Modifiable Areal Units: A *Problem* or a Matter of *Perception* in the Context of Residential Location Choice Modeling?

Jessica Y. Guo and Chandra R. Bhat
The University of Texas at Austin, Department of Civil Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744,
Email: jessica.guo@mail.utexas.edu, bhat@mail.utexas.edu

ABSTRACT

The sensitivity of spatial analytic results to the way in which the areal units are defined is known as the modifiable areal unit problem (MAUP). Although to date a general solution to the problem does not yet exist, it has been suggested in the literature that the effects of the problem may be controllable within specific application contexts. The current study pursues this line of inquiry and addresses the MAUP in the context of residential location choice modeling.

Previous residential location choice analysis typically involves the representation of alternative locations by areal units and the measurement of residential neighborhood characteristics based on these areal units. This study demonstrates the vulnerability of such an approach to effects of the MAUP. We contend that the fundamental issue is the inconsistency between the analyst's definition of areal units and the decision maker's perception of residential neighborhoods. An alternative approach of using a multi-scale modeling structure is proposed to mimic the notion of a neighborhood being a hierarchy of residential groupings. The proposed approach allows the spatial extent of choice factors to be determined endogenously. We show that the multi-scale approach produces richer and more interpretable results than its single scale counterpart.

1. INTRODUCTION

Generalization is an innate skill that we use all the time. We generalize about people, things and events. We generalize by filtering everything that we absorb with our five senses through our values, beliefs, attitudes and experiences. During the process, trivial details are deleted and attention is devoted to important features. Generalization is also an important consideration in the scientific analysis of data. As analysts, we collapse and aggregate observations in order to make the data more workable to the problem at hand, to gain understanding of the phenomenon in question, and to uncover patterns confounded by the noise typically found in observations. Filtering, in this case, is performed through what statisticians refer to as the data's *support* (*1*), that is, the units within which the aggregate measures of observations are computed. A data's support is characterized by its geometrical shape, size, and orientation. A change in any of these characteristics defines a new variable (*2*). For instance, when aggregating traffic counts observed on a link, we can use hours of a day, or days of a week, as the temporal units (supports) to arrive at hourly traffic volume, or daily traffic volume (variables). Different link volume variables derived from different choice of units will result in different interpretations about the observed traffic counts. This dependency of data interpretations on support is referred as the *support effect*.

The problems generated by support effects are ubiquitous. In studying spatial phenomenon, we often aggregate spatially scattered observations into predefined areal units, or spatial support. During the aggregation process, information is lost about the uniqueness of, and the variations between, the observations that fall within the same areal unit. As a study region can be segmented in different ways (in terms of shape, size, and orientation) to yield different spatial supports, the magnitude of information loss may vary. Consequently, the result of further analysis of the data will vary. This spatial version of the support effect has been known to spatial analysts as the modifiable areal unit problem (MAUP) (*3*). To date, a general workable solution to the problem does not yet exist and the MAUP remains one of the most stubborn problems in geography and spatial science (*4*). However, not all attempts in resolving the problem have been futile. As Miller (*5*) indicates in a review of recent MAUP-related studies, "it is clear that antecedent factors can be controlled and [the problem's] effects predicted, particularly within specific application contexts" (p.375). The current study pursues this line of inquiry and aims to address the issue of spatial support in the context of discrete choice modeling.

Discrete choice models have found considerable application in travel analysis. They are formulated to help analysts understand the behavioral process that leads to a decision maker's choice among a set of alternatives. The underlying behavioral assumption is that a decision maker evaluates every alternative before selecting the one yielding the maximum benefit. This concept is operationalized through a utility function consisting of factors that collectively determine the benefit of each alternative. The probability of an alternative being chosen is then derived from the corresponding utility value. The MAUP arises when there are spatial factors influencing the choice.

The presence of spatial determinants is common to many choice models, especially for activity-based travel modeling. This is because, as people travel to take part in activities distributed over space, their decisions regarding mobility, vehicle ownership, travel mode and activity participation location are influenced and/or constrained by the surrounding spatial structure and the urban environment. In cases where the decision is about locations in space, the

variability in ways of representing the alternative locations further compounds the effects of MAUP.

Despite the long recognition of the MAUP, past studies often involve aggregating spatial attributes over predefined areal units, such as census tracts or transport analysis zones (TAZ), prior to incorporating spatial measures into the utility function. Since the effects of spatial attributes on choice behavior are interpreted through the arbitrarily chosen spatial support, the logical questions to ask are: How accurate and interpretable are the parameter estimates obtained from such spatial measures? Can we rely on the modeling results to design effective spatial policies relating to the topic at hand? To what extent does, or does not, the MAUP affect the conclusions drawn from such studies?

The objective of this paper is to seek answers to the aforementioned questions through models developed for residential location choice. Using the multinomial logit structure, we show contradicting modeling results that suggest the vulnerability of discrete choice models to effects of the MAUP. We contend that the fundamental reason for the manifestation of the MAUP is the modeler's inability to relate the configuration of the spatial support to decision makers' *perception* of space. Had the characteristics of the space been measured in the same way as a decision maker filters spatial information, there would be no concern of the MAUP. In order to bring the spatial representation one step closer to the decision maker's perception of reality, we propose a multi-scale modeling structure which, in the context of residential location choice, mimics the notion of a neighborhood being a hierarchy of ecological groupings.

The remainder of this paper begins with a brief review of studies relating to the MAUP and of mechanisms previously proposed to resolve the problem. In Section 3, we discuss the behavioral foundations and measurement issues relating to residential location choice analysis. Section 4 surveys previous empirical studies on the subject and identifies their shortcomings. Section 5 describes our multi-scale approach to discrete location choice modeling. In Section 6, we present empirical results obtained from residential location choice models developed using different spatial supports for the San Francisco Bay area. The paper then concludes with a synthesis of the contribution of the current study.

2. THE MAUP
As mentioned earlier, the MAUP is essentially the spatial instance of the support effect. This effect has been found in a variety of spatial analysis and modeling studies, including univariate statistical analyses (*6*), bivariate regression (*6*), multivariate statistical analysis (*7*), spatial interaction models (*8,9*), and location-allocation modeling (*10,11*). Readers are referred to Openshaw (*3*) and Arbia (*6*) for more detailed reviews on the topic. The findings from the above studies raise our skepticism on the reliability of the outcome of any spatial study relying on the use of pre-defined areal units in the analysis. Though the degree of the impact has been found to vary from study to study, this unpredictability further complicates the problem and stresses the need for more insight, and solutions, to the problem.

While several research efforts have focused mostly on revealing the MAUP, the search for effective solutions has not been widely attempted, at least not with satisfactory results. According to Wong (*12*), past attempts may be categorized into three approaches: data manipulation, technique-oriented, and error modeling. The data manipulation approach entails constructing optimal areal units with respect to predefined objective functions (see *8*, *13*, *14*, *15*). The technique-oriented approach, on the other hand, involves replacing the classical statistical techniques with frame independent analyses (*16,17*). Another group of researchers (*18*) propose

the error modeling approach of explicitly documenting variations derived from changing scale, and incorporating these changes into modeling and analysis.

The reason that none of the efforts discussed above provide a general solution to the MAUP lies perhaps in the fact that "[t]he precise forms of generalization that prove successful with respect to any given phenomenon depend on the nature of that phenomenon" (*19*, p.5). To make generalizations work, whether over temporal, spatial or other domains, it is necessary to know something about the general nature of that phenomenon. In the temporal instances, there are often strong organizing principles associated with the observations that give rise to self-similarity, which analysts can exploit to pursue generalization. For example, it is intuitive that traffic volumes vary significantly between peak and off-peak hours. Based on this understanding, analysts produce level-of-service measures for peak and off-peak periods as opposed to some random temporal units. What often makes the spatial instances difficult is our lack of such intuition about the phenomenon at hand which, in turn, requires us to decide on the spatial support even before attempting to study the phenomenon. However, this is not always the case. As we will discuss in the next section, in residential location choice studies, knowledge about human perception may provide just what we need to devise successful forms of spatial generalization.

3. RESIDENTIAL LOCATION CHOICE
Residential location choice is a multidisciplinary topic of interest to sociologists, psychologists, urban economists, geographers and transportation planners. The substantial body of literature on the subject covers both theoretical and empirical investigations from different perspectives, including the relationship between life quality and location, market differentiation in housing demand, societal value of urban amenities and neighborhood quality, and effects of spatial policies. For transportation planning, the subject is of interest because residential land use occupies about two-thirds of all urban land, and because the association between the household and the rest of the urban environment influences the activity-travel patterns of individuals of the household.

3.1. Behavioral Foundations
The underlying decision mechanism behind residential location choice is an enormously complex process. First of all, it is intricately interrelated with other choices such as housing type, tenure, work location and vehicle ownership. For the purpose of the current study, the complete interplay among all these choice dimensions is not accommodated, so that we may focus our attention on spatial analysis issues specific to residential location choice. However, it should be noted that the concept and methodology proposed in this study can be applied to a comprehensive model system of these interrelated choices.

Another source of complexity underlying residential location choice is the notion of 'location'. In the broader sense, location refers to the neighborhood within which a property situates. It also refers to the exact land parcel that a building occupies. In the narrowest sense, location refers to the street address that uniquely identifies one dwelling unit from another. Therefore, the decision on location encompasses the choice of housing unit, property and neighborhood. As a consequence of this multi-faceted meaning of location, the decision making process involves a large number of factors. Based on past theoretical and empirical studies on residential location, we identified four groups of choice factors (see Table 1): dwelling, land parcel, neighborhood and accessibility. The dwelling factors are associated with the physical

structure of the housing unit and of the dwelling to which the unit belongs (in case of a multi-unit dwelling). The land parcel group of factors relates to the non-structural property attributes. The neighborhood group of variables includes considerations relating to the neighborhood, such as crime rates and housing density. The fourth group, the accessibility measures, can be viewed as part of the neighborhood factors. But since they are specifically travel related, we categorize them separately.

Naturally, our list in the Table 1 is by no means exhaustive nor do households necessarily consider all the factors listed. Further, some of the variables may be endogenous or co-determined with the residential location choice. For example, as indicated earlier, the dwelling type variables and work location choice (that determine the commute distance and time variables included in the accessibility category) are likely to be jointly determined with residential choice. We, like several other earlier studies (but see *20*), are ignoring their interdependency. Another point to note about the residential location choice factors is that some factors are objective and easy to quantify (such as number of bedrooms), while others are subjective and personal (such as the view from a property or the cleanliness of a street). The latter group of factors is obviously difficult to introduce in any quantitative analysis.

3.2. What is a Neighborhood?
In the context of the four groups of factors identified in the previous section, there is no ambiguity in defining the dwelling and land parcel attributes. However, the last two groups (neighborhood and accessibility) are tied to a neighborhood, which is an ambiguous term. In practice, very little attention is paid to what is meant by a 'neighborhood' – it is often implicitly assumed to be the same as the spatial level at which data is available. For example, if the primary source of data is at the TAZ level, then the TAZ boundaries are assumed as being coterminous with neighborhood boundaries. The validity of such an assumption is certainly dubious; we need a more precise definition of neighborhood for improved conceptual understanding as well as practical measurement of neighborhood factors.

The question of the definition of a 'neighborhood' has been raised in several other research fields, which we now turn to for insights on the term. One such research field that is closely related to residential location choice is real estate appraisal, which is concerned with the market value of properties. Real estate analysts use neighborhoods as the primary spatial cognitive aid to identify distinct markets. The concept of neighborhood is defined as a group of complementary land uses in terms of grouping of inhabitants, buildings, or business enterprises (*21*). The definition obviously does not translate directly into a practical means of defining neighborhood boundaries. For market analysis, administrative boundaries such as census tracts or ZIP codes are typically used as neighborhood definitions. In general, a specific neighborhood may include three-block groups or may be one-fourth of a census tract (*21*). However, the use of administrative boundaries to define neighborhoods in real estate appraisal has been criticized for not matching up with the actual market isochrones of housing quality.

Another arena in which neighborhood definition plays an important role is the study of *neighborhood effects*, which refers to the neighborhood influences on the well-being and behavior of families, and often children in particular. One of the pioneers in the field, sociologist Robert Park (*22*), identified local communities as natural areas that developed as a consequence of competition between businesses for land use and population groups for affordable housing. A neighborhood, based on this view, is a subsection of a larger community – a collection of both people and institutions occupying a spatially defined area influenced by

ecological, cultural, and sometimes political forces. Later, Suttles (*23*) argued that, in addition to being the result of free-market competition, some communities' identities and boundaries are imposed by outsiders. Suttles also suggests that the neighborhood is best thought of not as a single entity, but rather as a hierarchy of progressively more inclusive residential groupings. In practice, similar to real estate analysis, virtually all studies of neighborhood effects rely on administratively defined units as proxies for neighborhoods that, as the researchers themselves realize, do not necessarily reflect real neighborhoods (*24,25*). Furthermore, this approach is flawed in that neighborhood boundaries are not necessarily fixed. Rather, an individual's perceived neighborhood almost certainly depends on where she or he lives: "an individual living on the boundary of a census tract probably has more in common with residents of the adjoining tract than with residents on the far side of his own" (*24*, p. 435). This leads to the concept of 'sliding neighborhoods'.

Given the concern that administratively defined units are not commensurate with neighborhoods as perceived by residents, Coulton et al. (*25*) conducted a study in which 140 parents of minor children from the City of Cleveland were instructed to draw a map of what they considered as the boundaries of their neighborhoods. The study finds evident discrepancies between resident-defined neighborhoods and census geography, though the average size of residents' definition is close to the size of a census tract. The study also demonstrates that individuals residing in close proximity can differ markedly from one another in how they define the physical space of their neighborhood. This variability renders the task of defining resident-perceived neighborhoods a challenging proposition.

While the question of neighborhood definition remains to be further explored, the experiences of real estate appraisers and social scientists have shed some light on human perception of neighborhood boundaries. First, administratively defined units do not represent real neighborhoods and are thus imperfect operational definitions of neighborhoods for research and policy. However, census geography in terms of tracts, block groups and blocks are reasonably consistent with the notion of neighborhoods as nested ecological structures. Second, while neighborhoods may in part be defined by objectively recognizable boundaries such as major roads and geographical barriers, the definition is often subjective and depends on the characteristics and location of the residents. Third, and as a result of the previous two points, there is a lack of consensus on the size of a perceived neighborhood, which can range from multiple census blocks to multiple tracts.

4. EMPIRICAL MODELS OF RESIDENTIAL LOCATION CHOICE

Discrete choice theory has been widely used to model residential location choice, where the consumption decision is a discrete choice between alternative houses and/or neighborhoods. A household is assumed to evaluate the various factors characterizing each location and to choose the location that yields the highest utility. As analysts, we want to uncover how decision makers assign utilities to alternative locations. This is achieved by representing the utility of alternative i for decision maker n as:

$$U_i^n = V_i^n + \varepsilon_i^n, \tag{1}$$

where

$$V_i^n = \alpha' X^n H_i + \beta' X^n Z_i. \tag{2}$$

In Equation (1), V_i^n represents the utility factors observable to the analyst, while ε_i^n represents the unobserved factors that affect utility. As stated in Equation (2), it is common practice to specify V_i^n as a linear-in-parameters function of (a) observable aspatial attributes, H_i (such as floor area size), (b) observable spatial (neighborhood) attributes, Z_i (such as population density), and (c) observable characteristics of the household, X^n (such as household income). α and β are parameter vectors to be estimated. The assumption that the error term ε_i^n is independently and identically extreme value distributed across all attributes and decision leads to the multinomial logit (MNL) model, which has a very convenient form for the choice probability:

$$P_{ni} = \frac{e^{V_i^n}}{\sum_{i'} e^{V_{i'}^n}}. \tag{3}$$

In the case of residential location choice, however, the validity of the assumption of independence across spatial attributes is questionable. For example, a decision maker's preference toward one location for the cleanliness of its nearby streets (presumably unobservable) is probably equally applicable to other locations around those same streets. The presence of *spatial dependency*, that nearby things are more similar than distant things, often subject MNL models to criticisms in their application to spatial choices. This issue is certainly important but is beyond the scope of the current study. Interested readers are referred to Bhat and Guo (*26*) for a mechanism proposed to address this problem in a residential choice context.

Despite the criticisms surrounding its underlying assumptions, the MNL model has been used for studying residential location choice (see McFadden (*27*), Hunt, McMillan and Abraham (*28*), Sermons and Koppelman (*29*), Gabriel and Rosenthal (*30*), and Sermons (*31*)). There is also a body of literature that considers residential location choice jointly with other choices in a nested logit structure, including the works of Lerman (*32*), Anas and Chu (*33*), Quigley (*34*), Waddell (*20*), Abraham and Hunt (*35*), Ben-Akiva and Bowman (*36*), and Sermons and Seredich (*37*).

The aforementioned modeling efforts contribute to the field by considering different study areas, subpopulations of residents, and decision determinants. Common to almost all the studies is the representation of alternative locations by areal units (as opposed to individual dwelling units) and the measurement of spatial attributes based on these areal units. That is, the studies implicitly assume neighborhoods to be the same as the alternative residential zones. In light of our earlier discussion on neighborhood perception, this assumption can be flawed for at least a couple of reasons. First, since individuals' perceived neighborhoods are not observed, the estimates obtained for β in Equation (2) can be interpreted only with respect to the zones chosen for analysis and not as the effects of neighborhoods. Second, by measuring spatial attributes over a single definitional configuration of zones, one assumes that every spatial factor operates at one and the same spatial scale. The hierarchical nature of neighborhoods renders this assumption rather unrealistic, as previously observed by Orford (*38*) in the context of hedonic pricing. The conflicting and insignificant parameter estimates on spatial attributes in previous residential location choice models may be attributable to the naïve measurement of spatial effects and the ill representation of neighborhoods.

5. DISAGGREGATE MULTI-SCALE LOGIT (MSL) MODEL

In view of the limitations of earlier work on residential location choice with regard to neighborhood definition, and also to address the MAUP, we propose a choice model with neighborhood attributes measured at a hierarchy of spatial units. More specifically, let each alternative housing unit i be geographically described by (i, B_i, G_i, T_i), where B_i, G_i and T_i are the census block, block group and tract that the unit belongs to, respectively. Suppose also that spatial attributes, denoted as Z_{B_i}, Z_{G_i} and Z_{T_i}, are observed at these three levels. We then replace Equation (2) by:

$$V_i^n = \alpha' X^n H_i + \beta'_B X^n Z_{B_i} + \beta'_G X^n Z_{G_i} + \beta'_T X^n Z_{T_i}, \tag{4}$$

where H_i and X^n are defined as before. The parameter vectors α, β_B, β_G and β_T are then estimated using a MNL structure, as before.

It should be noted that the above utility structure can be generalized to any number of spatial scales and is not limited to spatial units defined by census geography. In fact, the use of census boundaries may be undesirable in view of the sliding neighborhood effect. A more ideal neighborhood representation might be, for instance, circular units of 1, 2 and 3 mile radii centered around each alternative housing unit. However, the current study does not adopt this spatial support because data is not readily available at these spatial support levels.

6. CASE STUDY

The study region used to empirically test the MSL model against the conventional single-level approach is the San Francisco Bay Area. The primary data source is the 2000 Bay Area Travel Survey (BATS) that collected detailed information on individual and household sociodemographic information, employment-related characteristics, and all activity and travel episodes for a two-day period from members of 15,064 households. The dataset also collected detailed geographic information, including the point geocodes of household residence from which the census block, block group and tract of residence can be identified.

For the current analysis, we randomly selected 50% of those households who lived in a single-family detached house, yielding a sample size of 4,791. The choice alternatives for each household are defined at the most disaggregate level, *i.e.* the individual dwelling. In theory, the universal choice set in this case comprises all the single-family detached houses in the Bay area. However, data on all such housing units in the area are unavailable. Therefore, we assume that the 4,791 residences observed in the sample are a random subset of all housing opportunities in the Bay area. We also assume an identically and independently distributed structure for the error terms across the choice alternatives, so that the model can be consistently estimated with only a subset of the choice alternatives (*27*). One straightforward way of drawing a choice subset from the set of 4,791 alternatives is the random sampling technique, which is the approach we use in estimation. Our final sample for estimation consists of the chosen alternative and nine randomly selected non-chosen alternatives for each of the 4,791 households. Attributes of the choice alternatives are derived from the 2000 Census data and the TAZ profile data obtained from the Bay Area Metropolitan Transportation Commission.

A complete listing of the variables considered, and the spatial scale at which the measurements are available for each variable, is provided in Table 2. A few points to note here. First, the ordering of the different spatial levels in terms of decreasing spatial resolution, or increasing spatial aggregation, is as follows: block level, block group level, tract level, and county level. While the TAZ are more disaggregate than the county level, they cannot be directly mapped to the census geography. Second, in our final specifications, the only commute-related variable that we used corresponded to commute time, though we had both commute time and commute cost variables. This was because of the high correlation between the time and cost variables, which resulted in statistically insignificant fit improvements when both variables were considered over when commute time was the only variable considered (between the commute time and commute cost variables, the former performed much better). Third, we considered county-specific dummy variables in the estimation process to capture the average effects of any unobserved attributes at the county level. However, no such effects were found to be statistically significant. Fourth, crime rates (measured at the county level) were also considered, but excluded from the final specifications due to lack of statistically significant impacts. This does not, however, necessarily reject the hypothesis that safety is not an influencing factor on residential utility. Rather, it is likely that the counties are too broad a spatial scale to reflect safety considerations. Crime rate statistics at a finer spatial resolution would be very helpful, but were not available.

6.1 The Single-Level Models

Four MNL models, as defined by Equations (1), (2) and (3), are developed to examine spatial effects at different scales. The parameter estimates and t-statistics are presented in Table 3. In model 1, the base model, only the variables listed under the TAZ and county columns from Table 2 are considered. The intension is to isolate the influence of census attributes we have available at three different spatial scales. Model 2, 3 and 4, on the other hand, represent the best specifications after considering all the variables in model 1 plus those variables available at the census block, block group and tract levels, respectively, as indicated in Table 3. Below, we discuss the similarities and differences in the model estimates.

As shown in Table 3, the best specification found for model 1 includes the commute times and generalized accessibility measures interacted with gender, employment level (full time versus part time), income, and race. The parameter estimates relating to commute indicate that households tend to locate themselves closer to, rather than farther from, the work locations of the workers in the household. In particular, households locate themselves close to the workplace of the female workers in the household. The gender disparity, which is found to be independent of employment level, is consistent with the household responsibility hypothesis; that is, the female partner invests less time in commuting relative to her male partner because she shoulders more of the household maintenance responsibility (*29,35,39*). A similar higher responsibility hypothesis may be the underlying cause for the greater commute time effect of part-time workers relative to full-time workers. The racial disparity in commute sensitivity indicates greater spatial job-housing mismatch for non-Caucasians compared to Caucasians (*40,41*). The positive sign associated with the interaction of commute time and income may be a reflection of the willingness of higher income earners to travel further in exchange for better housing quality.

The parameter on employment accessibility suggests households' general aversion to locations too close to substantial employment centers, after the direct accessibility to work locations is accounted for. However, taken together with the parameter on the interaction term

between employment accessibility and income, the results indicate that households earning an annual income greater than $101,818 tend to locate themselves near employment centers. The three parameters associated with shopping accessibility indicate that multi-person households whose annual incomes are less than $66,304 prefer locations closer to shopping opportunities; however, the reverse is true for multi-person households earning over $66,304. The attractiveness of good shopping access is clearly higher for single-person households. Finally, in the class of accessibility-related variables, the preference for good access to recreation is confined to nuclear-family households. The differential effects of the accessibility-related variables based on income earnings and household family type need further careful exploration and research.

Most of the observations made above regarding the base model also apply to models 2 through 4. However, in addition to the variables discussed above, models 2 through 4 also include several additional variables, whose effects we discuss next.

A consistent finding among models 2 through 4 is the evidence of substantial racial segregation, which may be attributed to one or more of the following factors: (a) racial discrimination in the housing market, (b) differences between racial groups in preferences for neighborhood attributes, or (c) a preference to be with others of the same ethnic background. Disparity is also found in households' preferences for residential density. All else being equal, African-American households are found more likely than households of other racial backgrounds to locate in census blocks of high density – a result also found in Waddell (*20*). On the other hand, nuclear family households have an aversion to high density at the census block group and tract level. The impact of annual household income in all three models indicates that high income households are less likely to locate in high density areas compared to low income households.

Models 3 and 4 are further comparable in terms of additional variables that are available only at the block group level and up. The income dissimilarity measure, captured by the absolute difference between the zonal median income and household income, confirms the income segregation hypothesis found in literature. In addition to segregation by race and socioeconomic status, households tend to cluster with other households of similar structure as suggested by the parameters associated with household size dissimilarity and the household type composition. For example, couple households tend to locate themselves in neighborhoods (block groups, tracts) populated by other couple households, while the clustering of nuclear-family households is also evident at the block group level. The most prominent difference between models 3 and 4 lies in their parameter estimates for zonal median housing value. Model 4 suggests that households will, in general, prefer lower housing values, while model 3 indicates that high-income households (earning over $125,000) prefer high housing values. This difference in modeling results appears to be a result of the MAUP.

6.2 The MSL Model
The MSL model considers all the variables listed in Table 2 jointly, as discussed in Section 5. Based on Horowitz's (*42*) adjusted likelihood ratio test, the goodness-of-fit of the MSL model is superior to the previously estimated single level models.

In the overall, the MSL model is consistent with the previously estimated single level models in the signs of the parameters (see Table 4). The parameter estimates relating to commute time and accessibility are of similar magnitudes as those found in the single level

models. The more interesting results correspond to the different spatial scales at which other variables are found significant or insignificant.

Across all the single level models in section 6.1, the four racial composition variables were statistically significant, suggesting the presence of racial segregation at the respective spatial scales. When the variables measured at all three spatial levels are examined at the same time, segregation among races is most prominent at the block level, as indicated in Table 4. When 'zoomed' out to the block group level, the preference of residing near households of the same race is significant only for the Asian and Hispanic populations. The reason that the segregation of these two racial groups carries from the block to the block group level is most likely related to their small shares represented in the sample (and in the population). Hence, for example, for a neighborhood to include enough Asians to be perceived and characterized as an Asian neighborhood, the spatial extent of this neighborhood likely needs to be larger than for a Caucasian neighborhood. If the spatial extent of reference is further expanded, perceptions of racial concentration are likely to decrease because of more heterogeneity in the population at the higher spatial level. This can explain the statistical insignificance of the race variables at the tract level.

Interestingly, the consistency of the density parameters between the MSL model and the single-level models is mixed. The block level density continues to have a positive influence on residential location of African American households, but shows no differential effects on households of different income levels. The relationship between density and household type (nuclear family households) is picked up by the model at the block group level, indicating that nuclear families are averse to high-density neighborhoods. The relationship between density and income, on the other hand, is significant only at the tract level. This variation in spatial extent of density effects may, again, be the product of aggregation.

Segregation by income level and household size are both significant at, and only at, the block group level, the finest scale at which measures for these variables are available. As for segregation by household type, we again observe variation in the spatial scale of significance. Couple households' preference for locating near similar households is identified at both block group and tract levels. On the other hand, the clustering of single person households and of nuclear-family households is significant at the tract level and the block group level, respectively. This is perhaps an indication of the size of clusters by household types. It may also suggest the hypothesis that nuclear families are more conscious of their immediate environment, possibly because of the presence of children in the household. Lastly, the parameters in Table 4 that relate to zonal housing value have similar interpretations as those given by model 3.

7. CONCLUSIONS

The success of a discrete choice modeling exercise, like for other modeling efforts, relies on correct model specifications, which are tied closely to accurate representation, or measurement, of relevant variables. Of particular interest to the current study is the measurement of spatial variables that is subject to the spatial support effect, or alternatively referred as the MAUP. In the context of residential location choice analysis, the literature points to a range of intuitive hypotheses about the impact of neighborhood characteristics on residential utility. Yet, the corresponding effects are not always empirically evident.   The problem is most likely attributable to the MAUP, or more specifically, the ill representation of perceived neighborhoods and the naïve measurement of spatial effects.

The innovation introduced in this paper is that spatial attributes are considered at multiple scales, as opposed to the conventional approach based on a single scale. The MSL model structure is a more realistic representation of how a neighborhood is perceived as a hierarchy of residential groupings. Although in the current study the census geography was used as the neighborhood hierarchy, the application of the MSL model is certainly not limited to administratively defined boundaries. In fact, the MSL model should be viewed as a valuable tool for future research in evaluating alternative hierarchical representations. The MSL structure is also advantageous in that it allows the spatial extent of each variable to be determined endogenously. By interpreting the parameters with reference to the spatial scale at which they are statistically significant, we gain insights about the underlying clustering of observations. We also demonstrate empirically how households of different characteristics may have different spatial cognitions of their neighborhood boundaries. In short, by accounting for the decision maker's perception of space, the proposed approach exploits the modifiability of areal units of measurement to produce richer and more interpretable results than with the conventional approach.

8. ACKNOWLEDGEMENT

REFERENCES
1. Olea, R.A. *Geostatistical Glossary and Multilingual Dictionary*. International Association of Mathematical Geology, Oxford University Press, New York, 1990.
2. Dungan, J.L. Scaling Up and Scaling Down: The Relevance of the Support Effect on Remote Sensing of Vegetation. In *Modeling Scale in Geographical Information Science*, Eds. Tate, N.J. and Atkinson, P.M., John Wiley & Sons, Ltd., 2001, pp. 3-10.
3. Openshaw, S. Concepts and Techniques in Modern Geography: Number 38 - The Modifiable Areal Unit Problem. Geo Books, Norwick, 1984.
4. Fotheringham, A.S., Brunsdon, C., and M. Charltom. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London, 2000.
5. Miller, H.J. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation. *Geographical* Analysis, Vol. 20, 1999, pp. 153-182.
6. Arbia, G. *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Kluwer Academic, Dordrecht, 1989.
7. Fotheringham, A.S., and D.W.S. Wong. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A*, Vol. 23, 1991, pp. 1025-1044.
8. Openshaw, S. Optimal Zoning Systems for Spatial Interaction Models. *Environment and Planning A*, Vol. 9, 1977, pp. 169-184.
9. Putman, S.H., and S.H. Chung. Effects of Spatial System Design on Spatial Interaction Models, 1: The Spatial Definition Problem. *Environment and Planning A*, Vol. 21, 1989, pp. 27-46.
10. Goodchild, M.F. The Aggregation Problem in Location-Allocation. *Geographical Analysis*, Vol. 11, No. 3, 1979, pp. 240-255.
11. Fotheringham, A.S., Densham, P.J., and A. Curtis. The Zone Definition Problem in Location-Allocation Modelling. *Geographical Analysis*, Vol. 27, No. 1, 1995, pp. 60-77.
12. Wong, D. Aggregation Effects in Geo-Referenced Data. Chapter 5 of *Practical Handbook of Spatial Statistics* Eds. Arlinghaus, S. et al., CRC Press, Boca Raton, Florida, 1996.
13. Ding, C. The GIS-Based Human-Interactive TAZ Design Algorithm: Examining the Impacts of Data Aggregation on Transportation-Planning Analysis. *Environment and Planning B*, Vol. 25, 1998, pp. 601-616.
14. Guo, J.Y. A Graph Partitioning Approach to Transport Analysis Zone Design in a Geographical Information System Environment. Master Thesis, Transport Research Center, RMIT University, 2000.
15. Alvanides, A., Openshaw, S., and J. Macgill. Zone Design as a Spatial Analysis Tool. Chapter 8 in *Modeling Scale in Geographical Information Science*, Eds. N.J. Tate and P.M. Atkinson, John Wiley & Sons, Ltd, 2001.
16. Amrhein, C.G., and R. Flowerdew. The Effect of Data Aggregation on a Poisson Regression Model of Canadian Migration. *Environment and Planning A*, Vol. 24, 1992, pp. 1381-1391.
17. Tobler, W. Frame Independent Spatial Analysis. In *Accuracy of Spatial Databases* Eds. Goodchild, M. and Gopal, S., Taylor and Francis, New York, 1991, pp. 115-122.
18. Steel, D.G., Holt, D., and M. Tranmer. Modelling and Adjusting Aggregation Effects. *Proceedings of the U.S. Bureau of the Census Annual Research Conference*, U.S. Department of Commerce, Washington D.C., 1994, pp. 382-408.
19. Goodchild, M. F. Models of Scale and Scales of Modeling. In *Modeling Scale in Geographical Information Science*, Eds. Tate, N.J. and Atkinson, P.M., John Wiley & Sons, Ltd., 2001, pp. 3-10.

20. Waddell, P. Exogenous Workplace Choice in Residential Location Models: Is the Assumption Valid?. *Geographical Analysis*, Vol. 25, 1993, pp. 65-82.

21. Price, S. Surface Interpolation of Apartment Rental Data: Can Surfaces Replace Neighborhood Mapping?. *Appraisal Journal*, July 2002, pp. 260-273.

22. Park, R. Suggestions for the Investigations of Human Behavior in the Urban Environment. *American Journal of Sociology*, Vol. 20, No. 5, 1916, pp. 577-612.

23. Suttles, G.D. *The Social Construction of Communities*. University of Chicago Press, Chicago, 1972.

24. Dubin, R.A. Spatial Autocorrelation and Neighborhood Quality. *Regional Science and Urban Economics*, Vol. 22, 1992, pp. 433-452.

25. Coulton, C.J., Korbin, J., Chan, T., and M. Su. Mapping Residents' Perceptions of Neighborhood Boundaries: A Methodological Note. *American Journal of Community Psychology*, Vol. 29, No. 2, 2001, pp. 371-383.

26. Bhat, C.R., and J.Y. Guo. A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. *Transportation Research Part B*, Vol. 38, 2004, pp. 147-168.

27. McFadden, D. Modeling the Choice of Residential Location. In *Spatial Interaction Theory and Planning Models*, Eds. A. Karlqvist et al., North Holland Publishers, Amsterdam, 1978.

28. Hunt, J.D., McMillan, J.D.P., and J.E. Abraham. Stated Preference Investigation of Influences on Attractiveness of Residential Locations. In *Transportation Research Record 1466*, TRB, National Research Council, Washington, D.C., 1994, pp. 79-87.

29. Sermons, M.W., and F.S. Koppelman. Representing the Differences Between Female and Male Commute Behavior in Residential Location Choice Models. *Journal of Transport Geography*, Vol. 9, No. 2, 2001, pp. 101-110.

30. Gabriel, S.A., and S.S. Rosenthal. Household Location and Race: Estimates of a Multinomial Logit Model. *The Review of Economics and Statistics*, Vol. 17, No. 2, 1989, pp. 240-249.

31. Sermons, M.W. Influence of Race on Household Residential Utility. *Geographical Analysis*, Vol. 32, No. 3, 2000, pp. 225-246.

32. Lerman, S.R. Location, Housing, Automobile Ownership and Mode to Work: A Joint Choice Model. In *Transportation Research Record 610*, TRB, National Research Council, Washington, D.C., 1976, pp. 6-11.

33. Anas, A., and C. Chu. Discrete Choice Models and the Housing Price and Travel to Work Elasticities of Location Demand. *Journal of Urban Economics*, Vol. 15, 1984, pp.107-123.

34. Quigley, J.M. Consumer Choice of Dwelling, Neighborhood and Public Services. *Regional Science and Urban Economics*, Vol. 15, 1985, pp. 41-63.

35. Abraham, J.E., and J.D. Hunt. Specification and Estimation of a Nested Logit Model of Home, Workplace and Commuter Mode Choice by Multiple Worker Households. In *Transportation Research Record 1606*, TRB, National Research Council, Washington, D.C., 1997, pp. 17-24.

36. Ben-Akiva, M., and J.L. Bowman. Integration of an Activity-Based Model System and a Residential Location Model. *Urban Studies*, Vol. 35, No. 7, 1998, pp. 1131-1153.

37. Sermons, M.W., and N. Seredich. Assessing Traveler Responsiveness to Land and Location Based Accessibility and Mobility Solutions. *Transport Research Part D*, Vol. 6, No. 6, 2001, pp. 417-428.

38. Orford, S. Valuing Locational Externalities: A GIS and Multilevel Modeling Approach. *Environment and Planning B: Planning and Design*, Vol. 29, 2002, pp. 105-127.

39. Freeman, O. and C.R. Kern. A model of workplace and residential choice in two-worker households. *Regional Science and Urban Economics*, Vol. 27, 1997, pp. 241-260.
40. Blumen, R. Gender Differences in the Journey to Work. *Urban Geography*, Vol. 15, 1994, pp. 223-245.
41. Turner, T., and D. Niemeier. Travel to Work and Household Responsibility: New Evidence. *Transportation*, Vol. 24, 1997, pp. 397-419.
42. Horowitz, J. Statistical comparison of non-nested probabilistic choice models. *Transportation Science*, Vol. 17, 1983, pp. 319-350.

LIST OF TABLES

TABLE 1 Factors Influencing Residential Location Choices

Dwelling	Land Parcel	Neighborhood	Accessibility
• Number of bed rooms	• Distance between houses	• Prestige	• Accessibility to CBD
• Number of rooms	• Backyard size	• Street quality	• Commute distance
• Structure type	• Spaciousness	• Pollution levels	• Commute time by different modes
• Architecture style	• Parking availability	• Housing density	• Accessibility to on-CBD centers
• Size of floor area	• Distance from house to sidewalk	• Amount of open space	• Proximity to local amenities (e.g. hospital)
• Sense of privacy	• View and other aesthetic attributes	• Non-residential land use	• Accessibility to shopping, recreational and employment opportunities
• Construction materials		• Quality of local amenities	
• Amenities		• Demographic and socio-economic status of neighbors	
• Cost / Rent		• School catchment areas and quality	
• Dwelling age and condition		• Pavement/Traffic condition	
• Number of units in dwelling		• Reputation and policies of local government	
• Number of stories in dwelling		• Security and safety (crime rates)	

TABLE 2 Exogenous Variables Considered in the Residential Choice Models for Bay Area

Variable Group	Variable	Property	TAZ	Block	Block Group	Tract	County
Socio-demographics, economic and housing compositions	Number of residents per square mile over total area			X	X	X	
	Number of residents per square mile over land area			X	X	X	
	Distribution of household income in zone				X	X	
	Average household size per occupied housing unit in zone				X	X	
	Share of single person households				X	X	
	Share of couple households				X	X	
	Share of nuclear households				X	X	
	Share of other types of households				X	X	
	Share of single detached unit housing				X	X	
	Median number of rooms per housing unit				X	X	
	Median value of owner occupied housing units in zone				X	X	
	Share of non-Hispanic Caucasian population			X	X	X	
	Share of non-Hispanic African American population			X	X	X	
	Share of non-Hispanic Asian / Pacific Islanders			X	X	X	
	Share of other non-Hispanic ethnicity			X	X	X	
	Share of Hispanic population			X	X	X	
Accessibility measures	Commute time between work zone and candidate residential zone		X				
	Commute cost between work zone and candidate residential zone		X				
	Accessibility to socio-recreational opportunities		X				
	Accessibility to shopping opportunities		X				
	Accessibility to employment opportunities		X				
Environmental	Crime rate						X
	Share of zonal area occupied by water			X	X	X	
	Share of zonal area occupied by land			X	X	X	
Housing	Housing type	X					

TABLE 3 Estimation Results for the Single-Level Models

Variables	Model 1 (Base)		Model 2 (Block)		Model 3 (Block group)		Model 4 (Tract)	
	Param.	t-stat.	Param.	t-stat.	Param.	t-stat.	Param.	t-stat.
Commute time								
Full-time male workers	-0.0422	-9.8	-0.0433	-10.0	-0.0419	-9.4	-0.0420	-9.4
Full-time female workers	-0.0565	-12.5	-0.0571	-12.4	-0.0553	-11.8	-0.0558	-11.9
Part-time male workers	-0.0753	-9.1	-0.0751	-8.9	-0.073	-8.6	-0.0735	-8.7
Part-time female workers	-0.0896	-16.1	-0.0911	-16.3	-0.0868	-15.2	-0.0872	-15.2
Interacted with Caucasian household	-0.0099	-3.3	-0.0094	-3.1	-0.0105	-3.4	-0.0105	-3.4
Interacted with household income (in $100,000)	0.0063	2.0	0.0066	2.1	0.0061	1.8	0.0061	1.9
Employment Accessibility	-0.0224	-5.1	-0.0220	-5.0	-0.0094	-5.2	-0.0099	-5.3
Interacted with household income (in $1,000)	0.0002	4.9	0.0002	4.7	0.0001	4.1	0.0001	4.5
Shopping Accessibility	0.0303	2.0	0.0294	2.0				
Interacted with household income (in $1,000)	-0.0005	-3.0	-0.0004	-2.6	-0.0002	-2.1	-0.0002	-2.2
Interacted with single-person household	0.0361	6.6	0.0366	6.6	0.0332	5.9	0.0300	5.3
Recreation Accessibility								
Interacted with nuclear-family household	0.0063	2.7	0.0059	2.5				
Racial composition								
Share of Caucasian population interacted with Caucasian household			0.6005	6.7	0.7321	5.8	0.6897	5.7
Share of African population interacted with African household			9.8462	9.7	11.4118	8.1	10.9941	7.1
Share of Asian population interacted with Asian household			4.6986	12.1	5.2321	11.7	5.2673	11.6
Share of Hispanic population interacted with Hispanic household			4.2477	6.7	4.6100	5.6	4.9068	5.9
Density over land area (per 10,000 mi^2)								
Interacted with African American household			0.2446	2.7				
Interacted with nuclear-family household					-0.1630	-2.3	-0.2535	-3.4
Interacted with household income between $25,000 and $45,000					0.1945	2.4	0.2517	2.9
Interacted with household income between $45,000 and $75,000					0.1198	1.9	0.1567	2.4
Interacted with household income > $75,000			-0.0824	-2.6				
Absolute difference between zonal median income and household income (in $1,000)					-0.0102	-12.4	-0.0103	-11.8
Absolute difference between zonal average household size and household size					-0.1975	-4.9	-0.2439	-5.8
Household type composition								
Share of couple-only household interacted with couple-only household					1.7116	5.7	2.5558	6.5
Share of nuclear-family household interacted with nuclear-family household					1.0571	3.1		
Zonal Median housing value (in $1,000)					0.0003	2.0		
Divided household income					-0.0376	-5.0	-0.0259	-4.7
Number of observations	4791		4791		4791		4791	
Log-likelihood at convergence	-9436.35		-9214.53		-9026.72		-9071.28	
Adjusted log-likelihood ratio	0.14353		0.16309		0.17966		0.17580	

TABLE 4 Estimation Results for the Multi-Scale Logit Model

Variables	Model 5 (MSL)					
	Param.			t-stat.		
Commute Time						
Full-time male workers	-0.0424			-9.5		
Full-time female workers	-0.0560			-11.9		
Part-time male workers	-0.0733			-8.7		
Part-time female workers	-0.0876			-15.3		
Interacted with Caucasian household	-0.0101			-3.3		
Interacted with household income (in $100,000)	0.0063			1.9		
Employment Accessibility	-0.0098			-5.3		
Interacted with household income (in $1,000)	0.0001			3.9		
Shopping Accessibility						
Interacted with household income (in $1,000)	-0.0001			-1.9		
Interacted with single-person household	0.0278			4.8		
Recreation Accessibility						
Interacted with nuclear-family household	0.0036			1.5		
	Block level		**Block group level**		**Tract level**	
	Param.	t-stat.	Param.	t-stat.	Param.	t-stat.
Racial composition						
Share of Caucasian population interacted with Caucasian household	0.6210	6.2				
Share of African population interacted with African household	9.8956	9.8				
Share of Asian population interacted with Asian household	2.8256	3.9	2.4318	3.0		
Share of Hispanic population interacted with Hispanic household	2.0197	2.0	2.8344	2.3		
Density over land area (per 10,000 mi^2)						
Interacted with African American household	0.1863	2.2				
Interacted with nuclear-family household			-0.1505	-2.1		
Interacted with household income between $25,000 and $45,000					0.2315	2.6
Interacted with household income between $45,000 and $75,000					0.1389	2.1
Absolute difference between zonal median income and household income (in $1,000)			-0.0100	-12.1		
Absolute difference between zonal average household size and household size			-0.1830	-4.2		
Household type composition						
Share of single-person household interacted with single-person household					0.8999	2.0
Share of couple-only household interacted with couple-only household			0.9197	2.2	1.6450	3.0
Share of nuclear-family household interacted with nuclear-family household			1.0477	3.0		
Zonal Median housing value (in $1,000)			0.0004	2.4		
Divided household income			-0.0383	-5.1		
Number of observations	4791					
Log-likelihood at convergence	-8999.41					
Adjusted log-likelihood ratio	0.18159					