Environment and Planning B: Planning and Design 2005, volume 32, pages 799 – 820

DOI:10.1068/b3071

Characterizing location preferences in an exurban population: implications for agent-based modeling

Luis E Fernandez¶
School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109, USA
Daniel G Brown
School of Natural Resources and Environment, and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA; e-mail: danbrown@umich.edu
Robert W Marans
Institute for Social Research, and Taubman College of Architecture and Urban Planning, University of Michigan, Ann Arbor, MI 48109, USA; e-mail: marans@umich.edu
Joan I Nassauer
School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109, USA; e-mail: nassauer@umich.edu
Received 10 September 2003; in revised form 22 March 2005

Abstract. Powerful computational tools are becoming available to represent the behavior of complex systems. Agent-based modeling, in particular, facilitates an examination of the system-level outcomes of the heterogeneous actions of a set of heterogeneous agents: for example, patterns of land-use and land-cover change, such as urban sprawl as a result of residential location decisions. These new tools create new demands for data, and empirical studies of the selection behavior of residents. Using resident responses from the 2001 Detroit Area Study survey, we compared two alternative approaches to characterizing the heterogeneous preferences of agents; both based on a factor analysis of resident responses to questions about their reasons for moving to their current location. We used cluster analysis to identify how many and what types of residents there are, grouped by similar preferences. We also evaluated the relationships between socioeconomic and demographic characteristics and location preferences using regression trees, and evaluated the fit of the relationship to determine the degree to which socioeconomic characteristics predict preferences. The results showed that the preferences of resident exurbans of single-family homes in the Detroit metropolitan area were heterogeneous and that distinct preference groups do exist in resident populations, but are not well characterized on the basis of simple socioeconomic and demographic variables. We conclude that, given the heterogeneous nature of preferences and a relatively limited number of preference groupings observed in the survey respondents, agent-based models simulating resident behavior should reflect this diversity in the population and incorporate distinct agent classes of empirically derived preference distributions.

Introduction

A variety of urban system models have been developed to understand the complex interactions that give rise to spatial patterns of development throughout a region. These models range in their realism and requirements for empirical support, from econometric and hedonic models (Bell and Irwin, 2002; Geoghegan, 2002; Irwin, 2002), which are based on individual, household, and price data, and discrete choice modeling (Landis, 1994) to cellular automata models with some degree of calibration of rules to data (Batty and Xie, 1997; Clarke et al, 1997), and physical analog models based on, for example, diffusion-limited aggregation and correlated percolation (Batty and Longley, 1994; Makse et al, 1998), which often involve little more use of data than comparing some aggregate characteristics of the patterns generated with observed patterns.

¶Corresponding author: US Environmental Protection Agency, Office of International Affairs, 1200 Pennsylvania Ave, Mail Code: 2650R, Washington, DC 20460, USA; e-mail: fernandez.luis@epa.gov.

Recent developments in, and applications of, agent-based simulation in land-use modeling have created new opportunities for simulating microlevel (that is, individual and household) behaviors and evaluating their implications at aggregate levels, but also new requirements for data (Parker et al, 2003). The multiple useful roles of empirical data in informing these models remain to be completely determined. Empirical support for agent-based models can generally take either or both of two forms: (a) using observations about aggregate outcomes to evaluate or calibrate the model or (b) using data on individual or agent decisions to generate agent behavior. In the case of aggregate outcomes, it might be desirable simply to observe that some phenomenon results from a simple model of behaviors, as in the case of Page's (1999) model of city formation. Alternatively, model output can be compared with maps of the observed pattern using some spatial descriptor, such as the power-law distribution of cluster sizes (Batty and Longley, 1994; Rand et al, 2003) or other summaries of patch characteristics (Parker and Meretsky, 2004). At the level of agent behaviors, role-playing games have been used to collect information about decisionmaking behaviors for use in models that simulate these behaviors (Barreteau, 2003; Barreteau et al, 2001). Alternatively, and less expensively, agent behaviors are often determined through deductive reasoning (for example, Brown et al, 2004; Otter et al, 2001). Though this approach can be satisfying in the case of very simple models, and the sensitivity of models to assumptions can be evaluated, the importance of diversity in agent behaviors in complex systems (for example, Rand et al, 2002) suggests that it is worth some effort to characterize empirically the heterogeneity in an agent population. Indeed, representing diversity among agents is one of the important strengths of agent-based modeling, but it may be important to understand not just that diversity of behaviors exists, but something about its distribution and relation to agent characteristics.

The research presented here supports a project that aims to develop agent-based models of land-use and land-cover change at the urban–rural fringe to help improve our understanding of human–environment interactions within this dynamic zone. The agent-based models developed in the project to date (Brown et al, 2002; Rand et al, 2002; 2003) model land-development patterns as a function of agent preferences for nearness to services and jobs, landscape aesthetic quality, and particular levels of density. The models have been populated with agents that have either identical preferences or preferences drawn from a single population, that is, characterized by a mean and standard deviation. The relative importance of preferences in determining the agent locations has been evaluated through parameter sweeping, that is, evaluating model behavior at different relative levels of preference. Alternative approaches might be to draw preferences from multiple populations, representing agent types, or to give each agent a unique set of characteristics and preferences. In the latter case, it might be possible to assign each agent a preference as a function of its socioeconomic and demographic characteristics (for example, age, marital status, race, and income). To judge the most realistic approach, we analyzed data collected through a survey of stated residential preferences conducted within the Detroit metropolitan area in 2001.

The analysis involves comparing two alternative approaches to characterizing the heterogeneous preferences of agents; both are based on a factor analysis of resident responses to questions about their reasons for moving to their current location. First, we used cluster analysis to identify how many and what types of residents there are, grouped by similar preferences. Second, we evaluated the relationship between socioeconomic and demographic characteristics and location preferences by using regression trees, and evaluated the fit of the relationship to determine the degree to which socioeconomic characteristics predict preferences. The results of the two analyses were compared to answer the basic question: to what degree are preferences independent of socioeconomic characteristics? The answer to this question will help us decide how to represent residents as agents, that is, whether to use preference groups or assign preferences to agents based on socioeconomic characteristics.

Following a summary of previous research into exurban residential location preferences in the United States we describe our dataset and outline the methods used. The cluster-analysis and regression-tree results are then presented and compared for interpretation of the patterns of residential preferences among residents and in relation to socioeconomic and demographic characteristics. We conclude with a discussion of the implications of such analyses of survey results for the conduct of agent-based modeling.

Background: exurban residents

Patel (1980, page xi) defined exurbia as "discrete, areally organized subdivisions[s] with an internal street pattern, located in a rural setting ... far enough beyond the frontier of suburban development that it will not be engulfed by the expanding city within the foreseeable future". For their studies in the Portland, Oregon, metropolitan area, Nelson and Dueker (1990), Davis (1993), and Davis et al (1994) used the Portland Metropolitan Area urban growth boundary to demarcate the inner edge of exurbia. Nelson and Sanchez (1997) defined exurbia as the portion of a metropolitan statistical area (regions classified as metropolitan) that is rural, and outside the urbanized area.

National surveys have consistently shown that about three quarters of Americans prefer to live outside of cities. A series of surveys administered to a national sample in 1972, 1988, and 1992 showed that preferences for size of place and proximity to urban areas have been consistent over twenty years. Across the three surveys, approximately 30% of respondents preferred to live in a city larger than 50 000 population, 50% preferred a smaller city, town, or rural area within 30 miles of a larger city, and approximately 20% preferred a small place more than 30 miles from a city (Brown et al, 1997). Similarly, a review of several national surveys conducted by interest groups in the late 1990s found that approximately 75 – 80% of respondents preferred a low-density suburban residential setting to a city environment (Malizia and Exline, 2000; Myers and Gearin, 2001). Furthermore, people who are not living in their preferred size of place are twice as likely to prefer a place that is less rather than more dense (Brown et al, 1997). This widespread desire for small town and rural residential settings is an important driver of sprawl.

National surveys of homebuyers have shown that the majority, including the majority of exurbanites, prefers single-family detached dwellings, low densities, and ease of automobile use (Malizia and Exline, 2000; Talen, 2001; Varady, 1990). Compared with suburbanites, exurban residents were slightly more likely on average to have children, have slightly more workers per household, and have a higher proportion of blue-collar workers (Nelson and Sanchez, 1997). Exurban residents tended to have long commuting times in common (Anas et al, 1998; Davis, 1993; Nelson and Sanchez, 1997; Patel, 1980) but, in most respects, exurban residents appear to be a heterogeneous group in most metrics (Davis, 1993; Davis et al, 1994; Nelson and Sanchez, 1997; Patel, 1980). For agent-based modeling, this heterogeneity may be more informative than their commonalities.

Aspects of exurban settings that draw people to live there may vary with income. One study found that income elasticity of land consumption and the elasticity of the cost of travel were very similar for high-income households, suggesting other factors may account for high-income households moving to exurbia (Wheaton, 1977). It may be that, in making residential location choices, higher income households show greater attention to 'pull factors' than lower income households, which are more subject to negative conditions at their origin (Spain, 1989). Davis (1993) used data from a survey of recent homebuyers in the Portland, Oregon region to develop the following typology of exurban residents:

Economy minded: 31% of total sample. These households earn less than $50 000 a year, buy smaller lots, buy the least expensive homes, live farther from downtown, and have fewer children. They rate large lots, owning instead of renting, and rural living highly. Davis characterizes them as "first-time home buyers wanting large lots and a rural environment" (page 11).
Family oriented: 29% of the sample. They always have children, are mainly two-earner families, live relatively close to downtown (for exurban dwellers), have home prices similar to suburbanites, and among exurbanites have average-size lots. They place strongest emphasis on owning a large lot, living in a rural area, and having a better place for their family. They value having a large lot more than having a 'different size' house.
Affluent: 24% of sample. Earn at least $50 000 year, own larger, more expensive properties (four acres is the median), are generally two-career managerial and professional couples without children, and live closer to the downtown than most exurbanites, probably because they tend to work in the central business district rather than the suburbs.
Long-distance commuters: 16% of the sample own the largest lots and paid the lowest prices for their lots. Their work trips are twice as long as the combined average (51 minutes as compared with 27 minutes), and they tend to have only one worker in the family. They often have children and strongly value living in a rural environment and having a large lot. They tend to have blue-collar, technical, or sales jobs and make $20 000 to $50 000 yearly.

Nelson and Sanchez (1997) identified two dominant types of exurban residents within their national sample as characterized by dual-wage professional families compared with nonprofessional or single-worker families (table 1). Their types tended to validate Davis's typology.

Table 1. Characteristics of exurban households (from Nelson and Sanchez, 1997; data source: American Housing Survey, 1984 – 85).

Suburban fringe or exurban clusters	Income	Persons per household	House size	Lot size	Commuting time
Dual-wage-earner or professional	above average	average	above average	above average	above average
Nonprofessional or single worker	average	average	below average	below average	above average

As Davis's topology suggests, exurban residents appear to place a high value on rural living. Lot size, housing price, living in a rural area, and having a better place to raise a family were the leading reasons given for choosing exurban residential settings (Davis et al, 1994). In survey responses, many (28%) also added comments about specific rural amenities they enjoy in their location. Small town and suburban movers gave a mix of reasons for moving, mostly related to housing and family, such as owning a home and living close to friends and relatives. Job-related reasons and public services were not as important to exurban movers. Studying people who move to settings outside of cities, Varady (1990) found that 'overall appearance of area', 'large lots', and 'suburban setting', were more important to suburban movers than to those who had moved to the city. Exurban residents of the Kentucky Bluegrass rated the most important aspects of their community to be (in order): (1) peaceful, quiet surroundings; (2) safety and low crime; (3) open space; (4) close to nature; and (5) cost of housing, low taxes, and friendly neighbors (three-way tie) (Patel, 1980). Geoghegan (2002) found that amounts and types of open space in the neighborhood affected the price of residential properties.

Paquette and Domon (2001a; 2001b) found a strong relationship between landscape attributes and sociodemographic profiles of residents in their study of exurban Montreal. Cluster analysis revealed four types of lots occupied by survey respondents: 'woodlot – closed view', 'upper hillside – panoramic view', 'agricultural lowland – limited view', and 'lower hillside – potential view'. 'Woodlot – closed view' lots were associated with migrants, individuals between 45 and 64 years of age, and second-home residents possibly indicative of a preference among these groups for privacy, a natural setting, and peacefulness. 'Upper hillsides – panoramic view' tended to be occupied by professionals and a greater proportion of urban-to-rural migrants than the whole population, possibly suggesting a link between social position and scenic lots. Part-time farmers and farm workers were significantly correlated with 'agricultural lowlands – limited views', but neither local nor migrant populations predominated. 'Lower hillside – potential view' lots tended to attract fewer newcomers and more full-time residents (Paquette and Domon, 2001a).

This summary of the literature provides an overview of the types of relationships we might expect to find between the location preferences of residential agents and the socioeconomic and demographic characteristics. At a minimum it suggests that residential agents who choose to live in exurban locations have varying preferences.

Methods

Data sources

The data used in this study were derived from household surveys conducted in the Detroit metropolitan area (figure 1) during the spring and summer of 2001. Issues related to quality of community life, including residential mobility, were addressed in the survey. The survey was conducted through face-to-face interviews and mail questionnaires. Interviewees were selected using a probability sample of households and

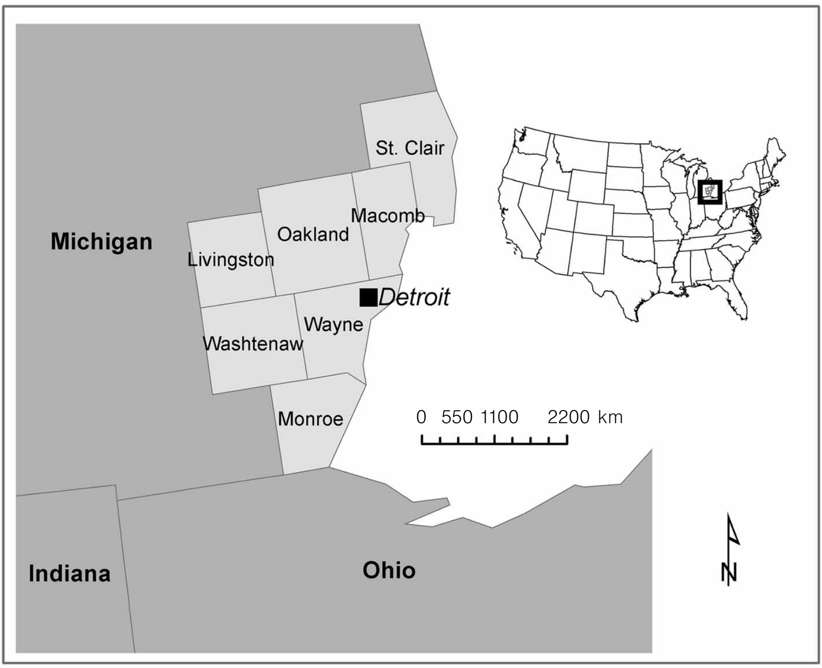

Figure 1. Map of the Detroit metropolitan area, including the locations of the seven counties included in this study.

adults 18 years and older in the City of Detroit and Wayne, Oakland, and Macomb Counties. Mail questionnaires were sent to a probability sample of adults living in Detroit, the same three counties plus four fringe counties (Livingston, Monroe, St. Clair, and Washtenaw). The sample of adults was drawn from a list frame that excluded addresses selected in the household sample. Of these, 315 people granted interviews, representing a response rate of 59.8%. The mail sample yielded 4392 questionnaires for a response rate of 56.7%. The analysis reported in this paper is based on a subset of the total survey response. We selected only those respondents who identified themselves as purchasing single-family detached houses during the past ten years and who lived in the exurban parts of the metropolitan area. We defined exurban locations to include towns with populations less than 15 000 and all unincorporated areas (that is, townships). The resulting subset consisted of 592 respondents.

In part of the survey, respondents were asked about the importance of factors influencing their decision to move to their current neighborhoods. Factor selections were guided by the literature, and by the interests of sponsoring governmental agencies. A four-point importance scale ranging from 'very important' to 'not at all important' was used for the following factors:

close to work,
good schools,
housing costs and good value,
convenient to shopping and schools,
lots of recreational opportunities,
attractive appearance of neighborhood,
community size,
people similar to me,
appearance and layout of the dwelling,
familiar with area,
close to natural areas (woods, ponds, streams, etc),
openness and spaciousness of area,
close to family and friends.

For use in the analyses presented here, the importance values were recoded as ordinal codes as: very important = 1, somewhat important = 2, not very important = 3, and not at all important = 4.

In addition to the residential preference responses, we had information on the following socioeconomic characteristics of all respondents: age (coded ordinally in the following groups: 18–29 years = 1, 30–39 years = 2, 40–49 years = 3, 50–65 years = 4, $\geqslant$ 65 years = 5); marital status (classed as married and not married); race (in the following five categories: White = 1, Black = 2, Hispanic = 3, Asian = 4, Other and American Indian = 5); education (coded ordinally in the following groups: less than high school = 1, high school graduate = 2, some college = 3, college graduate = 4, graduate or professional school = 5); whether the household includes children under 18 years (without children = 0, with children = 1); and income (coded ordinally in the following groups: less than \$20 000 = 1, \$20 000–\$29 999 = 2, \$30 000–\$39 999 = 3, \$40 000–\$49 999 = 4, \$50 000–\$74 999 = 5, \$75 000–\$99 999 = 6, \$100 000–\$125 000 = 7, >\$125 000 = 8).

Statistical analysis

We analyzed the scores of the thirteen-part resident preference question using factor analysis, with varimax rotation and a Kaiser normalization, to reduce the dimensionality of the dataset (SPSS version 11.5, SPSS, Inc., Chicago, IL). Factor analysis can be used for ordinal data if the assignment of ordinal categories to the data does not seriously distort the underlying metric scaling (Kim and Mueller, 1978).

The basic assumption of the technique is that it may be possible to explain the correlation between two or more variables in terms of some underlying factor(s) (Gorsuch, 1983). We suspected that there was a certain amount of collinearity between the preference-choice responses, and used the factor analysis to determine if there were certain factors that grouped together in an interpretable manner, and to explain the overall structure of the stated preferences better. Only factors with eigenvalues over 1 were interpreted and used for subsequent analysis. Loadings of preferences were used to interpret the meaning of the factors and the factor scores of each respondent were outputted for subsequent analysis. The factor scores were standardized to mean $= 0$ and standard deviation $= 1$.

To determine if and how survey respondents grouped by preferences, the resulting factor scores were further analyzed using a cluster analysis. A K-means cluster algorithm was used because of the large number of cases (Aldenderfer and Blashfield, 1984; Kaufmann and Rousseeuw, 1990). The K-means clustering technique calculates the n-dimensional Euclidian distance between each case and a predetermined number of cluster centroids, iteratively assigning membership, and recalculating the new cluster centroids. This process repeats until the variance between the cluster centroids is maximized (Pollard, 1981). To determine the number of clusters, we ran the analysis with the number of clusters set to all values between 5 and 20. To determine the optimal number of clusters, a scree plot (SPSS, version 11.5), which showed the mean distance of all respondents to the mean of their assigned clusters plotted against an increasing number of clusters, was examined, along with the numbers of respondents assigned to the clusters. A scree plot will typically have an inflection point at which the number of clusters changes significantly (Cattell, 1966). In our study, seven clusters were specified. Optimality of classes may be also assessed by comparison of the within-class variation as compared with the between-class variation (Milligan and Cooper, 1985). In the seven clusters derived, within-class variation was low compared with the between-class variation, further indicating an appropriate number of groups. Means and standard deviations of factor scores and the importances of each of the thirteen preference variables were generated for each cluster.

To begin our study of the relationships between socioeconomic factors and preferences, we analyzed the differences in the means of factor scores based on respondents with different socioeconomic characteristics by using an individual-samples t-test. Ordinal variables were recoded into binary groups of high and low, and t-tests compared scores of all factors for each pair of groups on each socioeconomic variable. The results suggested the kinds of relationships we might expect to find in the regression-tree analysis relating factor scores to socioeconomic characteristics.

Regression-tree models provide an alternative to linear regression models (Breiman et al, 1984). Regression-tree models are fit by successively splitting the data to form homogeneous subsets on the dependent variable, resulting in a hierarchical tree of decision rules based on the values and categories in the independent variables that can be used for prediction. The application of such methods in the delineation of heterogeneous agents in agent-based modeling is potentially very promising (Sengupta and Bennett, 2003). We applied this technique by building four regression-tree models. The dependent variables for the four models were the four preference factors. The independent variables were the six socioeconomic variables: age, education level, marital status, race, whether or not the household has children under 18 years old, and income. The regression trees were specified to include no fewer than thirty cases per node and 'pruned' using cost-complexity pruning to a goal of ten terminal nodes or 'leaves'. These specifications were made to approximate the size and distribution of the groups generated by the *K*-means cluster analysis, for comparison purposes. The graphic trees are presented to illustrate how the preferences related to socioeconomic factors. The fit of the trees to the preference levels was measured by using the residual mean deviance, which represents the proportion of the variation in preferences that was left unexplained by the trees (S-Plus 2000, Mathsoft Inc., Seattle, WA).

The terminal nodes on the tree, that is, those with no splits below them, represent groups of respondents that had similar socioeconomic characteristics and identical predicted preferences. Analysis of variance was used to evaluate differences between factor scores for respondents assigned to different terminal nodes and to compare these with differences in the factor scores between the preference clusters. This comparison was used to draw conclusions about the degree to which heterogeneity of agent preferences (grouped directly in the cluster analysis) were independent of socioeconomic factors (which were used to group preferences in the regression-tree analysis).

Results

Identification of preference groups

Factor analysis identified four preference factors with eigenvalue scores over 1, indicating significant explanatory power (table 2). The total variance accounted for by the four extracted factors was 52.2%, with each factor explaining over 10% of the total variance. None of the remaining factors accounted for more than 10% of the variance. Most of the individual questions loaded distinctly on one of the four factors and can be seen to group roughly into types of preferences. The preferences for 'familiar with area', 'close to family and friends', and 'people similar to me' loaded strongly on factor 1, which we labeled 'social comfort'.

Preferences for 'openness and spaciousness', 'close to natural areas', and 'lots of recreational opportunities' loaded predominantly on factor 2, earning it the label 'openness/naturalness'. The two questions concerning neighborhood and dwelling aesthetics (that is, 'attractive appearance of neighborhood' and 'appearance and layout

Table 2. Variable loadings of preference responses on four derived factors.

Preference dimension	Factor			
	1 social comfort	2 openness/ naturalness	3 residential aesthetics	4 schools/ work
Familiar with the area	**0.81**	−0.01	−0.01	−0.03
Close to family and friends	**0.75**	0.04	−0.06	0.02
People similar to me	**0.60**	0.12	0.23	0.13
Openness and spaciousness of area	−0.003	**0.85**	−0.03	−0.07
Close to natural areas	0.06	**0.85**	−0.05	−0.14
Lots of recreational opportunities	0.19	**0.50**	0.18	0.24
Attractive appearance of neighborhood	0.09	0.11	**0.75**	0.04
Appearance and layout of the dwelling	0.13	−0.02	**0.73**	−0.09
Good schools	0.02	0.16	−0.12	**0.74**
Close to work	0.05	−0.14	0.03	**0.56**
Housing costs and good value	−0.16	−0.03	0.50	0.47
Convenient to shopping and schools	0.45	−0.18	0.38	0.48
Community size	0.47	0.35	0.20	−0.01

Note: Total variance explained: factor 1—16.11%, factor 2—12.58%, factor 3—12.77%, factor 4—10.99%. High loading values for each preference response are given in bold type.

of the dwelling') loaded strongly on factor 3, which we therefore labeled 'residential aesthetics'. Questions about 'close to work' and 'good schools' loaded strongest on factor 4 (table 2), which we labeled 'schools and work'.

Three preferences did not load particularly strongly on any one of the factors. 'Housing costs and good value' loaded partially on both factors 3 and 4 (negligibly on factors 1 and 2), while 'convenient to shopping and schools' loaded weakly on factors 1, 3, and 4. 'Community size' did not load strongly on any of the factors (table 2).

Cluster analysis resulted in seven clusters that were defined by resident-preference responses (table 3, see over). Thanks to the response scale of the original questions (1 = most important to 4 = least important), a *lower* score represents a stronger affinity to a particular characteristic.

Individuals in cluster 1 assigned greater importance to social comfort, residential aesthetics, and schools and work issues, indicated by the low mean values for these factor scores (table 3). Residents in cluster 2 assigned greater importance to openness/naturalness, residential aesthetics, and schools and work, whereas those in cluster 3 assigned greater importance to social comfort, and openness/naturalness. Those in cluster 4 assigned greater importance to schools and work, and cluster 5 put more weight on residential aesthetics and, to a lesser extent, social comfort. Cluster 6 residents found openness/naturalness, and residential aesthetics most important, and cluster 7 residents assigned greater importance to openness/naturalness, and schools and work.

Relationships between preferences and socioeconomic factors

The analysis of the scores on the four preference factors in relation to individual and household social – economic characteristics revealed several significant effects due to the presence of children under 18 years old, marital status, and age (table 4, over). The only significant effect found for factor 1 (social comfort) indicated that households with annual incomes under \$75 000 rated this factor significantly stronger than did those making more than \$75 000. There were no significant differences between groups on socioeconomic variables and factor 2 (openness/naturalness). The presence in the household of children under 18 years old resulted in a significantly higher score (that is, less importance) on factor 3 (residential aesthetics) than when children were not present. Scores on factor 4 (schools and work) were significantly related to the presence of children, marital status, and being under 40 years old (table 4). Respondents in households with children or who were married or under 40 years old placed significantly more importance on schools and work than did their opposites on each variable.

The regression-tree models for each factor, built to explore further the relationship of preference to socioeconomic factors resulted in tree diagrams that illustrate how factor scores are related to socioeconomic factors (figures 2 – 5). Each graphic can be read as a sorting tree or, alternatively, as a set of decision rules for determining the terminal node or leaf a respondent occupies. The preference values predicted for respondents with the set of socioeconomic factors required to reach a given terminal node for a given factor are listed. The tree predicts all residents with these socioeconomic characteristics have the same preference values (that is, factor scores). It should be noted that predicted scores at the terminal node are standardized to a mean of 0, and a *negative* value represents a *stronger affinity* to a characteristic.

The residual mean deviance for the regression-tree models was used as a measure of model fit and is calculated based on the difference between the predicted and actual factor scores for each respondent. The residual mean deviance for factor 1 (social comfort) was 0.87, indicating that 87% of the variation in factor-1 scores is left

Table 3. Mean and standard deviations of factor scores (top section) and individual survey responses (bottom section) for each derived cluster group. To aid in interpretation, the bottom section breaks each factor into the individual constituent preferences that load most heavily on it.

Factor	Description	Cluster 1 $N = 148$		Cluster 2 $N = 108$		Cluster 3 $N = 35$	
		mean	SD	mean	SD	mean	SD
Factor 1	Social comfort	−0.80	0.59	1.12	0.59	−1.11	0.69
Factor 2	Openness/naturalness	0.00	0.77	−0.45	0.66	−0.27	0.81
Factor 3	Residential aesthetics	−0.49	0.55	−0.34	0.56	1.65	1.26
Factor 4	Schools and work	−0.68	0.50	−0.47	0.59	1.12	0.96
Factor 1: social comfort	Familiar with area	1.67	0.61	3.09	0.62	1.52	0.66
	Close to family and friends	1.64	0.81	3.06	0.74	1.39	0.55
	People similar to me	1.78	0.64	2.59	0.83	2.03	0.90
Factor 2: openness/ naturalness	Close to natural areas	2.08	0.74	1.76	0.66	1.58	0.71
	Openness and spaciousness of area	1.91	0.69	1.56	0.61	1.59	0.64
	Lots of recreational opportunities	1.89	0.65	2.18	0.71	2.44	0.97
Factor 3: residential aesthetics	Attractive appearance of neighborhood	1.24	0.48	1.41	0.55	2.15	0.94
	Appearance and layout of dwelling	1.28	0.45	1.52	0.55	2.28	0.84
Factor 4: schools and work	Close to work	1.74	0.69	2.08	0.88	2.80	0.94
	Good schools	1.21	0.44	1.25	0.50	2.06	1.11
Unassigned	Housing costs and good value	1.16	0.39	1.11	0.31	2.54	0.89
	Convenient to shopping and schools	1.44	0.52	2.18	0.72	2.56	0.89
	Community size	1.78	0.62	2.47	0.84	1.91	0.94

Note: The scales of the factors and the individual variables differ. Factor scores (top section) are standardized to a mean of 0.00, and a *negative* value represents a stronger affinity to the characteristic. Original preference response variables (bottom section) vary from 1 to 4, and a *score closer to 1* represents a stronger affinity to the characteristic.

unexplained by the tree (figure 2, over). This value for residual mean deviance suggests a relatively poor model fit. However, associations were found for age, income, and education (in order of decreasing importance and interpreted by the order in which they entered the tree). The model indicates individuals between 20 and 30 years of age earning \$50 000 – \$100 000 with some college education, and all those between 20 and 50 years old with a household income under \$50 000 have the highest preference for factor 1 (predicted factors scores were −0.39 and −0.36, respectively).

The regression-tree model for factor 2 (openness/naturalness) showed effects with income, age, and education (figure 3, over). The residual mean deviance was 0.93, again indicating poor model fit (figure 3). The highest preference for factor 2 was found for those with an annual household income greater than \$125 000 and without a college degree (predicted score is −0.76), and for those with incomes less than \$125 000, under 65 years of age, without a college education, earning less than \$50 000 per year (predicted score is −0.70).

Table 3 (continued).

Cluster 4 $N = 67$		Cluster 5 $N = 70$		Cluster 6 $N = 76$		Cluster 7 $N = 90$	
mean	SD	mean	SD	mean	SD	mean	SD
0.43	0.79	−0.17	0.99	0.14	0.85	0.13	0.67
1.29	0.64	1.21	0.66	−0.89	0.50	−0.52	0.67
0.77	0.68	−0.64	0.72	−0.66	0.64	1.04	0.50
−0.63	0.67	1.10	0.76	1.17	0.73	−0.12	0.57
2.60	0.70	2.18	0.89	2.42	0.90	2.39	0.77
2.50	0.86	2.13	0.87	2.28	0.98	2.18	0.80
2.71	0.63	2.33	0.81	2.22	0.85	2.50	0.74
2.97	0.58	2.73	0.75	1.27	0.47	1.63	0.57
2.74	0.67	2.70	0.57	1.16	0.37	1.49	0.50
2.87	0.69	2.73	0.72	2.12	0.78	2.20	0.72
1.97	0.58	2.35	0.48	1.23	0.43	1.93	0.53
2.08	0.55	1.18	0.43	1.22	0.45	2.21	0.60
1.79	0.63	2.59	0.94	2.81	0.80	2.26	0.75
1.32	0.65	2.69	1.07	2.44	1.12	1.36	0.64
1.38	0.55	1.53	0.54	1.53	0.61	1.69	0.61
2.02	0.65	2.04	0.53	2.46	0.67	2.43	0.75
2.77	0.65	2.32	0.70	1.80	0.77	2.17	0.70

Table 4. Relationships between socioeconomic variables and factors scores.

Factor	Children under 18 years		Married		Income >$75 000		College education		Minority		Age > 40 years	
	T	sig.	*T*	sig.	*T*	sig.	*T*	sig.	*T*	sig.	*T*	sig.
Factor 1: social comfort	−0.94	0.35	−0.14	0.99	**−2.15**	**0.03**	−1.79	0.07	−1.13	0.26	0.47	0.64
Factor 2: openness/ naturalness	1.03	0.30	0.16	0.87	1.38	0.16	0.65	0.52	3.21	0.75	0.58	0.56
Factor 3: residential aesthetics	**−2.04**	**0.04**	0.72	0.47	2.21	0.28	2.62	0.09	0.27	0.79	1.26	0.20
Factor 4:	**9.16**	**0.00**	**1.97**	**0.05**	0.69	0.50	1.36	0.17	1.68	0.09	**−3.86**	**0.00**

Note: bold numbers represent statistical significance (sig.) at the 0.05 level.

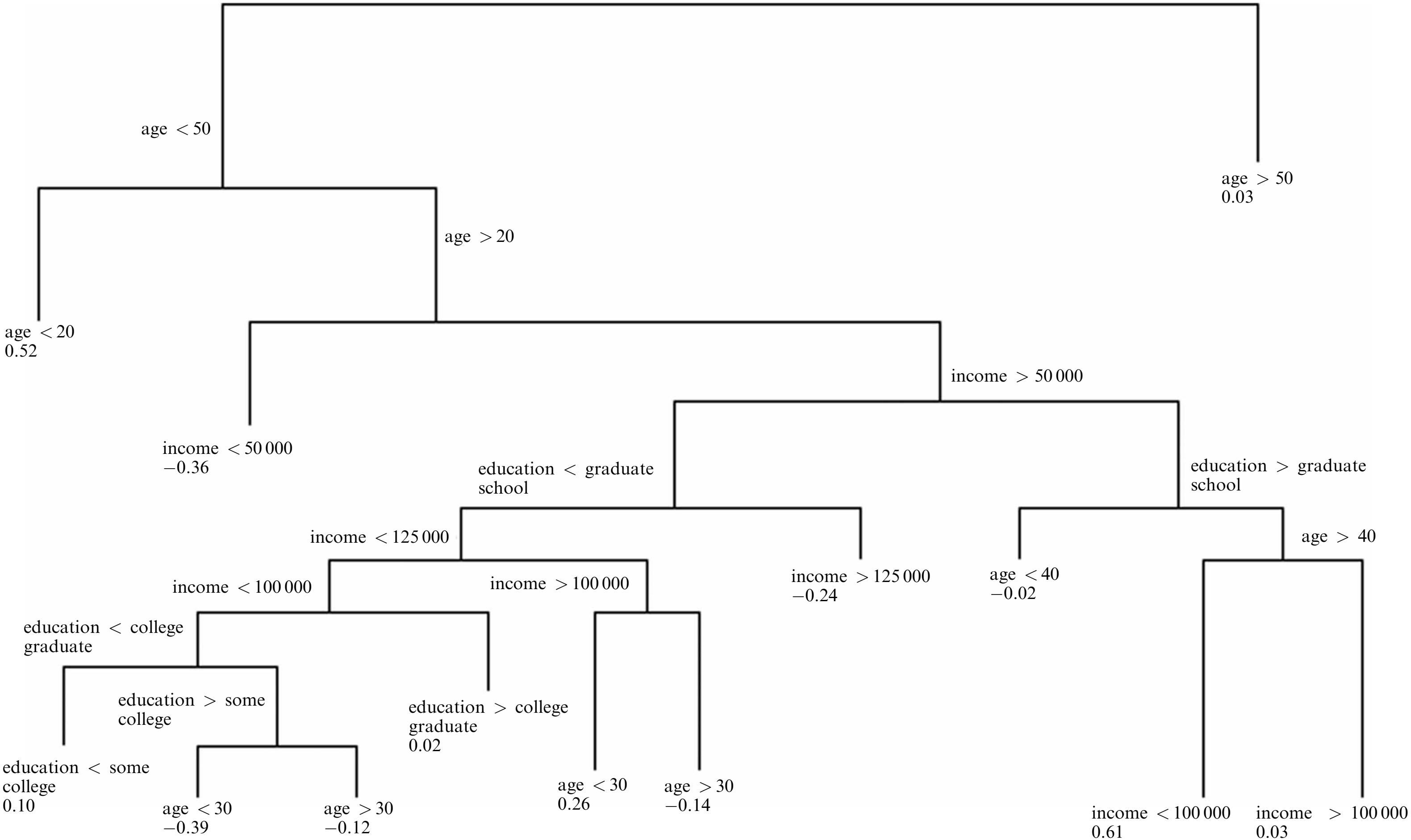

Figure 2. Regression tree: factor 1—social comfort.

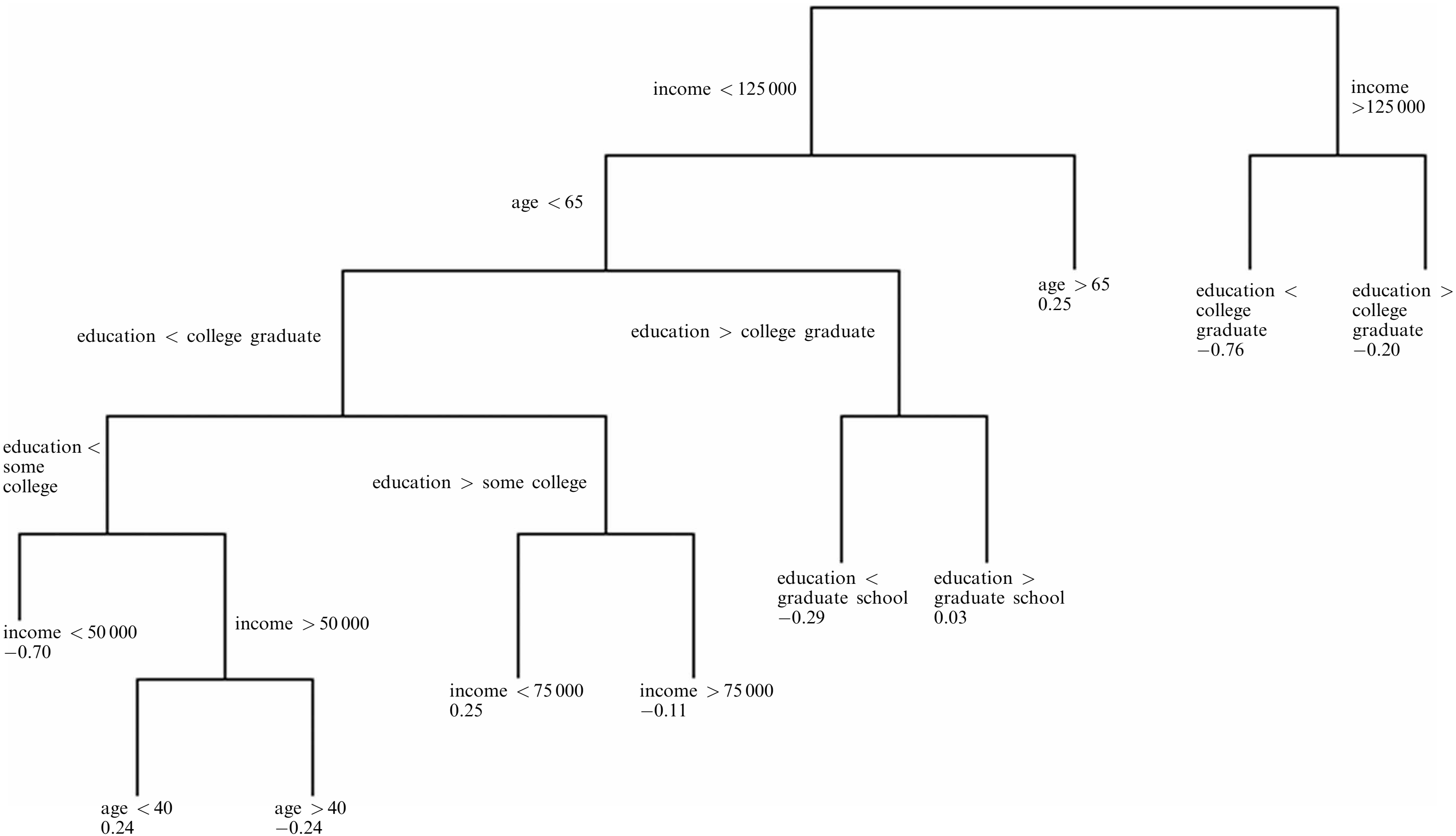

Figure 3. Regression tree: factor 2—openness/naturalness.

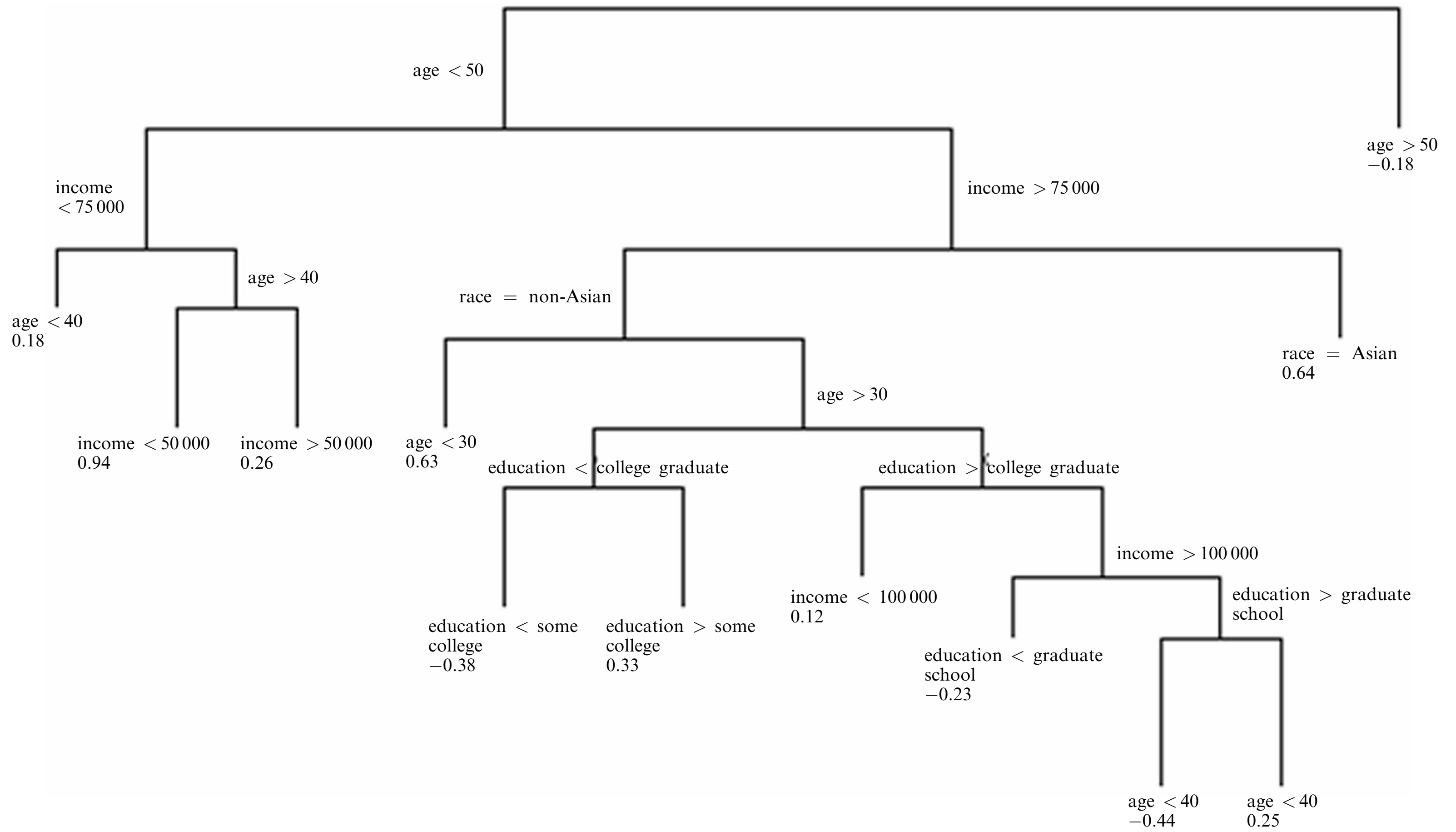

Figure 4. Regression tree: factor 3—residential aesthetics.

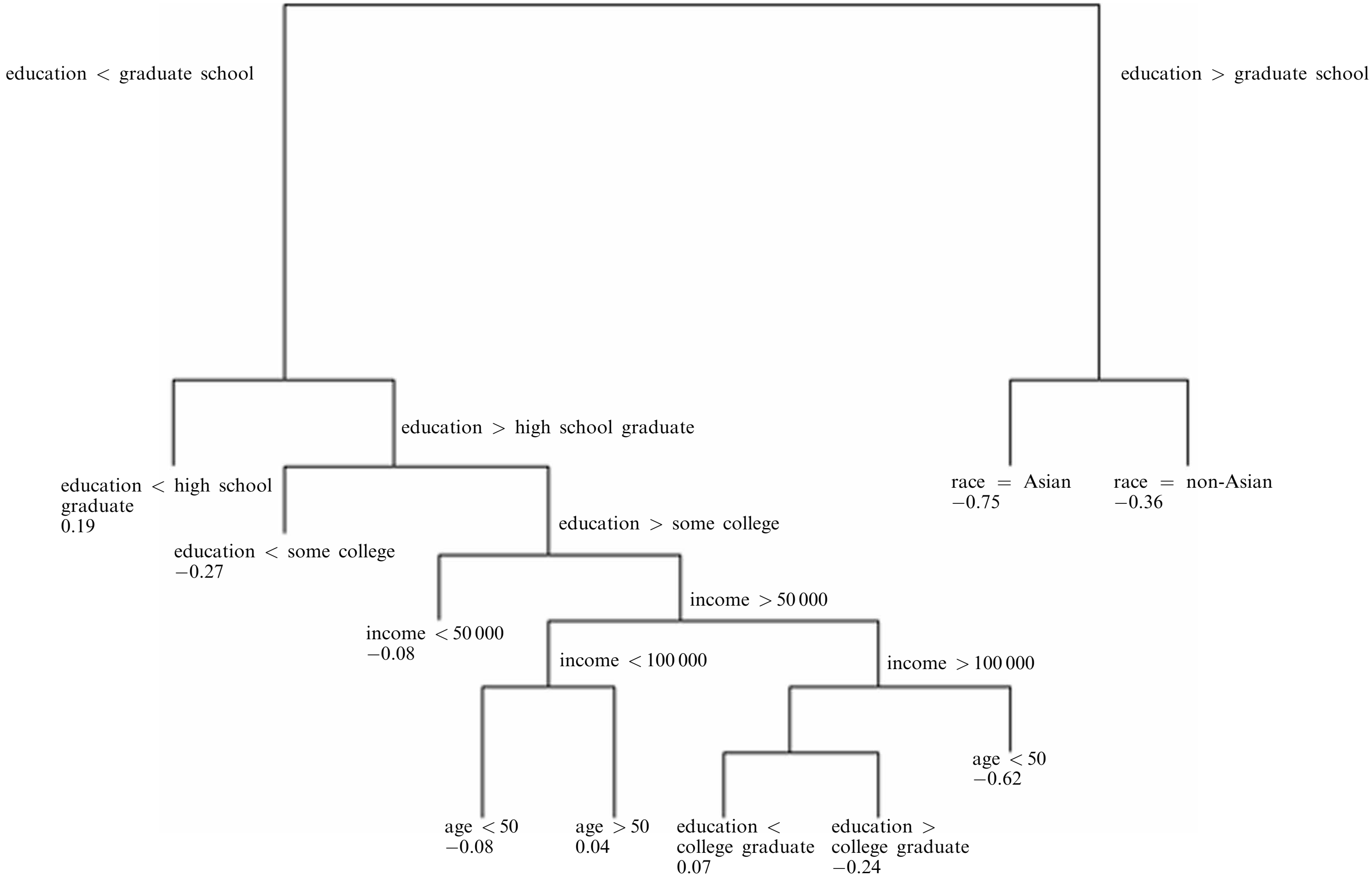

Figure 5. Regression tree: factor 4—schools and work.

The residual mean deviance for factor 3 (residential aesthetics) was 0.94, and age, income, race, and education were the associated independent variables (figure 4). The highest preference values for factor 3 were found in residents characterized as 30–50 years old, non-Asian college graduates, earning over \$100 000 a year (predicted score = −0.44), and non-Asians between 30 and 50 years old without any college education (predicted score = −0.38).

Education, income, age, and race were the variables associated with the variance in factor 4 (schools and work) (figure 5). The residual mean deviance of the model was 0.67, indicating the best fit of the four factor models. The groups that showed greatest affinity to factor 4 were Asians with a graduate school education (predicted score = −0.75), and those over 50 years old, with either some college education or a college degree, and earning more than \$100 000 per year (predicted score = −0.62).

The classification trees represent optimizations of the fit between socioeconomic characteristics and preference factors. By placing all respondents into their respective socioeconomic groups, according to the tree models, we had a basis for comparing the strict grouping of respondents by preferences (based on the cluster analysis) with a grouping based on socioeconomic characteristics (based on the tree models). The analysis of variance allowed us to test for the significance of the observed associations.

Comparison of models: preference clusters and socioeconomic groups

The results of the analyses of variance, which tested for significant differences between the factor scores of different clusters and between those of different terminal nodes on the regression trees (table 5), indicated that the preference clusters sorted residents according to preferences much better on the preference variables than did socioeconomic groups. We expected this result because the preference clusters were created by a statistical procedure that maximizes the variance between the cluster centroids on the preference factors. In fact, though there were significant differences between all the factor scores among preference clusters, there were no significant differences in any of the socioeconomic groups, defined by the regression trees.

Table 5. Analysis of variance results for clusters and regression-tree leaves.

Difference between	Factor 1: social comfort		Factor 2: openness/naturalness		Factor 3: residential aesthetics		Factor 4: schools and work	
	F	sig.	*F*	sig.	*F*	sig.	*F*	sig.
Clusters	**93.29**	**0.00**	**117.14**	**0.00**	**132.00**	**0.00**	**140.54**	**0.00**
Terminal nodes	0.89	0.56	1.54	0.13	1.36	0.19	1.59	0.11

Note: Bold numbers represent statistical significance (sig.) at the 0.05 level.

Discussion

Factor analysis of resident responses revealed four underlying types of preferences. The four types or classes of preference we found (social comfort, openness/naturalness, residential aesthetics, and schools and work) were similar to the preferences reported in previous studies of residential preference (Davis, 1993; Nelson and Sanchez, 1997; Patel, 1980; Varady, 1990). These four factors explained 52% of the variation in the resident preference responses.

The first, social comfort, was interpreted to represent the importance which residents indicated for living close to family and friends and living in areas familiar to them, possibly reflecting prior knowledge or experience. This factor also reflected a high importance placed on living near similar people.

The second factor, openness/naturalness, incorporates responses indicating the importance of openness of land and the availability of natural areas, as well as the abundance of recreational opportunities. Patel (1980) also found that these factors ranked high in exurban resident decisionmaking. Though recreational opportunities can feasibly be interpreted by residents in diverse ways (for example: more 'urban' types of recreation such as bowling alleys and pools, versus more 'natural' types such as parks and hunting areas), its association with 'openness' and 'natural areas' hints that it may have been interpreted more frequently to mean parks and open- or green-space opportunities than developed recreational facilities.

The residential aesthetics preference factor incorporated two aspects of aesthetics that homebuyers indicated were important: those of the neighborhood, and those of the dwelling itself. Varady (1980) reported similar factors ('overall appearance of area', 'suburban setting'), as did Patel (1980; 'peaceful, quiet surroundings'), as important to nonurban residents.

The schools and work preference factor straightforwardly represented the two strongest variables in residential preference 'close to work' and availability of 'good schools'. These two variables have been found to be important in many studies of residential preference, because of the prevalence of the two-wage-earner households and the desire to be closer to places of employment and the strong influence of child raising and education on resident behavior.

The responses that did not load dominantly on any of the factors were meaningful as well as those that did. All three preferences that were not clearly related to a single factor represented issues that affect all residents or that do not identify the predominant differences among exurban residents. 'Housing costs and good value', 'convenience to shopping and schools', and 'community size' may have been either a concern to all residents (costs) or generally did not distinguish among exurban locations (size of community). 'Housing costs and good value' did load moderately on factor 3 (residential aesthetics) and factor 4 (schools and work), suggesting that costs and value may be somewhat related to neighborhood and dwelling aesthetics and location of schools and work.

Given the above characterization of factors, we evaluated and compared two approaches to defining the heterogeneity of agents for representation within our agent-based models. The first approach involved simply clustering the agent preferences, assuming that preferences were distributed among agents in some way that could not be predicted with other data. Conversely, the second approach assumed that preferences were predictable, and that agents could be grouped on the basis of measures of socioeconomic status and demographics. We compared the degree to which preferences on each of the four factors were captured by each method.

The preference groups derived from the cluster analysis revealed differences among residents in general residential preferences. In an agent-based model, these different clusters could represent different agent types. Each agent could be assigned to a preference group in proportion to those found in the population (that is, based on the survey) and given a value for each of the preferences, to be used in a utility function to determine where these agents will choose to live, based on the mean and standard deviation for the importance of each major factor assigned to each group.

The first cluster represented residents who indicated that openness/naturalness was not as important in their residential choice as were the other three factors. Based on the individual questions they ranked highest (that is, lower score), good schools and housing cost were more important to these residents. Cluster 1 can be interpreted to include those residents for whom good schools, value for money, and social comfort were most important and who did not care strongly for openness/naturalness in their neighborhood. Cluster 2 was comprised of residents who indicated that openness/naturalness, schools and work, and residential aesthetics were most important, with very little importance given to social comfort. Costs and value were also very important to this group. Cluster 3 included residents who stated social comfort, specifically being close to family and friends, and openness/naturalness were of greatest importance (though recreational opportunities were not very important to this group). Those in cluster 4 generally indicated most concern for schools and work, and specifically for good schools and high value for cost, whereas residents in cluster 5 indicated that social comfort and dwelling and residential aesthetics were most important. Residents in cluster 6 rated openness and residential and dwelling aesthetics highly, with little importance given to schools and work, indicating a group primarily concerned with the aesthetic pleasure of their residences and the areas surrounding it. Alternatively, those in cluster 7 cited openness as important, but also considered schools and work important (particularly schools), and care relatively little for aesthetics or preferences related to social comfort.

In order to evaluate the predictability of preferences based on socioeconomic and demographic variables, we first examined the univariate relationships between socioeconomic variables and preference factor scores. We found several significant effects that were consistent with previous studies of location preferences (Feldman, 1996; Greenwood, 1985; Michelson, 1977; Sanchez and Dawkins, 2001). The importance of being near work and schools loomed large for married respondents, those with children, and those under 40 years old. Lower income respondents (that is, households with incomes less than $75 000) placed more importance on social comfort factors compared with their higher income counterparts, probably because of both the benefit one gains being near family and friends and the more common dislocation from family and friends, sometimes long distance, that is required of higher income professionals. Race and education had no significant effect on the preferences. These significant relationships confirm that life stage plays at least a part in determining how people make location decisions.

The use of the regression trees allowed us to examine the combined effects and significance of the six socioeconomic variables on each of the four preference factors. Each of the regression trees could be used to populate our agent-based model in the following way. Given information about the population of the region, we can generate computer agents that have socioeconomic status and demographic characteristics in the same proportions as the population. Then, the importance of each decision factor to each agent would be determined by calculating, on the basis of agent characteristics, to which terminal node the agent belongs. A mean and standard deviation of importance values can then be predicted for all the terminal nodes of the regression tree and used to simulate the particular preferences of an individual agent. This approach relies, of course, on the inherent predictability of the preferences on the basis of the agent characteristics.

The relationships observed in the univariate and regression-tree analyses suggest that there is some explanation of the preference factors by the six socioeconomic variables. These relationships are consistent with both intuition and previous research. However, our analysis of the fit of these models (residual mean deviance), and the comparison of the analysis of variance between the terminal nodes of the tree and the clusters indicated that the regression tree did a poor job of predicting preference values and of subsequently grouping agents. The implications of this finding are that the socioeconomic variables were incomplete predictors of stated residential preferences in the population, and that there is much more to residential preferences than can be predicted by socioeconomic and demographic variables. The clusters based on preferences were, therefore, more likely to capture the heterogeneity of agent behaviors than were the groups based on the six socioeconomic and demographic variables we used.

Previous attempts to explain location preferences and behaviors in terms of socio-demographic variables have found some very interesting and significant relationships. However, our findings suggest that these socioeconomic and demographic characteristics explain relatively little of variation in stated preferences between groups. The next question, then, may be: what does account for the variance in preference groups? Survey research on transportation mode and residential location selection has shown, for example, that attitudinal and lifestyle factors (for example, reading habits, the use of leisure time, and participation in a variety of outdoor activities and sports, entertainment and events, and hobbies) were significant predictors (Kitamura et al, 1994). Our survey did not ask questions of this sort, but our results suggest that it might be worth the effort to explore some of these other factors. However, if one needs to account for lifestyles in order to predict preferences adequately, it may make the most sense simply to use groupings of agents based directly on preferences.

It is important to consider that the four factors evaluated here explained only 52% of the variation in the resident preference responses. The emphasis of the analysis is necessarily on only the four factors we have extracted. Further, because we did not analyze the remaining portion of the variation, some important relationships between preference and socioeconomic and demographic variables may be missing.

It must also be pointed out that the basis of this study was an analysis of resident's *stated* preferences. Stated preferences, what people *say* they will consider or have considered when making a decision, can be different than revealed preferences, which is based on what people have *actually decided* (Ben-Akiva and Lerman, 1985). The differences between stated and revealed preference can vary considerably, especially when individuals are asked to speculate on future decisions. Reasons for differences between stated and revealed preferences can include changes in opinions, context, and experience of the individual from the time the decision was made; variability in respondent's memory of the reasons for the decision (we choose a ten-year decision window; that is, purchase of a home less than ten years ago, to help control for this), as well as the uncontrollable differences in responses when someone is asked to state an opinion (state preferences) versus when they actually have to make a decision that involves some cost or risk to them (revealed preferences). Though using methods that combine stated and revealed preferences may prove to be the most effective at determining resident preferences (Adamowicz et al, 1994), revealed preference data can be more difficult and expensive to collect and very difficult to interpret (Koppelman and Pas, 1980).

This survey captures the preferences of a large sample of residents across a geographic area, in the seven counties in southeastern Michigan. Though such a large and rich dataset facilitated the analysis presented here, useful data may be derived from smaller and more cost-effective datasets collected by surveys. Though statistical sampling rules would dictate how large a sample would have to be to provide statistical reliability, more focused survey data would provide information to derive resident preference clusters. The clusters derived from more focused surveys (that is, few respondents) would be more sensitive to the characteristics of the residents in the study area than a larger survey.

Alternatively, third-party information can be used to supplement results from approaches such as the one we present here. Surveys of real-estate agents can provide useful information on the decisionmaking processes of prospective residents. However, it is likely that the information of real-estate agents about resident preferences would differ from the actual stated preferences of the residents themselves. How they might differ is difficult to determine because it relies on professional experience of the agents, their position in the real-estate market, and variability of the agent's personal speculation. Given these caveats, third-party information should not be used as a substitute for surveys of resident preference, but can provide useful supporting information on resident preference patterns.

Conclusion

Powerful computational tools are becoming available to represent the behavior of complex systems. Agent-based modeling, in particular, facilitates an examination of the system-level outcomes of the heterogeneous actions of a set of heterogeneous agents. These new tools create new demands for empirical studies and data. In the context of application to agent-based modeling, our findings suggest that distinct preference groups do exist in resident populations, but are not well characterized on the basis of simple socioeconomic and demographic variables. If we want to model preferences well, we probably need much richer data on the agents' lifestyles and behaviors. Further, given the heterogeneous nature of preferences and a relatively limited number of clusters observed in the survey respondents, agent-based models simulating resident behavior should reflect this diversity in the population and incorporate distinct agent classes of empirically derived preference distributions.

This study showed that the preferences of exurban single-family home residents in the Detroit metropolitan area were heterogeneous and could be grouped into seven distinct clusters representing unique preference groups in resident choice. We also found that preferences themselves could be grouped into four general factors, related to social comfort, openness/naturalness, residential aesthetics, schools and work, and representing major topics considered by exurban residents in choosing a residence. Further issues related to cost and value, access to services, and community size were observed across all preference groups and did not distinguish among agent groups. By analyzing the relationships between the empirically derived preference clusters, and six socioeconomic and demographic variables, we found some significant univariate relationships between the two. However, we found that, overall, these variables were incomplete predictors of residential preference. We conclude that the delineation of agents might be most profitably done on the basis of agent types with separate distributions of preferences.

The methods we used to derive the preference groups relied on a large survey dataset covering several countries. Such datasets are obviously expensive, time-consuming, and relatively infrequent. The insufficiency of the six socioeconomic and demographic variables we used to explain preferences adequately suggests the need to explore additional factors. Lifestyle factors might be worth exploring in this context (Kitamura et al, 1994). Additionally, alternative approaches, such as experiments using role-playing games (Barreteau, 2003; Barreteau et al, 2001), may provide more directed information specific to the needs of the modeling effort than analysis of very large datasets collected through very general large-scale surveys. The advantages in using well-designed smaller scale experiments focused on preference and choice include: the need for fewer respondents to achieve suitable statistical power, a much more interpretable dataset, and the ability to design surveys which answer specific questions asked in the modeling effort.

Acknowledgements. We gratefully acknowledge support in the form of grants from the National Science Foundation Biocomplexity in the Environment Program (BCS-0119804) and the Office of the Vice President for Research at the University of Michigan.

References

Aldenderfer M S, Blashfield R K, 1984 *Cluster Analysis* (Sage, Newbury Park, CA)

Adamowicz W, Louviere J, Williams M, 1994, "Combining revealed and stated preference methods for valuing environmental amenities" *Journal of Environmental Economics and Management* **26** 271 – 292

Anas A, Arnott R, Small K, 1998, "Urban spatial structure" *Journal of Economic Literature* **36** 1426 – 1464

Barreteau O, 2003, "The joint use of role-playing games and models regarding negotiation processes: characterization of associations" *Journal of Artificial Societies and Social Simulation* **6**, http://jasss.soc.surrey.ac.uk/6/2/3.html

Barreteau O, Bousquet F, Attonaty J, 2001, "Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems" *Journal of Artificial Societies and Social Simulation* **4**, http://www.soc.surrey.ac.uk/JASSS/4/2/5.html

Batty M, Longley P, 1994 *Fractal Cities* (Academic, San Diego, CA)

Batty M, Xie Y, 1997, "Possible urban automata" *Environment and Planning B: Planning and Design* **24** 275 – 292

Bell K P, Irwin E G, 2002, "Spatially explicit micro-level modeling of land use change at the rural – urban interface" *Agricultural Economics* **27** 217 – 232

Ben-Akiva M, Lerman S R, 1985 *Discrete Choice Analysis* (MIT Press, Cambridge, MA)

Breiman L, Friedman J H, Olshen R A, Stone C J, 1984 *Classification and Regression Trees* (Wadsworth, Belmont, CA)

Brown D G, Page S E, Riolo R, Rand W, 2002, "Modeling the effects of greenbelts at urban rural fringe", in *Proceedings of IEMSS 2002. Integrated Assessment and Decision Support. Lugano, Switzerland, Volume 2* (iEMSs, Manno) pp 190 – 195

Brown D G, Page S E, Riolo R, Rand W, 2004, "Agent-based and analytical modeling to evaluate the effectiveness of greenbelts" *Environmental Modeling and Software* **19** 1097 – 1109

Brown D L, Fuguitt G V, Heaton T B, Waseem S, 1997, "Continuities in size of place preferences in the United States, 1972 – 1992" *Rural Sociology* **62** 408 – 428

Cattell R B (Ed.), 1966 *Handbook of Multivariate Experimental Psychology* (Rand McNally, Chicago, IL)

Clarke K C, Hoppen S, Gaydos L, 1997, "A self-modifying cellular automation model of historical urbanization in the San Francisco Bay area" *Environment and Planning B: Planning and Design* **24** 247 – 262

Davis J S, 1993, "The commuting experience of exurban home buyers" *Urban Geography* **14**(1) 7 – 29

Davis J S, Nelson A C, Dueker K J, 1994, "The new 'burbs': The exurbs and their implications for planning policy" *Journal of the American Planning Association* **60** 45

Feldman R M, 1996, "Constancy and change in attachments to types of settlements" *Environment and Behavior* **28** 419 – 445

Geoghegan J, 2002, "The value of open spaces in residential land use" *Land Use Policy* **19** 91 – 98

Gorsuch R L, 1983 *Factor Analysis* 2nd edition (Lawrence Erlbaum Associates, Hillside, NJ)

Greenwood M J, 1985, "Human migration – theory, models, empirical studies" *Journal of Regional Science* **25** 521 – 544

Irwin E G, 2002, "The effects of open space on residential property values" *Land Economics* **78** 465 – 481

Kaufmann L, Rousseeuw P J, 1990 *Finding Groups in Data* (John Wiley, New York)

Kim J-O, Mueller C W, 1978 *Factor Analysis: Statistical Methods and Practical Issues* (Sage, Beverly Hills, CA)

Kitamura R, Laidet L, Mokhtarian P, Buckinger C, Gianelli F, 1994, "Land use and travel behaviour", Report UCD-ITS-RR-94-27, Institute of Transportation Studies, University of California, Davis, CA

Koppelman F S, Pas E I, 1980, "Travel-choice behavior: models of perceptions, feelings, preference, and choice" *Transportation Research Record* number 765, 26 – 33

Landis J D, 1994, "The California Urban Futures Model: a new generation of metropolitan simulation models" *Environment and Planning B: Planning and Design* **21** 399 – 420

Makse H A, de Andrade J S, Batty M, Shlomo H, Stanley H E, 1998, "Modeling urban growth patterns with correlated percolation" *Physical Review E* **58** 7054 – 7062

Malizia E E, Exline S, 2000, "Consumer preferences for residential development alternatives", WP 2000-02, Center for Urban and Regional Studies, University of North Carolina, Chapel Hill, NC

Michelson W, 1977 *Environmental Choice, Human Behavior, and Residential Satisfaction* (Oxford University Press, New York)
Milligan G W, Cooper M C, 1985, "An examination of procedures for determining the number of clusters in a data set" *Psychometrika* **50** 159 – 179
Myers D, Gearin E, 2001, "Current preferences and future demand for denser residential environments" *Housing Policy Debate* **12** 633 – 659
Nelson A C, Dueker K J, 1990, "The exurbanization of America and its planning policy implications" *Journal of Planning Education and Research* **9**(2) 91 – 100
Nelson A C, Sanchez T W, 1997, "Exurban and suburban residents: a departure from traditional location theory?" *Journal of Housing Research* **8** 249 – 276
Otter H S, van der Veen A, Vriend H J, 2001, "Abloom: location behavior, spatial patterns, and agent-based modeling" *Journal of Artificial Societies and Social Simulation* **4**, http://www.soc.surrey.ac.uk/JASSS/4/4/2.html
Page S, 1999, "On the emergency of cities" *Journal of Urban Economics* **45** 184 – 208
Paquette S, Domon G, 2001a, "Trends in rural landscape development and sociodemographic recomposition in southern Quebec" *Landscape and Urban Planning* **55** 215 – 238
Paquette S, Domon G, 2001b, "Rural domestic landscape changes: a survey of the residential practices of local and migrant populations" *Landscape Research* **26** 367 – 395
Parker D C, Meretsky V, 2004, "Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics" *Agriculture, Ecosystems, and Environment* **101** 233 – 250
Parker D C, Manson S M, Janssen M A, Hoffman M J, Deadman P, 2003, "Multi-agent systems for the simulation of land-use and land-cover change: a review" *Annals of the Association of American Geographers* **93** 314 – 337
Patel D I, 1980 *Exurbs: Residential Development in the Countryside* (University Press of America, Washington, DC)
Pollard D, 1981, "Strong consistency of *k*-means clustering" *Annals of Statistics* **9** 135 – 140
Rand W, Zellner M, Page S E, Riolo R, Brown D G, Fernandez L E, 2002, "The complex interaction of agents and environments: an example in urban sprawl", in *Proceedings of Agent 2002, Chicago, IL* (Argonne National Laboratory, Argonne, IL) pp 149 – 161, http://www.agent2003.anl.gov/proceedings/2002.pdf
Rand W, Brown D G, Page S E, Riolo R, Fernandez L E, Zellner M, 2003, "Statistical validation of spatial patterns in agent-based models", in *Proceedings of Agent Based Simulation 4, Montpellier, France*
Sanchez T W, Dawkins C J, 2001, "Distinguishing city and suburban movers: evidence from the American housing survey" *Housing Policy Debate* **12** 607 – 631
Sengupta R, Bennett D A, 2003, "Agent-based modeling environment for spatial decision support" *International Journal of Geographical Information Science* **17**(2) 157 – 180
Spain D, 1989, "Why higher-income households move to central cities" *Journal of Urban Affairs* **11** 283 – 299
Talen E, 2001, "Traditional urbanism meets residential affluence: an analysis of the variability of suburban preference" *Journal of the American Planning Association* **67** 199 – 216
Varady D P, 1990, "Influences on the city – suburban choice: a study of Cincinnati homebuyers" *Journal of the American Planning Association* **56** 22 – 40
Wheaton W C, 1977, "Income and urban residence: an analysis of consumer demand for location" *The American Economic Review* **67** 620 – 631