J Geograph Syst (2007) 9:77–101
DOI 10.1007/s10109-006-0030-0

ORIGINAL ARTICLE

Latent lifestyle preferences and household location decisions

Joan L. Walker · Jieping Li

Published online: 26 September 2006

Abstract Lifestyle, indicating preferences towards a particular way of living, is a key driver of the decision of where to live. We employ latent class choice models to represent this behavior, where the latent classes are the lifestyles and the choice model is the choice of residential location. Thus, we simultaneously estimate lifestyle groups and how lifestyle impacts location decisions. Empirical results indicate three latent lifestyle segments: suburban dwellers, urban dwellers, and transit-riders. The suggested lifestyle segments have intriguing policy implications. Lifecycle characteristics are used to predict lifestyle preferences, although there remain significant aspects that cannot be explained by observable variables.

Keywords Lifestyle · Residential location · Latent class choice models · Mixture models · Error components · Neighborhood preferences

1 Introduction

The Oxford English Dictionary defines lifestyle as ''of or relating to a particular way of living''; it credits the term to the Austrian Psychologist Alfred Adler who first used it in 1929 to denote a person's basic character. Since then, the term has entered our common vocabulary and has been used in numerous fields including psychology, sociology, anthropology, health, politics, and, our

J. L. Walker (✉) · J. Li
Department of Geography and Environment, Boston University,
675 Commonwealth Ave, Boston, MA 02215, USA
e-mail: joanw@bu.edu

J. Li
e-mail: jpli@bu.edu

topic, transportation and geography. The term is amorphous as the description varies based on context. In this paper we take a spatial- and transport-centric focus. One of the earlier definitions of lifestyle in this area was introduced by Salomon and Ben-Akiva (1983): ‘‘A pattern of behavior which conforms to the individual's orientation toward the three major roles of: household member, a worker, and a consumer of leisure, and which conforms to the constrained resources available''. Aeroe (2001) in an investigation of household preferences described lifestyle as the ‘‘deep-rooted and embedded, prevalent attitudes towards different types of residential areas.'' In this context of residential location choices, such deep-rooted lifestyle differences lead to differences in considerations, criterion, and preferences for location. We exploit this fact and employ the framework of latent class choice models to infer both lifestyle groups (the latent variable) and how lifestyle preferences impact residential location decisions (the choice model). The latent class choice model allows us to estimate both phenomena simultaneously from observed household location decisions. Latent class choice models have been applied in various literatures. The method is particularly popular in the marketing research arena (see, for example, Grover and Srinivasan 1987; Kamakura and Russell 1989; Dillon et al. 1993; Swait 1994; Louviere et al. 2000, Chap. 10; Swait and Sweeney 2000) and has now permeated other fields including economics, transport, and geography (see, for example, Gopinath 1995; Boxall and Adamowicz 2002; Greene and Hensher 2003; Scarpa and Thiene 2005; Milon and Scrogin 2006). As far as we are aware, this is the first application using latent class choice modeling for residential location. It is also the first of which we are aware that explicitly determines the lifestyle groups and resulting behavior simultaneously rather than in a two stage process.

A note on terminology and contribution: In this work, we focus on lifestyle as reflected by preferences in the built environment. In the literature this is sometimes referred to as ‘attitudes' (for example towards residential settings) where ‘lifestyle' reflects a broader view of life orientation. For example, Bagley and Mokhtarian (1999) defined eleven ‘lifestyle' dimensions (culture-lover, altruist, nest-builder, relaxer, traveler, adventurer, fun-seeker, homebody, outdoor enthusiast, athlete, and hobbyist); or Hojrup (2003) defined ‘life-modes' of self-employed, wage earner, and career life. Such broad lifestyles are certainly relevant to our analysis, because housing and residential choices are one mechanism through which one attempts to realize lifestyle preferences (Aeroe 2001). Our objective is not to model lifestyle per se, but to segment households such that those with similar lifestyles as reflected by preferences in residential location choices are grouped together. Furthermore, the contribution of our paper is empirical (namely, using a model structure that has not been used before in this area), and we do not aim to contribute to behavioral theory. We use our proposed model structure to infer the best lifestyle measure that we can from the data that we have. With appropriate data, the method can be expanded to incorporate the other definitions of lifestyle and attitudes.

The paper is organized as follows. Section 2 discusses the background, including motivation and literature review. Section 3 provides an introduction to latent class choice models. Section 4 presents empirical results of a latent class residential location choice model estimated from stated preference data. Section 5 presents conclusions and further directions.

2 Background

The term lifestyle has been used for a long time in the transport and planning literature, for example qualitative references of *suburban living* or of *automobile-oriented households* or *transit-captive households*. There is now growing interest in better understanding attitudes and lifestyle and their roles as a driver for various activity, transport, and spatial behavior. For example, the topic has received prominent emphasis in national research reports such as TRB's Special Report 282 (2005) *Does the Build Environment Influence Physical Activity?*, which highlights as a 'knowledge gap' the lack of understanding regarding influences of lifestyle preferences and attitudes on behavior. Lifestyle is frequently referred to in the context of transport modeling frameworks, for example Ben-Akiva et al. (1996), Waddell (2000), and Moekel et al. (2003) referred to lifestyle in the context of the set of longer-term household choices; these lifestyle choices include residential ownership and housing type preferences and labor force and other activity participation, which condition patterns of daily activity and travel behavior.

There are several approaches with which researchers have aimed to capture lifestyle in the modeling process. First and foremost, for nearly as long as residential choice models have been developed, socio-economic variables have been used as explanatory variables to implicitly reflect heterogeneity of lifestyle preferences (see, for example, Lerman 1975). Related to this, albeit in a non-modeling context, there has been emphasis on the importance of understanding 'lifestyle trends' as reflected in shifts of demographics and travel statistics (for example, Ferrell and Deakin 2001, examined trends in California; Lyons et al. 2002, examined trends in the UK). A second approach to better reflect lifestyle preferences is to use correlation structures and error components to account for lifestyle preferences not reflected in the systematic portion of the utility. For example, Bhat and Guo (2006) simultaneously modeled residential location choice and auto ownership using a rich correlation structure, and they also include extensive socio-economic variables and measures of the built environment. The work here is built upon the third approach, expanded upon in the remainder of this literature review, which is to explicitly model lifestyle preferences and attitudes. The motivation for this approach is that the construct of lifestyle is richer in information than the conventional method of market segmentation using socio-economic variables (Salomon et al. 2002) or more recent methods focusing on correlation structures.

In terms of explicitly representing lifestyles, there have been some results from open format surveys such as a survey in Netanya, Israel (described in Salomon et al. 2002) in which respondents were requested to label and define four lifestyles that they perceived in the city. Interestingly, the 58 respondents to this survey referred to 41 different lifestyles, suggesting the difficulty of defining lifestyle groups. Below we focus on transport and activity literature related to defining lifestyles through quantitative methods and linking lifestyle with resulting spatial, activity, and transport behavior, with a specific focus on residential choice models.

There have been explicit attempts to quantify lifestyle orientation within the context of both spatial location choices and travel and activity choices. One of the earliest investigations was Salomon and Ben-Akiva (1983), who used K-means cluster analysis applied to a wide range of observable socio-economic variables (household structure, labor force participation, and education) to define five lifestyle segments (for example, one is upper socioeconomic, middle-aged and large households). In a second stage, they estimated separate mode and destination choice models for each lifestyle segment and demonstrated advantages over a traditional, single model approach. Krizek and Waddell (2003), rather than using observable socio-economic characteristics, used factor analysis and cluster analysis on observable travel and location choices made by households (travel characteristics, activity frequency, automobile ownership, and urban form at residential location) to identify nine lifestyle clusters, and then in a subsequent stage correlated the nine lifestyle clusters with socio-economic data. Their nine lifestyle segments were interpreted as retirees; single-busy urbanists; elderly homebodies; urbanists with higher income; transit users; suburban errand runners; family- and activity-oriented participants; suburban workaholics; and exurban, family commuters.

There is another line of research that applies a similar two stage approach, with the primary difference being the incorporation of responses to attitudinal surveys in the first stage to infer lifestyle segments. For example, Prevedouros (1992) applied factor analysis responses from a personality survey to define three personality factors (extroversion/introversion, materialism, and suburbanism), then applied cluster analysis on the resulting factor scores to define eight personality types, and finally correlated the personality types to observed residential location and travel behavior. Similarly, Lindberg et al. (1992) used principle component analysis on data regarding respondents' perceptions of neighborhoods at varying distances from the city center. Their extracted factors included 'neighborhood quality' and 'neighborhood centrality' as well as values of 'freedom', 'well-being', and 'togetherness'; through correlation of the factor scores with observable characteristics, they were able to show how these perceptions vary over the lifespan. Bagley and Mokhtarian (1999) used attitudinal surveys and factor analysis to identify ten attitudinal dimensions related to residential location preferences ('pro-high density'), travel mode preferences ('pro-driving', 'pro-drive alone', 'pro-transit'), views related to policies ('pro-environment', 'pro-pricing', 'pro-growth', 'pro-alternatives'), and activity and schedule orientation ('work-driven', 'time-satisfied'). In a second stage, they showed that extracted factor scores were significant in a binary logit model of the choice of a suburban versus an urban neighborhood (for example, pro-high density was a negative indicator for living in the suburbs and the nest-builder was positive). Bagley and Mokhtarian (2002) extended the second stage to a structural equation framework in which the endogenous variables included residential location type (represented as continuous factor scores measuring the degrees of 'suburbanness' and 'traditionalness' of each neighborhood), attitudinal variables, measures of travel demand, and job location. Cao and Mokhtarian (2005) used a similar approach to relate attitudes and lifestyle variables to a host of responses to survey questions regarding consideration of travel and land use related strategies, including major locational/lifestyle changes such as moving ones home closer to work. Targa and Clifton (2004) emphasized the importance of integrating lifestyle preferences and other psychological processes into urban land use and transport models, and also proposed similar two-stage approaches with factor analysis used to reduce responses from attitudinal surveys.

Similar two-stage approaches have been used in research focused on travel demand (as opposed to the residential choice emphasized above): Kitamura et al. (1997) found that attitudinal factors explained a higher proportion of travel demand than either land use factors or socioeconomics. Chliaoutakis et al. (2005) linked lifestyle factors (amusement, religious and family values, and sports) to aberrant driving behavior. Handy et al. (2005) were able to show that differences in attitudes largely explained observed differences in travel behavior between suburban and traditional neighborhoods. Ory and Mokhtarian (2005) used attitudes as explanatory variables for survey responses related to how much subjects like to travel for various purposes. Schwanen and Mokhtarian (2005) modeled commute mode choice accounting for 'cognitive dissonance' (the mismatch between ones current neighborhood type and ones preference for neighborhood type) as measured by attitudinal indicators.

A common factor in all of the studies reported above aimed at quantifying lifestyles (or attitudes) and the influence on behavior is the use of a two stage approach, in which there is separation between the extraction of lifestyle and the subsequent correlation of these lifestyle factors with residential location, travel, or activity behavior (note: in the case of Krizek and Waddell 2003, the order is reversed). The approach used in this paper is to model the lifestyle segmentation and the choice behavior of interest simultaneously. There are two key advantages to this approach:

1. Simultaneous estimation avoids any measurement error that may exist in the two-stage approach (Ben-Akiva et al. 2002), which is an issue because the lifestyle preferences are incorrectly assumed to be error-free in the second stage. This occurs, for example, when only the mean of a factor score is used in the second stage as an independent variable, and the

distribution of the factor score (i.e., the error from the first stage model) is discarded. Measurement error is a form of endogeneity that causes bias in parameter estimates. The issues of measurement error and endogeneity in regression are covered in many econometric textbooks (see, for example, Greene 2003), and issues of endogeneity and measurement error in discrete choice analysis are becoming more prominent in the literature (see, for example, Louviere et al. 2005, for a general discussion or Guevara and Ben-Akiva 2006, for an application to residential choice).

2. Simultaneous estimation directly employs the behavior of interest (in our case, residential location choice) as an indicator for the lifestyle construct, thereby using this information to define the lifestyle segmentation. There is significant and highly relevant information regarding lifestyle segmentation that can be inferred directly from residential location decisions, which is not exploited by a two-stage process.[1]

The empirical study presented below shows that latent class choice models can be used to infer latent lifestyle segments from the residential choice variable alone and simultaneously estimate the impact of household characteristics and lifecycle on lifestyle preferences. The latent class choice model will be explained below, followed by the empirical study, and a discussion of further extensions (including how the framework can be extended to incorporate simultaneously residential choice behavior and attitudinal indicators).

3 Methodology

Latent class choice models are appropriate for this analysis because our hypotheses are that discrete lifestyle preferences exist, that these lifestyles are not directly identifiable from the data, and that people with different lifestyles will exhibit different residential location choice behavior. This section provides a brief introduction to latent class choice models, see Gopinath (1995) or Magidson et al. (2003) for further information.

The latent class choice model is comprised of two components: a class membership model and a class-specific choice model as shown in Fig. 1. The class-specific choice model represents the choice behavior of each class and varies across latent classes. This class-specific choice probability is written as:

$$P(i|X_n, s),$$

which is the probability with which decision-maker n will select alternative i, conditional on the characteristics of the decision-maker and attributes of the

[1] Note that the exception is the 2-stage approach employed by Krizek and Waddell (2003). However, the two approaches are fundamentally different in that Krizek and Waddell assume that the lifestyle group implies homogeneous choices, whereas the latent class choice model approach assumes the lifestyle group implies homogeneous preferences.

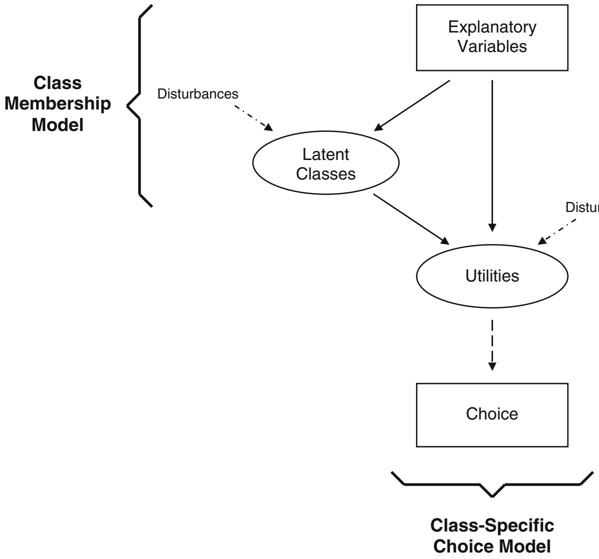

Fig. 1 Latent class choice model framework

alternative X_n and conditional on decision-maker n being a member of class s. The class-specific choice model may vary across classes on several dimensions including different parameter weights, different alternatives in the choice set, different model structure (for example, different nesting structures), or different decision protocols (for example, utility maximizing versus other).

While it cannot be deterministically identified to which latent class a decision-maker belongs from the observable variables, it is presumed that class membership probabilities can be estimated. These are denoted by the class membership model:

$$P(s|X_n),$$

which is the probability that decision-maker n with characteristics X_n belongs to latent class s.

Since the class of each decision-maker is unknown, neither of the above equations can be estimated individually. Rather, the two components are estimated simultaneously via a latent class choice model:

$$P(i|X_n) = \sum_{s=1}^{S} P(i|X_n, s)P(s|X_n), \tag{1}$$

where the probability of selecting a particular alternative i is equal to the sum over all latent classes s of the class-specific membership model conditional on class ($P(i|X_n,s)$) multiplied by the probability of belonging to that class ($P(s|X_n)$).

Using an example in the context of this paper, say the model of interest is residential location choice and there are two latent lifestyles: auto-oriented households and transit-oriented households. As in a traditional residential choice model, say the choice set consists of the set of transport analysis zones (TAZs) in an urban area, and each TAZ is described by its average household characteristics and its neighborhood and accessibility characteristics. The class-membership equation models the probability that household n with observable characteristics X_n (such as income and number of children) is auto-oriented ($P(\text{AutoOriented}|X_n)$) and the probability with which the household is transit-oriented ($P(\text{TransitOriented}|X_n) = 1 - P(\text{AutoOriented}|X_n)$). The class membership model can take on a number of forms, such as binary logit for this example. There is then a class-specific choice model for each latent class modeling the probability of choosing a particular TAZ i conditional on being in that class: $P(i|X_n,\text{AutoOriented})$ and $P(i|X_n,\text{TransitOriented})$, where X_n includes both characteristics of the household and attributes of the alternative zones (housing price, school quality, travel time to work, etc.). The class-specific choice behavior would vary across the two classes; for example an auto-oriented household may place more weight on travel time to work by auto and parking availability, whereas a transit-oriented household may place more weight on transit accessibility. The probability of a household choosing a particular TAZ i is then:

$$
\begin{aligned}
P(i|X_n) = {} & P(i|X_n, \text{AutoOriented})P(\text{AutoOriented}|X_n) \\
& + P(i|X_n, \text{TransitOriented})(1 - P(\text{AutoOriented}|X_n))
\end{aligned}
$$

The latent class choice model framework enables simultaneous estimation of the parameters of the class membership model and the class-specific choice models from observed household residential location decisions and without using variables that explicitly indicate lifestyle. The class membership model provides information as to who is likely to be in each class, whereas the class-specific choice models provide information on how each class behaves.

The primary modeling issues of a latent class choice model are the number of classes, the form of the class membership model, and the forms of the class-specific choice models. The number of classes is determined through a combination of statistical information (for example, the BIC, which will be discussed later) and interpretation of the model results. Typically the class-membership models are relatively straightforward logit equations, however Gopinath (1995) provides details on more complex relationships that can be introduced and why such complexity may be warranted. The class-specific choice model takes on whatever form is most appropriate for a non-class-specific choice model (for example, logit, nested logit, probit, random parameter logit, etc.) and can vary across classes.

There are several advantages of using a latent class choice model. First, it can capture underlying, unobservable *discrete* segmentation such as that theorized by our lifestyle hypothesis. Second, it estimates jointly both the parameters of the class membership model (linking observable household factors with the likelihood of being in a particular class) as well as the class-specific behavior (explaining how, conditional on a particular class, an individual will behave). Finally, it provides estimates of the size of each segment.

4 Empirical application

The hypothesis of latent lifestyle classes is tested using a stated preference survey of residential location choices. This section describes the data, model specification, model estimation results, and policy implications.

4.1 Data

The data were obtained from a household activity and travel behavior survey conducted in Portland, Oregon in 1994. The full data collection effort is described in Cambridge Systematics (1996). We make use of the stated preference survey of household residential location choice decisions that was administered to 611 individuals.[2] Each survey question asked for a preference among five hypothetical housing options, with alternatives varying across price, size, community amenities, accessibility, and several other factors influencing residential choices. An example choice experiment is provided in Fig. 2. Each choice experiment consisted of the five alternatives shown (buy single-family, buy multi-family, rent single-family, rent multi-family, and move out of the metro area) and the list of attributes shown (type of dwelling, residence size, etc.). The fifth alternative of moving out of the metro area is a so-called "opt-out" alternative, which is a technique typically employed in stated preference surveys.[3] Each individual was presented with eight different choice experiments of the format shown in Fig. 2, and the values of the attributes that describe each alternative varied across each choice experiment. There were 16 different versions (blocks) of the questionnaire, that is 16 different sets of 8 choice experiments, each appearing as in Fig. 2 but with differing attributes selected based on the experimental design. Each survey respondent was assigned randomly to 1 of the 16 blocks. This survey provides for a nice test case of the theory of latent lifestyle segmentation because of the well-defined choice problem (contrary to a revealed preference setting) and the extensiveness of the attributes. Persons with different lifestyle leanings will focus on different subsets of the attributes, weighing some more heavily than others. After data cleaning, the estimation results presented below are based on responses from 507 individuals, for a total of 4,056 choice experiments.

[2] See Louviere et al. (2000), for a review of stated preference survey methods and their use in behavioral modeling.

[3] For a literature review on the inclusion and form of the opt-out alternative in stated preference experiments, see Kontoleon and Yabe (2003).

	(Alternative 1)	(Alternative 2)	(Alternative 3)	(Alternative 4)	(Alt. 5)
	Buy Single Family	Buy Multi-Family	Rent Single Family	Rent Multi-Family	Move out of the Metro Area
Type of Dwelling:	*single house*	*apartment*	*duplex / row house*	*condominium*	
Residence Size:	*< 1,000 sq. ft.*	*500-1,000 sq. ft.*	*1,500 - 2,000 sq. ft.*	*< 500 sq. ft.*	
Lot Size:	*< 5,000 sq. ft.*	*n/a*	*5,000 - 7,500 sq. ft.*	*n/a*	
Parking:	*street parking only*	*street parking only*	*driveway, no garage*	*reserved, uncovered*	
Price or Monthly Rents:	*< $75K*	*$50K - $100K*	*> $1,200*	*$300 - $600*	
Community Type:	*mixed use*	*mixed use*	*rural*	*urban*	
Housing Mix:	*mostly single family*	*mostly multi-family*	*mostly multi-family*	*mostly multi-family*	
Age of Development:	*10-15 years*	*0-5 years*	*10-15 years*	*0 - 5 years*	
Mix of Residential Ownership:	*mostly own*	*mostly own*	*mostly rent*	*mostly own*	
Shops/Services/Entertainment:	*community square*	*basic shops*	*community square*	*basic, specialty shops*	
Local Parks:	*none*	*yes*	*none*	*none*	
Bicycle Paths:	*none*	*yes*	*yes*	*yes*	
School Quality:	*very good*	*very good*	*fair*	*fair*	
Neighborhood Safety:	*average*	*average*	*average*	*average*	
Shopping Prices Relative to Avg:	*20% more*	*20% more*	*same*	*10% more*	
Walking Time to Shops:	*20-30 minutes*	*20-30 minutes*	*< 10 minutes*	*10 - 20 minutes*	
Bus Fare, Travel Time to Shops:	*$1.00, 15-20 minutes*	*$1.00, > 20 minutes*	*$0.50, 5 - 10 minutes*	*$0.50, < 5 minutes*	
Travel Time to Work by Auto:	*> 20 minutes*	*15-20 minutes*	*15 - 20 minutes*	*< 10 minutes*	
Travel Time to Work by Transit:	*> 45 minutes*	*30-45 minutes*	*30 - 45 minutes*	*15 - 30 minutes*	

Fig. 2 Choice experiment example

4.2 Model specification

The latent class choice model requires specification of the class membership model and the class-specific choice model. The detailed specification will be discussed in the estimation results; this section covers the overall structure of the model.

The class-specific choice model consists of the choice among the five alternatives shown in Fig. 2 (i = 1,...,5): buy single-family (BSF), buy multi-family (BMF), rent single-family (RSF), rent multi-family (RMF) and move out of the metro area (MOVE). There are two specification issues with these data. The first is that logit may not be an appropriate model because of correlation of the error terms within the buy and rent alternatives. The second is that each household (represented by the individual responding for the household) responded to eight different choice experiments, and the responses across experiments from a single household are likely to be correlated. These two issues can be addressed by incorporating three error components in a logit mixture model for panel data (see Walker et al. 2006, for discussion of specification and identification issues). The utilities for each class specific choice model are then specified as follows:

$$\begin{array}{llllll} U_{nts}^{\mathrm{BSF}} & = \beta_s X_{nt}^{\mathrm{BSF}} & +\sigma^B \eta_n^B & & & + \varepsilon_{nts}^{\mathrm{BSF}} \\ U_{nts}^{\mathrm{BMF}} & = \beta_s X_{nt}^{\mathrm{BMF}} & +\sigma^B \eta_n^B & & & + \varepsilon_{nts}^{\mathrm{BMF}} \\ U_{nts}^{\mathrm{RSF}} & = \beta_s X_{nt}^{\mathrm{RSF}} & & +\sigma^R \eta_n^R & & + \varepsilon_{nts}^{\mathrm{RSF}} \\ U_{nts}^{RMF} & = \beta_s X_{nt}^{RMF} & & +\sigma^R \eta_n^R & & + \varepsilon_{nts}^{RMF} \\ U_{nts}^{\mathrm{MOVE}} & = \beta_s X_{nt}^{\mathrm{MOVE}} & & & +\sigma^M \eta_n^M & + \varepsilon_{nts}^{\mathrm{MOVE}} \end{array}, \qquad (2)$$

where $n = 1,...,N$ ($N = 507$ households), $t = 1,...,T$ ($T = 8$ choice experiments per household), and $s = 1,...,S$ (S = number of latent lifestyle classes). The correlation among alternatives (nesting structure) and correlation across responses from a single household (panel effect) are captured by the error components η_n^B, η_n^R, η_n^M, which are distributed *iid* Normal$(0, 1)$ across households n but remain constant within responses t from a given household. $\varepsilon_{nts}^{\mathrm{BSF}}$, ...,$\varepsilon_{nts}^{\mathrm{MOVE}}$ ~ *iid* Extreme Value across all households n, responses t, and classes s . The vectors $\eta(= \eta_n^B, \eta_n^R, \eta_n^M)$ and $\varepsilon(= \varepsilon_{nts}^{\mathrm{BSF}}, ...,\varepsilon_{nts}^{\mathrm{MOVE}})$ are independent (therefore the model is a logit mixture model for panel data, see, for example, Train 2003 or Walker et al. 2006). X_{nt}^{BSF}, ...,X_{nt}^{MOVE} are column vectors of observable characteristics of households and attributes of alternatives and are not a function of class s (variation in specification across classes can be captured through the specification of β_s). The estimated parameters are the row vectors β_s ($s = 1,...,S$) and scalars σ^B, σ^R, σ^M (denoted together as σ). While β_s varies across classes, the correlation parameters σ are assumed to be the same across classes; this imposes a parsimonious specification of the error structure that eases identification (see Chiou and Walker 2006, for more discussion). The likelihood *conditional on class s* for the *eight responses* of a given household is then:

$$P(i_1, \ldots, i_T | X_n, s; \beta_s, \sigma) = \int \prod_{t=1}^{T} P(i_t | X_{nt}, s, \eta; \beta_s, \sigma) f(\eta) \mathrm{d}\eta. \quad (3)$$

This is the product (over the T responses) of the logit probability of each individual response i_t conditional on unknown η, $P(i_t|X_{nt},s,\eta;\beta_s,\sigma)$, and the product is integrated over the distribution of η. By construction, $f(\eta)$ is a 3-dimentional multivariate normal with (3×1) mean vector of zeros and covariance matrix equal to a 3×3 identity matrix.

We take an exploratory approach to class-specific behavior in which each lifestyle class has the same specification, and β_s varies across classes (determined in estimation and inferred statistically from the choice behavior) denoting different trade-offs being made by the different lifestyle groups.

For the class membership model, we specify a logit equation, denoted $P(s|X_n;\gamma)$, where $s = 1,...,S$ (S = number of latent lifestyle classes), X_n are explanatory socio-economic characteristics, and γ are the estimated parameters. The explanatory variables include information on household structure (number of children by age group, number of adults, and family versus non-family), employment (number of employed persons, number of retired persons, whether in managerial or professional occupations, and maximum number of work hours), the age of the head of household, and resources available to the household (measured in terms of income).

Combining the class membership model discussed above with the class-specific choice model (Eq. 3), the joint likelihood function for each household is then as follows:

$$P(i_1,\ldots,i_T|X_n;\gamma,\beta_s,\sigma)=\sum_{s=1}^{S}P(s|X_n;\gamma)P(i_1,\ldots,i_T|X_n,s;\beta_s,\sigma). \quad (4)$$

This is the sum over all classes s of the product of the probability of belonging to class s multiplied by the product of the conditional probabilities of the eight chosen alternatives (conditional on belonging to class s). Thus, the class membership probability remains constant across all responses from a given household. Recall that since the class-specific choice probabilities are conditional on the unknown error components η, these must be integrated out over their distribution $f(\eta)$. It is interesting to note that in this empirical case, the models *without* the error components η are not able to uncover latent lifestyle segmentation in that the segmentation did not provide a statistical improvement over a model without segmentation.

The parameters are estimated through maximum likelihood estimation, and numerical integration is used to evaluate the three-dimensional integral (Eq. 4). The software Latent GOLD Choice 4.0 by Statistical Innovations Inc is used for estimation.

4.3 Determining the number of classes

The remaining aspect of the model specification is to determine the number of classes. This is not determined endogenously; rather, successive models are estimated with varying numbers of classes and statistics are used to compare different models. We estimated the models with 1–4 classes (all with the same specification other than the number of classes) and a summary of model results are shown in Table 1. There are numerous statistics that aid the selection of the number of latent classes, such as the BIC, AIC, and rho-bar-squared shown in Table 1. All such statistics are based on the same general principle of weighing the fit of the model (we denote as LL(β) or the log-likelihood calculated at the value of the fitted parameters) against the parsimony of the model or number of parameters (we denote as K). Various statistics place either more or less of a penalty on the number of parameters in the model. The AIC (Akaike Information Criterion) is often used for choice models, and is equal to 2*(LL(β)-K). The rho-bar-squared is a function of the AIC

Table 1 Overview of model estimation results

	Model without lifestyle segmentation	Models with lifestyle segmentation		
Number of classes	1	2	3	4
Number of parameters	37	76	115	155
BIC	– 10,319	– **10,281**	– 10,310	– 10,397
AIC	– 10,163	– 9,959	– 9,823	– **9,741**
Rho-bar-squared	0.222	0.237	0.248	**0.254**

All models estimated using 507 households with eight experiments per household, resulting in 4,056 choices (therefore applies the same discounting for each parameter) and is equal to 1-(LL(β)-K)/LL(0) where LL(0) is the log-likelihood of a naive model with no parameters. The BIC (Bayesian Information Criterion) is often used in latent class choice models and imposes a harsher penalty on the number of parameters than the AIC and rho-bar-squared; the BIC formula is (2*LL(β) – ln(N)*K) where N is the number of respondents. With 507 respondents (as in our case), the BIC penalizes each parameter with just over three log-likelihood points versus one log-likelihood points in the AIC, and therefore the BIC favors parsimonious structures. In general, the higher the AIC, rho-bar-squared, and BIC, the better is the model according to the statistics. While these statistics are informative, none should be used blindly without examination of the estimation results. In our model, all of the statistics reported in Table 1 indicate that a model with lifestyle segmentation is preferred over one without. However, the BIC suggests that the 2-class model is superior, whereas the AIC (and rho-bar squared) suggests that the 4-class model is superior. We have selected the 3-class model because, in examining the estimation results, it provides the most satisfying behavioral interpretation in terms of resulting lifestyle classes and class-specific choice models (primarily lack of anti-intuitive signs and interpretability of classes). The results of the 3-class model are discussed in detail in the next section.

4.4 Detailed estimation results for the 3-class model

The latent class choice model estimation results consist of parameter estimates for the class-specific choice models (Tables 3, 4) and the class membership model (Table 5 and Fig. 3). All of the parameters in these tables result from simultaneous estimation of the class-specific choice model and the class membership model. Each component of the model will be discussed in turn. Estimation results from the base choice model without latent segmentation are in Table 2. This single class choice model includes systematic segmentation via interaction between income and price. Further systematic segmentation in this base model was tested without resulting in significant improvement to the choice model.

4.4.1 Class-specific choice model: how does behavior vary across lifestyle classes?

Table 3 provides the parameter estimates of the class-specific choice models. First note that a subset of parameters is restricted to be equal across the three classes. The class independent parameters include the price variables (rent for the rent alternatives and purchase price for the buy alternatives), because we a priori focus on lifestyle segmentation that is a function of the built environment (land use and transportation) rather than segmentation based on price sensitivity. In order to reflect sensitivity to price based on resources, the pricing variables are interacted with income dummies based on annual household income. Low income is defined as less than \$30,000, medium in-

Table 2 Choice model without latent segmentation

	Variable	1-Class	
		Coefficients	*t*-Statistics
Housing attributes	Monthly rent ($00) - low/middle income	– 0.162	**– 13.8**
	Monthly rent ($00) - high income	– 0.051	**– 2.9**
	Monthly rent ($00) - income not available	– 0.160	**– 6.6**
	Purchase price ($000) - low income	– 1.192	**– 13.7**
	Purchase price ($000) - middle income	– 0.719	**– 10.7**
	Purchase price ($000) - high income	– 0.249	**– 3.0**
	Purchase price ($000) - income not available	– 0.644	**– 5.7**
	Single house (v. duplex)	0.382	**6.3**
	Condo (v. apartment)	0.170	**2.3**
	Residential size (square feet/1000)	0.440	**8.3**
	Lot size (square feet/1000)	0.008	1.2
Neighborhood attributes	Mostly owners (v. mostly renters)	0.181	**3.8**
	Mostly-multi-family housing (v. mostly single-family)	– 0.041	– 0.9
	Schools—75 percentile (v. below 60)	0.190	**2.8**
	Schools—60–75 percentile (v. below 60)	0.301	**5.1**
	Above average safety (v. average)	0.127	**2.8**
	Mixed use (v. rural)	0.073	1.1
	Urban (v. rural)	0.040	0.6
	Suburban (v. rural)	– 0.062	– 0.9
	Local bike path (v. no local bike path)	0.083	1.8
	Local park (v. no local park)	0.021	0.5
	Local community square (v. no shops)	0.213	**3.3**
	Basic plus specialty shops (v. no shops)	0.173	**2.6**
	Basic shops (v. no shops)	0.153	**2.3**
Transport/access attributes	Walk time to local shops (minutes)	– 0.009	**– 4.0**
	Travel time to work by auto (minutes)	– 0.001	– 0.2
	Travel time to work by transit (minutes)	– 0.006	**– 2.6**
	Off street parking available (v. no off street parking)	0.427	**6.9**
Correlation terms	Standard deviation on buy constant (σ^B)	0.840	**7.0**
	Standard deviation on rent constant (σ^R)	1.182	**10.8**
	Standard deviation on move out constant (σ^M)	2.178	**20.5**

come is greater than or equal to $30,000 and less than $60,000, and high income is greater than or equal to $60,000. Of those households who reported income (10% did not), 30% of the sample are low income, 50% are medium income, and 20% are high income. The other parameters that are fixed across the choice models are the standard deviations from the error components, as discussed with Eq. 2. All other parameters are allowed to vary across classes and there are no a priori exclusions of parameters for specific classes.

While the statistics in Table 1 indicate that the model with three latent classes is a significant improvement over a model without lifestyle segmentation, the detailed results in Table 3 allow us to see how the behavior varies across the classes. First note that many of the parameters are significant (denoted with a bold *t*-stat) at the 95% confidence level. Further, variables that vary significantly across classes (as determined by a Wald statistic at the

Table 3 Class-specific choice model estimation results

	Variable	Class independent		Class 1		Class 2		Class 3	
		Coefficients	*t*-Statistics	Coefficients	*t*-Statistics	Coefficients	*t*-Statistics	Coefficients	*t*-Statistics
Housing attributes	Monthly rent ($00) - low/middle income	– 0.152	**– 11.5**						
	Monthly rent ($00) - high income	– 0.079	**– 3.6**						
	Monthly rent ($00) - income not available	– 0.160	**– 5.0**						
	Purchase price ($000) - low income	– 1.007	**– 9.8**						
	Purchase price ($000) - middle income	– 0.928	**– 11.2**						
	Purchase price ($000) - high income	– 0.534	**– 4.8**						
	Purchase price ($000) - income not available	– 0.783	**– 4.4**						
	Single house (v. duplex)			*0.503*	***4.3***	*0.840*	***5.8***	*– 0.318*	***– 2.1***
	Condo (v. apartment)			0.302	1.7	0.468	**2.0**	0.036	0.3
	Residential size (square feet/1000)			*1.377*	***12.9***	*– 0.335*	***– 2.4***	*0.049*	*0.4*
	Lot size (square feet/1000)			*0.009*	*0.8*	*0.059*	***3.7***	*– 0.052*	***– 3.4***
Neighborhood attributes	*Mostly owners (v. mostly renters)*			*0.226*	***2.3***	*– 0.070*	*– 0.6*	*0.278*	***2.7***
	Mostly multi-family housing (v. mostly single-family)			*– 0.179*	*– 1.9*	*0.204*	*1.6*	*– 0.126*	*– 1.3*
	Schools—75 percentile (v. below 60)			*0.618*	***4.3***	*0.381*	***2.3***	*0.174*	*1.4*
	Schools—60–75 percentile (v. below 60)			0.336	**2.3**	0.294	1.6	0.029	0.2
	Above average safety (v. average)			*0.226*	***2.4***	*– 0.235*	*– 1.9*	*0.295*	***2.9***
	Mixed use (v. rural)			0.133	1.0	– 0.160	– 1.0	0.261	1.8
	Urban (v. rural)			*0.000*	*0.0*	*– 0.271*	*– 1.6*	*0.407*	***2.8***
	Suburban (v. rural)			– 0.199	– 1.4	– 0.128	– 0.7	0.106	0.7
	Local bike path (v. no local bike path)			*– 0.100*	*– 1.0*	*0.415*	***3.2***	*0.135*	*1.3*
	Local park (v. no local park)			0.073	0.8	0.154	1.2	– 0.110	– 1.2
	Local community square (v. no shops)			0.301	**2.1**	0.115	0.7	0.240	1.6
	Basic plus specialty shops (v. no shops)			*0.453*	***3.1***	*– 0.540*	***– 2.8***	*0.374*	***2.7***
	Basic shops (v. no shops)			*0.198*	*1.4*	*– 0.170*	*– 1.0*	*0.404*	***2.8***

Table 3 continued

	Variable	Class independent		Class 1		Class 2		Class 3	
		Coefficients	*t*-Statistics	Coefficients	*t*-Statistics	Coefficients	*t*-Statistics	Coefficients	*t*-Statistics
Transport/access attributes	*Walk time to local shops (minutes)*			*– 0.010*	***– 2.3***	*0.006*	*1.0*	*– 0.019*	***– 4.1***
	Travel time to work by auto (minutes)			*– 0.015*	*– 1.5*	*0.029*	***2.1***	*– 0.014*	*– 1.1*
	Travel time to work by transit (minutes)			*– 0.003*	*– 0.6*	*– 0.021*	***– 3.5***	*0.006*	*1.1*
	Off street parking available (v. no off street parking)			0.633	**5.2**	0.333	**2.1**	0.408	**3.0**
Correlation terms	Standard deviation on buy constant (σ^B)	0.932	**6.4**						
	Standard deviation on rent constant (σ^R)	1.251	**9.3**						
	Standard deviation on move out constant (σ^M)	2.166	**16.5**						

Italicized indicates parameters that vary siginificantly across classes (Wald statistic, 90% confidence)

90% confidence level) are italicized. These signs indicate significant heterogeneity across classes. For example, three of the four transport/access attribute parameters are statistically significantly different across the three classes (through the Wald test), with off-street parking being the only one without significant variability.

Another useful method of examining the class-specific choice model results is through an 'importance rating' of variables for each class, which is shown in Table 4. The top ten most important variables are listed for each class, with the first being the most important through to the tenth most important. This ranking is determined by taking the difference between the highest and lowest value of each variable as observed in the dataset and multiplying this difference by the coefficient for that variable. The variables are then rank ordered based on the absolute value of this product, which reflects the order of potential impact on the utility. Note that with analyst-designed SP data such as these in which there are relatively few levels (values) for each explanatory variable and no extreme values, this procedure is essentially the same as determining importance through "standardized" coefficients (i.e., standardizing each explanatory variable and then re-estimating the model) where the ranking is then based on the non-standardized estimated parameter times the standard deviation. While such methods are informative and helpful with

Table 4 Ten most important varibles for each class, rank ordered (1 is most important)

	Class 1 Suburban, auto, school orientation	Class 2 Transit, house orientation	Class 3 High density near urban activity, and auto orientation
1	Larger residence	Lower travel time to work by transit	Smaller lot size
2	Off street parking	Larger lot size	Shorter walk time to local shops
3	Schools in 75 percentile	Single house on lot	Urban setting
4	Single house on lot	Larger residential size	Off street parking
5	Basic plus specialty shops nearby	Longer travel time to work by auto	Basic shops nearby
6	Lower travel time to work by auto	Don't want basic shops nearby	Basic plus specialty shops nearby
7	Lower Walk time to local shops	Condo (rather than apartment)	Shorter travel time to work by auto
8	Schools in 60th percentile	Bike path nearby	Longer travel time to work by transit
9	Condo (rather than apartment)	Schools in 75th percentile	Single house on lot
10	Community square nearby	Off street parking	Above average safety

interpretation (as highlighted in the next section), it should be remembered that the importance rating results are a function of the underlying data.

Through examination of the estimation results and the variables that are important to each class, we can make inferences on the lifestyle of each class. Class 1 is oriented towards a suburban, auto-oriented lifestyle with a larger residence, off-street parking, single house on a lot, and lower travel time to work by auto serving prominently in their important variables. High quality schools are also important to this group as is availability of local high end shopping (specialty shops and community square).

Class 3 also indicates an auto orientation as they place importance on off street parking and shorter travel time to work by auto. However, members of Class 3 are drawn towards higher densities and urban activity as their top variables of importance are a smaller lot size, an urban setting, and shops within walking distance. Class 3 is also the only one for which safety appears in the top ten, presumably because safety is more of an issue in urban centers and therefore more on the mind of those in this lifestyle class.

Class 2 is the only transit-oriented lifestyle, which is indicated by the presence of travel time to work by transit as the most important variable. However, they seem to desire transit accessibility in a suburban setting as they indicate preferences for larger lot sizes, single houses on lots, larger residential sizes, no shops nearby, and being away from highways. Note that this finding could never result from revealed preference data, in which attributes of transit and the associated housing environment are confounded, making it impossible to identify the separate effects. Other attributes of importance to Class 2 are high quality schools and the presence of a local bike path.

To summarize, Class 1 is suburban, auto, and school oriented; Class 2 is transit and school oriented but in a suburban setting; and Class 3 is urban and auto oriented. Policy implications of this segmentation will be discussed later.

4.4.2 Class membership model: what are the predictors of lifestyle type?

Given the segmentation in preference for residential environments, now we use the class membership model to see if available socio-economic characteristics are good predictors for the latent lifestyle classes. The estimation results for the class membership model are shown in Table 5. This is a multinomial logit model of the probability with which each household belongs to each of the three classes. Explanatory variables include characteristics related to household structure, employment, age, and resources. As before, individual coefficients that are significant at the 95% confidence level are displayed with bold *t*-statistics, and variables that have significantly different effects across classes at the 90% confidence level are italicized.

Examining the parameters of the class membership model, those households in Class 1 (suburban, school, auto) tend to be affluent, more established, professional families; households in Class 2 (transit, suburban) tend to be less affluent, younger families; and households in Class 3 (urban, auto) tend to be older (middle-aged or retired), non-family, professionals. In general, these

Table 5 Class membership model estimation results

	Variable	Class 1 43%		Class 2 30%		Class 3 27%	
		Affluent, more established families		Less affluent, younger families and non-families		Older, non-family, professionals	
		Coefficients	*t*-Statistics	Coefficients	*t*-Statistics	Coefficients	*t*-Statistics
	Intercept	– 1.673	– 1.9	1.513	1.6	0.161	0.2
Household structure	Number of children under 5 years old	1.273	1.0	1.431	1.1	– 2.704	– 1.0
	Number of children from 5 to 11	*0.426*	***2.4***	*– 0.117*	*– 0.5*	*– 0.309*	*– 1.2*
	Number of children from 12 to 17	0.300	1.1	0.097	0.4	– 0.398	– 1.0
	Number of persons 18 and over	0.444	1.5	– 0.163	– 0.5	– 0.282	– 0.9
	Non-family dummy	– 0.706	– 1.7	0.124	0.3	0.582	1.7
Employment	Number of employed persons	– 0.096	– 0.5	– 0.048	– 0.2	0.145	0.5
	Number of retired persons	– 0.480	– 1.4	0.076	0.2	0.404	1.3
	Dummy if at least one "manger/professional"	*0.131*	*0.6*	*– 0.571*	***– 2.1***	*0.440*	*1.6*
	Maximum number of work hours	– 0.002	– 0.3	– 0.004	– 0.4	0.006	0.9
Age of the head of household	Piecewise linear age of HOH: age 20–35	– 0.030	– 0.6	0.040	0.9	– 0.010	– 0.2
	Piecewise linear age of HOH: age 36–60	*0.020*	*1.1*	*– 0.058*	***– 2.7***	*0.037*	*1.8*
	Piecewise linear age of HOH: age 60 plus	0.035	0.9	0.003	0.1	– 0.038	– 1.5
Resources	*Dummy for medium income*	*1.760*	***3.4***	*– 0.904*	*– 1.9*	*– 0.855*	***– 2.2***
	Dummy for high income	*1.607*	***4.4***	*– 1.026*	***– 3.7***	*– 0.581*	*– 1.9*
	Dummy for income not reported	*2.183*	***4.6***	*– 2.151*	***– 3.7***	*– 0.032*	*– 0.1*

Italicized indicates parameters that vary significantly across classes (Wald statistic, 90% confidence) lifecycle drivers from the class-membership model tend to match a priori hypothesis of who would be suburban, transit, and urban oriented. The model also indicates that there is a fairly even split of the surveyed households among these three classes: 43% in Class 1, 30% in Class 2, and 27% in Class 3. These percentages are the average probability over the sample of belonging to each class. (Note that statements on the split in the full population would require expansion of the survey to the characteristics of the full population.)

Figure 3 provides another form of examining the class membership model, which is through class profiling. The figure shows for a subset of variables, the average value of that variable for households within each of the three classes (equal to the weighted average of each variable, where the weight is the probability of being in a particular class). For example, households in Class 3 have relatively fewer children in all three of the child age categories and are more likely to be retired. The conclusion that can be drawn from Fig. 3 is that while there exists some profiling, it is not very strong. That is, there are things about lifestyle preferences that are unobserved, and that cannot be explained well by observable socio-economic characteristics. So while at first glance of the parameter estimates in Table 5, it appears that lifecycle characteristics are strong determinants of lifestyle preferences (validating research such as Lindberg et al. 1992, who emphasize that residential location preferences vary across the lifespan), it is clear that it is not deterministic and a probabilistic model is necessary. Evidence from literature in which both lifestyle variables and socio-demographic variables are significant in behavioral models also supports this point that lifecycle variables alone are not sufficient to capture behavior (see, for example, Mokhtarian and Salomon 1997; Bagley and Mokhtarian 1999). This is not surprising; when one looks at the population in the

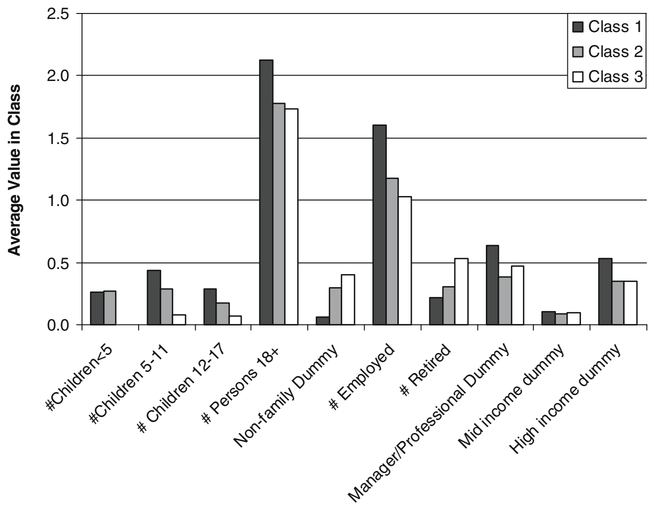

Fig. 3 Lifestyle class profiling

suburbs (or in the urban area) there are a variety of lifecycle points represented.

4.5 Potential policy implications of the inferred lifestyle classes

In terms of implications of these estimation results on policy analysis, one of the things to take away from this analysis is that people's preferences seem to be more complex than are typically hypothesized. For example, members in Class 1 have preferences for *both* large homes and an auto-oriented lifestyle (most often associated with segregated-use suburban neighborhoods) *and* for local high-end shopping (most often associated with smaller, multifamily residences in mixed use urban neighborhoods). Similarly, Class 2 is an interesting blend of simpler caricatures, perhaps suggesting the park-n-ride market that wants the suburban large-home lifestyle, but with the convenience of transit for commuting to their CBD jobs. Such complexity also extends to Class 3, which are the auto-oriented urbanites. To some extent these preferences may be unrealistic, suggesting that each in their own way want to "have it all". However, to the extent that such complex sets of preferences can be accommodated, they point to some potentially powerful ways to help people "have it all" while still encouraging more environmentally positive behavior. For example, supply the demand for lower-density residential neighborhoods, but support them with viable transit options for the "worst" (commuting) trips. Similarly, accommodate the need of urban dwellers to have and sometimes use a car (making that lifestyle more appealing to people who otherwise would reject it) while making it easier not to have to use the car.

5 Summary and future directions

The above empirical study supports our hypotheses that lifestyle preferences exist, that they are key determinants of residential location behavior, that they can be inferred from observed choices of residential location, and that they can be explained in part by observable socio-economic characteristics such as income, age, and household structure. The framework of latent class choice models was employed in the analysis, where the latent classes are the lifestyles and the choice model is the choice of residential location. This differs from previous literature in this area, because the lifestyle groups and the impact of lifestyle on residential location decisions are estimated simultaneously. This approach provides two key advantages. The first is that measurement error bias inherent in a two-stage procedure is avoided, and the second is that the behavior of interest is used as an indicator to determine the lifestyle segments. It is worthwhile to emphasize the significant amount of information about lifestyle segmentation that can be inferred from observed residential choices alone—information that is disregarded in the two-stage procedure common in the literature. The empirical results using the stated preferences survey suggest that there are three lifestyle segments: households that are suburban, auto, and school oriented; households that are transit oriented but want a suburban setting, and households that are urban and auto oriented. Such lifestyle groups suggest a 'wanting to have it all' attitude, which point to perhaps troubling, but also intriguing, potential directions for urban planning policies. Socio-economic variables related to lifecycle are shown to have significant explanatory power; however there remained significant aspects of lifestyle preferences that could not be explained by such observable explanatory variables.

The research demonstrated the potential of latent class choice models in uncovering discrete heterogeneity of lifestyle preferences. By making explicit connections between demographics, lifestyle, and residential location, we can better understand the manner in which an urban area develops, and the impact that changes in urban form will have on critical issues such as environment and health. For example, such research provides insight into the rapidly growing literature investigating the relationship between the built environment and activity. The literature has been successful in providing evidence of significant correlation between the built environment and behaviors such as automobile use and physical activity. However, it is well recognized that self-selection plays a critical role in that persons with greater interest in protecting the environment and/or remaining physically fit choose to live in areas that support such lifestyles and vice versa. It has also been noted that policies aimed at behavior modification may be limited by the large proportion of households who have strong preferences towards auto oriented lifestyles (Cao et al. 2006). The research presented here aims to determine the magnitude and characteristics of segments such as the 'auto oriented' households as well determine causal relationships between socio-economic variables and lifestyle choices.

This work will be extended in several directions. The use of the stated preference data allowed us to validate the hypothesis of latent lifestyle preferences without having to grapple with the full complexity of a residential choice model estimated with revealed preference data. One extension is to estimate a model that draws on both this stated preference dataset as well as revealed preference data provided in the same 1994 Portland household and activity survey. Another extension is to incorporate additional indicators of lifestyle preferences. This work thus far makes use of surveys that did not collect direct information on attitudes, perceptions, and lifestyles. The relatively weak explanatory power in the class membership model presented here suggests that the model could be strengthened by employing psychometric indicators. To obtain information on lifestyle preferences, indicators could include, for example, responses to the following type of survey questions:

How much do you agree with the following?

I am willing to travel longer to have a big house and a garden.
I like to live within walking distance to shops and restaurants.
I enjoy the hustle and bustle of the city.
Living in a multiple-family unit would not give me enough privacy.

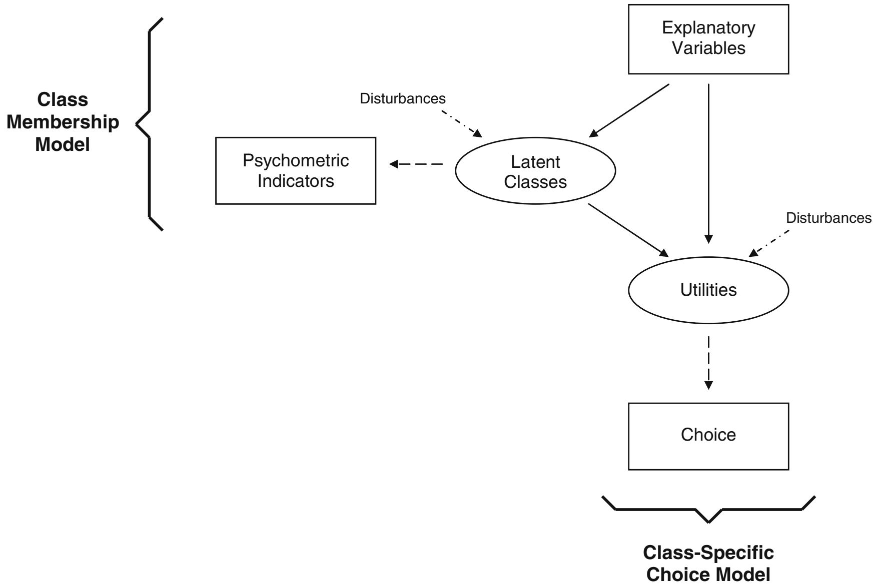

Fig. 4 Framework for latent class choice model with indicators

Such information can be introduced into a latent class framework as additional measurements for the latent classes as shown in Fig. 4, and the model can be estimated jointly. Such an approach represents a link between the method that is presented in this paper and the two-stage method often seen in the literature. Another extension is to add further structure to the lifestyle dimension. In this research, the classes were mutually exclusive and the class-membership model is multinomial logit. An important direction in which to improve the specification is to incorporate dimensions of lifestyle preferences for example those towards house, neighborhood, and transport characteristics. (See Gopinath 1995; Walker 2001; Ben-Akiva et al. 2002; Walker and Ben-Akiva 2002, for methods associated with these extensions).

In summary, the latent class choice modeling approach presented in this paper provides a powerful method for inferring meaningful latent lifestyle segments and resulting behavior in a way that has not been done before. Further, the empirical results point to intriguing policy implications suggested by the complexity of preferences towards residential housing characteristics. There are many directions in which to extend this work to provide richer and more robust inferences regarding lifestyle preferences.

Acknowledgments The authors gratefully acknowledge useful interactions with Pat Mokhtarian (who provided, among other things, particularly insightful comments on the policy implications), Antonio Páez, Darren Scott, two anonymous reviewers, and participants at a UC Davis seminar and the 52nd North American Regional Science Association International Conference in Las Vegas, Nevada.

References

Aeroe T (2001) Residential preferences, choice of housing, and lifestyle. PhD dissertation (English summary), Aalborg University

Bagley M, Mokhtarian P (1999) The role of lifestyle and attitudinal characteristics in residential neighborhood choice. Transportation and Traffic Theory, pp 735–758

Bagley M, Mokhtarian P (2002) The impact of residential neighborhood type on travel behavior: a structural equations modeling approach. Ann Reg Sci 36:279–297

Ben-Akiva M, Bowman JL, Gopinath D (1996) Travel demand model system for the information era. Transportation 23:241–266

Ben-Akiva M, Walker J, Bernardino A, Gopinath D, Morikawa T, Polydoropoulou A (2002) Integration of choice and latent variable models. In: Mahmassani H (ed) In perpetual motion: travel behaviour research opportunities and application challenges. Elsevier, Amsterdam, pp 431–470

Bhat C, Guo J (2006) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. In: Presented at the 85th Annual Meeting of the Transportation Research Board, Washington DC

Boxall PC, Adamowicz WL (2002) Understanding heterogeneous preferences in random utility models: a latent class approach. Environ Resour Econ 23:421–446

Cambridge Systematics, Inc. (1996) Data collection in the Portland, Oregon metropolitan area, Travel Model Improvement Program, Track D Data Research Program. Prepared for the U.S. Department of Transportation

Cao X, Mokhtarian P (2005) How do individuals adapt their personal travel? Objective and subjective influences on the consideration of travel-related strategies for San Francisco Bay Area commuters. Transp Policy 12:291–302

Cao X, Handy S, Mokhtarian P (2006) The influence of the built environment and residential self-selection on pedestrian behavior. Transportation 33(1):1–20

Chiou L, Walker JL (2006) Masking identification of discrete choice models under simulation methods. J Econom (forthcoming)

Chliaoutakis J, Koukouli S, Lajunen T, Tzamalouka G (2005) Lifestyle traits as predictors of driving behavior in urban areas of Greece. Transp Res Part F, 1–16

Dillon WR, Kumar A, de Borrero MS (1993) Capturing individual differences in paired comparisons: an extended BTL Model incorporating descriptor variables. J Mark Res 30(1):42–51

Ferrell C, Deakin E (2001) Changing California lifestyles: consequences for mobility. The University of California Transportation Center, Berkeley, available from http://www.uctc.net/trends

Gopinath DA (1995) Modeling heterogeneity in discrete choice processes: application to travel demand. PhD dissertation, Massachusetts Institute of Technology

Greene WH (2003) Econometric analysis, 5th edn. Prentice Hall, Englewood Cliffs

Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res Part B 37:681–698

Grover R, Srinivasan V (1987) A simultaneous approach to market segmentation and market structuring. J Mark Res XXIV, 139–153

Guevara CA, Ben-Akiva ME (2006) Endogeneity in residential location choice models. Transp Res Rec (forthcoming)

Handy S, Cao X, Mokhtarian P (2005) Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transp Res D 10:427–444

Hojrup T (2003) State, culture and life-modes: the foundations of life-mode analysis. Ashgate

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. J Mark Res 26(4):379–390

Kitamura R, Mokhtarian P, Laidet L (1997) A Micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. Transportation 24:125–158

Kontoleon A, Yabe M (2003) Assessing the impacts of alternative 'opt-out' formats in choice experiment studies: consumer preferences for genetically modified content and production information in food. J Agric Policy Res 5:1–43

Krizek K, Waddell P (2003) Analysis of lifestyles choices: neighborhood type, travel patterns, and activity participation. Transp Res Rec 1807:119–128

Lerman SR (1975) A disaggregate behavioral model of urban mobility decisions. PhD dissertation, Massachusetts Institute of Technology
Lindberg E, Harti T, Garvill J, Gärling T (1992) Residential-location preferences across the life span. J Environ Psychol 12:187–198
Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and applications. Cambridge University Press, Cambridge
Louviere J, Train K, Ben-Akiva M, Bhat C, Brownstone D, Cameron TA, Carson RT, Deshazo JR, Fiebig D, Greene W, Hensher D, Waldman D (2005) Recent progress on endogeneity in choice modeling. Mark Lett 16(3–4):255–265
Lyons G, Chatterjee K, Beecroft M, Marsden G (2002) Determinants of travel demand: exploring the future of society and lifestyles in the UK. Transp Policy 9:17–27
Magidson J, Eagle T, Vermunt J (2003) New developments in latent class choice models. In: Sawtooth Software Conference Proceedings, pp 89–112
Milon JW, Scrogin D (2006) Latent preferences and valuation of wetland ecosystem restoration. Ecol Econ 56:162–175
Moekel R, Spiekermann K, Schurmann C, Wegener M (2003) Microsimulation of land use. Int J Urban Sci 7(1):14–31
Mokhtarian PL, Salomon I (1997) Modeling the desire to telecommute: the importance of attitudinal factors in behavioral models. Transp Res A 31(1):35–50
Ory DT, Mokhtarian PL (2005) When is getting there half the fun? Modeling the liking for travel. Transp Res Part A 39:97–123
Prevedouros P (1992) Associations of personality characteristics with transport behavior and residence location decisions. Transp Res A 26:381–391
Salomon I, Ben-Akiva M (1983) The use of the lifestyle concept in travel demand models. Environ Plann A 15:623–638
Salomon I, Waddell P, Wegener M (2002) Sustainable lifestyles? Microsimulation of household formation, housing choice and travel behavior. In: Black WR, Nijkamp P (eds) Social Change and Sustainable Transport. Indiana University Press, pp 125–131
Scarpa R, Thiene M (2005) Destination choice models for rock climbing in the Northeastern Alps: a latent-class approach based on intensity of preferences. Land Econ 81(3):426–444
Schwanen T, Mokhtarian P (2005) What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? J Transp Geogr 13:83–99
Swait J (1994) A structural equation model of latent segmentation and product choice for cross-sectional preference choice data. J Retail Consum Serv 1(2):77–89
Swait J, Sweeney JC (2000) Perceived value and its impact on choice behavior in a retail setting. J Retail Consum Serv 7:77–88
Targa F, Clifton K (2004) Integrating social and psychological processes into the land use-travel behavior research agenda: theories, concepts and empirical study design. In: Prepared for the Seventh International Conference on Travel Survey Methods, Los Suenos, Costa Rica
Train KE (2003) Discrete choice models with simulation. Cambridge University Press, Cambridge
TRB Special Report 282 (2005) Does the built environment influence physical activity? Examining the evidence. (S Hanson, Chair) Transportation Research Board, Washington DC
Waddell P (2000) Towards a behavioral integration of land use and transportation modeling. In: Proceedings of the 9th International Association for Travel Behavior Research Conference, Queensland, Australia
Walker JL (2001) Extended discrete choice models: integrated framework, flexible error structures, and latent variables. PhD dissertation, Massachusetts Institute of Technology
Walker J, Ben-Akiva M (2002) Generalized random utility model. Math Soc Sci 43(3):303–343
Walker J, Ben-Akiva M, Bolduc D (2006) Identification of parameters in normal error components logit mixture models. Working paper available at http://www.people.bu.edu/joanw/papers.html