Explaining Household Location Choices Using a Spatial Probit Model*

Erdogan Ozturk and Elena G. Irwin[1]

Corresponding Author:
Erdogan Ozturk
Department of Agricultural, Environmental, and Development Economics,
The Ohio State University
2120 Fyffe Road Columbus, OH 43210
Phone: 614-292-9516
Fax: 614-202-0078
Email: ozturk.2@osu.edu

May 2001

Abstract

A spatial probit model is estimated to test the influence of public services and neighborhood characteristics on relocation decisions of homebuyers in the Columbus, Ohio region. The model explicitly accounts for the spatial error autocorrelation that arises due to unobserved similarities across alternatives, which, if uncontrolled, results in inconsistent parameter estimates.

* Paper prepared for presentation at the 2001 meeting of the AAEA.

[1] Authors are Ph.D. student and assistant professor at the Department of Agricultural, Environmental, and Development Economics at Ohio State University, respectively.

Introduction

The operation of a metropolitan area's housing market influences the well-being of residents and communities in a variety of ways. By moving from place to place within a metropolitan area, households make adjustments that may result in home ownership, a better location, or some other change that influences overall household utility. Although moving decisions are the result of private reasons, the collective decisions of individuals to move have public effects on both the places from where people come and on the places to which they move. Obvious examples include a declining tax base to fund minimal levels of public services and falling property values in places rapidly losing population. If people move away from an area, local businesses may lose money and may be forced to shut down their operations or move elsewhere. Alternatively, areas to which people move may experience congestion and struggle to provide adequate supplies of public infrastructure and services. Lastly, although the effects of household migration on places are more subtle, the impact of residential mobility on the size and composition of a neighborhood is a prime determinant of an area's economic and social well being.

Housing is a special kind of commodity for several reasons. First, housing is spatially immobile; its location is an intrinsic attribute of a dwelling. So, one component of the housing bundle is the residential location. This implies that the housing market is inherently affected by neighborhood externalities. With any purchase of a home, the buyer purchases both a dwelling and a set of neighborhood characteristics such as accessibility, local public services, environmental quality, and appearance of the neighborhood. A second particular feature of housing is that it is not a single commodity but a bundle of variously related commodities that are heterogeneous. Housing units differ in structural characteristics, lot features, the neighborhood, local public services, and in accessibility to desired destinations. Differences in all these features effect the location decision of homebuyers and make for a wide spectrum of degrees of substitutability among dwellings. The heterogeneity of housing also introduces important measurement problems. A third feature that distinguishes housing from most other goods is its expense. The typical household must borrow a large sum of money to acquire a house. So, the ownership of a housing asset represents a considerable portfolio management. Other features that distinguish housing are durability and transaction costs. Housing is extremely durable, but existing houses are modifiable in various ways. They can be physically changed in terms of their condition or various structural features. Finally, changes in ownership are very costly. The costs of buying a house include the search costs incurred from choosing among houses that are immobile and potentially quite heterogeneous; the complex legal and transactional services involved in purchasing a home; and the time and money costs associated with moving.

Although historically nearly two-thirds of moves occur within the same metropolitan area (Quigley and Weinberg, 1977; Simons, 1974), empirical studies have largely focused on interregional movements. This study takes advantage of survey data on repeat homebuyers to understand the relocation decision of households that moved within a local region. Repeat homebuyers are a major part of the residential real estate market in the United States (Morrow-Jones and Lipsetz, 2000). Having owned a home before gives these buyers some experience and better idea of what is important to them. The basic hypothesis of this study is that a household's decision of where to move will be influenced by a variety of factors that are associated both with the location that they are leaving and the location to which they are moving. Some of these factors may relate to the house itself e.g. size, age, and yard characteristics. Other factors relate to the location of the house, including neighborhood characteristics and the services that are available at that location. These include public services (e.g. police and fire protection and schools), private services (e.g. retail shopping and business services), and other amenities (e.g. cultural and recreational opportunities).

What distinguishes this research from previous analyses of residential location choice is the explicit consideration of the spatial dimension of the intrametropolitan location decision within a discrete choice framework. Because of the immobility of housing and the importance of its location relative to other features of a metropolitan area, the spatial dimension of household movements is important. For example, to what extent does the geographical location of a house matter vs. other locational features such as quality of public goods and environmental amenities? From a methodological perspective, this study considers the possibility that households perceive similarities among alternative location choices, which invalidates the assumption of independent errors previously maintained in discrete choice models of household migration. In reality, it is likely that similarities across neighboring locations create a spatially dependent error structure, which, if ignored, will lead to inconsistent estimates in a discrete choice setting. To allow for this possibility, this study estimates a spatial binary probit model of household relocation among school districts in Franklin County, Ohio that explicitly incorporates a spatially dependent error structure. Specifically, we model the choice of a household to stay within or move outside of their existing school district.

Determinants of Household Location Decisions

Earlier literature on local residential mobility has emphasized life-cycle factors as critical determinants of the decision to move. From this perspective, life-cycle changes in the size, age composition, and socioeconomic position of households creates dissatisfaction with the current residence, which influences the demand for a different type of housing or geographic location and ultimately leads to decision to move. Age has been found to be one of the most important of these life-cycle variables. The presence of children living in the household has been found to deter relocation whereas household crowding, in contrast, has been found to encourage relocation. Socioeconomic characteristics have also been linked to local residential mobility, although the evidence is less conclusive, e.g. education-level has only moderately been related to local residential mobility.

The spatial dimensions of the housing market have also provided an important theoretical tool for studying interactions between local government and household location choice during the last decades. The early work of Tiebout (1956) continues to prompt research based on his market model for explaining the behavior of both individuals and local governmental units operating within spatially defined political jurisdictions located in urban areas. Since Tiebout's article, many economists have viewed the decision of families to reside in a particular community as a conscious choice of one particular package of local public services over others (Friedman, 1981). Tiebout suggested that under certain conditions consumers might reveal their preferences for locally provided public goods. Tiebout proposed that a market-like mechanism might exist for local public goods because local public goods and the taxes used to finance those goods are specific to each jurisdiction. He further suggested that households would sort themselves such that all families in a given jurisdiction would derive the same marginal benefit from the local public goods, and marginal benefit would equal to marginal cost.

An alternative approach to modeling household location decisions is by considering the discrete set of location alternatives from which a household chooses. Discrete choice models of the housing decision were first estimated by Quigley (1976 and 1978), who was the first to combine McFadden's (1978) individual discrete choice model and the structure of the housing market into one approach. He extended the theoretical analysis of the demand for housing with the construction of a multinomial logit model for the combined choice of residential location and housing type. Williams (1979) extended Quigley's model by considering neighborhood quality as an important component of the housing bundle. Alternatively, Friedman (1975) constructed a model of residential location choice in which the objective of choice is a community. Households are assumed to survey all the communities within their feasible set of communities and to select the community that maximizes indirect utility function that includes local public services, the local property tax, neighborhood characteristics, commuting distance and housing services. Friedman finds that the housing service variable has the largest coefficient of any of his estimates. In fact other explanatory variables, including local public services, appear to have little or no affect on choice of community. Other multinomial estimations of housing decisions include Lerman (1976), in which he estimated a joint choice model of location, housing, and mode to work. In this model, households choose a housing bundle simultaneously with their choice of location, work trip, and spatial opportunities (defined as including accessibility to shopping and other non-work destinations). Their choices are also allowed to reflect their own characteristics, including income, race, household size, and number of drivers.

In a recent study, Nechyba and Strauss (1997) apply a discrete choice approach to estimate the impact of local fiscal and other variables on individual community choices using a micro data set of homeowners and information on local community characteristics. In their model, consumers choose among different house types in different communities, taking characteristics of the communities, local public good levels, tax rates and equilibrium house prices into account. Based on their estimation results, they conclude that information on local community characteristics play major part in explaining the location of individual households.

Spatial Autocorrelation in the Household Location Model

Autocorrelation refers to the similarity or dissimilarity of values of the same variable associated in different intervals. Although this notion is better known from the field of time series, analysis, the intervals (*lags*) may be of temporal or spatial nature. It is very natural for data observed in space to have similar values for adjacent spatial units. For example, if the location of the house influences its price, then the possibility arises that nearby houses will be affected by the same location factors. This clustering of similar values in geographical space is a form of spatial dependence or spatial autocorrelation (Cliff and Ord, 1973).

There are several reasons to expect spatial dependence in a discrete choice model of household location. First, in any situation with a large number of alternatives, it seems clear that individuals partition the choice set into clusters of alternatives. It is usually easier to identify clusters of alternatives in aspatial choice, such as brands of products, types of transit, etc. than it is in spatial choice. The difference arises because, whereas in aspatial choice the discriminating factor between alternatives is discrete (type of transit is either private or public), the discriminating factor between spatial alternatives is usually space itself, which is continuous. Boundaries of spatial clusters are more likely to be fuzzy than discrete, which causes problems in the application of choice models that need an *a priori* definition of alternatives.

Second, many spatial choice problems can be characterized as multi-dimensional. For example, the migration decision process can be broken down into the move/stay decision and, conditional on moving, the destination choice decision, which involves both the choice of a neighborhood and the choice of a particular dwelling unit. If this nested structure is ignored or if an incorrect nesting structure is assumed, spatial autocorrelation problems can arise. For example, in modeling intra-urban residential mobility, the error term is likely to include unmeasured characteristics of both neighborhood and dwelling types. Housing units within the same neighborhood will share the some similar characteristics, such as distance to the central business district (CBD), number of rooms, etc. So, unobserved similar characteristics of housing units across alternatives will affect the disturbance terms.

Third, availability of data at different spatial scales and different spatial distributions of data may also create this kind of correlation across alternatives. If the data for different variables are not measured within the same boundary, then there will be correlation across the alternatives. For example if the school district boundaries are chosen to represent alternatives within the choice set, other locational variables that enter the utility function (e.g.

distance to CBD, crime rate) need to be calculated by school district. These variables are generated by a different process than the process by which school district boundaries are determined and therefore will have heterogeneous values within the areas delineated by school districts. As a result, spatial association will exist among these variables across neighboring school districts.

The consequences of spatial autocorrelation are more severe in discrete choice models than in linear regression. Estimation results without incorporating spatial autocorrelation are not only inefficient, but also inconsistent because the model is heteroskedastic as well as autocorrelated. The heteroskedasticity problem arises because of the nature of spatial data as well as the spatially autocorrelated error structure. Because each error will be a function of surrounding errors, the error variance will be non-constant. To illustrate, assume the errors follow a simultaneous autoregressive spatial process:

$$Y = X b + u$$

$$u = l Wu + e$$

where Y is a vector of dependent variable, X is a vector of explanatory variables, b is vector of model parameters, l is the spatial autocorrelation coefficient to be estimated and W is an NxN weight matrix determined exogenously. The researcher takes W as *a priori*, and therefore all results are conditional upon the specification of W. The main diagonal of W consists of zeros and the off-diagonal elements, w_{ij}, represents the researcher's maintained assumption about the spatial relationship between observations i and j. The resulting error structure in this model is:

$$\mathrm{E}[ee'] = [(I - lW)'(I - lW)]^{-1} \tag{1}$$

As a consequence of the spatial weight matrix in the autoregressive specification, the random error at each location becomes a function of the random errors at all other locations as well. It is distribution is multivariate normal (for the probit model) with an NxN variance covariance matrix. Besides being non-diagonal, the diagonal elements of the variance covariance matrix are not constant and therefore the model is heteroskedastic. The probability of each alternative cannot be derived from the univariate standard normal distribution, but rather must be expressed explicitly as the marginal distribution of an N-dimensional multivariate normal vector, whose variance-covariance matrix contains the autoregressive parameter l. This is non-standard and typically not analytically tractable, which greatly complicates estimation and specification testing (Anselin, 1999).

Estimation Strategies

Unlike the linear regression case, for which large body of results already exists, spatial discrete choice models have also received much less attention in the spatial econometric literature. In general, there have been relatively few attempts incorporating a form of spatial dependence in discrete choice models largely due to the difficulty in adopting maximum likelihood estimation to models with limited dependent observations.[2] McMillen (1992) proposes and illustrates the estimation of a spatial probit model with spatial autocorrelation and heteroskedasticity using the Expectation/ Maximimization algorithm. He argues that spatial autocorrelation models imply heteroskedastic errors and that heteroskedasticity causes the standard probit model to be inconsistent. LeSage (2000) notes that McMillen's estimation procedure relies on asymptotic properties and requires large sample sizes for validity. He extends the non-

[2] For a full review of these models, see Fleming (2001).

spatial probit model introduced by Albert and Chib (1993) to a spatial probit model by applying the Markov Chain Monte Carlo (MCMC) method. He argues that this Bayesian method is more computationally demanding, but provides a more flexible framework for estimating models with small samples. Smith and LeSage (2000) extend LeSage's previous study by defining an additive error specification. They show that defining an additive error structure allows both spatial dependencies and general spatial heteroskedasticity to be treated simultaneously. Lastly, Pinkse and Slade (1997) develop a test for spatial-error correlation and methods of estimation in the presence of correlation for discrete-choice models using a generalized method of moments estimator.

In this paper, we use the MCMC method and the model developed by Smith and LeSage to estimate a binary discrete choice model of a household's decision to move within or outside of their original school district. Recent developments of Markov chain Monte Carlo (MCMC) techniques and Gibbs sampling have greatly facilitated many applications of Bayesian statistical techniques (LeSage, 1997; Bolduc, Fortin and Gordon,1997). MCMC methods are one such technique that have a Bayesian foundation and have proven to be a very useful approach to obtaining parameter estimates. Chib and Greenberg (1995) define the MCMC method as a simulation technique that generates a sample from the target distribution in the following way: the transition probability of a Markov process is specified with the property that its limiting distribution is the target distribution. The chain is then iterated a large number of times in a computer-generated simulation, and the output is a sample from the target distribution. In this sense, inclusion of the words Markov chain indicates that instead of generating a large sample of independent draws from the posterior distribution, each draw is related to the previous one.

There are three main MCMC algorithms: Metropolis algorithm (Metropolis et.al.1953), Hasting algorithm (Hasting, 1970) and Gibbs sampler (Geman and Geman, 1984). These algorithms differ according to the definition of acceptance probability in generated draws.[4] Because of its potential to fit a limitless range of models, Gibbs sampling has become a widely employed computational tool. The idea behind Gibbs sampling is to generate draws from a multivariate joint probability distribution by sequential sampling from a series of conditional probability distributions on lower-dimensional subsets of the random variables. In cases where it is difficult or impossible to generate random draws from the joint posterior distribution p(è|y, X), Gibbs sampling takes advantage of the fact that it is often possible to generate draws from a set of conditional probability distributions on these lower-dimensional subsets of random variables, due to simplifications that occur in the forms of the distributions (Dorfman, 1997). The steps to Gibbs sampling can be summarized as:

1. Begin with some initial values (guesses) for the parameter vectors $q_1, q_2 \ldots q_k$. Denote these initial values by $q_i^{0)}$.

2. Generate random draws in sequence from the conditional posterior distributions:

$$q_1^{(j+1)} \sim p(q_1 \,|\, y, X, q_2^{(j)}, \ldots, q_k^{(j)})$$

$$q_2^{(j+1)} \sim p(q_2 \,|\, y, X, q_1^{(j+1)}, \ldots, q_k^{(j)})$$

$$q_k^{(j+1)} \sim p(q_k \,|\, y, X, q_1^{(j+1)}, \ldots, q_{k-1}^{(j+1)})$$

3. Repeat step 2 many times, conditioning at each iteration on the most recently generated vectors for the other partitions. Discard the first J draws of parameters to avoid dependence on the initial values. Then save the remaining draws.

---

[4] Detailed information and comparison can be found in Besag (2000).

4. Check the chain for convergence to ensure it is safe to stop generating empirical observations on q .

LeSage (1997) and Smith and LeSage (2000) demonstrate that this approach proves useful for Bayesian estimation of spatial autoregressive models. When the data exhibits heterogeneity over space and spatial data contains outlying observations, the assumption of normality and the asymptotic arguments used to derive maximum likelihood estimates for the parameters are not met. Since Bayesian methods do not require a normality assumption, they have an advantage over other estimation methods that require the normality assumption. Lesage also argues that the Gibbs sampler is even more attractive for the case of probit and tobit models in which the likelihood function contains a number of integrals equal to the number of data observations, producing an intractable situation. Also, heteroskedasticity and outliers can be easily accommodated by applying the Gibbs sampler.

Empirical Model and Data

Following Smith and LeSage (2000), the choice model involving spatial agents is constructed as follows. It is assumed that decision makers are distributed within a system of spatial regions (in our case, school districts), i = 1,…,m,. The observed choice for each individual $k = 1,...,n_i$ in region i is the realization of random choice variable Y_{ik}. We define Y_{ik} as follows:

Y_{ik} = 1, if the household moves out of the school district and

Y_{ik} = 0, if the household move within same school district.

It is assumed that choices are based on utility maximizing behavior. Decision-maker k's utility for each of these two alternatives is assumed to be of the form:

$$U_{ik0} = gw_{ik0} + a_0' x_{ik} + q_{i0} + e_{ik0}$$
$$U_{ik1} = gw_{ik1} + a_1' x_{ik} + q_{i1} + e_{ik1} \qquad \text{i = 1,…,m and k = 1,…,}n_i \tag{2}$$

where w_{ika} is vector of observed attributes of alternative *a* and can vary across decision-makers and x_{ik} is vector of individual specific variables. The term $q_{ia} + e_{ika}$ is the error term and represents the unobserved variables that are not included in utility function. The error term is separated into a regional effect, q_{ia}, and an individual effect, e_{ika}. The utility difference for individual k can be denoted by

$$z_{ik} = U_{ik1} - U_{ik0}$$
$$z_{ik} = g'(w_{ik1} - w_{ik0}) + (a_1 - a_0)' x_{ik} + (q_{i1} - q_{i0}) + (e_{ik1} - e_{ik0})$$

or,

$$z_{ik} = \Psi_{ik}' b + q_i + e_{ik} \tag{3}$$

where $\Psi_{ik} = (w_{ik1}' - w_{ik0}', x_{ik}')$, b = *(g¢, (a₁ - a₀)¢)*, and $q_i = q_{i1} - q_{i0}$, $e_i = e_{i1} - e_{i0}$. It follows from the utility-maximization hypothesis that

$$\Pr(Y_{ik} = 1) = \Pr(U_{ik1} > U_{ik0}) = \Pr(z_{ik} > 0) \tag{4}$$

The specification of the error term in (1) assumes that the error can be decomposed into a regional, q, and individual-level, *e*, effect. For reasons outlined in the preceding section, we believe that spatial dependence exists across neighboring regions and that the choices and characteristics of individuals making housing choices do not exhibit spatial dependency. For these reasons, we model the spatial dependency among neighboring regions by assuming that q follows the following spatial autoregressive structure:

$$q_i = r \sum_{j=1}^{m} w_{ij} q_j + u_i \quad , \qquad \text{i = 1,…,m} \tag{5}$$

or in vector form,

$$q = rWq + u, \qquad u \sim N(0, s^2 I_m) \tag{6}$$

where W is an mxm spatial weight matrix and I_m is m-dimensional identity matrix. From Equation 6, q can be written as

$$\begin{aligned} & q = (I_m - rW)^{-1} u \Rightarrow \\ & q \,\big|\, (r, s^2) \sim N\left[0, s^2 \left((I_m - rW)'(I_m - rW)\right)^{-1}\right] \end{aligned} \tag{7}$$

The specification of the spatial weights matrix, W, represents the researcher's maintained assumption about the structure of the underlying spatial process that generates the spatial error. In specifying the weights matrix, we reasoned that the spatial error autocorrelation is most likely to arise between school districts that share a common border due to data measurement problems. Therefore, we define a contiguity weight matrix in which $w_{ij} = 1$ if *i* and *j* are nearest neighbors (i.e. they share a common border) and $w_{ij} = 0$ otherwise. For computational reasons, the matrix is then row standardized, so that $\sum_j w_{ij} = 1$.

The model is estimated using survey data of repeat homebuyers who bought and sold homes within Franklin County, Ohio in 1995.[3] Franklin County is the center of the Columbus metropolitan area and includes almost all of the City of Columbus as well as many of its suburbs. The dataset includes information on the characteristics of the origin

[3] Support for gathering these data was provided by The Ohio State University Urban Affairs Committee, The Ohio Housing Research Network, and The Ohio Center for Real Estate Education and Research. We are grateful to Hazel Morrow-Jones, the original PI for these grants, for her generosity in sharing these data.

and destination of individual household moves. Survey results provide information on household demographics and the stated reasons of 823 households for selling and buying a home. In this survey, only the households who sold and bought homes within Franklin County are included. The households who moved across county lines or moved to, from, or between rental units are not represented in the data set.

As mentioned earlier, the decision makers are distributed within a system of spatial regions, where we define a region as a school district within Franklin County. The only exception to this rule is that the Columbus City school district is divided into six sub-regions because of it is size and heterogeneity compared to the other school districts. In total, there are 22 regions that correspond to school districts within the county. The number of households moved from/to each alternative is given in Table 1.

Survey results show that one of the most important stated reasons for buying a home is school quality. School quality can be measured by different instruments including test scores, teachers' average salary, percentage of teachers with master degree, and expenditure per pupil. In a recent study by Brasington (1999), different instruments are explored to test the effect of school quality in housing studies. He finds that proficiency tests, expenditure per pupil and pupil/teacher ratio are consistently capitalized into housing prices. He also finds that variables such as teacher salary and student attendance rates are also valued, but that these results are sensitive the estimation technique employed. In this study, school quality is approximated by expenditure per pupil (STUDEX), the average score of ninth graders within each school district on the mathematics proficiency test (MATH), and the mean salary of teachers by school district (MEANSAL).[4]

In addition to school quality, the safety of the neighborhood is another main concerns of households. In order to capture this, the total estimated crime rate (CRIME) within the school district (available for 1998) is used as an explanatory variable. Lastly, we use the percentage of home built within the previous decade, 1980-90, as a proxy for the relative newness of housing stock within the school district. In addition to these community specific variables,[5] the following variables are included as individual specific variables: age of the head of household (AGE), the price difference between the house bought and the house sold (PRICEDIFF), and a dummy variable that takes one if household has a school age child (DSCKID).

Estimation Results

We use the MCMC method to estimate the spatial probit model of household location specified in (3) to test the influence of alternative specific and individual specific variables on the probability that a household moves out of their school district. In addition, we estimate a traditional probit model and compare the results between the spatial and non-spatial probit models. Table 2 reports the results of both estimations.

To the extent that errors are spatially correlated, the non-spatial model will fail to control for this autocorrelation and inconsistent parameter estimates will result. In addition, the estimates will be inefficient. The question of whether or not the data exhibit

[4] The data for these variables are obtained from Ohio Department of Education.

[5] Estimates on crime and data on year builts were available to us at a block group level. Because block group boundaries do not correspond with the location choice set boundaries, a Geographic Information System (GIS) was used to construct weighted values for each of the twenty-two locations based on the percent of the area of each block group that falls within each location.

significant spatial dependence of the form specified in (3) is confirmed. The spatial error autocorrelation parameter is estimated to be 0.47 and is found to be very statistically significant. Therefore the parameter estimates from the non-spatial model will be both inconsistent and inefficient. In comparing the estimates across the two models, the sign of the coefficients are same across the models, but the non-spatial probit produces mostly statistically insignificant coefficients. While coefficients of STDEXP, YB8090, DSCHKID, and PRICEDIFF are statistically significant in the spatial probit model, they are insignificant in traditional probit.

The estimated coefficients on the explanatory variables are found to be significant and most are found to be of the expected sign. The coefficient associated with CRIME is found to be negative and significant, suggesting that the probability of moving out of a school district decreases with high crime rate in the alternative school district. The coefficient on PRICEDIFF is also negative and significant. This implies that higher housing prices in an alternative school district decreases the household's probability of moving out of their current school district. The variables that capture the quality of school districts are STDEXP, MATH and MEANSAL. As expected, the estimated coefficients for STDEXP and MATH are found to be positive and significant. On the other hand, the coefficient for MEANSAL is found to be negative. It is possible that the unexpected sign for the parameter is due to colinearity between STDEXP and MEANSAL. Lastly, the coefficients of individual specific variables, AGE and DSCHKID, have the expected signs. The age of the head of the household, while not significant in the spatial probit estimation, is positive. The existence of school age children is significant and negatively influences a household's probability of moving to a different school district.

Table 2. Estimation Results

	Spatial Probit Model			Probit Model		
Variable	Coefficient	Standard Deviation	P-level*	Coefficient	t-statistics	P-value**
Constant	-1.103528	0.906721	0.10	-0.349478	-1.132011	0.28
CRIME	-0.326550	0.133859	0.00	-0.142516	2.754949	0.00
STDEXP	0.489405	0.250538	0.02	0.059825	0.839177	0.40
YB8090	0.204614	0.152203	0.07	0.085590	1.468834	0.14
MATH	0.034853	0.010190	0.00	0.015226	4.874070	0.00
MEANSAL	-1.239328	0.260293	0.00	-0.456119	-6.319876	0.00
AGE	0.005041	0.010685	0.33	0.001636	0.442101	0.65
DSCHKID	-0.074636	0.054693	0.04	-0.033890	-1.475880	014
PRICEDIFF	-0.158758	0.129146	0.09	-0.052990	-1.039857	0.29
RHO	0.479349	0.225099	0.00	-	-	-

As part of the model specified in (3), the regional effect component of the error structure is estimated for each region. These estimates are mapped in Figure 1. The map exhibits a fair amount of spatial clustering. As pointed out by Smith and LeSage (2001), this is consistent with the positive spatial error dependence exhibited by the positive and significant estimate of the spatial autocorrelation parameter, r.

Conclusion

Using survey and housing data from a unique dataset of household moves within Franklin County, Ohio, we estimate the probability that a household decides to stay within or move outside of their existing school district. In doing so, we use Bayesian techniques to estimate a binary probit model that allows spatial dependencies in underlying error structure, as developed by Smith and LeSage (2000). We find that data exhibits statistically significant spatial dependence, implying that estimates from a non-spatial probit model are inconsistent and inefficient and therefore invalid for hypothesis testing. In examining the most influential determinants of household location decisions in our model, we find that school quality and crime rate are both important factors that determine whether or not the household moves out of their existing school district. In addition, the presence of children in the household is an important determinant. Excluded from the current model are other individual variables that we hypothesize would also matter, namely household income, age of children, and race. In addition, other neighborhood variables, including tax rates and other measures of public and private services are hypothesized to matter. We plan to include these variables in subsequent models.

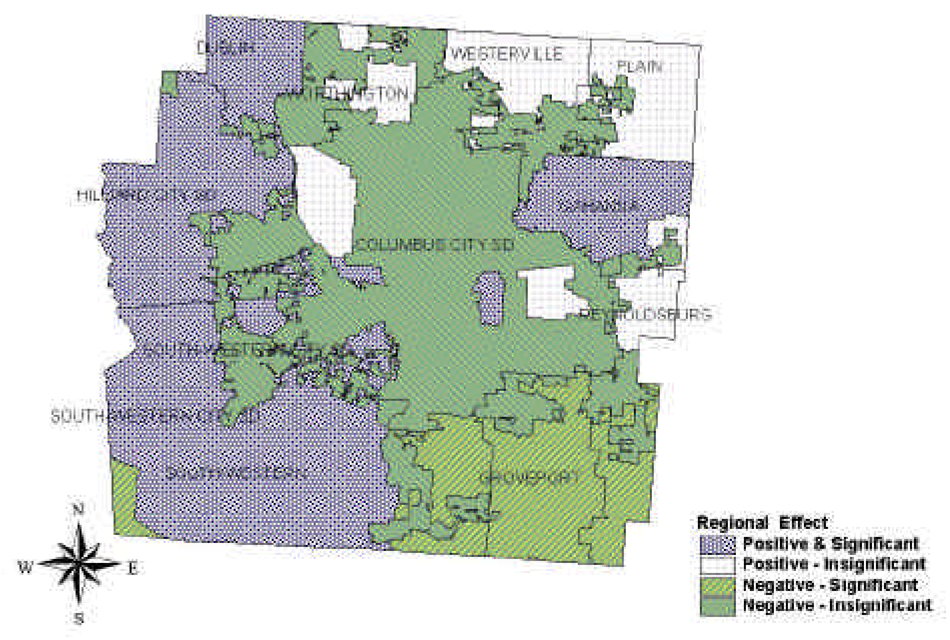

Figure 1: Regional Effect Estimates across School Districts

References

Alba, R.D., and J.R.Logan(1991),” Variations on Two Themes: Racial and Ethnic Patterns in the Attainment of Suburban Residence”, Demography, 28, pp. 431-453.

Albert, James H. and Siddhartha Chib (1993), “Bayesian Analysis of Binary and Polychotomous Response Data”, Journal of the American Statistical Association, Volume 88, number 422, pp. 669-679.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht.

Anselin, L. and A. Bera (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: A. Ullah and D. E. A. Giles, Eds., Handbook of Applied Economic Statistics, pp. 237–289. New York: Marcel Dekker.

Arias, Carlos and Thomas L. Cox (1999), “Maximum Simulated Likelihood: A Brief Introduction for Practitioners, Staff Paper Series, No.421, university of Wisconsin-Madison.

Besag, Julian (2000), “An introduction to Markov Chain Monte Carlo, Working Paper, University of Washington.

Bolduc, Denis, Bernard Fortin and Stephen Gordon (1997). “Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of two New Techniques”, International Regional Science Review 20, 77–101.

Brasington, D.M (1998), “Which Measures of School Quality Does the Housing Market Value? Spatial and Non-Spatial Evidence, Department of Economics Working Paper No. 1998-05, Tulane University.

Cliff, A. and J. K. Ord (1973). Spatial Autocorrelation. London: Pion.

Dorfman, Jeffrey H. (1997) Bayesian Economics Through Numerical Methods, New York, Springer-Verlag.

Dowding, K., P. John, and S. Biggs (1994), “Tiebout: A Survey of the Emprical Literature”, Urban Studies, vol.31number 4/5, pp. 767-797

Edel M., and E. Sclar (1974), “Taxes, Spending and Property Values: Supply adjustment in Tiebout-Oates Model”, Journal of Political Economy, 82, pp.941-954.

Fleming, M. (2001), “ A Review of the Techniques for Estimating Spatially Dependent Discrete Choice Models”, Working Paper, Fannie Mae.

Friedman, J. (1981).”A Conditional Logit Model of the Role of Local Public Services in Residential Choices”, Urban Studies, 18, pp.347-358.

Geman, S., and D. Geman (1984), “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, pp. 609-628.

Goodspeed, T.J., “ On the Importance of Public Choice In Migration Models”, Economics Letters, 59, pp.373-379.

Goldstein, G.S., and M.V. Pauly, “Tiebout Bias on the Demand for Local Public Goods”, Journal of Political Economy, 16, pp. 131-143.

Hamilton, B. (1976), “The Effect of Property Taxes and Local Public Spending on Property Values: A Theoretical Comment”, Journal of Political Economy, 84, pp. 647-650.

Hasting, W.K. (1970), “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”, Biometrika, 57, pp. 97-109.

Hoyt, W.H., and S.S. Rosenthal (1997),”Household Location and Tiebout: Do Families Sort According to Preference for Locational Amenities?” Journal of Urban Economics, 42, pp. 159-178.

Irwin, Elena G. " Moving Up and Out or Flight from Blight? Evidence of the Causes of Suburbanization." Paper presented at the 1999 American Agricultural Economics Association Annual Meeting, Nashville, TN, August, 1999.

Lerman, S.R. (1976),”Location, Housing, Automobile Ownership and Mode to Work: A Joint Choice Model”, Transportation Research, 610, pp. 6-11.

LeSage, James P. (1999) The Theory and Practice of Spatial Econometrics, unpublished manuscript available at: http://www.spatial-econometrics.com.

LeSage, James P. (2000) “Bayesian Estimation of Limited Dependent variable Spatial Autoregressive Models”, Geographical Analysis, 2000 Volume 32, number 1, pp. 19-35.

Logan, J.R, and R.D. Alba,(1993),” Locational Returns to Human Capital: Minority Access to Suburban Community Resources”, Demography, 30, pp.243-368.

McFadden, D. (1973), Conditional Logit Analysis of Qualitative Choice Behaviour, in P.Zarembka, ed., Frontiers of Econometrics, Academic Press, New York.

McMillen, Daniel P. (1992). "Probit with Spatial Autocorrelation." Journal of Regional Science 32:3, 335-348.

McMillen, Daniel P. (1995). "Spatial Effects in Probit Models: A Monte Carlo Investigation." In Luc Anselin and Raymond J.G.M. Florax, eds., New Directions in Spatial Econometrics, Heidelberg: Springer, 189-228.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) “Equations of State Calculations by Fast Computing Machines, Journal of Chemical Physics, 21, pp. 1087-1092.

Morrow-Jones, H and D. Lipsetz (2000),” Climbing the Housing Ladder: Repeat Homebuyers and Their Satisfaction. Report to the Center for Real Estate Education and Research”, Center for Urban and Regional Analysis Working Paper no.00-06, The Ohio State University.

Nechyba, T., and R.P. Strauss (1998), “Community Choice and Local Public Services: A Discrete Choice Approach”, Regional Science and Urban Economics, 28, pp 51-73.

Oates, W.E. (1969),”The Effect of Property Taxes and Local Public Spending on Property Values: An Emprical Study of Tax Capitalization and the Tiebout Hypothesis”, Journal of Political Economy, 77, pp. 957-971.

Pinkse, Joris and Margaret E. Slade (1998). “Contracting in space: an application of spatial statistics to discrete-choice models”. Journal of Econometrics, 85, pp.125–54.

Quigley, J.M., and D.H. Weinberg (1977), “ Intra-urban Residential Mobility: A Review and Synthesis”, International Regional Science Review, 2, pp.41-66.

Rapaport, Carl (1997), “Housing Demand and Community Choice: An Emprical Analysis”, Journal of Urban Economics, 42, pp. 243-260

Rubinfield, D.L., P.Shapiro, and J. Roberts (1987), “Tiebout Bias and the Demand for Local Public Schooling”, The Review of Economics and Statistics, 69:3, pp. 426-437.

Smith, T.E., and J.P. LeSage (2000), “A Bayesian Probit Model with Spatial Dependencies”, Working Paper, University of Toledo.

Simons, J., “”Changing Residence in the City”, in Greenburg, M.R, ed.,Readings in Urban Economics and Spatial Patterns, Center for Urban Policy Research, New Brunswick, 1974.

Tiebout, C..M (1956), “A Pure Theory of Local Expenditures”, Journal of Political Economy, 64, pp. 416-424.

Table 1: 1995 Moves of Surveyed Household across School Districts within Franklin County, Ohio

DESTINATION → / ORIGIN ↓	Bexley	Dublin	Hilliard	Gahanna	Grandview	S.Western	Reynoldsburg	U.Arlington	Westerville	Whitehall	Worthington	Plain	Hamilton	Groveport	C.Wincheste	Madison	Columbus	W.Columbus	NW.Columbus	NE.Columbus	SE.Columbus	SW.Columbus	TOTAL
Bexley	9			3			1					3		2	1		5	1					25
Dublin		12	1		1			2				1	1				2		2				22
Hilliard		4	22			2		1			1						6	10	2			2	50
Gahanna	1			36					1					1			6		1	4	1		51
Grandview			1	1	4	1		1									1	1	2			1	13
S.Western		2	8	1		45		2							1	1	2	2	3	1		9	77
Reynoldsburg			1	1			8	1									5			1	1		18
U.Arlington		3	8	2	6	1		35			2	1					6	2	11	1			78
Westerville		1				2			35		1						2	1	6	5			53
Whitehall	1			4			2		1	2										2	2		14
Worthington		3		2				3	1		13						5		9	2		1	39
Plain				2					2											1			5
Hamilton				1		2							2		1								6
Groveport	1			4			4							7	2		1			1	4		24
C.Winchester															1					1			2
Madison																							0
Columbus	3	3	7	9	2	13	5	3	12	1	8	2	4	5			63	8	15	12	3	10	188
W.Columbus		2	18			3	1	1									1	8	4				38
NW.Columbus		20	8	2				2	3		3	1					8	1	20	1		1	70
NE.Columbus				7			1	1	5								4	1	1	1			21
SE.Columbus		1				2									1		2		1	1	1	1	10
SW.Columbus			1			8										1	2		4		1	2	19
TOTAL	15	51	75	75	13	79	22	52	60	3	28	8	7	15	7	2	121	35	81	34	13	27	823