International Journal of Economics and Finance; Vol. 4, No. 9; 2012
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

Residential Location Choice: The Role of a Taste for Similarity

Hua Kiefer

Correspondence: Hua Kiefer, Economics Department, The Office of the Comptroller of the Currency, 250 E Street SW, Room No. 3153, Washington DC, 20219-0001, USA. Tel: 1-202-874-3917. E-mail: hua.kiefer@occ.treas.gov

Received: June 25, 2012 Accepted: July 19, 2012 Online Published: July 27, 2012

doi:10.5539/ijef.v4n9p34 URL: http://dx.doi.org/10.5539/ijef.v4n9p34

Abstract

This paper examines the importance of social interactions on a household's location decision. The theory argues that individuals' utility will be greater when socially interacting with similar others. The hypothesis that a household desires to find a good community match is tested through the application of a discrete residential location choice model. In addition, this paper also tests Tiebout's hypothesis that households search for a community where their benefits from local public goods will exceed their local tax costs. The findings tend to support both hypotheses, indicating that a household prefers neighbors with a similar socio-economic background and somewhat larger houses.

Keywords: location choice, nested logit, hedonic, tiebout's hypothesis, social interactions

JEL classification: R21, R23, D10

1. Introduction

For many households, purchasing a home is one of the most significant economic decisions that they will make. The choice of an optimal level of housing consumption requires that a household gather information regarding the features of potential residences. The characteristics of the housing structure itself will be an important determinant of a household's choice of residence (Quigley, 1976), but other factors also will influence this decision, such as households' individual characteristics (Gabriel and Rosenthal, 1989), and neighborhood's quality (Friedman, 1980; Quigley, 1985; and Nechyba & Strauss, 1997). All of these additional determinants have been verified to be significant in a household's location choice. For example, Nechyba and Strauss (1997) showed that an individual household's location decision is significantly affected by community attributes. The findings of Gabriel and Rosenthal (1989) indicate the influential impact of household socioeconomic and demographic characteristics on community choice. Previous studies in this literature place emphasis on the impact of households' characteristics and neighborhood's attributes on residence choice respectively and/or jointly. There is no research taking into account of the influence of the overall match of a community and a household on the household's community choice. This effect cannot be simply captured by a cross effect. It involves how the degree of match is measured.

In this study, location determinants are grouped into two subsets. The first captures the nature of the community by focusing on factors, which describe the natural and sociopolitical environment, such as local tax rate, crime rate, and parks. The second set describes the match of a potential household and the existing inhabitants of the community, capturing the compatibility of a household and its socioeconomic environment, such as education level, income, age, etc,.

Neighborhood variables in the first category, representing the intrinsic features of a community, however, might not fully capture all of the features that might attract a potential incoming household. The attraction is based on not only the intrinsic features but also the compatibility of the community with a potential household's own interests. Therefore, the value of the household's neighbors' attributes relative to the household matters as well. The inclusion of the second category of location determinants explores this possibility. These relative values may directly affect the household's social interactions with its neighbors.

A great deal of research in sociology and psychology points to the benefits of social interaction, as it promotes emotional and physical health (Diose and Mugny, 1984), affects learning and work-related skills (Rogoff, 1990), intensifies social cohesion to achieve collective efficacy, and generates collective efficacy for children (Sampson, Raudenbush, and Earls (1997). The extent of social contact between two individuals is determined by the perceived cost and expected reward derived from that interaction. The cost is the effort and expense that must be spent on the social interaction, which can be measured by propinquity (proximity in physical space). The reward can be based on the similarity of attitudes and values between the two persons in some instances (Sampson, Morenoff, and Earls, 1999; Clark, 1986; Hagestad and Uhlenberg, 2005) and it can also be based on complementary needs in other cases (Winch, 1952). (Note 1) In economic terms, the benefits and costs from social interactions may affect an individual's utility function through expectations and constraints (Manski, 2000). Because of the important facilitative roles played by propinquity and similarity in the utility function, a household will tend to searches for similarities between him/herself and potential neighbors in order to maximize its utility, which leads to one of the foci of this paper: the test of households' searching for similarity. (Note 2) Although the literature of social interaction in sociology and psychology provides verbal reasoning on households' preferences of similar neighbors, no mathematical argument or empirical analysis are offered. In the surge of economic research on social interactions, no studies that have explored the possibility of households' voluntarily search for similarity (in some economically relevant dimension) for the purpose of beneficial social interactions. This study serves to fill a gap in the empirical studies on social interactions, offering mathematical analysis on households' preferences of similarities.

This paper explores the impact of similarities and differences between a household and the community's other residents along the dimensions of race, income, education and family size (i.e. whether or not the household has children) (Note 3) upon a discrete choice econometric model. The degree of similarity between a household and his/her neighbors takes the value of the absolute difference between a household's characteristics and its potential neighbors' characteristics.

The second focus of this study is to test Tiebout's hypothesis (1956): a households search for fiscal surplus. The conventional perspective suggested by Tiebout's hypothesis that households vote with their feet, implies that households have preferences toward public services, local taxes, and community amenities that influence their choice of housing location. Hamilton (1975) extended Tiebout's model by allowing a property tax instead of the community-specific head tax in Tiebout's paper. In Hamilton's heterogeneous community model, high income households and low income households could co-exist only if the community had very strong zoning rules. (Note 4) Accordingly, the fiscal deficit and the fiscal surplus will be capitalized in house prices. The studies of Tiebout and Hamilton reveal that both high and low income households prefer to live with high income households in order to avoid a fiscal deficit and to obtain fiscal surplus. (Note 5) Many previous studies have confirmed the search for affluence through a hedonic capitalization process (Yinger, 1982). The confirmation of the capitalization process only weakly corroborates Tiebout's hypothesis, while this paper takes a more direct approach. Households' preference for living in a high income neighborhood is tested by finding a migrant household' responses to the difference between their likely house value and its potential neighbors' house values.

A nested logit (NL) regression is employed to analyze a household's residential location decision within Franklin County, OH. The application of the discrete choice model enables the measurement of the impact of similar and affluent neighbors as well as public services and community attributes on a household's community choice. The NL calculates the probability of a household choosing a certain community; conditional on the set of attributes determining utility and budget constraint. Through the budget constraint, the community's "entry price" becomes a necessary input into the household's indirect utility function. The community entry price is defined as the premium paid by households in order to locate within this community. It is represented by a set of housing price indices which is constructed through a hedonic price function (Rosen, 1974). To circumvent the difficulties of omitted variables and measurement errors, a spatial geostatistical model is used to control the correlations among the disturbances for more precise estimates of the price indices.

This paper is organized as follows. Section 2 outlines the basic model specifications. In Section 3, the data is described and a statistical summary is provided. The empirical results are presented in Section 4. Section 5 provides detailed interpretations of the empirical results. Finally, the last section offers some conclusions from the analysis.

2. Model Specification

This section describes the nested logit model that is employed to analyze a housing market in which households sort themselves among the set of available residential locations on the basis of utility derived from given alternatives (McFadden, 1974) and their budget constraints.

2.1 Nested Logit Model

The nested logit model is an extension of the conditional logit model or multinomial logit model. (Note 6) It groups the substitutable alternatives into subgroups that allow the variance to differ across subgroups while maintaining the independence of irrelevant alternatives (IIA) assumption (Note 7) within each group. Suppose, then, the J alternatives can be divided into L subgroups such that the choice set can be written as

$$[C_1, \cdots, C_J] = (C_{1|1,} \cdots, C_{J1|1}), \cdots, (C_{1|l,} \cdots, C_{Jl|l}), \cdots, (C_{1|L,} \cdots, C_{JL|L})$$

with a choice of subgroups indexed by $l=1,2, \cdots L$, and alternatives $j=1,2, \cdots, Jl$ in subgroup l. Logically, the choice process can be thought as that of choosing among the L choice sets and then making the specific choice within the chosen set. One way to express the utility that a household derives from choosing a location j in subgroup l is to use the random utility model (RUM). It is defined as:

$$U_{jl} = V_{jl} + \varepsilon_{jl} \tag{1}$$

where U_{jl} represents a household's conditional utility function (conditional on the decision) in choosing location j in subgroup l, V_{jl} is the deterministic component of the household's conditional utility, and ε_{jl} is the random element due to unobserved attributes.

If the household is rational, the location he/she choses must have maximized his/her utility. Therefore, U_{jl} becomes the unconditional utility function:

$$U_{jl} = max(U_k) \text{ for } k = 1, \cdots, Jl \tag{2}$$

Consequently, the probability of an individual choosing community j among the total available alternatives Jl in choice set l is,

$$Pr_{j|l} = Pr(U_{jl} = max(U_k)) = pr(U_{jl} \geq U_k, \forall k \neq j) \tag{3}$$

Equation (3) means that the conditional probability of alternative j being chosen given choice set l is equal to the probability that the utility obtained by the individual from option j exceeds the utility obtained from any of the other alternatives in the given choice set. Substitution of (1) into (3) yields:

$$Pr_{j|l} = Pr\left(V_{jl} + \varepsilon_{jl} \geq V_k + \varepsilon_k, \forall k \neq j\right)$$
$$= Pr\left(\varepsilon_k - \varepsilon_{jl} \leq V_{jl} - V_k, \forall k \neq j\right)$$

The nested logit model can then be derived from the assumption that the residuals

$$[(\varepsilon_{11}, \cdots, \varepsilon_{J1}), \cdots, (\varepsilon_{1l}, \cdots, \varepsilon_{Jl}), \cdots, (\varepsilon_{1L}, \cdots, \varepsilon_{JL})]$$

have a generalized extreme-value distribution (McFadden, 1974). (Note 8) Thus the conditional probability of choosing community j given the l subgroup, $Pr_{j|l}$, can then be written as,

$$Pr_{j|l} = \frac{\exp(V_{jl})}{\sum_{k=1}^{Jl} \exp(V_k)}. \tag{4}$$

For estimation purpose, we further assume that the deterministic component of the total utility, V_{jl}, has a linear form as,

$$V_{jl} = \beta' x_{jl} + \alpha' Z_l, \tag{5}$$

where x_{jl} denotes the vector of observed attributes of choice $C_{j|l,}$, Z_l represents the vector of observed attributes of the choice set l, α and β are the vectors of the unknown parameters.

Substitution of (5) into (4) yields

$$Pr_{j|l} = \frac{\exp(\beta' x_{jl} + \alpha' Z_l)}{\sum_{k=1}^{Jl} \exp(\beta' x_k + \alpha' Z_l)} = \frac{\exp(\beta' x_{jl})}{\sum_{k=1}^{Jl} \exp(\beta' x_k)} = \frac{\exp(\beta' x_{jl})}{\exp(I_l)} \tag{6}$$

where $I_l = \log(\sum_{k=1}^{Jl} \exp(\beta' x_k))$ is defined as an inclusive value for nest l, which is the expected utility for the choice of alternatives within nest l. The probability of choosing nest l is derived as,

$$Pr_l = \frac{\exp(\alpha' Z_l + (1-\sigma_l) I_l)}{\sum_{m=1}^{L} \exp(\alpha' Z_m + (1-\sigma_l) I_m)} \tag{7}$$

where, $1 - \sigma_l$ is a measure of the correlation in unobserved factors within nest l. (Note 9)

By the law of probability, the unconditional probability of the observed choice made by a household has the form of

$$Pr_{jl} = Pr_{j|l} \cdot Pr_l$$

$$= \frac{\exp(\beta' x_{jl})}{\exp(I_l)} \cdot \frac{\exp(\alpha' Z_l + (1-\sigma_l) I_l)}{\sum_{m=1}^{L} \exp(\alpha' Z_m + (1-\sigma_l) I_m)}$$

$$= \frac{\exp(\beta' x_{jl} + \alpha' Z_l - \sigma_l I_l)}{\sum_{m=1}^{L} \exp(\alpha' Z_m + (1-\sigma_l) I_m)} \tag{8}$$

2.2 Model Parameterization

In order to parameterize the nested logit model, further assumptions about the form of the underling indirect utility function are required. Specifically, a household allocates its income between non-housing and housing consumptions, as well as deciding the community to achieve the maximum level of utility. If we describe the direct utility a household receives by choosing to live in a particular community as a function of housing consumption, non-housing consumption, and his/her tastes, the indirect utility of household i in choosing community *jl* (i.e., community *j* in subgroup *l*) can be written in the following form: (Note 10)

$$U_{i,jl} = U_{i,jl}(h_{i,jl}, c_{i,jl}, z_i) \quad s.t. \, y_i = c_{i,jl} + P_{i,jl}, \tag{9}$$

where z_i denotes household *i*'s demographic characteristics which determine the individual's preferences, y_i represents individual *i*'s income, $c_{i,jl}$ is the non-housing consumption of household *i* in community *jl*, (Note 11) $h_{i,jl}$ is the housing services consumed by household *i* in community *jl*, and $P_{i,jl}$ denotes household *i*'s housing expenditure in community *jl*. (Note 12)

Because the housing expenditure in community *jl* can be artificially decomposed into two categories including the spending on housing structural attributes, $p_s s_i$, (where p_s is the composite price of housing structure bundle, and s_i is the vector of housing structural variables), and the spending on community characteristics, $p_Z Z_{jl}$, (where p_Z is the composite price of community feature bundle, and Z_{jl} is the vector of community variables, which includes both the specific attributes relevant to community j, and the common attributes relevant to subgroup *l*), (Note 13) the utility function can therefore be written as,

$$U_{i,jl} = U_{i,jl}(c_{i,jl}, s_i, Z_{jl}, z_i) \quad s.t. \, y_i = c_{i,jl} + p_s s_i + p_Z Z_{jl},$$

where s_i is household *i*'s consumption bundle of the housing structural attributes, Z_{jl} denotes community *jl*'s attributes, such as local public service, socio-economic characteristics composition, median income and so on.

To maximize its utility, household *i* can choose the quantity of $c_{i,jl}$, the value of s_i, and the value of the product of p_Z and Z_{jl}, but not the quantity of Z_{jl}. This is because the value of $Z_{jl}$is constant for all households who choose community *jl* in which to reside. Therefore household *i*'s budget constraint is rewritten as,

$$U_{i,jl} = U_{i,jl}(c_{i,jl}, s_i, Z_{jl}, z_i) \quad s.t. \, y_i - p_Z Z_{jl} = c_{i,jl} + p_s s_i.$$

The available income that household i can allocate between non-housing consumption and housing structure bundle is the difference of real income and the expenditure on community attributes of location *jl*.

After maximizing direct utility subject to the budget constraint, and substituting the derived demand functions into the direct utility function, the indirect utility function can be expressed as, (Note 14)

$$U_{i,jl} = V_{i,jl}(y_i, p_Z Z_{jl}, p_s, Z_{jl}, z_i) + \varepsilon_{i,jl}, \tag{10}$$

where $p_Z Z_{jl}$ is the product of p_Z and Z_{jl}(expenditure on the community attributes), $\varepsilon_{i,jl}$ denotes the random component of the total utility, and $V_{i,jl}$ is the deterministic component of the total utility, which has a linear function form according to the assumption for equation (5).

The community variables, Z_{jl}, entering the indirect utility function through the product of Z_{jl} and p_Z, include both community amenities and socio-economic composition of the neighborhood.

The economic interpretation of the relevance of neighbors' socio-economic characteristics comes from the independent preferences suggested by Pollak (1976). (Note 15) This is the so called exogenous or contextual effect (Manski, 1993) in the neighborhood effects literature, and was used to estimate households' demand in some previous studies.

While the inclusion of Z_{jl} in the indirect utility function reveals the role that is played by the potential neighbors' characteristics on the household's total utility, it does not describe household's preference for homogeneous neighbors. If the household values a homogenous community, the homogeneity of the community can be thought as a special good that should also be priced. Consequently, the price of homogeneity should also enter the household's utility function. However, this paper takes a different approach than pricing homogeneity as a special good; rather, it tests a household's preference for a homogeneous community directly through the inclusion of dissimilarity in the household's utility function. (Note 16) Specially, the amount of dissimilarity is included in the utility function to indicate the disutility of a household resulting from living with heterogeneous neighbors. The quantity of dissimilarity is measured by the absolute difference between the household's and the neighbors' amount of a particular characteristics. The utility function (10) is extended as following,

$$U_{i,jl} = V_{i,jl}\left(p_Z Z_{jl}, p_s, Z_{jl}, z_i, y_i, |y_i - y_{jl}|, |z_i - z_{jl}|, |w_i - w_{jl}|, (w_i - w_{jl})\right) + \varepsilon_{i,jl}, \tag{11}$$

where z_{jl} represents community *jl*'s demographic characteristics, y_{jl} represents community *jl*'s median income level, w_i is household *i*'s house value, w_{jl} is community *jl*'s median house value. Thus, $|z_i - z_{jl}|$ denotes the absolute difference of household *i*'s demographic characteristics and the median characteristics of the neighbors in community jl, $|y_i - y_{jl}|$ denotes the absolute difference of household *i*'s income and the median income of the neighbors in community jl, $|w_i - w_{jl}|$ denotes the absolute difference of household *i*'s house value and the median value of the houses in community jl, and $(w_i - w_{jl})$ denotes the difference of household *i*'s house value and the median value of the houses in community jl.

Three absolute difference terms are created to test a household's preferences for similarity. The use of them captures the desires of a household to match the neighborhood in multiple dimensions. The absolute values of the differences are constructed for the purpose of avoiding the confusion arising from the signs of the estimated coefficients. This study focuses on households' preferences over similarity, therefore; only the absolute value of the difference of a household and its neighbors is relevant for this paper.

Tiebout's hypothesis is tested by including $(w_i - w_{jl})$, the difference of household *i*'s house value and the median house value in community *jl*. The smaller the value of $(w_i - w_{jl})$ is, the more likely that household *i* has a relatively smaller house than most of the houses in community jl. If the estimated coefficient of $(w_i - w_{jl})$ has a negative sign, Tiebout's hypothesis is confirmed. In other words, households prefer to live in an affluent community. (Note 17)

The probability of choosing community jl can be computed for the NL model by substituting the linear function of $V_{i,jl}$, which is indicated by equation (11), into equation (8). Because z_i, y_i, and p_s are constant across alternative communities, they will no longer remain in the probability function. Therefore, the vector of the explanatory attributes in the probability function is composed of $p_Z Z_{jl}$, Z_{jl}, $|y_i - y_{jl}|$, $|z_i - z_{jl}||, w_i - w_{jl}|$, and $(w_i - w_{jl})$.

The community entry price, $p_Z Z_{jl}$, in the indirect utility function needs to be constructed for each community in the choice set. The hedonic house price estimation method is employed to handle this issue (Rosen, 1974).

2.3 Hedonic House Price Index

2.3.1 Hedonic Price Function

According to Rosen, the conceptual form of the hedonic function is a relationship between the bidding function from consumers and the offering function from the suppliers in the housing market. In other words, consumers bid for structural components of housing units and packages of neighborhood amenities in order to maximize their utility; suppliers maximize their profits by offering different housing unit packages to the market. Considering the market process described above, the variation in transaction prices can be explained as a function of property characteristics. In this paper, the structural characteristics of housing units, neighborhood socioeconomic attributes, community features and public services enter the hedonic price function: (Note 18)

$$w_i = w\left(school_j, s_i, Z_{j(i)}\right) + v_i, \tag{12}$$

where $school_j$ is a dummy variable indicating school district j, w_i reflects the transaction price (Note 19) of household *i*'s house, $Z_{j(i)}$ is the vector of the community variables (Note 20), s_i is the vector of the structural attributes of household i's house, and v_i represents the independent and identically distributed (IID) random error term. (Note 21)

Due to the durability and the fixed position of a house, it has been broadly accepted in the housing literature that the geographic location of a house has nontrivial impact on its economic value. The inclusion of $Z_{j(i)}$ and $school_j$ is to test for capitalization of the geographic locations on house values. According to Can (1992), (Note 22) the elements of $Z_{j(i)}$ that denote the neighbors' socio-economic characteristics test for the capitalization of the adjacency effects. Neighbors' socio-economic characteristics can be interpreted as the signals of maintenance decisions that may affect the market value of a given house. Other elements of $Z_{j(i)}$ and variable $school_j$, capture the neighborhood effects of a house's geographic location.

To parameterize the hedonic function, (Note 23) a linear form is assumed and equation (12) is rewritten as,

$$w_i = \alpha_0 + \sum_{j=1}^{J-1} \alpha_{0j}\, school_j + \alpha_1 Z_{j(i)} + \beta s_i + v_i, \tag{13}$$

where, α_0, α_{0j}, α_1, and β are the coefficients to be estimated, while we use $\hat{\alpha}_0$, $\hat{\alpha}_{0j}$, $\hat{\alpha}_1$, and $\hat{\beta}$ to indicate the estimated coefficients.

The hedonic regression reveals the marginal price of each attribute given the linear hedonic form, but tells nothing about the demand function or the supply function; therefore, it is only a market clearing function determined by the equilibrium market-clearing price.

2.3.2 Hedonic Price Index

Upon the selection of the functional form and the appropriate determinants of housing values, the hedonic function is estimated, and the estimated coefficients are applied to construct the house price index for each school district. The price index is,

$$\hat{w}_j = \sum_{j=1}^{J-1} \hat{\alpha}_{0j}\, school_j + \hat{\alpha}_1 \bar{Z}_j, \tag{14}$$

where $\hat{w}_j$ denotes the price index of school district *j*, (Note 24) and $\bar{Z}_j$ is the average of neighborhood characteristics in community *j*.

In this formulation, the price index, $\hat{w}_j$, reflects spending on community attributes, which is one part of the entire expenditure on the house. This corresponds to the community premium denoted before. The price index defined in this paper is different from the conventional ones, because it excludes the spending on the constant-quality housing structures. The reason for making this change is to be consistent with the input element in the indirect utility function. Because the excluded part of the conventional price index is a constant number, using $\hat{w}_j$ instead of the standard measure in the NL model does not make any difference to the results.

2.3.3 Spatially Autocorrelated Error Terms in the Hedonic Regression

The precision of the price indices can be affected by a number of factors, such as the functional form of the hedonic regression and the set of influential explanatory variables excluded from the function. This section addresses the issue of the spatially dependent error terms in order to achieve more efficient coefficient estimates and unbiased estimates of the standard errors. The IID assumption ofthe stochastic error term in the hedonic function does not hold because of the possibility of measurement errors and omitted variables.

Among the determinants of house values, housing structure variables are usually measured with little error, but it is likely that many of them are omitted. The location variables may not be fully observed, and even they can be observed, some measurement errors are likely to occur, which will leave the residuals produced by the hedonic equations spatially correlated. It is reasonable to expect that the correlation between residuals are determined by the proximity of observations, given that nearby house units share the same neighborhood, which tend to create similar errors in measuring the attributes of the neighborhood. The strength of the relationship diminishes as the distance between the observations increases. To address this problem, this study models the spatial correlation of the error terms explicitly through the use of a geostatistical model (Dubin, 1988). The covariance matrix of the error terms is defined as,

$$E\{vv'\} = \sigma^2 K = \Omega, \tag{15}$$

where *K* is the correlation matrix, and Ω is the covariance matrix with nonzero off-diagonal elements.

To estimate the covariance matrix, Ω, a semivariogram model can be employed, which expresses the variance of the difference between the values of the regionalized variables as a function of separation distance. The process is,

$$\begin{aligned} \gamma(L_i - L_j) &= 0.5var\{v(L_i) - v(L_j)\} \\ &= C(0) - C(L_i - L_j) \\ &= \sigma^2(1 - K_{ij}, \end{aligned} \tag{16}$$

where $L_i = (x_i, y_i)$ denoting the location of property *i* (x_i indicates the latitude, and y_i indicates the longitude of property *i*), $C(L_i - L_j) = cov\{v(L_i), v(L_j)\}$ denoting the covariance of residual v_i and residual v_j, and $C(0)$ denotes the variance.

To explicitly model the semivariogram process, the functional form of the semivariogram needs to be specified. This paper applies the spherical model, which is,

$$\gamma_{ij} = \begin{Bmatrix} c_0(\frac{3d_{ij}}{2a_0} - \frac{{d_{ij}}^3}{{a_0}^3}) & if\ i \neq j\ and\ 0 < d_{ij} < a_0 \\ 0 & if\ i = j \\ c_0 = (\sigma^2) & if\ d_{ij} > a_0 \end{Bmatrix} \tag{17}$$

where c_0 and a_0 are parameters to be estimated in the spherical function, and d_{ij} is the distance between property i and property j.

The parameters in the semivariogram function and the coefficients of the hedonic regression are estimated simultaneously through a maximum likelihood method. The price indices can then be constructed using the more efficient estimates.

3. Data Description

There are several sources for the data used in this essay. The Center for Urban and Regional Analysis (CURA) provides two surveys, which matches housing transaction data and households' characteristic variables. (Note 25) National Center for Education Statistics (NCES) provides neighborhood socio-economic characteristics at the school district level, and census tract level variables come from the U.S. Census. The Ohio Department of Education (ODE) provides community school quality data. The Bureau of Justice Statistics (BJS) provides the crime index. The Franklin County Treasurer provides property tax data. All monetary values are converted to a base period value (2001 quarter 1) for the control of inflation. The combinations of the data sets are used to estimate the residential location choice model and the hedonic price model. The two surveys from CURA are the 2001 Housing Survey in Franklin County (pre-September 11), and the 2001 Housing Survey in Franklin County (post-September 11). There are three reasons for combining these two data sets. First, these two surveys were collected in very close time periods. Second, there is no overlap of the sample objects. Third, the questionnaires of these two surveys are same, except for some additional questions about the opinions of households' attitudes towards terrorists that are not the concern of this paper. The conjunction of these two data sets increases the number of total observations substantially. (Note 26)

The observations in the CURA data set were recorded in 2001, but most transactions happened before that time, which means that the variables describing housing structures and households' characteristics may not be the same as when the houses were bought. In other words, the residents in the data set are not necessarily recent movers. (Note 27) This suggests that a household's past choices of a community may not be the same as what his/her current choice would be if they relocated, but they have not relocated due to transaction costs. To circumvent this incompatibility, a subset of the recent movers is employed to retest the hypothesis. The date that is used for the nested logit model is described as following. (Note 28)

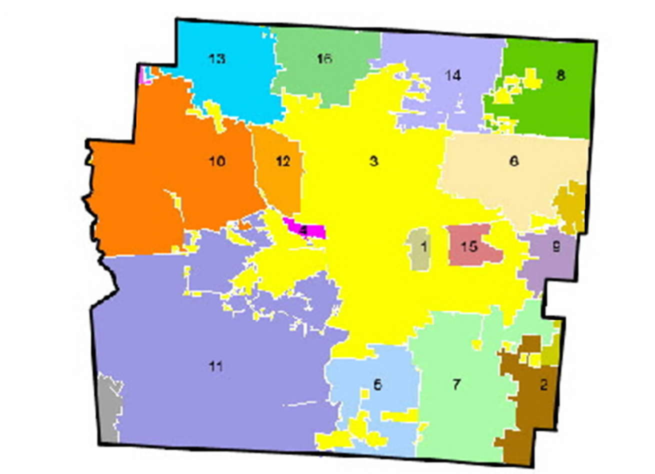

Figure 1. School District Map of Franklin County, OH

Table 1. School District Index

Index Number	School District
1	Bexley city school district
2	Canal Winchester local school district
3	Columbus city school district
4	Grandview heights school district
5	Hamilton local school district
6	Gahanna Jefferson city school district
7	Groveport Madison local school district
8	Plain local school district
9	Reynoldsburg city school district
10	Hilliard city school district
11	Southwestern city school district
12	Upper Arlington city school district
13	Dublin city school district
14	Westerville city school district
15	Whitehall city school district
16	Worthington city school district

Although Franklin County contains 16 major school districts, (Note 29) only 11 of them have sufficient observations to be used for the estimation of the residential location choices. These eleven school districts are: the Bexley city school district, the Columbus city school district, the Gahanna Jefferson city school district, the Groveport Madison local school district, the Reynoldsburg city school district, the Hilliard city school district, the Southwestern city school district, the Upper Arlington city school district, the Dublin city school district, the Westerville city school district, and the Worthington city school district. Consistent estimates are still available according to McFadden (1978). (Note 30) See Figure (1) for the map of Franklin County indexed by school districts. A legend for the index numbers is reported in Table (1).

The data set that is used to estimate the NL model contains 1,467 observations in total. (Note 31) The definitions and statistical descriptions of the variables in the estimation of the location choice model are provided in Table (2) and Table (3). Table (2) contains some community variables as well as the set of price indices constructed from ML estimates (GEOPIND). These estimates were constructed from the hedonic regression, which is explained in detail in Appendix A2. Table (3) displays the key variables of this study, representing the dissimilarity of a household and its neighbors. A household's preference over a homogenous community is studied through five categories: (Note 32) the difference in education background, the difference in number of children, the difference in race, the difference in income, and the difference in house value. (Note 33)

Table 2. Descriptive Statistics for Community Variables in NL

Panel A: Definition of Variables	
GEOPIND	Price index of geostatistical model (not standardized)
TAX	Property tax rate, in percentage, in the township
PUPILEXP	Education expenditure per pupil in 10,000 dollars in school district
CRIME	Crime index divided by population in the police jurisdiction
CBD	Distance of centroids of school districts to the CBD in miles
AVGSC9	Average of math, reading, and writing of the Ohio 9th-grade proficiency test in the school district
MHVALUE	Median house value in 10,000 dollars (from census)

Panel B: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
GEOPIND	15.33	3.04	12.33	23.83
TAX	1.75	0.23	1.48	2.44
PUPILEXP	0.62	0.10	0.49	0.85
CRIME	5.33	1.54	1.34	11.81
CBD	8.38	2.78	0.90	14.41
AVGSC9	87.87	8.95	64.10	98.10
MHVALUE	6.77	1.72	4.54	10.11

Regarding variable, GEOPIND, Table (10) reports the standardized price indices. The summary report of GEOPIND in this table, which are the entry prices in the NL model, are not the standardized price index, but the original fitted values of the hedonic model. The reason of doing this is to avoid the errors caused by the standardization procedure.

The indicators of dissimilarity are used to measure how different a household is from its neighbors. If households prefer to group themselves into homogenous neighborhoods, the signs of those coefficients on these indicators are expected to be negative. The coefficient on DFMASTER measures how the location decision of a household with a Masters degree is affected by the percentage of households whose highest level of education is not a Masters degree. The variable DFCHILD is included to test whether a household with children is more attracted to a community where most of his/her neighbors also have children. The variable DFINC gives a measure of the dissimilarity of household and its neighbors in regards to income. In the same vein, the coefficients on DFWHITE and on DFBLACK capture the disutility caused by moving into a community where the composition of ethnicities is different from a household's own race.

In order to distinguish the hypothesis of households' preferences for similarity from Tiebout's hypothesis that households prefer to live in a community with large houses, a pair of variables are created: DFHVAL and DFHVALTIE. The former is the absolute difference of a household's house value and its median neighbor's, while the latter is the simple difference of house values. A negative sign of DFHVAL is expected if households have a preference for similarity. According to Tiebout (1956) and Hamilton (1975), households prefer to buy a small house in a community with many relatively large houses, for the purpose of capturing a fiscal surplus. The variable DFHVALTIE is expected to have a negative sign because positive values of DFHVALTIE imply the household would be subsidizing others in the neighborhood through property tax payments. Because of the high correlation of DFINC and DFHVAL (rich households intend to purchase more expensive houses, while poor households can only afford cheap houses), we expect it be difficult to separate the effects of these two variables.

Table 3. Descriptive Statistics for Characteristic Difference between Household and Neighbors in NL

Difference Variables				
	Sample size: 1,467		Sample size: 731	
Variable	Mean	Std.Dev.	Mean	Std.Dev.
DFLESSHS	0.79	0.26	0.08	0.26
DFHS	0.12	0.28	0.11	0.28
DFCOLLEGE	0.19	0.32	0.20	0.33
DFBACHELOR	0.22	0.34	0.23	0.35
DFMASTER	0.16	0.33	0.15	0.33
DIFCHILD	0.26	0.32	0.33	0.32
DFWHITE	0.12	0.09	0.12	0.09
DFBLACK	0.06	0.23	0.07	0.24
DFINC	3.74	3.59	3.67	3.61
DFHVAL	-	-	3.13	3.70
DFHVALTIE	-	-	-0.07	4.85

(a) A household’s highest education achievement is categorized into less than high school, high school, some college degree, bachelor’s degree, and master’s degree. For whatever level of education the household achieves, we subtract the percentage of the population with that level from 1 and assign it to the household. All others are assigned zero. For instance, if a household has attained some college, and 45% of the population has this level, the DFCOLLEGE will be assigned .55, while DFLESSHS, DFHS, DFBACHELOR and DFMASTER will be assigned zero. By construction, these variables are always between zero and one.

(b) For a household that has children, DFICHILD is the absolute difference of one and the percentage of households that have children in the school district. For a household that has no children, DFCHILD is zero.

(c) For a white household, DFWHITE is the difference between 1 and the percentage population that is white in the neighborhood and DFBLACK is zero; for a black household, n DFBLACK is the difference between 1 and the percentage of population who is black and DFWHITE is zero.

(d) DFINC is the absolute value of the income difference between the household’s and the median neighbor’s (scaled by $10,000), and is always positive.

(e) DFHVAL is always positive, but DFHVALTIE can be negative or positive.

The mean values of community attributes and selected dissimilarity variables (Note 34) of each school district are plotted in Figure (2) and Figure (3). (Note 35) Figure (2) reports the distribution of the large sample used for the restricted model, and Figure (3) contains the distribution of the small sample for the full model. Both of these figures reveal that Upper Arlington school district (school 12) has the highest housing price index (GEOPIND), the highest average proficiency test score (AVGSC9/10), and the highest per pupil education spending (PUPILEXP). This school district is always considered as the most desirable one among all of the options for families with school age children. Grandview School District (school 4), whcih has the best accessibility and the lowest crime rate, is another highly demanded school district. Since Plain Local school district (school 8) is most far away from the city center (CBD), lands are relatively cheap, which results in big houses in average. This contributes to the highest median house value (MHVALUE) in Plain Local. It is also suggested that Bexley school district (school 1) has the least variation on hosuehold income among its residents, while the income variation is the highest in Columbus school district (school 3). In Figure (3), Plain Local school district (school 8) has the smallest DFHVALTIE and highest DFHVAL. This suggests that the values of a lot of houses in this district are lower than the average, and the downward deviations are big. The lowest value of DFHVAL in Canal Winchester local school district (school 2) implies that the house values there are mostly concentrated around the average value.

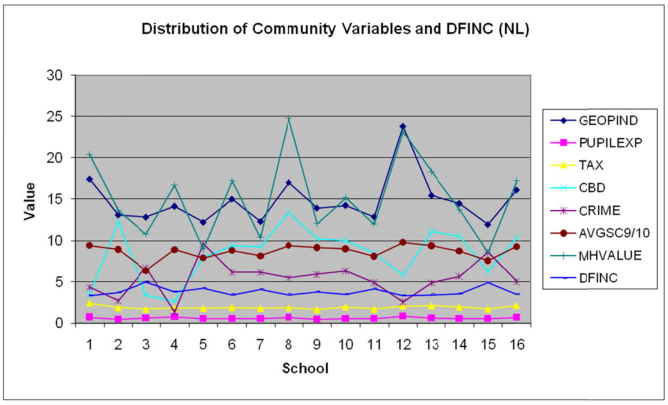

Figure 2. Distribution of Community Variables and DFINC (NL_Restricted Model)

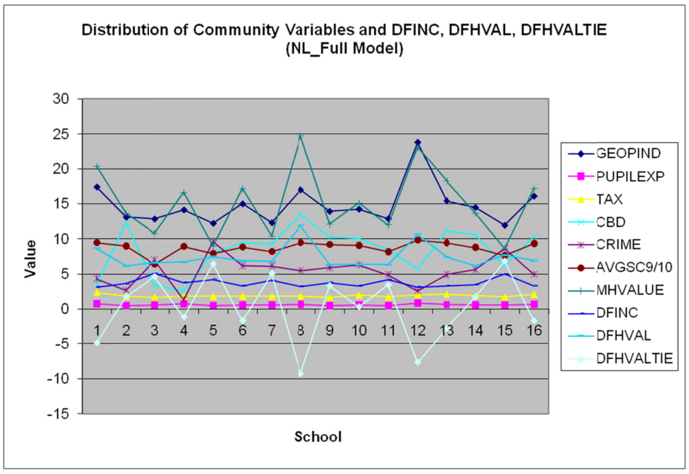

Figure 3. Distribution of Community Variables DFINC, DFHVAL, and DFHVALTIE (NL_Full Model)

There are a few competing hypotheses that might explain why households tend to group with similar neighbors. One of these competing hypotheses is based on the effect of zoning and limitations to the supply of housing. If a uniform zoning rule is imposed in a school district for residential lots, a homogenous community should be expected. Because there are multiple types of zoning rules governing residential housing in each community in the study area, none of the communities are homogenous. As a result, the composition of households within each community is not homogenous. This can be seen from the census data. Figure (4), Figure (5), and Figure (6) provide some evidence of the diversity in several sample school districts. While income levels tend to be greater in the Worthington and Southwestern school districts than in Columbus, there is a wide distribution in all areas. A similar results is observed for house values and family sizes.

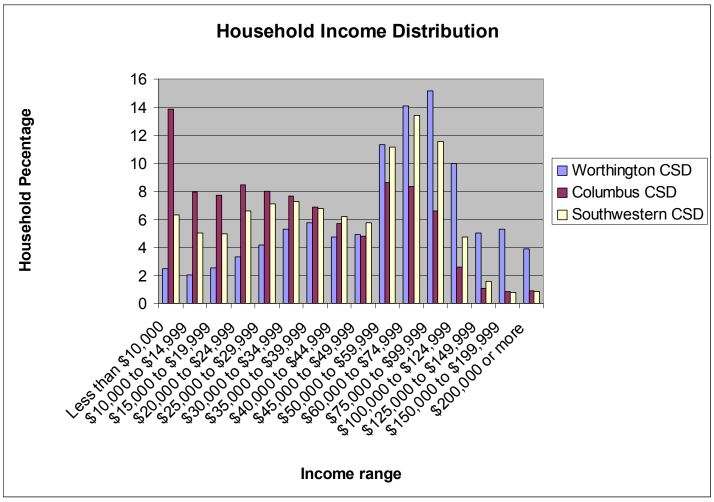

Figure 4. Household Income Distribution

Another competing hypothesis is that homogeneity is due to income segregation, which is predicted by urban economic models. This hypothesis predicts that households with similar incomes tend to group together, but it predicts nothing about their preferences along the other socio-economic dimensions. If the hypothesis of similarity search is corroborated in more than just income dimension, then the income segregation hypothesis does not completely describe the results.

If black households select to live with other black households, it either suggests the pure preferences of blacks for black culture, or it might be caused by a constrained choice set faced by black households, due to racial discrimination. For example, Yinger (1998) documents real estate agents engage in racial steering. This study cannot separate discrimination from preferences as the cause of racial segregation.

4. Estimation: Results

In this model, the eleven possible community choice alternatives are divided into two subgroups that comprise the central city nest, and the suburbs. Given the geography of Franklin County, the nest of the central city consists of the Columbus City School District, while the nest of the suburbs consists of the rest of the school districts. (Note 36) The results of the NL regression are reported in Table (4). (Note 37)

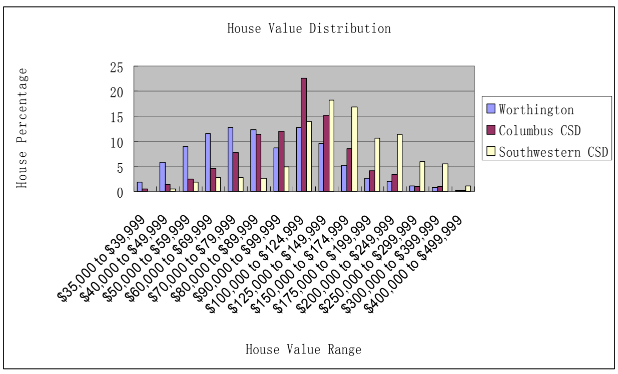

Figure 5. House Value Distribution

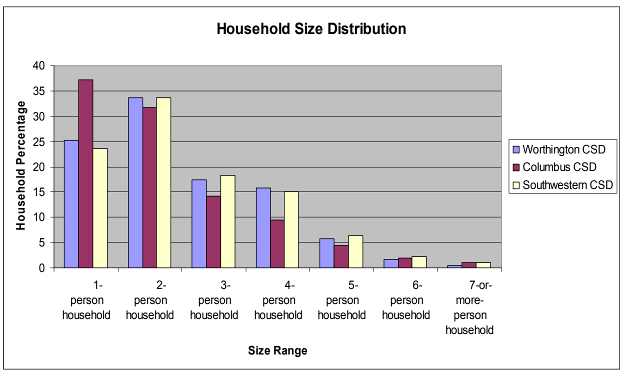

Figure 6. Household Size Distribution

Table 4. Estimates of the Coefficient in the NL Model

Specification Explanatory Variables	I Restricted Model Sample size:1467	II Full Model Sample: 731
Panel A: Dissimilarity Variables		
DFLESSHS (< high school absolute difference)	4.43** (1.81)	-
DFHS (high school absolute difference)	2.57** (1.07)	-
DFCOLLEGE (college absolute difference)	6.89** (2.94)	-
DFBACHELOR (bachelor absolute difference)	**-7.38***** **(0.91)**	**-4.35***** **(1.03)**
DFMASTER (master absolute difference)	**-9.95***** **(1.17)**	-1.77 (1.50)
DIFCHILD (child absolute difference)	**-12.05***** **(1.41)**	-1.85 (1.36)
DFINC (income absolute difference)	**-0.12***** **(0.03)**	-0.04 (0.05)
DFWHITE (white absolute difference)	-1.40 (0.95)	-1.72 (1.15)
DFBLACK (black absolute difference)	**-3.15*** **(1.71)**	-1.36 (1.88)
DFHVAL (house value absolute difference)	-	**-0.15***** **(0.01)**
DFVALTIE (house value difference)	-	**-0.46***** **(0.04)**
Panel B: Community Variables		
GEOPIND (price index)	**-1.84***** **(0.09)**	**-0.37***** **(0.10)**
PUPILEXP (per pupil education expenditure)	**47.38***** **(2.13)**	**6.46**** **(2.56)**
CBD (distance to CBD)	-0.012 (0.03)	0.05 (0.04)
PTAX (property tax)	**-17.59***** **(0.59)**	**-8.55***** **(0.56)**
PCRIME (crime rate)	**-0.25***** **(0.03)**	**-0.13***** **(0.03)**
AVGSC9 (average score)	**0.14***** **(0.02)**	-0.08*** (0.02)
MHVALUE (median house value)	**1.28***** **(0.08)**	-
Panel C: IV Parameters($1-\sigma_l$)		
Central City	**1.00** **(Fixed Parameter)**	**1.00** **(Fixed Parameter)**
Suburbs	**0.94***** **(0.01)**	**1.03***** **(0.01)**

* indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level or less

The dependent variable of the NL model is the choice of school district. For example, if a household buys a house in Bexley school district, we assign "1" to Bexley and "0"s to the other school districts.

Column I shows the results of the restricted model on a large sample, while column II displays the results of the full model with the house value difference variables. Due to the smaller sample size of the full model, column II in Table (4) excludes some insignificant variables to enable the convergence of the iterations. Because most of the excluded variables are the difference variables, our discussion that follows is mainly based on the restricted model (column I). The report on the results of the full model here is for the purpose of discovering the effects of variable DFHVAL and variable DFHVALTIE. (Note 38)

Recall that when the IV parameter is 1 a nested logit model becomes a conditional logit model. The estimated IV parameters of the NL model found in Table (4) are very close to 1, which explains why Hausman-McFadden Test is only partially passed for the CL model. (Note 39)

In column I of Table (4), more than half of the dissimilarity variables perform well with the expected negative sign, which confirms the hypotheses in this paper and suggests that households are more attracted to a similar neighborhood. The community variables also perform well. Other estimation results are consistent with previous studies.

4.1 Preference on the Similarity of Education Background

Among the five categories of education levels, the dissimilarity measures of the two highest education level, DFMASTER and DFBACHELOR have the expected negative coefficients, but the other three have significant positive estimates. The negative signs of DFMASTER and DFBACHELOR suggest that a household with a master's degree or a bachelor's degree tends to group with others who have a similar educational background, while the positive signs of DFLESSHS, DFHS, and DFCOLLEGE indicate that a household without a bachelor's degree would be discouraged in living with someone else who is equally educated as he/she is. The former finding is consistent with the hypothesis in this paper. But the latter one seems to contradict the similarity hypotheses. To test whether the households in the three lower level educational categories would prefer an educationally heterogeneous neighborhood or would be more interested in choosing better educated neighbors because of the positive externality of education, two other variables that indicate the neighborhood education level (variable SDBA is the percentage of the population in the school district that has a Bachelor's Degree and the variable SDMA is the percentage of the population in the school district that has a Master's Degree) are added in the NL model. After the inclusion of these two new variables, DFLESSHS, DFHS, and DFCOLLEGE are no longer significant, while SDBA, and SDMA are significantly positive, and DFMASTER and DFBACHELOR keep their negative signs. These new results suggest that a household is in general more attracted by a better educated neighborhood, but discouraged by a poorly educated one, no matter whether he/she is well educated or not. Additionally the unchanged negative coefficients on DFMASTER and DFBACHELOR suggest that the general preference of a household over a better educated neighborhood does not dominate the preference of a highly educated household over a residency with homogenously educated residents. In other words, a household with a bachelor's degree or a master's degree is not only attracted by a well educated neighborhood area, but also interested in living with others who are equally educated as he/she is, while a household with little education would prefer to choose better educated neighbors.

For those well educated households, similar high education background implies more common interests for those residents, which in turn creates a more interactive community. These households enjoy a higher utility in a highly interactive community in contrast to a segmented one. In contrast, relatively poorly educated households, appear to prefer trading a homogenous neighborhood for a positive externality flow of education.

4.2 Preference for Households with Children

The negative sign of the coefficient on DFCHILD indicates that a household that has children is less likely to choose a community with few children per household. (Note 40) The needs of reciprocated exchange of childcare could be one of the motivations of households with children for similar neighbors. When a family chooses a residence, the parents come up with a question naturally: Will there be other children nearby for my children to play with? This consideration should lead to the preference of the with-children households for a neighborhood with a lot of other with-children residents. Residing in this kind of community, parents can exchange advice and information about childbearing and help each other with the task of controlling children, while children can socialize with each other in their social network.

4.3 Preference for the Similarity of Income

The negative sign of the coefficient on DFINC is consistent with the hypothesis that households prefer to group themselves within a neighborhood of similar income households. Because income works as a signal, indicating hobbies and interests of households, seeking similar income neighbors is another way to approximate an interactive community and therefore create greater gains from the social interaction in the neighborhood.

4.4 Preference for the Similarity of Races

The coefficient on DFBLACK is negative and significant, which implies that a black household is more likely to live with blacks. The coefficient on DFWHITE is negative but not significant, which implies that white households do not have a particular preference regarding their neighbors' races. The insignificance of DFWHITE might result from the approximation of the percentages of white households across the sample school districts. (Note 41) These results are partially consistent with the hypothesis of searching for similarities.

The negative coefficient on DFBLACK may suggest another explanation for the racial segregation phenomenon. Given results of previous studies that suggested black households' preference over racially integrated neighborhoods, the findings in this paper point to another possible explanation. The segregation of black households and white households might be not caused by racial discrimination, but households' own preferences over the neighbors' ethnicities. This self-selection might result from comfort with the culture, which implies that a racially segregated community is more interactive than an integrated one for the members in it. Of course, black households' choice might not necessarily suggest their preference, but their inability to locate in a white neighborhood, which is caused by racial discrimination. Both of the explanations are consistent with the findings in this paper, thus further study will be required.

4.5 Preference on the Community Attributes

All of the community variables have the expected significant signs. The positive sign of the coefficient on PUPILEXP suggests the greater educational expenditures per pupil; the more attractive is a community. The positive sign of AVGSC9 implies that the better the scholarly performance of students, the more desirable is a community. For parents who have children, greater school spending and a high proficiency score signal better student performance. For households without school age children, greater school spending and high proficiency score imply a better investment value of houses in the school district. (Note 42) The negative coefficients on TAX, CRIME and GEOPIND indicate that high tax rate, high crime rate, or high entry price of a school district make a community less attractive. The positive sign of MHVALUE suggests that households are attracted to a community with a relatively high median house value, which is consistent with Tiebout's hypothesis. The coefficient on CBD is not significant, implying the distance between the property and the CBD does not affect a household`s choice.

The fact that the coefficients of DFMASTER, DFBACHELOR, DFCHILD, and DFBLACK are significant even when DFINC is included in the regression suggests that a household's preference toward neighborhood with a similar socio-economic background is not swamped by housing affordability. The high correlation between income and education/ethnicity/age might suggest that the observed homogeneous neighborhood (i.e., similar education background/ethnic group/family structure) is likely the result of households with similar income and therefore can afford houses in the same neighborhood. Our regression result indicates that households' preference for similarity persists when income level is controlled. In other words, though households with similar income can afford houses in the similar price range, because of their preference for similarity they do not necessarily end up buying houses in the same community.

Column II of Table (4) provides the results of complete model that includes the house value differences variables, most of which are consistent with the results in column I of Table (4), except some variables that are no longer significant. This is partially because of the reduced sample size. Another reason is that the variable DFHVAL picks up the effects of DFINC and DFMASTER. Because households' income differences and education differences are highly correlated to their house value differences, it is not surprising to see that the coefficients of DFINC and DFMASTER are no more significant after the inclusion of DFHVAL. The negative coefficient on DFHVAL confirms the similarity hypothesis, while the negative sign of DFHVALTIE confirms Tiebout's hypothesis. If household i's house value, w_i, is greater than the average house value in the community jl, w_{jl}, \$10,000 increase of $w_i$will decrease i's total utility by 0.61. (Note 43) If household i's house value, w_i, is less than the average house value in the community jl, w_{jl}, \$10,000 increase of w_i will decrease i's total utility by 0.31. (Note 44) The dominant effect of DFHVALTIE suggests that a household would always like to buy a cheaper house than his/her neighbors' house values. This effect adds to the effect of a household's preference on similarities, which creates an asymmetry of preferences for downward deviations over upward deviations from the average house value in the neighborhood. This finding extracts the influence of DFHVAL from the prominent influence of DFHVALTIE.

5. Estimation: Discussion

A case study of the Columbus city school district provides more insight about the estimated results. This case study is based on the NL results in column I of Table (4).

To aid in interpretation of the findings, Table (5) contains the estimated marginal effects on the probabilities with respect to the explanatory variables in the Columbus city school district, while Table (6) reports the elasticity effects on the probabilities with respect to the explanatory variables in the Columbus city school district. (Note 45) All marginal effects and elasticity effects are evaluated for each individual household separately, and then averaged across all households of the appropriate subgroup.

Table 5. Variable Marginal Effects, the Columbus City School District

School District	Marginal Effects				
Panel A: Dissimilarity Variables					
	DFBACHELOR	DFMASTER	DIFCHILD	DFINC	DFBLACK
Gahanna	6.61	8.91	10.79	0.10	2.83
Bexley	1.00	1.35	1.64	0.02	0.43
Upper Arlington	10.27	13.85	16.78	0.16	4.39
Hilliard	2.66	3.58	4.34	0.04	1.14
Southwestern	1.36	1.84	2.22	0.02	0.58
Grove City	5.99	8.07	9.77	0.09	2.56
Reynoldsburg	3.90	5.25	6.34	0.06	1.66
Dublin	6.06	8.16	9.89	0.10	2.59
Westerville	5.87	7.91	9.59	0.10	2.51
Worthington	2.06	2.78	3.36	0.03	0.88
*Columbus	-45.78	-61.69	-74.73	-0.72	-19.56

Panel B: Community Variables						
	GEOPIND	PTAX	PCRIME	AVGSC9	MHVALUE	PUPILEXP
Gahanna	1.65	15.75	0.22	-0.12	-1.15	-42.43
Bexley	0.25	2.39	0.03	-0.02	-0.17	-6.44
Upper Arlington	2.56	24.47	0.35	-0.19	-1.78	-65.93
Hilliard	0.66	6.33	0.09	-0.05	-0.46	-17.06
Southwestern	0.34	3.24	0.05	-0.03	-0.24	-8.74
Grove City	1.49	14.26	0.20	-0.11	-1.04	-38.42
Reynoldsburg	0.97	9.28	0.13	-0.07	-0.68	-25.00
Dublin	1.51	14.42	0.20	-0.11	-1.05	-38.86
Westerville	1.46	13.99	0.20	-0.11	-1.02	-37.68
Worthington	0.51	4.91	0.07	-0.04	-0.36	-13.22
*Columbus	-11.39	-109.04	-1.54	0.85	7.95	293.77

* indicates direct marginal effect of the attribute. The responses are for a one unit change in the direct effect, which would be a 100% change.

Table 6. Variable Elasticity Effects, the Columbus City School District

School District	Elasticity Effects				
Panel A: Dissimilarity Variables					
	DFBACHELOR	DFMASTER	DIFCHILD	DFINC	DFBLACK
Gahanna	0.169	0.133	0.277	0.039	0.005
Bexley	0.169	0.133	0.277	0.039	0.005
Upper Arlington	0.169	0.133	0.277	0.039	0.005
Hilliard	0.169	0.133	0.277	0.039	0.005
Southwestern	0.169	0.133	0.277	0.039	0.005
Grove City	0.169	0.133	0.277	0.039	0.005
Reynoldsburg	0.169	0.133	0.277	0.039	0.005
Dublin	0.169	0.133	0.277	0.039	0.005
Westerville	0.169	0.133	0.277	0.039	0.005
Worthington	0.169	0.133	0.277	0.039	0.005
*Columbus	-1.291	-1.342	-2.805	-0.366	-0.193

Panel B: Community Variables						
	GEOPIND	PTAX	PCRIME	AVGSC9	MHVALUE	PUPILEXP
Gahanna	2.605	3.234	0.098	-1.125	-0.848	-2.917
Bexley	2.605	3.234	0.098	-1.125	-0.848	-2.917
Upper Arlington	2.605	3.234	0.098	-1.125	-0.848	-2.917
Hilliard	2.605	3.234	0.098	-1.125	-0.848	-2.917
Southwestern	2.605	3.234	0.098	-1.125	-0.848	-2.917
Grove City	2.605	3.234	0.098	-1.125	-0.848	-2.917
Reynoldsburg	2.605	3.234	0.098	-1.125	-0.848	-2.917
Dublin	2.605	3.234	0.098	-1.125	-0.848	-2.917
Westerville	2.605	3.234	0.098	-1.125	-0.848	-2.917
Worthington	2.605	3.234	0.098	-1.125	-0.848	-2.917
*Columbus	-27.000	-34.359	-1.148	11.665	8.793	30.238

The own marginal effect of locating in the Columbus city school district for the attribute DFCHILD is negative and has the largest absolute value among all the dissimilarity variables, which implies that households are more responsive to the change of DFCHILD than the other difference variables. This is also confirmed by the elasticity of DFCHILD. For a household with children, a 1 percentage point of decrease on the number of households who also have children in the Columbus city school district will decrease the probability of choosing the Columbus city school district by 0.75 percentage point, at the same time increase the probability of the Upper Arlington city school district being chosen by 0.17 percentage points, and increase the probability of the Worthington city school district being chosen by 0.03 percentage points. The high values of the own marginal effect and the estimated elasticity of PUPILEXP in Columbus show how significant the expenditures on schools are in influencing households' location decisions. (Note 46) A $1,000 increase of per pupil expenditure in Columbus leads to a 29.38 percentage points increase in the probability of moving into Columbus. (Note 47) A 1 percent increase of per pupil expenditure in Columbus causes a 30.24 percent increase in the likelihood of choosing Columbus. Because per pupil expenditure can be used to measure school quality, the large value of the own marginal effect of PUPILEXP implies households' concern for school quality is a major determinant of location choice when they migrate.

6. Conclusions

The focus of this paper has been to test the effects of dissimilarity on households' residential location choice. This was tested by a two step procedure. First a location choice model requires inclusion of a house price index and thus the construction of an estimate of that index. Next, with an estimate of the house price index in hand, we estimate a nested logit regression model. Several data sets containing both households' characteristics and community attributes are utilized. The location decision of a household is modeled as a searching process for a matching community along the dimensions of households' socio-economic characteristics. The dissimilarity variables have been grouped into five categories: educational background, with school age children or not, race, income, and house value.

The findings reveal that a household prefers neighbors who are like herself/himself with the exception that poorly educated households prefer better educated neighbors. It is hypothesized that the reason for these preferences over similarity comes from the desirability of social interaction, while the preferences of households with little education over better educated neighbors suggests the positive externality flow of education. The revealed preference of white households in our sample suggests that they are indifferent about their neighbors' races, which might be explained by the small variations on the percentages of white households across the sample school districts. These results suggest an alternative explanation for the widely observed phenomena of homogeneous communities. The concentration of like households in many communities may be the result of the households' own volunteer selection for the purpose of beneficial social interactions among neighbors.

The introduction of households' preferences on similarities in this study raises additional questions, whcih provides fertile ground for future research. For example, in this paper, the preferences for similarity are tested along the dimensions that are constrained by the data in hand. The similarity of households can also be measured on other charateristics, such as religion, and social-political attitude. Furthermore, the analysis here is based on the benificial social interactions that are derived from similar individuals. An extension of of current study is to introduce heterogeneous preference structures among like households, allowing beneficial social interactions to arise from individuals with complementary needs in addition to individuals with similar characteristics. There may exist variations of households' utility determination processes based upon on their characteristics. Some households gain positive utility from bounded heterogeneous communities, whereas others lose utility on that. For instance, better educated households may respond differently than less educated households to a racially homogeneous community. The possible findings based on alternative preference settings might strengthen the robustness of the resutls in this paper and provide suggestions on policies about neighborhood integration.

In addition, Tiebout's hypothesis has been tested and has been corroborated by the data in this paper. Households would like to buy a somewhat cheaper house as compared with the values of the neighbors' houses in order to gain a fiscal surplus. Meanwhile, the house values difference of a household and his/her neighbors' should be small enough for the household to stay in a relatively homogeneous community.

Acknowledgements

The author is grateful for comments from Donald Haurin, Lung-Fei Lee, the participants at Halle Forum on Urban Economic Growth in Germany, and the participants at the section of Migration, Residential Mobility and Choice in the 53rd North American Regional Science Association International Conference. The author retains responsibility for all errors.

References

Can, A. (1992). Specification and Estimation of Hedonic Housing Price Models. *Regional Science and Urban Economics, 22*, 453-474. http://dx.doi.org/10.1016/0166-0462(92)90039-4

Center for Urban and Regional Analysis, Ohio State University. (1995). Housing Survey in Franklin County, and 2001 Housing Survey for Franklin County (pre-September 11 and post-September 11).

Clark, William A. V. (1986). Residential Segregation in American Cities. *Population Research and Policy Review, 5*, 95-127. http://dx.doi.org/10.1007/BF00137176

Clark, William A. V. (1991). Residential Preference and Neighborhood Racial Segregation: A test of the Schelling Segregation Model. *Demography, 28*, 1-19. http://dx.doi.org/10.2307/2061333

Coleman, J. S. (1990). *Foundations of Social Theory.* Cambridge, MA: Harvard University press.

Diose, W., & Mugny, G. (1984). *The Social Development of Intellect.* Oxford, England: Pergamen.

Dubin, Robin A. (1988). Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *Review of Economics and Statistics, 70*, 466-474. http://dx.doi.org/10.2307/1926785

Friedman, J. (1980). A Conditional Logit Model of the Role of Local Public Services in Residential Choice. *Urban Studies, 18*, 347-258. http://dx.doi.org/10.1080/00420988120080641

Gabriel, Stuart, & Rosenthal, Stuart. (1989). Housing Location and Race: Estimates of a Multinomial Logit Model. *The Review of Economics and Statistics, 71*, 240-249. http://dx.doi.org/10.2307/1926969

Greene, W. (1997). *Econometric Analysis*, 3rd edition. Upper Saddle River, NJ: Prentice Hall.

Greene, W. (2004). *NLOGIT 3.0 Reference Guide*.

Hagestad, G. O., & Uhlenberg, P. (2005). The Social Separation of Old and Young: A root of Ageism. *Journal of Social Issue, 61*(2), 343-360. http://dx.doi.org/10.1111/j.1540-4560.2005.00409.x

Hamilton, B. W. (1975). Zoning and Property Taxation in a System of Local Governments. *Urban Studies 12*(2), 205-11. http://dx.doi.org/10.1080/00420987520080301

Hamilton, B. W. (1976). Capitalization of Intrajurisdictional Differences in Local Tax Prices. *The American Economic Review, 66*, 743-754.

Hausman, Jerry, & Mcfadden, Daniel 1984. Specification Tests for the Multinomial Logit Model. *Econometrica, 52*(5), 1219-40. http://dx.doi.org/10.2307/1910997

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The review of Economic Studies, 60*, 531-542. http://dx.doi.org/10.2307/2298123

Manski, C. (2000). Economic Analysis of Social Interactions. *Journal of Economic Perspectives, 14*(3), 115-36. http://dx.doi.org/10.1257/jep.14.3.115

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology, 58*, 1246-1266. http://dx.doi.org/10.2113/gsecongeo.58.8.1246

Mcfadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economy, 33*, 303-28. http://dx.doi.org/10.1016/0047-2727(74)90003-6

National Center for Education Statistics. Census. (2000). School District Demographics.

Nechyba, T., & Strauss, R. (1997). Community Choice and Local Public Services: A Discrete Choice Approach. *Regional Science and Urban Economics, 28*, 51-73. http://dx.doi.org/10.1016/S0166-0462(97)00013-6

Pollak, A. (1976). Interdependent Preferences. *The American Economic Review, 66*, 309-320.

Priest, R. F., & Sawyer, J. (1967). Proximity and Peership: Bases of Balance in Interpersonal Attraction. *The American Journal of Sociology, 72*(6), 633-649. http://dx.doi.org/10.1086/224400

Quigley, John M. (1976). Housing Demand in the Short Run: An Analysis of Polytomous Choice. *Exploration in Economics Research, 3*, 76-102.

Quigley, John M. (1985). Consumer Choice of Dwelling, Neighborhood and Public Services. *Regional Science and Urban Economics, 15*, 41-63. http://dx.doi.org/10.1016/0166-0462(85)90031-6

Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. Oxford, England: Oxford University Press.

Rosen, Sherwin. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy, 82*, 34-55. http://dx.doi.org/10.1086/260169

Sampson, J. R., Raudenbush, S. W., & Earls, Felton. (1997). Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science, 277*, 918-924. http://dx.doi.org/10.1126/science.277.5328.918

Sampson, R. J., Morenoff, J. D., & Earls, F. (1999). Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children. *American Sociological Review, 64*(5), 633-660. http://dx.doi.org/10.2307/2657367

Stutz, F. P. (1973). Distance and Network Effects on Urban Social Travel Fields. *Economic Geography, 49*(2), 134-144. http://dx.doi.org/10.2307/143082

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy, 64*, 416-424. http://dx.doi.org/10.1086/257839

US Bureau of Census. (2000). Census (SF 3).

Winch, R. F. (1952). *The Modern Family*. New York: Holt.

Yinger, John. (1982). Capitalization and the Theory of Local Public Finance. *The Journal of Political Economy, 90*, 917-943.

Yinger, John. (1998). Housing Discrimination Is Still Worth Worrying About. *Housing Policy Debate, 9*(4), 917-943.

Notes

Note 1. The rewards of complementary needs are not the focus of this paper.

Note 2. The similarity search hypothesis is also associated with the relative deprivation theory in sociology. Relative deprivation refers to the discontent people feel when they compare their positions to those of other similarly situated and find out that they have less than they think they deserve. This theory implies that households tend to feel less happy if they do not do well as their neighbors, which is consistent with the idea of searching for similarities.

Note 3. In the sociology literature, the characteristics that determine households' similarities are measured or proxied by age, race, sex, income, occupation, education, and family size (for the potential reciprocated exchange of childcare). See Priest and Sawyer (1967), Stutz (1973), Sampson, Morenoff, and Earls (1999), Clark (1986, 1991), Hagestad and Uhlenberg (2005), and Coleman (1990).

Note 4. Zoning rules are the classification of allowable land use by a government, such as the classification for single-family residential district, the classification for multi-family residential district, the classification for business district, etc.

Note 5. Fiscal surplus is defined as the excess of local service benefits over local property tax payments.(Hamilton, 1976) Fiscal deficit is defined as the shortage of local service benefits over local property tax payments.

Note 6. A conditional logit (CL) model has been applied before the application of the NL model in this paper. Because the data failed to pass Hausman-McFadden Test (1984), the results from the CL are not presented.

Note 7. This IIA property can be stated as follows: The ratio of the probabilities of any two alternatives is independent of the choice set. The IIA derives from the assumption that the stochastic disturbance terms are independent and identically distributed. In the location choice model, the IIA assumption is likely to be violated if households perceive the destination alternatives as close substitutes. When the IIA property fails to hold, the nested logit (NL) model is an appropriate method of estimation.

Note 8. A generalized extreme-value distribution allows the correlation of error terms within a subgroup, which partially relaxes the IIA assumption.

Note 9. When $1-\sigma_l$ equals 1 for all $l=1,2,\ldots,L$ (no correlation within nests), the nested logit model collapses to a simple conditional logit model.

Note 10. The price of the non-housing goods is normalized to 1.

Note 11. Because of the different community entry prices, household i's non-housing consumption may differ across communities. The same reason holds for the variable $h_{i,jl}$. Both the non-housing consumption and housing consumption are indexed by the household and the community.

Note 12. The local taxes are not in the budget constraint, but appear as an attribute of housing and are capitalized in house price, $P_{i,jl}$.

Note 13. The community variables, Z_{jl}, represents the combination of x_{jl} and $Z_l$in section 2.1.

Note 14. In the indirect utility function, the prices, p_s and 1, substitute for the quantities, s_i and $c_{i,jl}$. Because the quantities of s_i and $c_{i,jl}$ are under control of individual i, but not the quantity of Z_{jl}, the price of Z_{jl} does not substitute for the quantity in the indirect utility function.

Note 15. Pollak (1976) stated that one' preference may be affected by others' consumption. Therefore, from a household's point of view, the characteristics of neighbors may be taken as signals of their consumption.

Note 16. Pricing homogeneity is different than the match of a particular household with a community. The former is the price of a measure of the similarity or diversity of a community that would permit a match. The latter measures the similarity or diversity between a particular household and a given community. To develop a price for homogeneity or diversity would require fully capturing the aspects of homogeneity or diversity that are included in the market price. But these aspects are not very clear. For example, a perfectly homogeneous community would be attractive to a limited group of similar people. Thus the market price could be low. Alternatively, a heterogeneous community would provide matches to a lot of people, but perhaps not good matches. Thus the market price could be low. In either case, it would be difficult to measure and difficult to develop this price. Therefore, this paper does not take the former approach. The omission of pricing of homogeneity is likely a minor omission.

Note 17. The negative sign of $(w_i - w_{jl})$ indicates that affluent community is in general attractive to potential households, while Tiebout hypothesized that poor households prefer to live in a rich community in order to capture the fiscal surplus, and rich households also like to live in a rich community to avoid a fiscal deficit.

Note 18. In the approach of equation (12), $school_j$ and $Z_{j(i)}$ are employed to indicate the community quality that is capitalized into the house price. The author also tried another alternative approach that leaves out $Z_{j(i)}$, keeping only the dummy variables, $school_j$, to capture the differences of non-housing-structural attributes across communities. The results show that the first approach has a better fit. This may result from the utilization of more complete information in the first approach. The result of the second approach is not presented in this paper, but can be provided upon request.

Note 19. All the houses in the data set used here are owned by households, therefore, rents are not considered.

Note 20. In the hedonic regression, the neighbors' socio-economic characteristics and community amenities are not distinguished from each other. They are all expressed by $Z_{j(i)}$. Due to the merger of three data sets, some variables of $Z_{j(i)}$ are on the census tract level, varying across census tracts within a school district; while some other variables of $Z_{j(i)}$are on the school district level that is constant within a school district. Therefore, Z is indexed by both the school district j and household i.

Note 21. The IID assumption is relaxed in the section of Spatially Autocorrelated Error Terms in the Hedonic Regression.

Note 22. Can (1992) separated location effects into two categories: The first one is called neighborhood effects, which is the effects from the absolute geographic location of a house, such as the public services and accessibilities, and can be capitalized into the house value as a premium or through varying implicit prices of attributes. The other one is adjacency effects, which are associated with the externalities from the surroundings of the housing unit.

Note 23. For an easy interpretation of the parameters, a linear function is employed. A semi log form has also been tested by the author. It produces highly consistent results as the linear regression. Furthermore, both the price indices created from the semi log form and the linear form are applied in the NL model, and similar estimates are generated. The results of the semi log form are not presented in this paper, but it is available from the author under request.

Note 24. If we define $\widehat{W}_j$ as the fitted values of hedonic regression given the estimated coefficients and the average values of characteristics on the school district level, then $\widehat{W}_j = \hat{\alpha}_0 + \sum_{j=1}^{J-1} \hat{\alpha}_{0j}\, school_j + \hat{\alpha}_1 \bar{Z}_j + \hat{\beta}\bar{s}$. The price index, $\widehat{w}_j$, is the value of $\widehat{W}_j$deducting a constant term, $\hat{\alpha}_0 + \hat{\beta}\bar{s}$, which represents the spending on a standard housing bundle of structural attributes. According to the indirect utility function, the price index, $\widehat{w}_j$, is the entry price of community j, and will be used in the logit models.

Note 25. The data from CURA contains both households' socio-economic information and the housing structural characteristics. The former is useful in the estimation of the NL model; the latter is useful in the estimation of hedonic model.

Note 26. Originally, there are 1258 observations in the pre-September 11 survey, and 803 observations in the post-September 11 survey. The combination of them gives 2,061 observations. (These are the raw numbers before the data cleaning process.)

Note 27. This feature of our data has both pros and cons. As a random sample of population, sample selection bias is eliminated. On the other hand, being observations of new movers, the observed characteristics of households may not be consistent with their choices that were made before.

Note 28. Data for the hedonic price function is reported in Appendix A1.

Note 29. There are 22 school districts listed under Franklin County. Because some school districts are primarily located outside Franklin County, only 16 school districts are considered when the price indices are created. Due to insufficient observations, 5 school districts are left out in the NL model. They are the Canal Winchester local school district, Grandview heights school district, Hamilton local school district, Plain local school district, and Whitehall city school district.

Note 30. McFadden (1978) suggested that the discrete choice model may be estimated by estimating a sub-sample of the alternatives from the full choice set. As long as an alternative has the logical possibility of being an observed choice, given it is included in the assigned set, consistency holds.

Note 31. The discrepancy of the observation numbers for the hedonic and the logit models results from the missing variables in the CURA survey. Among the 2,061 observations in the CURA survey, only 962 of them have complete information about house structural variables needed for the hedonic estimation.; however, 1,467 observations have complete information of the household' socio-economic characteristics, which are needed in the estimation of location choice.

Note 32. The household head's socio-economic characteristics are used to compute the difference. To use the household head' value instead of the family average is a very common approach in the empirical studies of the housing market. This occurs primarily due to data limitations.

Note 33. The construction of the difference in house value requires every observation to have a transaction price variable. Therefore, due to the missing variables, some observations need to be deleted whenever a transaction price variable is included in the estimation, which gives us 731 total observations. The large sample with 1,467 observations is used for a restricted model with the exclusion of house value difference variables. The small sample with 731 observations is used for a full model including the house value difference variables. The statistical description of the data with the variable of the housing value difference and the data without it are presented separately, as well as the regression results.

Note 34. DFINC, DFHVAL, and DFHVALTIE are the only difference variables presented in these figures. Because the mean value of any other difference variable in each school district does not have a clear economic meaning, the distribution of them are not reported here. For example, a low value of DFWHITE can be interpreted as either a high concentration of white population or a low concentration of whites. If there are a lot black residents, many zeros will be assign to the variable of DFWHITE. On the contrary, if there are a lot white residents, the percentage of whites will be high, which means the value of DFWHITE (one minus the percentage of whites) will be low for the white households.

Note 35. Although 5 schools are excluded from the NL estimation, the values of their community attributes and the selected difference variable are still included in the figure. To better fit in the figure, AVGSC9 is scaled by 10. It is renamed to AVGSC9/10 in the figure.

Note 36. Other nesting structures have been tested besides the one that is reported in this paper. Comparing the estimated parameters of the inclusive values, $(1-\sigma_l)$, only the nesting structure reported provides the most appropriate value, which is between 0 and 1 (Greene 1997).

Note 37. Results of the hedonic price function are reported in Appendix A2. The set of price indices (GEOPIND) constructed using the estimates of the hedonic function enter the NL regression as one of the explanatory variables.

Note 38. Because variable DFHVALTIE, $(w_i - w_{jl})$, is not available in the restricted model, we include MHVALE, which measures w_{jl}, to explore a household's preference for communities with large houses.

Note 39. Actually, the point estimates of all other coefficients as well as their standard errors in both the CL model and the NL model are nearly indistinguishable. The estimates of the CL model are available from the author upon request.

Note 40. The opposite also holds. The author replaced DFCHILD by the variable, DFSINGLE, whose coefficient measures the preference of single families, and the results provided the similar conclusion: a single prefers a community with more singles.

Note 41. The percentage of white households ranges between 72.8 percent and 94.7 percent for the 11 sample school districts. Meanwhile, the percentage of black households ranges between 2.1 percent and 21.2 percent. Obviously, the variation of the latter is much higher than the former. This might explain the statistical significance of DFBLACK, and the statistical insignificance of DFWHITE.

Note 42. The households with no children can easily extract the benefits of the greater school spending through selling the house to some family that have school age children.

Note 43. Household i's total utility is affected by the both DFHVALTIE $(w_i - w_{jl})$, and DFVAL $(|w_i - w_{jl}|)$, while the coefficient of $(w_i - w_{jl})$, is -0.46, and the coefficient of $|w_i - w_{jl}|$ is -0.15. Thererfore, the total effect is -0.61.

Note 44. Given the estimated coefficients, the marginal effect of DFHVAL and the marginal effect of DFHVALTIE have been computed. For example, the direct marginal effect of DFHVAL in Columbus school district is -2.01%, and the direct marginal effect of DFHVALTIE in Columbus school is -6.58%. These values suggest that if household i's house value is less than the average house value in the community jl, \$10,000 increase of w_i will decrease the probability of i choosing Columbus school district by 4.57 percentage points.

Note 45. The marginal effect can be interpreted as the percentage points change of the probability of moving into school district j (where j can be any school district of the 11 school districts in the choice set) with response to a unit change in the attributes of the Columbus City School District. The elasticity effect can be interpreted as the percentage change of the probability of moving into school district j (where j can be any school district of the 11 school districts in the choice set) with response to one percentage change in the attributes of the Columbus City School District.

Note 46. A large value of PUPILEXP signals not only a school district with good public schools, but also the availability of a group of high quality children. For those families with children, this kind of school districts are particularly attractive because of the potential beneficial peer effects on their children.

Note 47. Because the variable PUPILEXP measures per pupil expenditure in $10,000, the big value of the marginal effect of PUPILEXP (293.77%) can be interpreted as 29.38 percent increase on a $1,000 change in per pupil expenditure.

Note 48. Some estimates of the community variables are statistically insignificant, but they are still left in the hedonic function when constructing the price indices.

Note 49. The "Chi-Square" value is -2 times the log likelihood from the null model minus -2 times the log likelihood from the fitted model, where the null model is the one with the same explanatory variables and iid distributed errors. Because the ML estimates of a regression with iid disturbances equal the OLS estimates of such regression, the likelihood ratio test can be interpreted as the test for the MLE of geostatistical model and the OLSE of the hedonic regression. In addition, the out-sample prediction power of geostatistical model relative to OLS has been widely confirmed by the literature. Because of the small sample size, this paper skips the prediction test.

Note 50. OLS price indices are tested in the NL framework by the author and the results are indeed not different from ML estimates. Therefore, they are not reported here.

Appendix A1. Data for the Hedonic Price Function

The recorded transaction dates (yy/mm/dd) in the 2001 surveys make it possible to adjust for inflation across time. The transaction prices are converted to a base period (1st quarter, 2001) dollar value using a quarterly house price index for Columbus Metropolitan Statistics Area (MSA) from Freddie Mac.

The variables included in the estimation of hedonic prices are categorized into two subgroups corresponding to the hedonic function (13). Table (7) provides the definition and summary statistics of the housing structural characteristics. Each house is assigned values of the selected community's attributes and neighbors' socio-economic characteristics, which are elements of $Z_{j(i)}$. The definition and descriptive statistics of the community variables are provided in Table (8). The merger of these two data sets leaves 962 observations available. The variation of these housing characteristics and community attributes across school districts are plotted in Figure (7) and Figure (8).

Table 7. Descriptive Statistics on the Housing Structural Variables

Panel A: Definition of Variables

Variable	Definition
PRICE	Transaction price in $2001(Q1) and divided by 10,000
ROOMS	Number of total rooms
ROOMS2	Number of total rooms squared
HOUSEAGE	Age of the structure
HOUSEAGE2	Squared age of the structure
BATHTOT	Number of full bathrooms+0.5*number of half bathrooms
LOTS	Lot size in thousands of square feet
LOTS2	Lot size squared
LOTS3	Lot size cubed
FIREPLACE	Dummy variable that indicates the existence of a fireplace
NOSTORIES	Number of stories and times 10
CONDITION	Dummy variable that indicates the condition of a house (high number means better condition)
SQFT	Building area in thousands of square feet
CROSS1	Cross term of SQFT and NOSTORIES
CROSS2	Cross term of SQFT2 and NOSTORIES
CROSS3	Cross term of SQFT3 and NOSTORIES
X_COORD	X coordinate of the observation in square feet
Y_COORD	Y coordinate of the observation in square feet

Panel B: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
PRICE	14.65	9.55	1.01	152.07
ROOMS	6.50	1.38	3.00	16.00
ROOMS2	44.18	20.05	9.00	256.00
HOUSEAGE	39.26	24.37	3.00	162.00
HOUSEAGE2	2134.68	2552.31	9.00	26244.00
BATHTOT	1.85	0.69	1.00	5.50
LOTS	12.63	23.59	0.87	306.66
LOTS2	715.30	5582.33	0.76	94042.58
LOTS3	133044.31	1434433.35	0.66	28839438.10
FIREPLACE	0.57	0.50	0.00	1.00
NOSTORIES	14.64	4.93	10.00	25.00
CONDITION	2.27	0.58	1.00	4.00
CROSS1	26.30	16.08	6.72	128.54
CROSS2	53.71	61.53	4.52	826.13
CROSS3	127.58	293.81	3.03	5309.52
X_COORD	1827731.25	27982.74	1760808.20	1893377.37
Y_COORD	731159.24	27193.99	661905.00	779187.00

Table 8. Descriptive Statistics on the Community Variables in the Hedonic Model

Panel A: Definition of Variables

Variable	Definition
MEDHHINC	Median household income in 10,000 dollars
PERCOL_25	Percentage of population (age>=25) with at least some college
PERWHITE	Percentage of Caucasians in census track
MEDAGE	Median age of households in census track
PUPILEXP	Education expenditure per pupil in 10,000 dollars in school district
CRIME	Crime index divided by population in the police jurisdiction
TAX	Property tax rate, in percentage, in the township
CBD	Distance of property to the central business district (CBD) in miles
AVGSC9	Average of math, reading, and writing of the Ohio 9th-grade proficiency test in the school district

Panel B: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
MEDHHINC	5.39	1.88	1.24	11.60
PERCOL_25	0.64	0.20	0.13	0.97
PERWHITE	0.83	0.18	0.03	0.98
MEDAGE	35.59	5.63	22.30	60.30
PUPILEXP	0.61	0.08	0.49	0.85
CRIME	6.18	3.80	1.34	11.81
TAX	1.75	0.26	1.48	2.44
CBD	7.39	3.02	0.89	15.27
AVGSC9	78.00	12.89	64.10	98.10

The crime index is provided by the Bureau of Justice Statistics (BJS). It is the total number of murders, non-negligent manslaughters, forcible rapes, robberies, aggravated assaults, burglaries, larceny-thefts, motor vehicle thefts, and arsons. The property tax rate is provided by the Franklin County Treasurer. The expenditure/pupil and efficiency score are provide by the Ohio Education Department. MEDHHINC, PERCOL_25, PERWHITE, and MEDAGE are from census data. AVGSC9 is provided by the Ohio Department of Education.

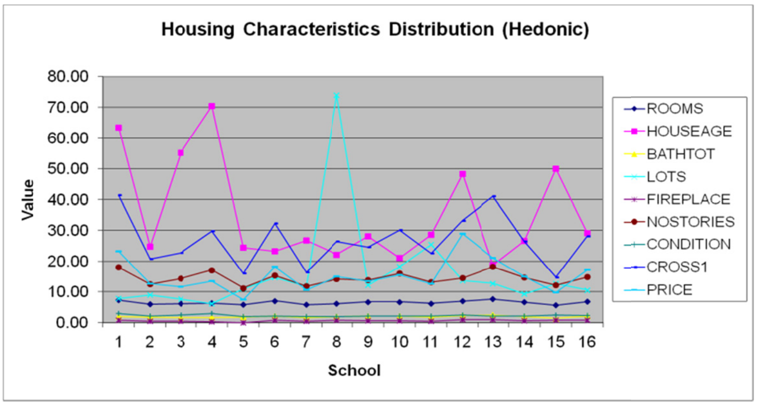

Figure 7. Housing Characteristics Distribution (Hedonic)

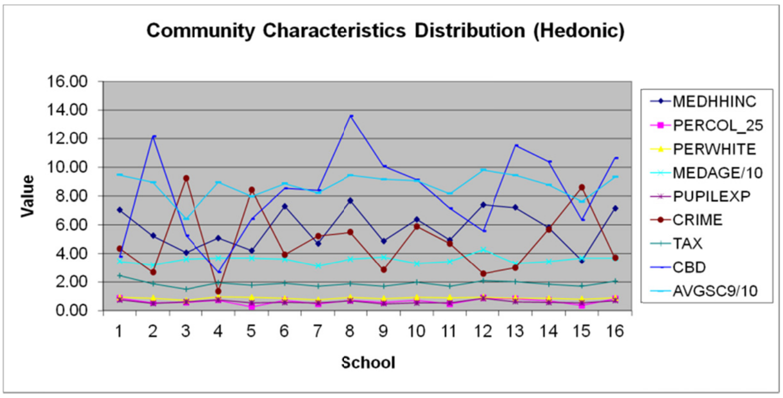

Figure 8. Community Characteristics Distribution (Hedonic)

The distribution figures above suggest that Grandview School District (school 4) is worth notice. It is the oldest and the closest district to the CBD. Grandview also has the lowest crime rate, the smallest average lot size, and the greatest concentration of white. In fact, Grandview is among the most desirable school districts in Franklin County given its convenient location and good environment. It is located right in the heart of Columbus city, which results in the limited lot size of most houses there. Plain Local school district (school 8) is most far away from the city center. Because there is plenty of land, houses there are usually huge. People who bought those big houses tend to be relatively rich, which results in the highest average household income level in Plain Local, Upper Arlington school district (school 12) is always considered the best option for families with school age children. It has the highest average proficiency test score and the highest per pupil education spending. In addition to these, the percentage of residents who have attained at least some college education is higher in Upper Arlington than in any other school district within Franklin County.

Appendix A2. The Hedonic Price Indices

Equation (13) can be written as the following equation according to the selected variables in Table (7) and Table (8), (Note 48)

$$\begin{aligned} Price = \alpha_0 + \textstyle\sum_{j=1}^{J-1} \alpha_{0j}\, school_j + \alpha_1 MEDAGE + \alpha_2 PROWHITE \quad (18) \\ +\alpha_3 PROCOL25 + \alpha_4 MEDHHINC + \alpha_5 EXPUPIL \\ +\alpha_6 CRIME + \alpha_7 TAX + \alpha_8 CBD + \alpha_9 AVGSC9 \\ +\beta_1 ROOMS + \beta_2 ROOMS2 + \beta_3 HOUSEAGE \\ +\beta_4 HOUSEAGE2 + \beta_5 BATHTOT + \beta_6 ACRE \\ +\beta_7 ACRE2 + \beta_8 ACRE3 + \beta_9 FIREPLACE + \beta_{10} NOSTORIES \\ +\beta_{11} CONDITION + \beta_{12} SQFT * NOSTORIES \\ +\beta_{13} SQFT2 * NOSTORIES + \beta_{14} SQFT3 * NOSTORIES + v_i \end{aligned}$$

Although results of previous studies suggest that the marginal prices of housing structural attributes may vary across geographic locations, the parameters of housing characteristics in Equation (18), which are represented by *βs*, are restricted to be invariant across school districts. Given the small size of Franklin County, a high variation on those coefficients is not very likely. Moreover, estimating *βs* for each school district respectively will incur a substantial loss on the reliability of the estimation results concerning our sample size.

Specification I in Table (9) reports the OLS coefficients estimates along with their standard errors. Given the potential spatial correlations among the error terms, Table (9) also provides the maximum likelihood estimates of the spherical semivariogram function in the second column.

Before estimating the spatial model, the correlation pattern of error terms is tested by fitting the residuals from an OLS regression into a spherical semivariogram function. This process is called the method of moment estimate of the semivariogram (Matheron 1963). The results confirm the existence of auto-correlated disturbances, which recommends the use of the ML method to control for the spatial correlations of the residuals and improve the precision of the predictions. The values of the estimated parameters of the spherical semivariogram function (c_0 and a_0) using the method of moment estimation are very close to the values of the estimates using the ML method. This fact validates the robustness of the ML estimates. The chi-square value of 18.58 in Table (9) indicates that the geostatistical model is superior to the OLS model. (Note 49) Because the indices created from OLS are so close to the indices created from ML, we do not expect that the results of the discrete choice model will be will significantly different using OLS indices or ML indices. Only the ML price indices (Note 50) are reported in Table (10).

Table 9. Coefficient Estimates of the Hedonic House Price Model

Specification	I		II	
Variable	OLS		ML	
INTERCEPT	-20.40***	(5.63)	-21.98***	(5.81)
SCHOOL1	3.19*	(1.83)	3.20	(2.09)
SCHOOL12	6.21***	(1.66)	6.49***	(1.80)
ROOMS	3.85***	(0.92)	4.21***	(0.89)
ROOMS2	-0.31***	(0.06)	-0.33***	(0.06)
HOUSEAGE	-0.05	(0.03)	-0.06	(0.04)
HOUSEAGE2	4.091E-04	(3.03E-04)	5.29E-04*	(3.04E-04)
BATHTOT	1.81***	(0.48)	1.87***	(0.47)
LOTS	0.10**	(0.05)	0.11**	(0.05)
LOTS2	-1.39E-03***	(4.97E-04)	-1.46E-03***	(5.11E-04)
LOTS3	3.59E-06***	(1.35E-06)	3.66E-6***	(1.40E-6)
FIREPLACE	0.80	(0.49)	0.79	(0.49)
NOSTORIES	0.64***	(0.23)	0.65***	(0.23)
CONDITION	1.57***	(0.42)	1.28***	(0.41)
CROSS1	-1.08***	(0.24)	-1.04***	(0.24)
CROSS2	0.53***	(0.08)	0.50***	(0.08)
CROSS3	-0.05***	(0.01)	-0.04***	(0.01)
MEDHHINC	0.47**	(0.22)	0.53**	(0.24)
PERCOL_25	2.41	(1.87)	2.68	(2.02)
PERWHITE	2.87**	(1.44)	2.50	(1.54)
MEDAGE	0.13**	(0.05)	0.14***	(0.05)
PUPILEXP	4.76	(5.05)	4.50	(5.49)
CRIME	0.05	(0.08)	0.05	(0.08)
TAX	-2.72	(2.14)	-2.58	(2.32)
CBD	0.12	(0.12)	0.13	(0.13)
AVGSC9	0.04	(0.04)	0.04	(0.05)
R^2		0.64		-
Adjusted R^2		0.63		-
a_0 (in the spherical semivariogram)		-	1146.87	
c_0 (in the spherical semivariogram)		-	39.49	
LikelihoodRatio(χ^2)			18.58	
Observations	1467		1467	

Numbers in parentheses correspond to standard errors.

* indicates significance at 10% level

** indicates significance at 5% level

*** indicates significance at 1% level or less

Table 10. Price Indices from the ML Estimates

Index Number	School District	GEOPIND
1	Bexley city school district	100
2	Canal Winchester local school district	75.40
3	Columbus city school district	73.90
4	Grandview heights school district	81.36
5	Hamilton local school district	70.35
6	Gahanna Jefferson city school district	86.24
7	Groveport Madison local school district	70.85
8	Plain local school district	97.64
9	Reynoldsburg city school district	80.13
10	Hilliard city school district	81.96
11	Southwestern city school district	74.15
12	Upper Arlington city school district	136.87
13	Dublin city school district	88.71
14	Westerville city school district	83.29
15	Whitehall city school district	68.76
16	Worthington city school district	92.54