more likely to be closer to a major road in urban-dominated communities. On the other hand, the environmental amenities of proximity to a lake, higher elevation, greater access to streams and greater access to open spaces draw a substantial number of households to rural communities.

According to the community model estimates, policy makers could build programmes which encourage or discourage housing development in their community depending on the characteristics or preferences of the community. For example, policy makers could encourage more housing developments in the community by developing a programme lowering crime rate and increasing the stability of the communities. A programme improving air quality would be helpful for encouraging development in rural communities. They could also use the estimates from the housing demand model to predict future housing density for different types of communities given projected factors for housing demand. For example, an increase in population density increases housing demand more in the urban communities than it does in rural communities. Accordingly, policy makers in urban communities would need to reserve a greater budget for infrastructure expansion resulting from the anticipated increase in housing demand when compared with rural communities with the same increase in population density.

Based on the results of this study, growth drivers play out in distinctive ways in different community types. These distinctively different growth drivers imply that growth of an area has to be managed differently according to community type. These findings indicate that as development proceeds, shifts between community types will bring changes in their social structures. These changes will probably give rise to conflict as development proceeds and will have implications for how subsequent development might be organized across a landscape.

The next logical step of the analysis is in the resolution of the block level in the site-specific housing choice model. Housing choices at an individual level could be used for a better analysis of more-fine-scale units if the individual housing data were readily available. This dataset could be built using a database of individual houses from county tax assessors' offices, the census dataset of block levels, and the GIS database that could be created using information about individual houses. While collecting a dataset from the 98 counties of the Southern Appalachian region would be extremely expensive, a sample study for some selected counties in which all the types of communities are contained might be feasible.

Another extension of this research would be to develop predictive models of land-use choice that incorporate socio-economic and environmental influences at the micro level. Another direction for further research would be to address the conflict between long-time residents and newcomers to the region. This region is increasingly divided into social structures of long-time residents and newcomers who move to this area mainly in pursuit of retirement, vacation homes and second homes. The interests of these two groups conflict in many ways, including in the area of housing decisions. The models used in this study can be modified to investigate the heterogeneity and potential dynamics of these two groups in the area.

Acknowledgements

The authors would like to thank Barrie Collins and Tripp Lowe for their data analysis assistance, and Sara Cho, Dawn Black, Carol Hyldahl, and two anonymous reviewers for their useful comments and suggestions.

This material is based upon work supported by the NSF funded Coweeta Long Term Ecological Research, the USDA Forest Service, and the Warnell School of Forest Resources of the University of Georgia.

References

Alig, R. (1986) Econometric analysis of the factors influencing acreage trends in the southeast, *Forest Science*, 32, pp. 119–134.

Can, A. (1998) GIS and spatial analysis of housing and mortgage markets, *Journal of Housing Research*, 9, pp. 161–186.

Cao, T. & Cory, D. (1981) *Mixed Land Uses and Residential Property Values in the Tucson Metropolitan Region: Implication for Public Policy*. Technical Bulletin No. 241 (Arizona: Agricultural Experiment Station, University of Arizona).

Chomitz, K. & Gray, D. (1996) Roads, land use, deforestation: a spatial model applied to Belitz, *World Bank Economic Review*, 10, pp. 487–512.

Dale, V., O'Neill, R. & Pedlowski, M. (1993) Causes and effects of land-use change in Central Rondonia, Brazil, *Photogrammetric Engineering and Remote Sensing*, 59, pp. 997–1005.

Ding, C. (2001) An empirical model of urban spatial development, *Review of Urban and Regional Development Studies*, 13, pp. 173–186.

Feridhanusetyawan, T. & Kilkenny, M. (1996) *Rural/Urban Residence Location Choice*. Report 96-WP 157 (Iowa: Center for Agricultural and Rural Development, Iowa State University).

Geoghegan, J., Weinger, L. & Bockstael, N. (1997) Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS, *Ecological Economics*, 23, pp. 251–264.

Goodman, A. (1977) A comparison of block group and census tract data in a hedonic housing price model, *Land Economics*, 53, pp. 483–487.

Greene, W. (1997) *Econometric Analysis* (Upper Saddle River, NJ: Prentice Hall).

Hardie, I. & Parks, P. (1997) Land use with heterogeneous land quality: an application of an area base model, *American Journal of Agricultural Economics*, 79, pp. 299–310.

Hardie, I., Parks, P., Gottleib, P. & Wear, D. (2000) Responsiveness of rural and urban land uses to land rent determinants in the US south, *Land Economics*, 76, pp. 659–673.

Ioannides, Y. & Zabel, J. (2003) Neighborhood effects and housing demand, *Journal of Applied Econometrics*, 18, pp. 563–584.

Judge, G., Hill, C., Griffiths, W., Lütkepohl, H. & Lee, T. (1982) *Introduction to the Theory and Practice of Econometrics* (New York: John Wiley and Sons).

Kmenta, J. (1986) *Elements of Econometrics* (New York: Macmillan).

Lake, I., Bateman, I., Day, B. & Lovett, A. (2000) Improving land compensation procedures via GIS and hedonic pricing, *Environment and Planning C: Government and Policy*, 18, pp. 681–696.

Lee, L. (1983) Generalized econometric models with selectivity, *Econometrica*, 51, pp. 507–512.

Maddala, G. (1983) *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press).

Mahan, B., Polasky, S. & Adams, R. (2000) Valuing urban wetlands: a property price approach, *Land Economics*, 76, pp. 100–113.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior, in: P. Zarembka (Ed.) *Frontier of Econometrics* (New York: Academic Press).

Meen, G. & Andrew, M. (1998) On the aggregate housing market implications of labour market change, *Scottish Journal of Political Economy*, 45, pp. 393–419.

Miller, D. & Plantinga, A. (1999) Modeling land use decisions with aggregate data, *American Journal of Agricultural Economics*, 81, pp. 180–194.

Nechyba, T. & Strauss, R. (1998) Community choice and local public services: a discrete choice approach, *Regional Science and Urban Economics*, 28, pp. 51–73.

Nelson, G. & Hellerstein, D. (1997) Do roads cause deforestation? Using satellite images in econometric analysis of land use, *American Journal of Agricultural Economics*, 79, pp. 80–88.

Plantinga, A. (1996) The effect of agricultural policies on land use and environmental quality, *American Journal of Agricultural Economics*, 78, pp. 1082–1091.

Rapaport, C. (1997) Housing demand and community choice: an empirical analysis, *Journal of Urban Economics*, 42, pp. 243–261.

Spies, T., Ripple, W. & Bradshaw, G. (1994) Dynamics and pattern of a managed coniferous forest landscape in Oregon, *Ecological Applications*, 4, pp. 555–568.
Turner, M., Wear, D. & Flamm, R. (1996) Land ownership and land-cover change in the Southern Appalachian Highlands and the Olympic Peninsula, *Ecological Applications*, 6, pp. 1150–1172.
US Census Bureau (1990) Technical Documentation for Census of Population and Housing 1990 (Washington DC).
US Forest Service (1996) The Southern Appalachian Assessment Online Reports, coordinated through Southern Appalachian Man and Biosphere Program.
Wear, D. & Bolstad, P. (1998) Land-use changes in Southern Appalachian landscapes: spatial analysis and forecast evaluation, *Ecosystems*, 1, pp. 575–594.
Wear, D. & Flamm, R. (1993) Public and private disturbance regimes in the Southern Appalachians, *Natural Resource Modeling*, 7, pp. 379–397.

Appendix

Table A1. Descriptive statistics of predicted-normalized value of community characteristics, $\Delta Z_b = Z_b^* - \hat{Z}_b^*$, used in the multinomial logit model

	Mean	Standard Deviation	Min	Max
Education	0.0291	5.2126	−68.6611	31.0084
Political view	0.0001	10.3254	−39.8422	59.7255
Population density	0.0073	1.7475	−15.1089	88.9008
Crime rate	−0.0104	11.0954	−67.4531	82.3242
Stability	−0.0996	11.7744	−63.9010	36.0742
Pollution	0.0298	9.4986	−26.1140	31.2493

Table A2. Parameter estimates for the housing price equations for alternative community choices

	Urban-dominant	Urban-moderate	Rural-moderate	Rural-dominant
Constant	− 38.4639*	− 66.8771**	− 111.7560*	− 96.9941**
Socio-economic Variables				
Housing density	− 26.2025**	− 20.9760**	− 83.7765**	− 180.6060**
Income	3.6826**	3.0769**	2.8065**	3.2846**
Population density	13.5963**	9.5812**	27.3733*	38.3529**
Crime rate	− 0.0014	− 0.0027	0.0787	0.0111
Stability	− 21.0089**	− 39.9708**	− 43.6609**	− 39.5169**
Education	6.1074**	− 7.7764**	11.9162	9.4891**
Political view	2.7665	10.0521	− 11.6179	1.6797
Travel time to work	0.1611	0.4277	0.5126*	0.0926
Distance to any city	− 0.6631	− 0.3542	0.0129	0.3754
Distance to major city	− 0.0689**	− 0.0314	− 0.0204	0.0183*
Distance to major road	0.2051**	0.0933	0.0053	0.0577**
Road index	99.0902	209.9210**	97.7989	724.6633**
Environmental Variables				
Distance to major open spaces	0.0937	0.0831	− 0.0594	− 0.0840**
Distance to lakes	− 0.1798	− 0.0225	0.0467	− 0.0985**
Air pollution level	− 0.1257	0.0644	0.0497	0.1324
Elevation	4.0689	9.3571**	3.5778	8.3901**
Stream index	− 234.3440	74.5373	420.2009	791.0950**
Open space index	− 8.0645	48.8744**	52.3405	3.9760**
Self-selection Variable				
λ	1.1595*	1.4014*	3.5834**	1.8354**
System weighted R^2	0.79	0.75	0.79	0.70
Sample size	554	1027	495	1611

**indicates statistical significance at the 1% level; *indicates statistical significance at the 5% level.