Modeling of Household Location: A Statistical Approach

RAYMOND H. ELLIS, Department of Civil Engineering and The Transportation Center, Northwestern University

The purpose of this study is to simulate the behavior of households in choosing their homes. A two-phase model is proposed. During the first phase, the housing preferences of the locating family are determined. In the second phase, the search process by which the household picks a location possessing these characteristics is simulated. A number of multivariate statistical techniques are employed in the partial calibration of the model.

RECENT progress in land-use modeling has included the implementation of macro-level procedures for estimating future spatial distributions of urban activity. Although these techniques have fulfilled the initial requirement for land-use projections as inputs to the transportation planning process, they do not meet the general planning requirement for a policy oriented land-use planning methodology. Such a methodology should furnish planners with the information necessary to evaluate alternative policy sets. Further, it should be adaptable to changes in behavior, technology, resource availability, and policy.

At present, progress towards this goal can best be achieved by research leading to the development of a more micro-level model of residential location behavior. Advances in sociology and economics have resulted in a substantial body of theory upon which such work can be based. These developments, in conjunction with the increasing quality and quantity of survey findings, provide a firm foundation for modeling efforts. In addition to leading toward a desired goal, a concentration of effort on modeling residential location has intrinsic merit. The largest single use of land in any metropolitan area is for residential purposes. Since the total amount of land is fixed, this allocation has important implications for the structure of land prices. Further, the individual is vitally affected by the residential land-use consequences of alternative policies, and any difference in the welfare level attained is therefore an important criterion for evaluating plans.

This study was designed as an exploratory investigation into the location behavior of households. The development is based on the premise that the family's choice of a home is an essentially rational decision which is reached through a consideration of preferences, financial resources, and the market. A prototypical micro-level model is developed and partially calibrated using data from Tucson, Arizona. Ultimately, of course, such an analysis must be integrated with an overall simulation, although not necessarily a Monte Carlo simulation, of the urban development process.

FORMULATION OF MODEL

The proposed model of residential location behavior operates in two steps. First, a description of the environment desired by a given household is determined through consideration of its socioeconomic characteristics. In effect, the quality and quantity

Paper sponsored by Committee on Land Use Evaluation and presented at the 46th Annual Meeting.

of the housing which the household will select is fixed in an environmental space. Numerous sites meeting the environmental requirements specified by the household normally exist throughout a metropolitan area. Second, a search process is conducted to select one of these sites as the new location for the family.

The choice of a house usually involves a simultaneous decision about the future trip set of the family and the style of life which the family will follow. Since each location is characterized by a set of accessibilities to all trip destinations, the household, in choosing a site, is also making a decision about desired levels of these accessibilities. By choosing the job site as the origin of the search process, the oft-cited relationship between the location of the work site and the residential location choice is introduced into the model. Alternatively, if the work trip does not appear to strongly influence the housing choice, the influence of information about opportunities can be entered into the model by choosing the existing homesite as the origin for the search. In practice, either of these origins may be chosen depending on the circumstances of the searching family.

Mathematically, the premise that the housing environment which a locator selects is a function of his socioeconomic characteristics may be expressed as

$$E^i = f\left(x_1^i, x_2^i, \ldots, x_j^i, \ldots, x_n^i\right) \tag{1}$$

where E^i is a scalar representing the environment selected by the ith household, and x_j^i is the jth socioeconomic characteristic influencing the location behavior of the ith household.

For purposes of clarity, the environment was characterized in Eq. 1 by a scalar value. In reality, the environment must be characterized by a vector whose elements completely describe the quality of the housing and the area in which the housing is located. For example, the environmental vector might include measures of housing cost, number of persons per room, and type of house to characterize the housing, and measures of social status, racial composition, and quality of educational and recreational facilities to characterize the area in which the housing is located. Eq. 1 may now be written as

$$\begin{aligned} y_1^i &= f\left(x_1^i, \ldots, x_j^i, \ldots, x_n^i\right) \\ &\vdots \\ y_k^i &= f\left(x_1^i, \ldots, x_j^i, \ldots, x_n^i\right) \\ &\vdots \\ y_m^i &= f\left(x_1^i, \ldots, x_j^i, \ldots, x_n^i\right) \end{aligned} \tag{2}$$

where y_k^i is the kth element of the environment which is considered by household i.

Any given household characteristic j does not necessarily influence the level of the measure of the environment, mathematically:

$$\frac{\partial y_k^i}{\partial x_j^i} \gtreqless 0 \tag{3}$$

Eq. 2 may be written in matrix form as

$$[Y] = f[X] \tag{4}$$

where Y is an $m \times 1$ matrix characterizing the environment, and X is an $n \times 1$ matrix characterizing the household.

The functional relationship may take any of a number of forms. The form adopted for this analysis involves a linear transformation of elements. Rewriting Eq. 2:

$$\begin{aligned} y_1^i &= a_{11} x_1^i + \ldots + a_{1j} x_j^i + \ldots + a_{1n} x_n^i \\ &\vdots \\ y_k^i &= a_{k1} x_1^i + \ldots + a_{kj} x_j^i + \ldots + a_{kn} x_n^i \\ &\vdots \\ y_m^i &= a_{m1} x_1^i + \ldots + a_{mj} x_j^i + \ldots + a_{mn} x_n^i \end{aligned} \tag{5}$$

This formulation may be represented in matrix form as

$$[Y] = [A]\,[X] \tag{6}$$

where A is an $m \times n$ matrix of coefficients.

Eq. 3 states that an element a_{kj} of matrix A can assume any real value. To summarize, the model may be characterized as follows:

$$\underset{\text{environmental vector}}{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}} = \underset{\text{desire coefficients matrix}}{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}} \underset{\text{household vector}}{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}} \tag{7}$$

Estimates of the parameters of the coefficients matrix are obtained by an analysis of the current location structure. The model, therefore, does not predict what a household desires, but rather what the household must settle for under current conditions. These conditions may change, and it is implicitly assumed in this development that the values in the coefficients matrix are not fixed. The values of the matrix may be altered to take account of changing group norms, as revealed by studies of historical trends, consumer studies of the housing desires of various groups, and other types of analyses. For example, the group characterized by the vector (blue-collar, white, income less than $5000, stage three in the family life cycle, six members in the family) might be placing an increasingly important emphasis on recreation relative to the general population. This change in group behavior could be taken into account through the alteration of the appropriate coefficients.

An environmental bundle which will be demanded by each group of families has been developed, but the households have not been located in physical space nor has the decision component of work site accessibility been considered. At each time stage, the number of opportunities of each environment in each subarea can be estimated. In effect, an opportunity surface for each environment is constructed for the metropolitan area. It is assumed that a finite number of employment centers have been located within the metropolitan region in the industrial allocation phase of the overall land-use simulation model. The characteristics of families whose principal wage earner works at each work site could then be determined, and the residential demand relative to each employment center estimated using the demand equations.

The process by which each family uses its employment center as an origin to search the environmental opportunity surface for a homesite could then be simulated through models analogous to those used in the distribution phase of a synthetic traffic study. Trip distribution models use one or more of the following factors to estimate the probability of an interaction between zone k and zone n:

1. The intensity of activity at zone n, being in the residential location model the number of opportunities for family i at zone n;
2. The number of opportunities between zone k and zone n which the family must pass up to locate in zone n; and
3. The cost of interaction between zone k and zone n, in the residential location model this being the time distance from the homesite to the work site.

In the absence of further information relating to the relative importance of any one of these factors, all three will be included in the allocation model. The probability that a given zone will be accepted is, therefore, a function of both the cost of interaction with the zone and the opportunity surface for the given housing environment. Mathematically, this may be characterized as:

$$P_{ikn} = K \left[\frac{O_{mn}}{\sum_{j=1}^{n} O_{mj}} \right] t_{kn}^{b} \qquad (8)$$

where

P_{ikn} is the probability that members of socioeconomic group i working in zone k will locate in subdivision n,
O_{mn} is the number of opportunities of environmental set m in subdivision n,
t_{kn}^{b} is the time distance from employment zone k to subdivision n,
b is a parameter which must be calibrated, and
K is a normalizing constant.

After the coefficients matrix has been estimated, the parameter b is approximated by an analysis of the current pattern of residential location. Note that the model implicitly assumes that everyone desiring a certain type of environment can find a homesite with this environment within the metropolitan area and within a commuting range which the head of the household will accept. There is no feedback, in this sense, from the accessibility of the desired environments to the actual desire for the environments. The feedback is considered by recalibrating the coefficients matrix for each metropolitan area and calibrating b for each zone. The effect of accessibility on the choice of environment is therefore implicitly but crudely considered in the desire coefficients.

The model can be operated in either an iterative or a single-pass mode, depending on the degree of accuracy desired. If only a reasonable degree of accuracy is desired, the employment centers from which locators are distributed would be randomly selected, the locators assigned, and the opportunity surface appropriately adjusted. The alternative procedure would involve distributing the locators from all employment zones, determining those zones in which the number of assigned locators exceeds the available supply, appropriately reducing the available supply at the zones in which the supply was exceeded so as to reduce their attractiveness, and iterating until a stable situation develops. It would appear that the first procedure would be adequate in most situations.

To recapitulate, the location model operates in two stages (Fig. 1). In the first stage, the household vector is manipulated to yield the housing environments which the family will desire. Prototypical vectors for representing the environment and the household are developed in the next section. In the second stage, households are distributed to available housing sites using the work site of the principal worker as the

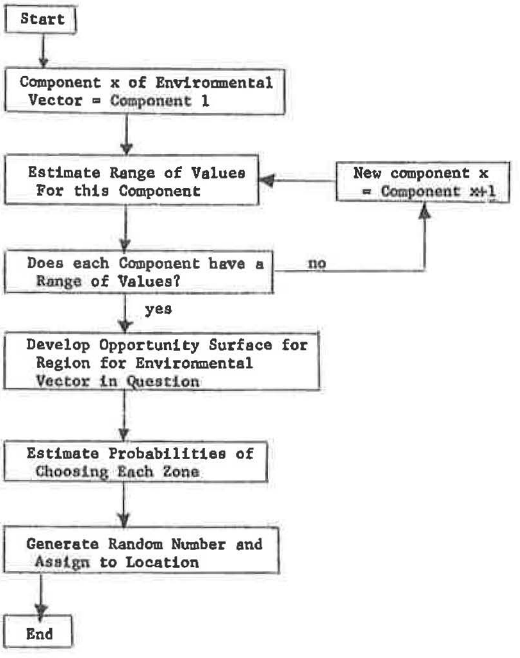

Figure 1. Flow chart for location process for an individual household.

origin and a distribution procedure which considers both the opportunity surface and the actual distance from work site to homesite.

DEVELOPMENT OF ENVIRONMENTAL AND HOUSEHOLD VECTORS

Prior to the calibration of the desire matrix, it is necessary to define the environmental and household vectors. Fulfillment of the following criteria is suggested as necessary for the development of an effective environmental vector:

1. The components of the model must be definable at the level at which the model is to be applied, the level of the travel analysis zone;
2. It is necessary that the variables form a relatively small group, since they will be frequently manipulated and interpreted;
3. As a set, the variables must exhaust the possible variation in environment among alternate sites;
4. The variables used at the census tract or travel analysis zone level must be redefinable at the level of the individual household; and
5. The variables used at the travel analysis zone level of aggregation must be related to all and exhaust at least the major factors actually considered by a given household in choosing a site.

This study will focus at the level of aggregation—the travel analysis zone—and determine if a defining environmental vector for each area can be developed. The analysis is concerned with developing measures which fulfill the first three criteria. The limited resources available for the study precluded demonstrating that the suggested measures of physical differentiation do indeed correspond to the individual's perception of site amenities, or that the proposed measures "exhaust" the individual's perception of that environment. The first goal, definition at the census tract-analysis zone level, is achieved by defining the variables at that level. The other two criteria will be achieved through operations on the proposed indices.

Measures of the environment are proposed below which are defined at the level of the census tract-travel analysis zone. Difficulties in obtaining the necessary data resulted in the exclusion of measures of shopping facilities. The following indices of environment are suggested:

1. The population density of the tract (persons per acre);
2. The proportion of land in parks;
3. The proportion of land in open space;
4. The proportion of the units built before 1939;
5. The proportion of the units built after 1950;
6. The proportion of the units in sound condition and containing all of the standard plumbing facilities;
7. The median value of the homes;
8. The median gross rent;
9. The proportion of single family dwelling units;
10. The median rooms per dwelling unit;
11. The proportion of units with 0. 50 persons per room or less;

TABLE 1

LISTING OF ORIGINAL ENVIRONMENTAL INDICES

Variable Number	Mnemonic	Definition
1	DENSIT	Density of census tract in persons per acre
2	PCPARK	Proportion of area of tract which is park
3	PCOPSP	Proportion of area of tract which is open space
4	PCOLDH	Proportion of homes in tract built before 1939
5	PCNEWH	Proportion of homes in tract built after 1950
6	PCGOOD	Proportion of homes in good condition with all plumbing
7	MEDVAL	Median value of homes in tract in dollars
8	MEDRNT	Median gross rent in dollars
9	PCSFDU	Proportion of units in tract which are single-family dwelling units
10	MEDRMS	Median number of rooms per house
11	LOWDEN	Proportion of units in tract with less than 0. 50 persons per room
12	HGHDEN	Proportion of units in tract with more than 1. 01 persons per room
13	PCWHIT	Proportion of the tract's population which is white
14	SBSOCR	Shevky and Bell social rank index
15	SBOCCI	Shevky and Bell occupation index
16	SBEDUI	Shevky and Bell education index
17	MEDYRE	Median years of education of people in tract
18	MEDINC	Median income of inhabitants of the tract

12. The proportion of units with 1. 01 persons per room or more;
13. The proportion of the population which is white;
14. The social rank of the area as defined by Shevky and Bell (1);
15. The Shevky and Bell occupation ratio for the area;
16. The Shevky and Bell education ratio;
17. The median years of education of the residents; and
18. The median family income.

Table 1 contains a listing of the assigned variable number, a mnemonic and a description of the variable.

Tucson, Arizona, was chosen as the study city. Those census tracts outside the legal limits of the city contain areas which are functionally unrelated to the city. Two

TABLE 2

ENVIRONMENTAL VECTOR: NORMAL VARIMAX ROTATED FACTOR MATRIX

Variables	Loadings on Factor[a]							Communality
	1	2	3	4	5	6	7	
1 DENSIT	–	–	–	–	-0. 938	–	–	0. 959
2 PCPARK	–	–	0. 967	–	–	–	–	0. 973
3 PCOPSP	–	-0. 631	–	–	–	–	0. 522	0. 941
4 PCOLDH	–	0. 811	–	–	–	–	–	0. 938
5 PCNEWH	–	-0. 808	–	–	–	–	–	0. 941
6 PCGOOD	0. 777	–	–	–	–	0. 491	–	0. 952
7 MEDVAL	0. 852	–	–	–	–	–	–	0. 963
8 MEDRNT	0. 781	–	–	–	–	–	–	0. 922
9 PCSFDU	–	-0. 894	–	–	–	–	–	0. 849
10 MEDRMS	0. 750	–	–	–	–	–	–	0. 881
11 LOWDEN	0. 811	–	–	–	–	–	–	0. 933
12 HGHDEN	-0. 957	–	–	–	–	–	–	0. 962
13 PCWHIT	–	–	–	-0. 935	–	–	–	0. 996
14 SBSOCR	0. 978	–	–	–	–	–	–	0. 980
15 SBOCCI	0. 970	–	–	–	–	–	–	0. 966
16 SBEDUI	0. 916	–	–	–	–	–	–	0. 941
17 MEDYRE	0. 885	–	–	–	–	–	–	0. 927
18 MEDINC	0. 621	–	–	–	–	–	–	0. 946
Eigenvalue	8. 342	3. 565	1. 443	1. 255	1. 316	0. 502	0. 546	
Cumulative percent common variance	49. 2	70. 2	78. 7	86. 1	93. 8	96. 8	100. 0	
Interpretation of factor	Socio-economic	Single-family dwelling units	Percent parks	Racial composition	Density	Percent good	Percent open space	

[a]Factor loadings between + 0.49 and - 0.49 have been omitted to ease reading.

TABLE 3

LISTING OF INDEPENDENT VARIABLES USED IN CALIBRATION OF COEFFICIENTS MATRIX

Variable Number	Mnemonic	Description	Dummy (D) or Continuous (C)
1	ZERCAR	Household owns no cars	D
2	ONECAR	Household owns one car	D
3	TWOCAR	Household owns two cars	D
4	ONEPER	One person in household	D
5	TWOPER	Two persons in household	D
6	THFRPR	Three or four persons in household	D
7	FVSXPR	Five or six persons in household	D
8	NOPEMP	No persons employed	D
9	ONEEMP	One person employed	D
10	TWOEMP	Two persons employed	D
11	RACEHH	Race of head of household	D
12	OCCPHH	Occupation of head of household	D
13	LENRES	Length of residence at this site	C
14	PCEMPD	Proportion of household employed	C

of the five tracts, for example, consist of Indian reservations. The observations were therefore limited to those census tracts within the city limits. For each of these tracts, the previously described set of observations was obtained from the 1960 Census report for Tucson and from Volume One of the Final Study Report of the Tucson Area Transportation Study.

Following preliminary analysis, a seven-component environmental vector was hypothesized which contained the following factors:

1. A socioeconomic status factor;
2. A factor pertaining to proportion of single family dwelling units;
3. A recreational facilities factor;
4. A racial composition factor;
5. A density (of population) factor;
6. An age of housing factor; and
7. A proportion of open space factor.

A factor analysis was carried out to test the power of this model. The rotated factor matrix is given in Table 2. Since the results generally confirmed the hypothesized model, a prototypical environmental vector containing seven components was adopted for use in the analysis.

The requirement for a vector whose mutually independent components precisely define the household is implicit in the presentation of the model. A common-sense approach suggests that the following factors which influence the location behavior of the household should be considered for inclusion in the vector:

1. Income;
2. Number of years of education of the head of the household;
3. Proportion of the household which is employed;
4. Size of the household;
5. Number of children in the household;
6. Stage in the family life cycle;
7. Race of the head of the household;
8. Occupation of the head of the household (blue-collar or white-collar); and
9. Sex of the head of the household.

Unfortunately, data limitations precluded the development and analysis of these indices. Based on available data, and following an analysis similar to that performed for the environmental vector, a seven-component household vector was developed. The following indices were used to measure socioeconomic differentiation among households:

1. The number of cars owned by the family, a surrogate for income;
2. The total number of persons in the family;
3. The number of persons employed;

4. The proportion of the household employed;
5. The length of residence in the area, introduced because of its potential relevance to the ex post facto analysis of the household's location behavior;
6. The race of the head of the household; and
7. The occupation of the head of the household.

CALIBRATION OF PREFERENCE MODEL

The calibration of the desire coefficients matrix could have been performed at either the household level, which is essentially non-aggregated data, or at the level of the census tract. Operating at the census tract level implies the concept of a single composite family representing all the families in the tract by the average value within the tract for each parameter. The principal argument against the use of such data involves the conceptual difficulty which emerges in defining an average family for a census tract. A tract family does not exist and it is difficult to make statements of a behavioral nature about a nonexistant entity. Further, the desire coefficients become a function of the artificial set of boundaries which are used to define a census tract.

The alternate and chosen approach focuses at an appropriate level of behavior, the household. Problems of a different sort emerge in operating at this level. The race and the occupation of the head of the household are measured on ordinal scales, but the technique of linear regression which will be used to calibrate the coefficients matrix considers the relationship between interval scales. Utilizing the technique of dummy variables (2), race and occupation can be included in a regression equation. Further difficulties result from the nonlinear effect of other variables, particularly number of cars owned, number of people in the family, and number of people employed. These nonlinearities may be taken into account by coding these variables as dummy variables.

Converting the noncontinuous and nonlinear variables identified previously into dummy variables results in 14 independent variables, which are given in Table 3. The criterion for choosing among the infinite number of possible combinations of the elements a_{kj} is the maximization of the explained variance of the dependent variable y_k. Since the equation will not be forced through the origin, an m × 1 vector of coefficients must be added to take account of the y intercept. Rewriting Eq. 7:

$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_7 \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,14} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{7,1} & \cdots & a_{7,14} \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_{14} \end{bmatrix} + \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_7 \end{bmatrix} \tag{9}$$

TABLE 4

LISTING OF DEPENDENT VARIABLES

Variable Number	Name	Description
1	Social rank	Shevky and Bell social rank index
2	Proportion of single-family dwelling units	Proportion of dwelling units in tract which house only one household
3	Recreational facilities	Proportion of land area of tract taken up by parks
4	Racial composition	Proportion of population of tract which is white
5	Density	Density of tract in person per acre
6	Age	Proportion of dwelling units in tract which were built between 1950 and 1960
7	Open space	Proportion of land area of tract taken up by open space

TA

REGRESSION E

Variable	Constant	ZERCAR	ONECAR	TWOCAR	ONEPER	TWOPER	THFRPR	FVSXI
Social rank	47. 2	-15. 4 (4. 0)	-4. 8 (2. 4)	—	—	—	—	—
Proportion of single-family homes	86. 2	-8. 6 (1. 4)	—	—	—	—	—	3. 1 (1. 0)
Recreational facilities	0. 2	—	—	0. 3 (0. 1)	—	—	—	—
Racial composition	72. 3	-3. 0 (1. 3)	—	—	—	—	—	—
Density	6. 7	—	—	—	1. 7 (0. 5)	0. 8 (0. 4)	—	—
Age of housing	51. 3	-20. 8 (5. 1)	—	—	—	-6. 6 (3. 3)	—	—
Open space	39. 1	-8. 5 (3. 0)	—	—	—	-6. 2 (2. 0)	—	—

All coefficients are significant at the 0.05 level.

The calibration of the coefficients matrix is equivalent to a series of seven multiple regressions, one regression for each of the seven components of the vector Y. Since there is no reason to assume that every variable enters into every equation, a stepwise form of multiple regression was employed with the conditions that a variable must attain a level of significance of 0. 05 or greater to be included in the equation, and that the variable be removed from the equation if the addition of subsequent variables causes it to fall below a 0. 05 level of significance. Table 4 lists the chosen dependent variables.

Table 5 summarizes the equations which were developed. The standard error for each regression coefficient is shown in parentheses beneath the coefficient. The correlation coefficient (r), the standard error of estimate, and the level of significance of each equation are shown to the right. It is observed that the correlation coefficients are low; this implies that the independent variables used in the calibration phase of the study do not give sufficient insight into the housing choices made by the household. The availability and utilization of other measures of household characteristics, such as income, stage in the family life cycle, education, and number of children, might have improved the results obtained from the calibration of the model. Nonetheless, the results are encouraging, particularly in view of the disaggregate nature of the observations.

Several generalizations can be based on the results. Each of the independent variables entered into at least one of the equations. In this sense, the hypothesis that these factors do relate to the household's location behavior is supported. Only four independent variables enter into three or more equations. Since these four factors could be interpreted as influences which cause location desires to deviate from the norm, each will be discussed separately.

The dummy variable "household owns no cars" entered into five of the seven equations. It was previously suggested that car ownership is a surrogate for income. The analysis supports this contention by showing that households not owning a car tend to live in older, nonwhite areas containing fewer single-family homes and having less open space and a lower social rank. An ex post facto hypothesis is advanced that the dummy variable "two persons in household" represents the influence of retired couples. Having recently moved to Tucson and not owning a car, these people would tend to live in older areas with a higher density and less open space.

The dummy variable "race of the head of the household" entered strongly into four of the seven equations. The dummy variable was coded one if the head of the household was white and zero otherwise. Those coded one tended to live in newer, less dense, and white areas which had a considerably higher social rank. The length of residence variable entered into four equations. Not surprisingly, those who have lived at a site for a longer period of time live in older areas and have less open space.

In order to implement the model developed here, the future attributes of each site in the metropolitan area must be known. It is observed that all of the environmental variables except social rank are either predictable, e. g., age and racial composition, or

ONS DEVELOPED

NOPEMP	ONEEMP	TWOEMP	RACEHH	OCCPHH	LENRES	PCEMPD	r	Standard Error	Significant at Level
—	—	-6.1 (2.4)	27.7 (2.6)	—	—	—	0.57	18.4	0.0005
15.2 (7.0)	—	—	—	4.5 (2.0)	-0.3 (0.1)	—	0.46	6.8	0.0005
—	—	-0.2 (0.1)	—	—	—	—	0.21	0.65	0.001
—	—	—	6.5 (1.0)	21.0 (2.0)	-3.0 (1.3)	—	0.69	6.5	0.0005
—	—	—	-1.0 (0.4)	—	—	—	0.25	2.9	0.0005
57.4 (25.6)	—	—	23.6 (3.6)	—	-1.1 (0.3)	-22.9 (5.2)	0.54	24.8	0.0005
—	-5.0 (2.0)	—	—	—	-0.5 (0.2)	-11.5 (3.3)	0.40	15.2	0.0005

can be planned for, e. g., density level, proportion of single-family homes, recreational facilities, and open space. Although an effort could be made to predict social rank, the alternative strategy of substituting a plannable factor, median value of housing, appears to be more feasible, particularly since these measures are highly correlated. A multiple regression was performed using median value of housing as the dependent variable, and approximately the same variables and levels of quality were obtained.

CONCLUSIONS AND RESEARCH IMPLICATIONS

Two tentative conclusions can be drawn from this analysis. Three factors, income, stage in the family life cycle, and race, appear to be correlated with variations in the environments selected by different households. These factors, and additional factors identified in future work, should be considered in future efforts to predict the environmental preferences of households.

The second conclusion is based on the observed relationship between length of residence and characteristics of the environment. As households change through time, their housing preferences also vary. This relationship suggests that, in lieu of constantly moving to obtain environments which satisfy their existing preferences, housholds accept certain gaps between their preferred and their existing environment.

It has been the purpose of this paper to explore selected issues relating to the micro-level simulation of residential location behavior. The most important research implication of this work is the need for more sophisticated data on both consumer behavior and consumer preferences in the housing market. For example, there is presently little data available concerning the effect of variations in the level of information available to the consumer on the consumer's behavior in the housing market. The amount of information on which households actually base their location decisions is unknown. It is therefore impossible to build these considerations into new location models. Behavioral data on the levels of information achieved by different locators could be obtained through in-depth surveys of households which have recently selected a new location.

Surveys of consumer preferences based on data about actual behavior may handle the preference-reality gap emphasized previously in either of two ways. They may sample only those households which have recently moved or they may attempt to measure dissatisfaction with the existing environment. Neither of these approaches, however, explicitly probes the vital question of the play-offs among preferred attributes which households make in selecting an environment and a site. Such information can best be achieved through the development of games in which households would be asked to choose among locations with varying attributes and to explain their choices. While the implementation of such games is undoubtedly difficult, the additional empirical insight thereby gained should be considerable.

REFERENCES

1. Shevky, E., and Bell, W. Social Area Analysis. Stanford Univ. Press, 1955.
2. Suits, Daniel B. Use of Dummy Variables in Regression Equations. Jour. Amer. Stat. Assoc., Vol. 52, pp. 548-551, 1957.