Preference Tradeoffs Across Spatial Scales: Developing a Micro Level Sorting Model

Mitchell R. Livy
The Ohio State University

H. Allen Klaiber
The Ohio State University

Preference Tradeoffs Across Spatial Scales: Developing a Micro Level Sorting Model*

PRELIMINARY DRAFT: PLEASE DO NOT CITE

Abstract

This paper investigates the role of spatial scale in residential location choice. While the current residential sorting literature has largely focused on a single spatial unit, we expect that homeowners face different tradeoffs across the spatial spectrum, and that these tradeoffs interact across space in different ways to shape observed outcomes. To investigate these phenomena, we implement a nested logit discrete choice model of residential household sorting. With this model, we examine residential location choice at the school attendance boundary and residential neighborhood levels, and we find that the influence of environmental amenities on the location choice of households is complex, often having different implications depending on spatial scale.

Keywords: Location choice; nested logit; environmental amenities

* This research was partially supported by the National Science Foundation award CBET-1160961.
Preference Tradeoffs Across Spatial Scales: Developing a Micro Level Sorting Model

1. Introduction

The location choice of households reveals tradeoffs individuals make to obtain public goods, including environmental amenities, across a range of spatial scales. At a regional level, these tradeoffs take the form of differences in employment opportunities, environmental quality, and overall provision of public goods. In contrast, at a metropolitan level, households have already chosen a labor market and important differences in school quality, public safety, and recreational opportunities become apparent. Even more locally, tradeoffs in housing attributes, proximity to recreation, and surrounding land use are shown to play an important role in the decision of where to locate. Despite handling different spatial location decisions separately in the literature, there is growing recognition that tradeoffs from one spatial scale may play an important role in other spatial scales of decision making (Hamilton and Phaneuf, 2012). This paper adds to this emerging literature by estimating multiple spatial scales of sorting in a unified economic framework to examine these spatial linkages.

The model we develop in the current paper extends the analysis in Hamilton and Phaneuf (2012) by estimating a full micro level component of location choice at a neighborhood level. We deviate from their estimation approach by focusing on a single metropolitan area, capturing key elements of local public policy such as school quality, while exploiting rich micro level data to characterize location choice tradeoffs at the neighborhood and school attendance boundary scales in the Baltimore, MD region. Applying a nested logit model of household location choice that captures location specific unobservables across spatial scales allows us to measure the impact of land composition and environmental amenities on residential sorting as
they vary across space, while relaxing the independence of irrelevant alternatives (IIA) assumption present in virtually all of the micro level residential sorting literature to date.

Residential sorting models have become popular as an alternative to reduced form estimation due to their flexibility in policy simulation following non-marginal changes, the ability to capture endogenous payoffs and attributes, and the option to include initial conditions and stickiness in the sorting process. These models have been applied to environmental valuation scenarios, where household sorting potentially influences the quality and quantity of the underlying environmental good. We hypothesize that the IIA property of the existing micro level models is likely to result in predicting greater responses to environmental quality changes than one would expect if households are reluctant to substitute across schools or other public goods at a broader spatial level of sorting.

In this paper we develop and implement a nested logit sorting framework that allows us to explicitly model preferences for location specific amenities at multiple spatial scales. By linking these levels in one model, we are able to consider spatial specific and universal amenity measures. To implement our empirical model, we investigate amenities at the neighborhood and elementary school attendance boundary spatial scales in Baltimore County, MD. The amenities of interest include public parks, school quality, agricultural preserves, green infrastructure, and land use composition, among others, and were selected because they have been well-studied in the literature. For example, Irwin (2002) determines that land use values depend on density and preservation status, and Geoghegan et al. (2003) find that the value of forest and agricultural land can change based on the spatial scale measured. In our initial results we find that correlations in utility between the spatial scales exist, and that preferences for environmental amenities and land use compositions are variable across space. The paper is organized as follows: Section 2
provides a review of residential location choice and nested logit literature; Section 3 discusses our theoretical nested logit sorting model and the empirical implementation of this model; Section 4 describes our data on school quality, housing attributes, environmental amenities, and land use; Section 5 presents our results; and Section 6 concludes.

2. Spatial scale and residential location choice

The empirical literature on household location choice has largely chosen a single spatial scale for analysis. The pure characteristics model developed by Epple and Sieg (1999) is typically applied to household location choice at a metropolitan level using fairly aggregate “neighborhoods” at the scale of school districts. The class of sorting models characterized by their inclusion of a random disturbance term leading to random utility models (RUM) of location choice are largely divided between very micro level analyses of individual house or neighborhood choice (Bayer et al, 2004; Klaiber and Phaneuf, 2010) and more macro-oriented choice sets defined as regions or counties (Bayer et al, 2009). As a result of these differences in spatial scale, the amenities and tradeoffs captured by each of the above approaches are largely dissimilar with little overlap in the suite of covariates included in the analyses.

Regardless of differences in spatial scale, virtually all sorting models exploit the structural modeling approach to extend their analysis in ways that traditional reduced form approaches, such as the first stage hedonic, have difficulty capturing. For example, the structure of the sorting model permits the inclusion of endogenous feedbacks, which are likely to occur if local tax monies are used to maintain public goods and are themselves a function of housing values. The ability to capture this endogenous behavior is at the center of general equilibrium in the sorting model, as characteristics of each choice are often determined by the agents who make
that choice (e.g. Bayer and Timmins, 2005; Bayer and Timmins, 2007). These feedbacks are captured as households sort and determine new prices, which may feed into an auxiliary function of public goods funding that alters the quality of those local public goods.

Despite innovations in residential sorting models, challenges remain. One of the biggest impediments to the integration of models across spatial scale is the reliance on convenient computational advantages derived from the IIA property (Klaiber and Kuminoff, 2014). When evaluating responses to non-marginal policy changes, this property is likely to constrain the sorting behavior of households. A potential solution to this problem is an extension of the basic logit RUM model to a nested logit framework. However, this extension alters many of the models’ properties that are used to facilitate estimation, including the ease of incorporating a large number of location specific unobservables that if left unaccounted would confound estimation results.

Nested logit models have been used in the discrete choice literature since Daly and Zachary (1978), McFadden (1978), and Williams (1977). These models are primarily implemented when the IIA property is expected to have a significant effect on the estimates of the logit model. For example, Hausman et al. (1995) use empirical data from Alaska to estimate recreation use losses from the Valdez oil spill. They find that the multinomial logit and the nested logit models result in significantly different estimates due to the IIA substitution restrictions in the multinomial logit model. When implementing a nested logit, it is important that the nesting structure be logically defined. While different nesting structure specifications may lead to similar welfare estimates (Hauber and Parsons 2000), the choice to use a nested logit is driven by the belief that a certain nesting structure is representative of an actual decision making process. In this case, the structure of a school attendance boundary and neighborhood is
logical for most home buyers, especially those with children. For residents without children, differences in public goods across school boundaries may factor into the home buyers’ decision through property value expectations.

Hamilton and Phaneuf (2012) is the only study to date that implements a nested logit model in a residential sorting framework. In their paper, a census tract serves as the micro level choice, which feeds into the macro level choice, a metropolitan statistical area. However, the attributes of the micro choice are not investigated. That is, the nested logit is only implemented to integrate the cumulative micro level attributes into the macro level model. Following the work in Blackorby and Russell (1997), who prove that two-stage budgeting is consistent with utility maximization, Hamilton and Phaneuf also show that two-stage budgeting holds in the nested logit framework. In this paper, we alter and extend the nested logit framework by including a fully specified micro level of sorting. Using an upper level of elementary school attendance boundary and a lower level of neighborhood, we examine how micro level policies interact and influence the sorting of households at more aggregate spatial scales.

3. A nested logit model of residential sorting

Following the previous sorting literature, our model of household location choice begins by assuming a household chooses a certain housing type such that it maximizes utility. The housing types make up the choice set, which we define as a neighborhood within a certain school attendance boundary. In this model, individual $n$ chooses neighborhood $j$ within the nest of school attendance boundary $k$. The utility for choosing a housing type $h$ is

$$U_{jk}^n = U_h^n = V(H_h, N_j, S_k, I^n, p_h, b, \xi_j, \eta_k) + \epsilon_h^n$$

where $H$ is a vector of housing type characteristics, $N$ is a vector of neighborhood attributes, and
$S$ is a vector of school level characteristics. $I$ is a vector individual attributes, $p$ is the price of the housing type, and $b$ is numeraire consumption. $\xi_j$ are unobserved (to the econometrician) neighborhood level characteristics while $\eta_k$ are unobserved (to the econometrician) school level characteristics.

To estimate the model, utility is assumed to take a linear form and is given by

$$U_{jk}^n = \alpha^n_h H_h + \alpha^n_N N_j + \alpha^n_S S_k + \alpha^n_p p_h + \xi_j + \eta_k + \epsilon_h^n$$

The above utility is calculated using a separate model for each spatial level of nested choice. In addition, each spatial level of this model is estimated as a combination of two distinct stages. In the first stage, a logit model is used to calculate the individual heterogeneity parameters and mean indirect utilities. Then, the second stage decomposes the mean indirect utilities using ordinary least squares to reveal the preference parameters common to all individuals. As with much of the existing literature (see Bayer et al, 2005), this decomposition is made difficult due to the likely endogeneity of price. That is, observed prices are likely to reflect attributes of locations that are not accounted for in the model, yet partially determine the sorting behavior of households.

### 3.1 Nested logit implementation

While the commonly used multinomial logit assumes an extreme value distribution for the unobserved portion of utility, the nested logit assumes a more general distribution to capture the correlation between alternatives. For the nested logit used in this paper, a generalized extreme value (GEV) error structure is assumed. The vector of unobserved utility is given by

$$\exp \left( -\sum_{k=1}^K \left( \sum_{j \in B_k} \frac{\epsilon_{nj}}{\lambda_k} \right)^{\lambda_k} \right)$$
In equation (3), \( \lambda_k \) is the coefficient that indicates the independence between nests, and 
\((1-\lambda_k)\) approximates the degree of correlation. When \( \lambda_k = 1 \) for each of the nests, the nested logit collapses to a standard multinomial logit. The probability of choosing a certain choice from the choice set is given as

\[
P_{ni} = e^{\lambda_k \sum_{j \in B_k} e^{\lambda_k} \frac{V_{nj}}{V_{ni}}} / \sum_{l=1}^{K} (\sum_{j \in B_k} e^{\lambda_l})
\]

In equation (4), IIA is only present when choices are within the same nest. As before, when \( \lambda_k = 1 \) there is no correlation between the unobserved within-nest utility, and the model reduces to a multinomial logit.

Equation (4) can also be calculated as the product of two logit models, an approach we use in our empirical analysis; the nested logit equation is calculated as the product of a logit model for the micro level and a logit model for the macro level, with an inclusive value representing the expected utility for choosing an alternative within a nest linking the two models. For this, utility is divided into the variables, \( W \), which describe the nests, and \( Y \), which describe the alternatives within the nests. It follows that individual utility and individual choice probabilities are represented as

\[
U_{jk}^n = W_{jk}^n + Y_{jk}^n + \varepsilon_{jk}^n
\]

\[
P_{nB_k} = \frac{e^{\varepsilon_{nk}^n+\lambda_k^l V_{nk}^n}}{\sum_{l=1}^{K} e^{\varepsilon_{nk}^n+\lambda_k^l V_{nk}^n}}, \text{ where } IV_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}^n/\lambda_k}
\]

\[
P_{n|B_k} = \frac{e^{\lambda_k^l V_{ni}^n}}{\sum_{j \in B_k} e^{Y_{nj}^n/\lambda_k}}
\]

\[
P_{ni} = P_{n|B_k} P_{nB_k}
\]
3.2 Empirical estimation

Through separating the utility into nest specific and alternative specific portions, as shown in equations (5)-(8), two steps emerge that can be estimated sequentially. For our implementation, this sequential estimation occurs spatially such that the school level parameters are contained in $W$, and the neighborhood land use and environmental attributes are contained in $Y$. We are able to estimate each spatial level separately, with only the inclusive values connecting the equations, because of the absence of cross-nest parameters and restrictions in our model. Specifically, since we do not estimate any parameters in the micro level of the model that are present in multiple nests, we are able to fully separate the estimation of each nest. This restriction on the micro level parameters constrains our ability to estimate individual heterogeneity parameters at the micro level. However, estimating the nested logit as two logit models decreases the computational burden, allowing us to use more data, and aids in the interpretation of the estimates.

We estimate the school attendance boundary and neighborhood levels separate, each in two distinct stages. In the first stage of each level, only the mean indirect utilities and individual interaction parameters, where applicable, are recovered. Equations (9) and (10) show this stage using two logit models, at the school level and neighborhood level respectively.

\[
P_{nB_k} = \frac{\exp(\alpha_n^S S_l + \theta_l)}{\sum_{k=1}^K \exp(\alpha_n^S s_k + \theta_k)}
\]

\[
P_{ni|B_k} = \frac{\exp(\delta_{ik})}{\sum_{j \in B_k} \exp(\delta_{jk})}
\]

The above characterization of the nested logit as two individual logit models is used to estimate the mean indirect utilities needed for the second stage of the empirical model. Equation (10) is estimated using the effects coding strategy outlined in Hamilton and Phaneuf (2012) where mean indirect utilities for each nest sum to zero. Therefore, the mean indirect utilities are interpreted as deviations from the mean nest level of utility. Equation (11) presents this
calculation, with \( s_{ik} \) representing the share of individuals in school attendance boundary \( k \) who chose to live in neighborhood \( i \).

(11) \[
\delta_{jk} = \frac{1}{n} \sum_{j \in B_k} \ln \left( \frac{s_{ik}}{S_{jk}} \right)
\]

In the second stage of each estimation step, the recovered mean indirect utilities are decomposed to recover the mean preference parameter for each amenity, while also allowing for instrumentation for the endogenous price variable according to

(11) \[
\theta_k = \alpha_0S + \alpha_S^0S_k + \lambda_{IV} IV_n + \eta_k
\]

(12) \[
\delta_{jk} = \alpha_h^0H + \alpha_N^0N_j + \alpha_p^0p + \psi_k + \xi_j
\]

The absence of an intercept in equation (12) is due to the inclusion of fixed effects for all nests to normalize the mean indirect utility calculations from equation (10). In equation (12) \( \psi_k \) is a fixed effect for each nest, allowing us to compare estimates across nests, and the inclusive value, \( IV_n \), is given as

(13) \[
IV_n = \ln \sum_{j \in B_k} \exp(\delta_{jk})
\]

Due to the effects coding strategy implemented here, the inclusive value term is increasing in the variability within the choice set and the number of elements in the choice set. Therefore, expected utility for choosing a nest increases with the number of neighborhoods within a nest. In this paper only one correlation coefficient is estimated. In addition to improving the intuition of the model, the decomposition of the nested logit into two logit models allows the use of contraction mapping to decrease the amount of time for the numerical estimation of the model (Berry, 1994). While we lose some of the flexibility in a fully implemented nested logit model, notably the ability to have individual heterogeneity parameters at the neighborhood level, decomposing the nested logit into two models allows us to model choice sets that would be numerically infeasible with a fully implemented nested logit model.
The effects coding strategy uses few computational resources to model the smaller spatial scale, allowing us to consider larger choice sets while maintaining the IIA relaxation and interactions between the two spatial scales.

To address the endogeneity concern associated with the price variable in equation (12), we follow the market equilibrium logic of Bayer et al (2005) and exploit relationships between locations that are linked through a market equilibrium. This instrument approach relies on distant attributes which influence the price in a local area through equilibrium, but are unlikely to directly influence the location choice of individuals. In this paper, this instrument is composed of attributes of block groups surrounding each neighborhood.

4. Data

Our area of study is Baltimore County, MD, and the composition of the county is largely suburban and exurban. Most of the residents of Baltimore County are employed within the county or commute to Baltimore City, and most of the residential development has taken place close to Baltimore City. The Chesapeake Bay borders the southeastern portion of the county and is a major environmental amenity for residents and visitors. Baltimore County is an ideal area for study due to the absence of incorporated municipalities within the county. Therefore, the schools, parks, and other amenities are provisioned at the county level and share a similar set of unobserved policy idiosyncrasies.

Because our sorting model operates at both a neighborhood and elementary school attendance boundary level, it is necessary to form covariates that enter the utility specification in equation (2) at each spatial scale. To do this, we rely on a variety of data sources including property transactions data obtained from MD Property View and Corelogic, a private data
vendor; school quality data from the MD Department of Education; park data obtained from the Baltimore County Department of Recreation and Parks, and land use information from the MD Department of Natural Resources (DNR) and parcel level data provided by the county. To incorporate heterogeneity in the model, we use rich micro level data of individual incomes obtained by matching housing transactions with data from the Home Mortgage Disclosure Act (HMDA), which requires the collection of individual information from all mortgage applicants. The sub-sections that follow briefly describe the choice set creation process and each major component of data needed to estimate the model.

4.1 Choice set creation

Asymptotic requirements of the two-stage estimation strategy we employ require the number of households to grow large in relation to the number of housing alternatives (Berry, Linton and Pakes, 2004). Therefore, the choice set we create takes the form of a spatial entity that has similar housing, environmental, and other attributes, but is an aggregation of individual homes. Since we are investigating multiple spatial scales, our choice set has two components: elementary school attendance boundaries and neighborhoods. Elementary school attendance boundaries were chosen because they are numerous and there is a large gradient in location and quality across the county. School attendance boundaries used are derived from 2002 attendance boundary maps, which have changed little over time and are shown in Figure 1. The concentration of boundaries in the region nearest the center of the county is expected since the majority of the population resides in those areas due to proximity to Baltimore City. There have also been a series of development policies by Baltimore County to restrict exurban growth.

While other papers have used census tract, census block groups, and other larger
definitions of neighborhoods, we use a much smaller definition for our neighborhood choice set. Neighborhoods are defined using subdivision fields in our parcel level data. When the subdivision field is missing, a designation for the block group the parcel resides in is given as the neighborhood variable. If a neighborhood is within multiple elementary attendance boundaries, then a choice is created for that neighborhood within each attendance boundary. Figure 2 presents a map of sample neighborhoods within Baltimore County. From the map, it is apparent the choice sets are relatively small and are consistent in grouping homes in the same spatial area. This close proximity of homes within each neighborhood ensures that there are similar internal and external characteristics for the homes that makeup that neighborhood. In total, there are 6,141 neighborhoods and 99 elementary school attendance boundaries observed in the data. Table 1 presents the summary statistics for the neighborhoods.

4.2 Property transactions and individual demographics

Since the sorting model requires us to create covariates that align to our definition of the choice set, we aggregate individual housing sales data to form median, or representative, homes for each neighborhood. We use arms-length single family home sales from 2001 through 2006, and each neighborhood is given the median attributes of the homes within its boundaries. In addition to the MD Property View database, Corelogic and HMDA data are used to clean and complement our housing data. Specifically, the HMDA data are used to gain information on the residents of each neighborhood. In total, there were 44,360 individuals in the HMDA data purchasing homes during the study time period that we were able to link to individual parcel locations. These individuals had a mean income of $79,891.

The price index variable presented in Table 1 represents the mean sales price for homes
within a neighborhood net the average attributes of the homes, and is estimated by regressing the sales price of each home on the year of sale, the attributes of the house, and a fixed effect for each neighborhood. The fixed effects estimates from this model represent a location varying price index corresponding to each neighborhood that is net of the other attributes contained in the regression. The price index can be interpreted as the cost of housing for a neighborhood after controlling for the attributes of the median house within that neighborhood.

Investigating the other housing attributes in Table 1, we observe expected values for each of the measures. The average age of homes is 35 years, with a spread of 0 to 100 due to data cleaning, and there is an average of two bathrooms per home. The garage variable indicates that half of the homes within the study area have a garage.

4.3 School quality

The elementary school attendance boundary and performance data were acquired from the MD Department of Education. In this paper, school data from 2002 are used. The map of elementary school attendance boundaries presented in Figure 1 is evidence of the dispersion of elementary schools within Baltimore County. Since the county oversees all public elementary schools, the potential for unobserved shocks due to public funding changes across our sample area is minimized.

Despite the county level governance for all schools, there is still a large degree of variance in the school system. The school summary statistics in Table 1 indicate the differences in school quality and size across Baltimore County. The second grade standardized tests scores for reading have a mean score of 0.65 with a range between 0.36 and 0.91, and the scores for all three tested subjects have similar means and standard deviations. In our data, these scores are
also highly correlated within each attendance boundary, such that schools with high scores in one subject are likely to have high scores in other subjects. The number of enrolled students per elementary school ranges from 245 to 984.

4.4 Park data
The data on park amenities were obtained from the Baltimore County Department of Recreation and Parks, and are calculated at the neighborhood level. These amenities are only calculated at the neighborhood level because they are similar across school attendance boundaries; this is because the provision of parks, like the provision of schools, is proportional to population. The park dataset includes all public parks, playgrounds, and community centers in the county and is from a 2012 list of Baltimore County parks and their associated amenities. These variables are attributed to all neighborhoods within one half mile of each park, and each amenity is aggregated by count when multiples of the same amenity are present.

Table 1 reports the park summary statistics, and 65% of the neighborhoods in the county are within one half mile of a park, with an average of almost 2.5 acres of park land. Community centers are within one half mile of approximately 7% of neighborhoods. There are an average of 1.69 playgrounds per neighborhood. In related first-stage hedonic work, Livy and Klaiber (2014) disaggregate park amenities and find evidence that playgrounds in this area are likely associated with a negative value, due to their age. However, following playground replacements the authors find a rising value, but this increase attenuates rapidly over time.

4.5 Land use and preservation
Land use and preservation measures are included at both the neighborhood level and the
elementary attendance boundary level to determine preferences for surrounding land use characteristics at each spatial scale. Agricultural preserves and green infrastructure land use measures were obtained from the MD DNR and are matched to our parcel data. The agricultural preserves are land that has been preserved for agricultural use through various government programs, and the green infrastructure parcels are identified by the DNR as large hubs of forest, wetland, and other natural support system lands, and the corridors connecting these hubs. Aside from these two land use categories, the land use statistics are calculated from county provided parcel level data, and include residential, agricultural, industrial, commercial, and other land use types. School level measures are calculated as the ratio of acreage for each land use type within the attendance boundary. The neighborhood level land use variables are calculated as the ratio of acreage for each land use type within one half mile of the neighborhood centroid.

Summary statistics for the elementary attendance boundary and neighborhood level land use measures are presented in Table 1. At the school level and neighborhood level residential land use is the most common, consuming 51% and 64% of parcels on average, respectively, while country club land is the least common. The high number of residential parcels near neighborhoods is driven by the clustering of residential development. There is a similar percentage of agricultural land and commercial land at each spatial scale, suggesting that there is a near even divide across the county of these land uses. The large spread for each of the land use measures at the neighborhood and school levels is evidence of the diverse range of development across the county. Green infrastructure is more common than agricultural preserve land at both spatial scales, and agricultural preserves compose almost one third of agricultural land throughout the county.
5. Estimation results

As previously noted in the empirical implementation section, estimation occurs in two stages for each spatial scale. The logit model is estimated first, and the mean indirect utilities are then decomposed using ordinary least squares. The neighborhood level model is estimated initially so the inclusive values necessary for the elementary attendance boundary model can be calculated; then estimation at the school level is performed.

5.1 Neighborhood choice

To estimate preferences at the neighborhood level, we recover the first stage mean indirect utility parameters at the neighborhood level using the effects coding strategy implemented by Hamilton and Phaneuf (2012). Due to restrictions in our estimation procedure, no individual heterogeneity parameters are estimated at this level. The mean indirect utility results are not shown here due to the large amount of neighborhoods, but are available upon request.

Our second stage of estimation at the neighborhood level decomposes the mean indirect utility values and these results are shown in Table 2. There is no intercept in this model because of the presence of a full suite of nest fixed effects. The inclusion of these fixed effects, not presented due to space concerns, accounts for arbitrary differences in the normalized mean indirect utilities across school attendance boundaries. Price is included here as a house attribute demeaned index; that is, we have regressed the sales price on the housing attributes to obtain a price index for each of the neighborhoods. Since these prices are endogenous, we create an instrument using data on demographics and land use measures three miles from each neighborhood which follows the procedure in Klaiber and Phaneuf (2010). As expected, the coefficient on price is negative, revealing that homeowners prefer less expensive homes, holding
other attributes constant.

The land use and preservation coefficients presented in Table 2 are interpreted as preferences relative to the omitted category of industrial land use. Investigating the coefficients, residential land use is the most preferred, while condominium and apartment land use is the least preferred. The coefficients for green infrastructure and country club are positive, suggesting preferences for wildlife and land with minimal development. Despite the positive and large coefficient on agricultural land, agricultural preserves have a negative, but statistically insignificant estimate. Collectively, these results indicate preferences for low density development in the areas surrounding a neighborhood.

Examining the park attribute estimates, measured within one half mile of each neighborhood, the coefficient for the existence of a park near a neighborhood is negative; however, the coefficient on park acres is positive. Together, these results suggest preferences against parks, but for larger parks when a park is present. The coefficient for being near a playground is negative and significant, and we conclude that the negative externalities from public playgrounds outweigh their benefits to nearby households. Community centers also have a negative coefficient. The positive coefficient for park acres is evidence that residents may prefer parks for their open space and not for the playground and community center amenities they contain.

5.2 School choice

At the school level, we first estimate the logit model of heterogeneous effects and mean indirect utilities. Income is the source of heterogeneity and is interacted with second grade reading standardized test scores and green infrastructure land use within the attendance boundary.
Results are presented in Table 3, with the mean indirect utility measures suppressed due to space concerns. Interpreting the estimates reveals that individuals with higher incomes prefer to reside in school attendance boundaries with higher mean reading tests scores and higher percentages of green infrastructure. The positive coefficient on reading test scores is expected, as the existing literature suggests wealthier individuals typically prefer to locate in better performing school districts. Green infrastructure is estimated positive, possibly because wealthier individuals prefer less developed areas, and are able to afford transportation to distant commercial and industrial zones. The significance of the two estimates in this model also confirms the existence of heterogeneity in our sample.

We decompose the mean indirect utilities to determine the mean impacts of land use measures and school attributes on household utility, and these results are reported in Table 4. The significant, positive, and less than one estimate for the inclusive value term indicates that our model is consistent with utility maximization. Further, this is evidence of a positive correlation between the elementary school attendance boundary nests, confirming our use of a nested logit model. Investigating the school quality estimate, the negative sign on reading standardized test score is unintuitive. However, the interaction of reading test score and income is positive, leading to an overall positive effect at the mean income level of $79,891 for our sample.

As before, the land use measures are interpreted against the omitted category of industrial land use. At the school level, country club land is the most positive coefficient, and agricultural land is the most negative coefficient. Agricultural preserves have a positive estimate, revealing that residents prefer agricultural land that is permanently preserved at this spatial scale. The coefficient on green infrastructure is negative, and the overall effect at the mean income
level is negative. However, at higher income levels, preferences become positive for green infrastructure hubs and corridors.

5.3 Comparing the spatial scales

While the estimate on the inclusive value term substantiates our use of the nested logit model through correlations between across and within-nest utility, the differing coefficients between the neighborhood and school level models also confirm our use of this model. Overall, these results expose preferences for environmental amenities and low development near neighborhoods, and for commercial and other developed lands in the larger surrounding areas.

Contrasting the land use estimates between the elementary school attendance boundary level and neighborhood level models, there are many changes in relative preferences. For example, commercial land use is preferred to agricultural and green infrastructure land use at the school level but not at the neighborhood level. This result is expected because residents may prefer commercial conveniences within the larger spatial area near their residence, but not within close proximity to their home due to negative externalities. Similarly, preferences at the elementary attendance boundary level are generally higher for industrial land use than they are at the neighborhood level, suggesting that residents prefer to be near developed areas without having them in immediate proximity to their home.

6. Conclusion

Estimating a nested logit micro level sorting model allows us to more accurately describe the residential sorting framework. In this paper, we have outlined and implemented a procedure in which multiple spatial scales can be considered in one model. In addition to estimating different
spatial levels together, we are able to relax IIA and consider choice sets that would otherwise be infeasible to estimate in a logit framework.

Through implementing a nested logit sorting model using elementary school attendance boundary and neighborhood levels, we are able to realistically measure the location decision of homeowners. Our results have demonstrated differing preferences for land use attributes across space, while also considering spatially unique covariates at each level. Specifically, we find that residents’ preferences toward developed land vary greatly between the two spatial scales. This paper is the first step in creating more accurate sorting models for determining household preferences and in measuring the effects of proposed policy changes.

While estimation results presented here are preliminary and incomplete, we feel that these results demonstrate promise for this approach to modeling the household location choice process. Work is ongoing to expand the suite of covariates included in the model and to demonstrate the tradeoffs between a traditional one-spatial scale approach and the nested approach used here in terms of both marginal and non-marginal welfare measures. We also intend to further our analysis of the improvements made by the inclusion of a nested logit through determining the differences in outcomes from proposed policy scenarios.
References


Daly, Andrew and Zachary, Stanley. "Improved multiple choice models." *Determinants of Travel Choice.* 335 (1978): 357.


Figures

Figure 1: Baltimore County elementary attendance boundaries (2002)

Figure 2: Baltimore County sample neighborhoods
### Table 1: Summary statistics for choice set*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>price index (dollars)</td>
<td>240325.3</td>
<td>89995.94</td>
<td>39387.8</td>
<td>834901</td>
</tr>
<tr>
<td>age</td>
<td>35.48323</td>
<td>20.33732</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>baths</td>
<td>2.167074</td>
<td>0.8544726</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>square feet</td>
<td>1852.63</td>
<td>854.5957</td>
<td>561</td>
<td>7830</td>
</tr>
<tr>
<td>stories</td>
<td>1.709331</td>
<td>0.4070891</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>garage</td>
<td>0.4576616</td>
<td>0.4852871</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>lot size (acres)</td>
<td>0.5987602</td>
<td>1.108588</td>
<td>0.02</td>
<td>16.44</td>
</tr>
<tr>
<td>reading (2nd gr.)</td>
<td>0.6453688</td>
<td>0.1263989</td>
<td>0.36</td>
<td>0.91</td>
</tr>
<tr>
<td>lang. (2nd gr.)</td>
<td>0.6650187</td>
<td>0.1460636</td>
<td>0.33</td>
<td>0.97</td>
</tr>
<tr>
<td>math (2nd gr.)</td>
<td>0.5864696</td>
<td>0.160041</td>
<td>0.18</td>
<td>0.9</td>
</tr>
<tr>
<td>school enroll. (1k)</td>
<td>495.0871</td>
<td>117.5991</td>
<td>245</td>
<td>984</td>
</tr>
<tr>
<td>agricultural land school</td>
<td>0.1035738</td>
<td>0.185829</td>
<td>0</td>
<td>0.74177</td>
</tr>
<tr>
<td>commercial land school</td>
<td>0.0885159</td>
<td>0.1289632</td>
<td>0.00019</td>
<td>0.84017</td>
</tr>
<tr>
<td>residential land school</td>
<td>0.5131976</td>
<td>0.2401944</td>
<td>0.00266</td>
<td>0.99822</td>
</tr>
<tr>
<td>condo./apartment land school</td>
<td>0.0439201</td>
<td>0.0818576</td>
<td>0</td>
<td>0.66812</td>
</tr>
<tr>
<td>country club land school</td>
<td>0.0046601</td>
<td>0.025929</td>
<td>0</td>
<td>0.24992</td>
</tr>
<tr>
<td>public/exempt land school</td>
<td>0.1981408</td>
<td>0.2004822</td>
<td>0.0006</td>
<td>0.97457</td>
</tr>
<tr>
<td>industrial land school</td>
<td>0.0479919</td>
<td>0.1065838</td>
<td>0</td>
<td>0.99526</td>
</tr>
<tr>
<td>agricultural preserves school</td>
<td>0.0380764</td>
<td>0.1048294</td>
<td>0</td>
<td>0.53648</td>
</tr>
<tr>
<td>green infrastructure school</td>
<td>0.1719719</td>
<td>0.2131689</td>
<td>0</td>
<td>0.84633</td>
</tr>
<tr>
<td>agricultural land neigh.</td>
<td>0.0653281</td>
<td>0.155252</td>
<td>0</td>
<td>0.97942</td>
</tr>
<tr>
<td>commercial land neigh.</td>
<td>0.0778442</td>
<td>0.1147788</td>
<td>0</td>
<td>0.98653</td>
</tr>
<tr>
<td>residential land neigh.</td>
<td>0.6367774</td>
<td>0.2035012</td>
<td>0.00288</td>
<td>1</td>
</tr>
<tr>
<td>condo./apartment land neigh.</td>
<td>0.0425218</td>
<td>0.0684354</td>
<td>0</td>
<td>0.68098</td>
</tr>
<tr>
<td>country club land neigh.</td>
<td>0.0036131</td>
<td>0.0362792</td>
<td>0</td>
<td>0.54083</td>
</tr>
<tr>
<td>public/exempt land neigh.</td>
<td>0.1421281</td>
<td>0.1363231</td>
<td>0</td>
<td>0.991</td>
</tr>
<tr>
<td>industrial land neigh.</td>
<td>0.0317874</td>
<td>0.0784303</td>
<td>0</td>
<td>0.75745</td>
</tr>
<tr>
<td>agricultural preserves neigh.</td>
<td>0.0154985</td>
<td>0.0822603</td>
<td>0</td>
<td>0.98408</td>
</tr>
<tr>
<td>green infrastructure neigh.</td>
<td>0.1041455</td>
<td>0.2173401</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>park</td>
<td>0.6464745</td>
<td>0.478103</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>park size (acres)</td>
<td>2.418094</td>
<td>3.542278</td>
<td>0</td>
<td>43.5</td>
</tr>
<tr>
<td>playgrounds</td>
<td>1.69321</td>
<td>1.983975</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>community centers</td>
<td>0.0734408</td>
<td>0.3030459</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

*Land use measures are presented at the ratio of acres for that land type to the total amount of acres at the respective spatial scale.*
### Table 2: Second stage results (neighborhood)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Error</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price ($100,000)</td>
<td>-5.9103</td>
<td>0.2829</td>
<td>-20.892</td>
</tr>
<tr>
<td>agricultural land</td>
<td>3.1818</td>
<td>0.8878</td>
<td>3.584</td>
</tr>
<tr>
<td>commercial land</td>
<td>0.3631</td>
<td>0.7619</td>
<td>0.4765</td>
</tr>
<tr>
<td>residential land</td>
<td>4.3057</td>
<td>0.6789</td>
<td>6.342</td>
</tr>
<tr>
<td>condo./apartment land</td>
<td>-5.9014</td>
<td>0.9788</td>
<td>-6.0293</td>
</tr>
<tr>
<td>country club land</td>
<td>2.8107</td>
<td>1.6734</td>
<td>1.6796</td>
</tr>
<tr>
<td>public/exempt land</td>
<td>2.7172</td>
<td>0.7648</td>
<td>3.5526</td>
</tr>
<tr>
<td>agricultural preserves</td>
<td>-1.3029</td>
<td>0.8816</td>
<td>-1.4778</td>
</tr>
<tr>
<td>green infrastructure</td>
<td>1.1592</td>
<td>0.3223</td>
<td>3.5962</td>
</tr>
<tr>
<td>park</td>
<td>-0.8392</td>
<td>0.156</td>
<td>-5.379</td>
</tr>
<tr>
<td>park size (acres)</td>
<td>0.0286</td>
<td>0.0242</td>
<td>1.1849</td>
</tr>
<tr>
<td>playgrounds</td>
<td>-0.2429</td>
<td>0.0348</td>
<td>-6.9837</td>
</tr>
<tr>
<td>community centers</td>
<td>-0.4357</td>
<td>0.1345</td>
<td>-3.238</td>
</tr>
<tr>
<td>Nest FE</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: First stage heterogeneity parameters (school attendance boundary)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Error</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>reading (2nd gr.) x income</td>
<td>4.0423</td>
<td>0.0839</td>
<td>48.1881</td>
</tr>
<tr>
<td>green infrastructure x income</td>
<td>1.0156</td>
<td>0.0412</td>
<td>24.6704</td>
</tr>
</tbody>
</table>

### Table 4: Second stage results (school attendance boundary)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Error</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-1.6895</td>
<td>0.0328</td>
<td>-51.536</td>
</tr>
<tr>
<td>read. (2nd gr.)</td>
<td>-3.2126</td>
<td>0.0307</td>
<td>-104.57</td>
</tr>
<tr>
<td>agricultural land</td>
<td>-1.2019</td>
<td>0.0538</td>
<td>-22.321</td>
</tr>
<tr>
<td>commercial land</td>
<td>0.139</td>
<td>0.032</td>
<td>4.3432</td>
</tr>
<tr>
<td>residential land</td>
<td>0.1579</td>
<td>0.017</td>
<td>9.269</td>
</tr>
<tr>
<td>condo./apartment land</td>
<td>-0.3369</td>
<td>0.0286</td>
<td>-11.765</td>
</tr>
<tr>
<td>country club land</td>
<td>1.4772</td>
<td>0.1263</td>
<td>11.695</td>
</tr>
<tr>
<td>public/exempt land</td>
<td>0.4073</td>
<td>0.0191</td>
<td>21.3454</td>
</tr>
<tr>
<td>agricultural preserves</td>
<td>0.3983</td>
<td>0.0405</td>
<td>9.842</td>
</tr>
<tr>
<td>green infrastructure</td>
<td>-1.0559</td>
<td>0.0165</td>
<td>-64.071</td>
</tr>
<tr>
<td>inclusive value</td>
<td>0.6851</td>
<td>0.0007</td>
<td>940.406</td>
</tr>
</tbody>
</table>