UNDERSTANDING THE MULTIPLE DIMENSIONS OF RESIDENTIAL CHOICE

Xuemei Fu
Institute of Transportation Studies
Department of Management Science
Antai College of Economics & Management
Shanghai Jiao Tong University
535 Fahua Zhen Rd., Shanghai 200052 P. R. China
Tel: 1-881-731-2792
E-mail: shuiyeyuzi@sjtu.edu.cn

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu
and
King Abdulaziz University, Jeddah 21589, Saudi Arabia

Ram M. Pendyala
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-894-2201; Fax: 404-894-2278
Email: ram.pendyala@ce.gatech.edu

Sravani Vadlamani
Arizona State University
School of Sustainable Engineering and the Built Environment
Tempe, AZ 85287-3005
Tel:(480-965-3589; Fax: 480-965-0557
Email: sravani.vadlamani@asu.edu

Venu M. Garikapati
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 480-522-8067; Fax: 404-894-2278
Email: venu.garikapati@gatech.edu

ABSTRACT

Residential choice may be characterized as a household's simultaneous decisions of location, neighborhood, and dwelling. Traditional models do not account for the latent unmeasured constructs which capture individuals' preferences for and attitudes towards residence and mode choice. This paper employs Bhat's (2015) Generalized Heterogeneous Data Model (GHDM) to accommodate five inter-related residential choice dimensions, including residential location, neighborhood land-use pattern, public transportation availability, housing type, and dwelling ownership. Four latent variables including pro-driving, pro-public transportation, facility availability, and residential spaciousness are constructed to capture individuals' attitudes towards travel modes and preferences for residential features. The inclusion of these latent constructs helps account for self-selection effects in residential choice processes. The determination of relationships among multiple dimensions of residential choice behavior, socio-demographics, and latent attitudes and preferences is critical to integrated land use – transport modeling and the formulation of policies as well as urban residential and neighborhood environments that cater to individual preferences and enhance quality of life.

KEYWORDS

Residential choice, latent variables, integrated choice and latent variable model, MACML.

1. INTRODUCTION

Residential land use occupies about two-thirds of all urban land (Guo and Bhat, 2007), indicating its central role in land-use planning. As an anchor point where individuals live with their families and start out-of-home activities (such as working, shopping, and recreation), residential location has an important effect on people's well-being, social status, and access to jobs, schools, and social networks (Mulder, 2007). Due to its multidisciplinary nature, residential choice has been the focus of study for engineers and planners, environmental designers, urban geographers, economists, architects, sociologists, and psychologists. From an activity-travel demand modeling perspective, it is essential for transportation planners to fully capture the decision mechanism underlying residential choice because of its long-standing influence on travel behavior (Srinivasan and Ferreira, 2002). For instance, individuals in a residential location with no public transit accessibility tend to use private vehicles more frequently than those who live in a neighborhood with convenient public transportation service. Given the important role residential location plays in the spatial distribution of people's activities and travel, it is conceivable that changes in travel behavior (towards more sustainable activity-travel patterns and choices) may be brought about through appropriate designs of the built environment and residential landscape. The recognition of the interactions between residential environment and transportation systems is fundamental to the application of integrated land-use and transportation modeling approaches in the metropolitan planning process (Waddell *et al*., 2007).

The issue with many extant residential choice models is that they singularly focus on the choice of spatial unit, *i.e.*, the location expressed as a dwelling unit, parcel, block, tract, or zone. However, it is conceivable that households, when making residential location choices, are choosing a bundle of attributes related to the environment in which they intend to reside. There is considerable evidence in the literature that alludes to the bundled nature of the residential choice phenomenon. Waddell (2001) notes that residential choice is a conglomeration of related dimensions including the location type, dwelling ownership (own or rent), neighborhood land use pattern, and type of housing. Harold and Leonard (1991) suggest that households make a simultaneous determination of the type of housing unit and residential location in the context of residential choice. Studies in the field of microeconomics also emphasize the necessity of simultaneously analyzing residence-related decisions (Barrios-García and Rodríguez-Hernández, 2008). According to Dieleman and Mulder (2002), residence selection includes both choice of a certain residential environment and type of dwelling. Jansen (2012) pointed out that residential choice involves multiple aspects, including the physical characteristics of available homes (*e.g.*, housing type, number of bathrooms, and number of bedrooms) and the regional or social characteristics of a neighborhood (*e.g.*, proximity to a workplace).

Previous studies have contributed to enhancing the conceptual understanding of factors influencing the dimensions of residential choice, and advancing the methodological approaches to residential choice modeling. However, the multi-faceted nature of residential choice processes has been relatively under-developed because of the inherent computational challenges associated with modeling multiple choice dimensions in an integrated simultaneous equations framework. In particular, many earlier studies either focus solely on the residential location dimension or examine one or two non-spatial dwelling unit dimensions (see, for example, Rashidi *et al*., 2012, Coulombel, 2010, Flavin and Nakagawa, 2008, and Frenkel and Kaplan, 2014; Zolfagiri *et al*., 2013 provides an extensive and recent review of this literature). The nested model structure and combinations of feasible alternatives of each choice dimension are the most common approaches used in these earlier studies when two or more residential choice dimensions are considered (see, for example, Quigley, 1976, Lerman, 1977, Boheim and Taylor, 1999, and Frenkel and Kaplan, 2014). But an increase in choice dimensions beyond two to three makes it difficult to define the choice set as the structure of the nested model becomes rather complex and the number of combinations of alternatives will be extremely large which may result in a computationally intractable model.[1] This paper aims to make a contribution to the simultaneous modeling of multiple residential choice dimensions using a novel integrated choice modeling approach that offers computational tractability.

Another important aspect related to residential choice is that housing choice is a lifestyle choice. That is, traditional socio-economic characteristics such as income (Lee and Waddell, 2010) and lifecycle stage (Chen *et al*., 2013) are insufficient to explain housing choice behavior (see Bhat and Guo, 2007, Van Wee, 2009, and Bhat and Eluru, 2009). For example, Fleischer (2007) reinforces the notion that "to choose a house means to choose a lifestyle" in his investigations based on qualitative data from ethnographic fieldwork. Aeroe (2001) also notes that housing and residential choices are a mechanism through which one attempts to realize lifestyle preferences. Many earlier studies have explicitly acknowledged the presence of these intrinsic psycho-social effects (see Van Acker *et al*., 2011, Bohte *et al*., 2009, and Bhat *et al*., 2014 for extensive reviews), though these earlier studies consider lifestyle and attitude-related variables in modeling only the location dimension of residence. For example, Handy and Clifton (2001) found that individuals who prefer walking to stores tend to choose residential neighborhoods with higher accessibility. Schwanen and Mokhtarian, 2005 and Pinjari *et al*., 2009 suggest that households that intend to drive less and be physically active are more likely to live in neighborhoods with abundant recreational facilities and sidewalks. Schwanen and Mokhtarian 2007 also point out that the choice of a suburban neighborhood could be attributed to an individual's enjoyment of fast, flexible, and comfortable car travel, or the perception of cars as status symbols. In other words, the literature provides evidence of attitudes, preferences, and lifestyle desires playing a significant role in influencing residential location choice. Yet, virtually all earlier studies consider such intrinsic lifestyle considerations only in modeling the location dimension, ignoring the impacts of such considerations on other non-spatial dimensions of the housing decision. In many ways, this is because few studies examine location and non-location dimensions simultaneously, but even the few that jointly model a limited number of non-location dimensions do not explicitly accommodate the effects of underlying attitudinal and lifestyle preferences. Indeed, we believe that the jointness in the many dimensions of the housing decision originates in such underlying lifestyle and attitudinal preferences. For example, families that have a "green lifestyle" preference with a favorable perception of public transportation may locate in high density neighborhoods, while also preferring transit-friendly, mixed land-use, and rented apartment living. This paper intends to address this issue through the incorporation of

[1] For example, in the empirical study of this paper, there are 72 alternatives based on the combinations of the five dimensions under study. The number of possible nested structures in the traditional approach explodes to over 200. Further, with the 1300 observations available, this leads to an average sample size of 18 per alternative, which indicates how the explosion in alternatives can lead also to statistical estimation problems given the small average sample size per alternative. Some other studies side-step the issue of multiple dimensions by examining a single discrete variable (such as housing tenure) and quantity of housing demand (continuous choice), but these studies use hedonic relationships to estimate the quantity of housing demand as the market value of a dwelling unit divided by a constructed price of a standardized unit of the flow of housing services. However, the demand for housing services in such studies is rather abstract and does not correspond to individual dimensions of the dwelling unit. Examples of this literature include Lee and Trost (1978), Rosen (1979), Dubin and McFadden (1984), Rouwendal and Meijer (2001), Barrios-García and Rodríguez-Hernández (2008), and Chen and Jin (2014).

latent constructs that reflect the lifestyle preferences and modal attitudes of households and individuals in residential choice analysis.

To summarize, the specific objective of this study is to simultaneously model the relationships between multi-dimensions of residential choice behavior, observable socio-demographic characteristics and individuals' latent attitudes and preferences. A comprehensive framework built on the multinomial probit (MNP)-kernel Generalized Heterogeneous Data Model (GHDM) proposed by Bhat (2015) is employed to jointly model the five dimensions of residential choice including location, neighborhood land-use pattern, public transportation availability in the neighborhood, housing type, and dwelling ownership. The data set used in this study is derived from the 2013 Housing, Transportation and Community survey conducted in the US.

The following section presents the data and sample used. The third section provides an overview of the modeling framework. The estimation and modeling results are presented in the fourth section. The concluding remarks and future research directions are discussed in the final and fifth section.

2. DATA

The data for the current study is derived from the 2013 Housing, Transportation and Community Survey, conducted nationwide by the Urban Land Institute (ULI) to obtain information about household preferences and satisfaction related to residential choice. The survey includes a series of questions on respondent level of satisfaction with the current home, neighborhood, and transportation facilities. The survey questions also ask the respondents to specify their future desired features for neighborhoods, homes, and transportation facilities. The survey also collects detailed socio-demographic information. Each respondent belongs to a different household (that is, only one individual is sampled per household).

The present study assumes the respondent's travel attitudes and residential preferences to represent those of the entire household of which they are a part. The residential choice behavior of all respondent types are of interest and hence specific survey questions pertaining to commuters only were excluded from the analysis. The survey sample following extensive data processing included 1300 respondents (households).

The model considers five dimensions of residential choice that are combined to reflect a household's residential choice bundle. In the modeling effort of this paper, the five dimensions of residential choice are jointly considered as dependent variables of interest. The descriptive characteristics of the choice dimensions (dependent variables) are provided in Table 1 and it is to be noted that these statistics represent information about the respondent's current residence and not their stated preference for future residence features. Within the survey sample used for this modeling effort, 21.5% of households (respondents) live in a rural area/small town, 43.5% live in a suburban area, and 35.0% live in an urban area. The majority of the households (63.4%) live in a single-family detached house followed by 24.1% in an apartment/condominium and 12.5% in a single-family attached/townhome. The proportion of households situated in a mixed land-use neighborhood versus a residential neighborhood is somewhat similar, with the former at 44.6% and the latter at 55.4%. The proportion of households that live in a neighborhood with access to public transportation (66.5%) is almost twice that of the proportion not having public transportation access (33.5%). It is found that 63.3% of the households own their home while 36.7% rent their property.

3. METHODOLOGY

This section provides an overview of the modeling process built on Bhat's (2015) Generalized Heterogeneous Data Model (GHDM) approach. This model enables the consideration of multiple ordinal, multiple count, multiple continuous, and multiple nominal variables jointly using a latent variable structural equation model that ties latent constructs to exogenous variables, and a measurement model that links the latent variables and possibly other explanatory variables to a set of different types of outcomes. The approach uses a multinomial probit kernel for the discrete (nominal, binary, and ordinal outcomes) and explains the covariance relationship among a large set of mixed data outcomes through a much smaller number of unobservable latent factors. The adoption of the MNP kernel for the nominal outcomes allows for correlations across error components of the utilities of different alternatives, and also enables the estimation of the model with relative ease using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach. In particular, in this approach, the dimensionality of integration in the composite marginal likelihood (CML) function that needs to be maximized to obtain a consistent estimator (under standard regularity conditions) for the GHDM parameters is independent of the number of latent factors and easily accommodates general covariance structures for the structural equation and for the utilities of the discrete alternatives for each nominal outcome. Further, the use of the analytic approximation in the MACML approach to evaluate the multivariate cumulative normal distribution (MVNCD) function in the CML function simplifies the estimation procedure even further so that the proposed MACML procedure requires the maximization of a function that has no more than bivariate normal cumulative distribution functions to be evaluated.

In the rest of this section, we briefly present the GHDM methodology, customized to the case of multiple ordinal indicators and multiple nominal dependent variables (the empirical analysis in this paper includes thirteen ordinal dependent indicators, two nominal dependent variables and three binary dependent variables, but the latter binary dependent variables may be considered as special cases of nominal variables with only two categories).

3.1. The GHDM Model Formulation

Let q be the index for households $(q = 1, 2, ..., Q)$, which we will suppress in parts of the presentation below. Assume that all error terms in the GHDM model for a household are independent of other household error terms.

3.1.1. Structural Equation Model

Let z_l^* be the lth latent variable (l = 1, 2,…, L) for a specific household. Write z_l^* as a linear function of covariates:

$$z_l^* = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(\tilde{D} \times 1)$ vector of observed covariates (excluding a constant), $\boldsymbol{\alpha_l}$ is a corresponding $(\tilde{D} \times 1)$ vector of coefficients, and η_l is a random error term assumed to be standard normally distributed for identification purposes (see Stapleton, 1978). Next, define the $(L \times \tilde{D})$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_L)'$, and the $(L \times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, ..., z_L^*)'$ and

$\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$. Let $\boldsymbol{\eta} \sim MVN_L[\mathbf{0}_L, \boldsymbol{\Gamma}]$, where $\mathbf{0}_L$ is an $(L \times 1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is an $(L \times L)$ correlation matrix. In matrix form, we may write Equation (1) as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha w} + \boldsymbol{\eta}. \tag{2}$$

3.1.2. Measurement Equation Model Components

Consider N ordinal outcomes (indicator variables) for the individual, and let n be the index for the ordinal outcomes $(n = 1, 2, ..., N)$. Also, let J_n be the number of categories for the n^{th} ordinal outcome $(J_n \geq 2)$ and let the corresponding index be j_n $(j_n = 1, 2, ..., J_n)$. Let y_n^* be the latent underlying continuous variable whose horizontal partitioning leads to the observed outcome for the n^{th} ordinal variable. Assume that the individual under consideration chooses the a_n^{th} ordinal category. Then, in the usual ordered response formulation, for the individual, we may write:

$$y_n^* = \gamma_n + \boldsymbol{d}_n' \boldsymbol{z}^* + \varepsilon_n, \text{ and } \psi_{n,a_n-1} < y_n^* < \psi_{n,a_n}, \tag{3}$$

where γ_n is a scalar constant, $\boldsymbol{d}_n$ is an $(L \times 1)$ vector of latent variable loadings on the n^{th} continuous outcome, the ψ terms represent thresholds, and ε_n is the standard normal random error for the n^{th} ordinal outcome. For each ordinal outcome, $\psi_{n,0} < \psi_{n,1} < \psi_{n,2} ... < \psi_{n,J_n-1} < \psi_{n,J_n}$; $\psi_{n,0} = -\infty$, $\psi_{n,1} = 0$, and $\psi_{n,J_n} = +\infty$. For later use, let $\boldsymbol{\psi}_n = (\psi_{n,2}, \psi_{n,3} ..., \psi_{n,J_n-1})'$ and $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', ..., \boldsymbol{\psi}_N')'$. Stack the N underlying continuous variables y_n^* into an $(N \times 1)$ vector $\boldsymbol{y}^*$, and the N error terms ε_n into another $(N \times 1)$ vector $\boldsymbol{\varepsilon}$. Define $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_N)'$ [$(N \times 1)$ matrix] and $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_N)$ [$(N \times L)$ matrix], and let $\mathbf{IDEN_N}$ be the identity matrix of dimension N representing the correlation matrix of $\boldsymbol{\varepsilon}$ (so, $\boldsymbol{\varepsilon} \sim MVN_N(\mathbf{0}_N, \mathbf{IDEN_N})$; again, this is for identification purposes, given the presence of the unobserved $\boldsymbol{z}^*$ vector to generate covariance. Finally, stack the lower thresholds for the decision-maker $\psi_{n,a_n-1}(n = 1, 2, ..., N)$ into an $(N \times 1)$ vector $\boldsymbol{\psi}_{low}$ and the upper thresholds $\psi_{n,a_n}(n = 1, 2, ..., N)$ into another vector $\boldsymbol{\psi}_{up}$. Then, in matrix form, the measurement equation for the ordinal outcomes (indicators) for the decision-maker may be written as:

$$\boldsymbol{y}^* = \boldsymbol{\gamma} + \boldsymbol{d z}^* + \boldsymbol{\varepsilon}, \quad \boldsymbol{\psi}_{low} < \boldsymbol{y}^* < \boldsymbol{\psi}_{up}. \tag{4}$$

Consider G nominal (unordered-response) variables $(g = 1, 2, 3, ..., G)$, with I_g being the number of alternatives corresponding to the g^{th} nominal variable $(I_g \geq 2)$ and i_g being the corresponding index $(i_g = 1, 2, 3, ..., I_g)$. Let the household under consideration choose the alternative m_g for the g^{th} nominal variable and assume the usual random utility structure for each alternative i_g:

$$\boldsymbol{U}_{gi_g} = \boldsymbol{b}_{gi_g}' \boldsymbol{x} + \boldsymbol{\vartheta}_{gi_g}' (\boldsymbol{\beta}_{gi_g} \boldsymbol{z}^*) + \varsigma_{gi_g}, \tag{5}$$

where $\boldsymbol{x}$ is a fixed $(A \times 1)$ vector of exogenous variables (including a constant), $\boldsymbol{b}_{gi_g}$ is an $(A \times 1)$ column vector of corresponding coefficients, and ς_{gi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is a

$(N_{gi_g} \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and $\boldsymbol{\vartheta}'_{gi_g}$ is an $(N_{gi_g} \times 1)$-column vector of coefficients capturing the effects of latent variables and its interaction effects with other exogenous variables (see Bhat and Dubey, 2014). Let $\boldsymbol{\varsigma_g} = (\varsigma_{g1}, \varsigma_{g2}, ..., \varsigma_{gI_g})'$ ($I_g \times 1$ vector), and $\boldsymbol{\varsigma_g} \sim MVN_{I_g}(0, \boldsymbol{\Lambda_g})$. Define $\boldsymbol{U}_g = (U_{g1}, U_{g2}, ..., U_{gI_g})'$ ($I_g \times 1$ vector), $\boldsymbol{b}_g = (\boldsymbol{b}_{g1}, \boldsymbol{b}_{g2}, \boldsymbol{b}_{g3}, ..., \boldsymbol{b}_{gI_g})'$ ($I_g \times A$ matrix), and $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, ..., \boldsymbol{\beta}'_{gI_g})'$ $\left(\sum_{i_g=1}^{I_g} N_{gi_g} \times L\right)$ matrix. Also, define the $\left(I_g \times \sum_{i_g=1}^{I_g} N_{gi_g}\right)$ matrix $\boldsymbol{\vartheta}_g$, which is initially filled with all zero values. Then, position the $(1 \times N_{g1})$ row vector $\boldsymbol{\vartheta}'_{g1}$ in the first row to occupy columns 1 to N_{g1}, position the $(1 \times N_{g2})$ row vector $\boldsymbol{\vartheta}'_{g2}$ in the second row to occupy columns $N_{g1}+1$ to $N_{g1}+N_{g2}$, and so on until the $(1 \times N_{gI_g})$ row vector $\boldsymbol{\vartheta}'_{gI_g}$ is appropriately positioned. Further, define $\boldsymbol{\varpi}_g = (\boldsymbol{\vartheta}_g \boldsymbol{\beta}_g)$ ($I_g \times L$ matrix), $\vec{G} = \sum_{g=1}^{G} I_g$, $\widetilde{G} = \sum_{g=1}^{G} (I_g - 1)$, $\boldsymbol{U} = (\boldsymbol{U}'_1, \boldsymbol{U}'_2, ..., \boldsymbol{U}'_{\boldsymbol{G}})'$ ($\vec{G} \times 1$ vector), $\boldsymbol{\varsigma} = (\boldsymbol{\varsigma_1}, \boldsymbol{\varsigma_2}, ..., \boldsymbol{\varsigma_G})'$ ($\vec{G} \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_{\boldsymbol{G}})'$ ($\vec{G} \times A$ matrix), $\boldsymbol{\varpi} = (\boldsymbol{\varpi}'_1, \boldsymbol{\varpi}'_2, ..., \boldsymbol{\varpi}'_G)'$ ($\vec{G} \times L$ matrix), and $\boldsymbol{\vartheta}_{vec} = \text{Vech}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G)$ (that is, $\boldsymbol{\vartheta}_{vec}$ is a column vector that includes all elements of the matrices $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G$). Then, in matrix form, we may write Equation (6) as:

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi}\, \boldsymbol{z}^* + \boldsymbol{\varsigma}, \tag{6}$$

where $\boldsymbol{\varsigma} \sim MVN_{\vec{G}}(\mathbf{0}_{\vec{G}}, \boldsymbol{\Lambda})$, with $\boldsymbol{\Lambda}$ as follows:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 & 0 & 0. \cdots & 0 \\ 0 & \boldsymbol{\Lambda}_2 & 0 & 0 \cdots & 0 \\ 0 & 0 & \boldsymbol{\Lambda}_3 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \dots & \vdots \\ 0 & 0 & 0 & 0 \dots & \boldsymbol{\Lambda}_G \end{bmatrix} (\vec{G} \times \vec{G} \text{ matrix}), \tag{7}$$

Let $\boldsymbol{\delta}$ be the collection of parameters to be estimated: $\boldsymbol{\delta} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vech}(\boldsymbol{\Gamma}), \boldsymbol{\gamma}, \text{Vech}(\boldsymbol{d}), \boldsymbol{\psi}, \boldsymbol{\vartheta}_{vec}, \text{Vech}(\boldsymbol{\Lambda})]$, where the operator "Vech(.)" vectorizes all the non-zero unique elements of the matrix/vector on which it operates. We will assume that the error vectors $\boldsymbol{\eta}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\varsigma}$ are independent of each other. Additional details on the GHDM formulation, including sufficiency conditions for identification of model parameters, and the MACML estimation approach for the formulation may be found in Bhat (2015). These details are suppressed here with a focus on the empirical analysis.

3.2. Behavioral Framework

The behavioral framework used in this paper to model the multiple dimensions of residential choice based on the GHDM methodology proposed by Bhat (2015) is presented in Figure 1. The framework presented in Figure 1 is merely a graphical representation of the directions of relationships between variables in place of a full path diagram, and for purposes of brevity, does not list specific indicators for the latent variables.

We developed four latent variables to characterize attitudes and lifestyles, as shown in Figure 1. In doing so, we used earlier studies investigating (directly or indirectly) general modal attitudes and lifestyle-related characteristics that affect residential choice decisions. In particular, based on the extensive principal components study of over 18 attitudinal statements regarding housing choices undertaken by Schwanen and Mokhtarian (2007), we identified *pro-driving* and *pro-public transportation* as the two household attitudes towards specific transportation modes. Similarly, based on the qualitative studies undertaken by Handy and Clifton (2001) and Fleischer (2007), we identified *facility accessibility* and the preference for privacy, spaciousness, and exclusivity (or *luxury living* for short) to represent household desires and preferences for residential features. Our expectation is that households with a pro-driving modal disposition will prefer residing in single family owned houses in exclusive non-urban residential enclaves, while those with a pro-public transportation disposition (which is also tied closely with a pro-environmental lifestyle) will prefer urban apartment rented living in urban and mixed land-use areas. Similarly, households with a high preference for facility accessibility (such as proximity to shops, parks, and other public services) may be expected to prefer a mixed land-use and high-density urban environment, while households that place a premium on luxury living should prefer relatively rural, single family detached housing in exclusive residential enclaves.

The attitudinal and the perceptual ordinal indicators of the participants that were used for the construction of the four latent variables just discussed are presented in Table 2, along with the likert scale used for each indicator and the mean/standard deviation for each indicator. As the table indicates, the likert response scale is either a five point scale or a three point scale. The descriptive statistics indicate that, in general, individuals (a) believe that it is most important to maintain local streets and roads, (b) are about equally willing to pay taxes for new roads as well as better quality public transportation services, and (c) value privacy from neighbors more so than residential spaciousness.

In our analysis, we tested different loadings of the latent constructs onto the indicators in the measurement equation system, while also ensuring sufficiency conditions for econometric identification. At the same time, intuition also suggests that some latent constructs are naturally likely to load strongly on specific indicators (which formed part of the basis in the first place, along with the results from earlier studies, to identify the four latent constructs used in the analysis).[2] The final specification in the measurement equations for the indicators and for the choice model was based on statistical testing using standard predictive likelihood ratio tests.

Following the construction of the latent variables (representing lifestyle preference and attitudes towards modes), the dependent variables are modeled as a bundle of choices. The latent constructs and socio-economic and demographic variables are used as explanatory variables in the discrete choice model set. All five choices are discrete variables with choice alternatives as shown earlier in Table 1.

4. MODEL ESTIMATION RESULTS

Estimation results for different components of the GHDM model are presented in this section.

[2] Indeed, almost all applications in the transportation literature that collect a handful of indicators use such a combination of intuitiveness, judgment, and earlier studies to identify the latent constructs (see, for example, Daly *et al*., 2012, Bolduc *et al*., 2008, de Abreu e Silva *et al*., 2014, La Paix *et al*., 2013, Temme *et al*., 2008).

4.1. Latent Variable Structural Model

Table 3 presents the results of the latent variable structural equations model. Higher income individuals have a greater propensity to be pro-driving and men tend to have stronger pro-driving attitudes compared to women. These findings are consistent with those reported by Ory and Mokhtarian (2005). Individuals with lower incomes tend to be pro-public transportation possibly due to their lower levels of auto affordability and dependency on transit services. Higher education levels are associated with a pro-public transportation stance, suggesting that individuals with higher levels of education are interested in supporting and using transit if the opportunity presents itself. Older individuals, on the other hand, are less likely to be pro-public transportation. It is likely that older individuals are not as environmentally conscious as younger individuals and are therefore not as pro-public transportation as their younger counterparts. The age effect may also be reflecting physical challenges with age that make older individuals less likely to be pro-public transportation.

The latent variable on facility accessibility reflects households' desire for proximity to various facilities (*e.g.*, shops and parks) within the neighborhoods. Individuals with low income (<$50,000) exhibit a greater desire for facility accessibility, presumably because of their desire to access opportunities at low transportation costs. Men are less concerned about facility accessibility when compared with women, likely reflecting the activity-travel needs and desires of women who continue to shoulder a greater share of household obligations and responsibilities. Individuals aged between 30 and 49 years show a diminished level of need for facility accessibility, perhaps because they enjoy a high level of mobility and are able to access destinations and opportunities even if they are at a farther distance.

High-income households/individuals and households with children are likely to prefer spacious and exclusive housing units. These findings are consistent with expectations as one would expect higher income households to be interested in the luxury that larger dwelling units afford and households with children appreciate the space and capacity that larger housing units provide. It is found that men prefer spacious residences more so than women. The elderly, who may not be all that interested in maintaining a large home, and are likely to be retired on a fixed income and have smaller household sizes (with the children having moved out), express a preference for smaller housing units.

A positive correlation (0.425) was estimated between the latent variables *pro-driving* and *luxury living*. This is behaviorally intuitive as large exclusive houses with considerable privacy from neighbors are usually built in lower density suburban areas where the transportation system attributes are auto-centric and favor driving. In other words those who favor large houses are also likely to be pro-driving in nature. An urban area with a mixed land use pattern is typically more easily served by public transportation, and this seems to be a plausible explanation for the positive correlation (0.149) between the latent variables *facility availability* and *pro-public transportation*. In other words, those who consider access to facilities important are likely to favor neighborhoods that are dense, have mixed land use, and are well served by transit.

4.2. Latent Variable Measurement Model for Ordinal Indicators

The estimation results for the latent variable measurement equation model for the ordinal indicators are presented in Table 4. In the presentation, we do not provide the threshold parameters that govern the mapping of the underlying latent propensity of the ordinal indicators to the actual observed ordinal categories. These thresholds do not have a substantive interpretation, and are available on request from the authors.

As indicated earlier, thirteen outcomes/indicators that contribute to the latent attitudes and preferences of the households are included in the model to provide measurement scales for the four latent variables. The results of the measurement equation model are quite intuitive and consistent with expectations. The measured attitudinal indicators that contribute to the pro-driving construct include the importance that individuals attach to expanding highways and maintaining local streets and roads, and willingness to pay for new roads. All of the constants are positive and factor loadings are significant, suggesting that individuals who score high on these variables are clearly auto-centric pro-driving in their attitude. On the other hand, those who are pro-public transportation indicate a greater willingness to pay taxes for expanding public transportation, consider the expansion of public transit important, and are willing to pay taxes to improve existing bus and rail services. The desire for accessibility to facilities is a latent construct that is represented by indicators representative of an individual's preference for a neighborhood that provides easy access to various amenities such shops and restaurants, parks and playgrounds, places to walk or exercise, and large discount and warehouse stores. Proximity to shops, restaurants, and large discount warehouse stores appear to contribute more strongly to the facility accessibility construct. The luxury living lifestyle is captured by the importance that an individual attaches to having a large house, the importance of having privacy from neighbors, and the importance of buying as large a house as possible. The measurement equations provide a basis to use measured attitudinal indicators to construct a parsimonious set of latent constructs (as estimated in the structural equation system) that may be used as explanatory variables in residential type choice models. In particular, typical household travel surveys do not collect information about attitudes and latent constructs (lifestyle preferences and perceptions of different modes). It is therefore essential to have a structural equations model system that relates the unmeasured latent constructs to observed and measured explanatory variables typically available in travel surveys. Through such a structural equations model system, the latent constructs can be estimated for each individual as a function of socio-economic and demographic characteristics. The estimated constructs, together with socio-economic variables, built environment attributes, network level of service variables, and accessibility indicators may then be included in residential type choice model specifications thus providing a mechanism to account for latent constructs in residential choice behavior models.

4.3. Latent Variable Measurement Model Results of Discrete Choice Models

The estimation results of the discrete choice models with five dependent variables (including two nominal and three binary) are displayed in Table 5. The "Rural Area" is selected as the base alternative for the residential location choice model. The positive and negative constants for suburban and urban alternatives indicate there is a baseline preference for residing in a suburban neighborhood characterized by larger homes, auto-centric transportation systems, and low-medium density of land use. Individuals aged between 30 and 49 years and those with higher education levels (Bachelors or Post Graduation) have a higher propensity to live in suburban areas. Minority households (black and Asian) tend to locate in urban areas more so than white households, a finding that has been well established in the literature. Single individuals have a greater inclination to reside in urban areas, probably due to the accessibility to a number of activity opportunities that such locations offer, a finding that is also reported by Bagley and Mokhtarian (1999). Employed individuals opt for urban areas, presumably to keep commute durations low and access a variety of employment opportunities (Schwanen and Mokhtarian, 2007), while retired people tend to prefer suburban and rural areas over urban areas possibly seeking a quieter lifestyle. As expected, the latent variables significantly influence choice of residential location. Those who are pro-driving prefer suburban areas, those who are pro-public transportation prefer urban areas and are less likely to reside in rural areas, and those favoring a luxury lifestyle are likely to seek the space they desire in rural areas. All of these findings are intuitive and consistent with expectations.

For the housing type model, single-family detached house is the chosen base alternative. The negative constants for both apartment and single-family attached house alternatives indicate that the baseline preference is in favor of the choice of a detached single-family dwelling. Individuals with higher levels of education are less likely to reside in single family attached housing units while those with just a high school diploma are more likely to reside in such units. Relative to households in the highest income category, households in the lower income category are more likely to live in apartments, possibly due to cost considerations. Blacks and Asians are more likely to reside in apartments. As expected, single and separated individuals are more likely to reside in apartments presumably because of the smaller sample sizes and lower incomes of these household types and a desire to increase the social opportunities to meet other people. Married individuals are also less likely to reside in single family attached housing units. All of these findings are consistent with expectations and residential patterns of choice observed in the real world and reflect the role played by ethnic, income, lifecycle stage, and educational attributes on housing choice. Once again, latent variables play a significant and important role. Pro-driving individuals favor single-family detached housing and single-family attached housing (over apartments), suggesting that these individuals seek the lower density environments with this type of housing. Those who are pro-public transportation prefer apartments, which are likely to be located in higher density areas served by transit. Those with a luxury living predisposition prefer to reside in single-family detached and single-family attached housing units over apartments, a finding that is consistent with expectations.

In both the residential location and housing type models just discussed, covariance across utilities of alternatives within each model is engendered by the latent constructs. For example, a negative covariance is engendered between the urban and rural location utilities by the stochastic pro-transportation latent construct. Similarly, there is a positive covariance between the suburban and rural utilities because the suburban utility is positively influenced by the pro-driving latent variable, while the rural utility is positively influenced by luxury living lifestyle, and there is a positive correlation between the pro-driving and luxury living latent variables (see Section 4.1). In addition to these latent construct-generated covariances, we also allowed a general covariance structure for the utility differences (taken with respect to the base alternative) of the three alternatives in the residential location model as well as in the housing type model. But the resulting 2×2 covariance matrix in each of these models provided estimates that could not be statistically distinguished from a matrix with the value of 1.0 on the diagonal and the value of 0.5 on the off-diagonal. Thus, we fixed the 2×2 covariance matrix with 1.0 on the diagonals and 0.5 on the off-diagonal. This is equivalent, of course, to an IID error structure for the original three alternatives with a variance of 0.5 for each alternative. That is, after accommodating for the error heteroscedasticity and correlation in the utilities of the location alternatives and in the utilities of the housing type alternatives due to the stochastic latent constructs, there is no remaining heteroscedasticity and correlation across the utilities within each model.

The neighborhood land-use pattern is a binary choice of residential (base) or mixed land-use. Minority groups including African-American, Hispanic, and Asian individuals show a preference for mixed land-use neighborhoods, presumably due to cultural differences and social preferences. Individuals who are married or have children are less likely to choose neighborhoods with mixed land-use; these households are more likely to choose suburban housing enclaves that are more homogeneous in nature. Employed individuals prefer mixed land use environments, presumably to take advantage of the opportunities that such locations offer. On the other hand, retired individuals who may not have the same need for diverse employment and destination opportunities prefer to reside in housing-only neighborhoods. Recent movers are found to prefer mixed land use environments, possibly because they want to have easy access to various opportunities in a new and unfamiliar location. Those who are pro-public transportation prefer mixed land use environments while those who crave residential spaciousness are likely to choose housing-only enclaves (likely located in suburban areas that offer larger housing options).

In terms of public transportation availability, it is found that minority groups prefer neighborhoods with good transit service, once again reflecting cultural differences (a greater propensity and willingness to use transit) and income considerations. The absence of public transportation is chosen the base alternative in the model. Similarly, those who are single or separated prefer neighborhoods with public transportation. Married households, on the other hand, are likely to favor areas not well served by transit (suburban areas, for example). Also, individuals who moved in the past three years prefer to reside in neighborhoods with access to public transportation, a finding that reinforces the result reported previously where recent movers prefer to reside in areas with mixed land use. As expected, the latent construct depicting a pro-public transportation attitude is found to significantly favor the choice of a neighborhood that is well served by transit.

In the dwelling ownership model, it is found that older individuals tend to own a home when compared with younger counterparts possibly due to household size and financial security effects. Those with a lower level of education are less likely to own their home. Minority groups are less likely to own a home, reflecting financial credit disparities and also market discrimination that may be contributing to differential levels of home ownership. These results are consistent with those reported by Harold and Leonard (1991). Married individuals, and employed and retired individuals (as opposed to homemakers, students, and unemployed individuals) are more likely to own a home. Recent movers are more likely to rent, a finding consistent with expectations as individuals may choose to explore an area for a while before purchasing a home. Pro-driving individuals are likely to own a home (possibly in an auto-centric suburban area), while pro-transit individuals and individuals who desire facility accessibility are likely to rent (likely in transit-friendly mixed land use areas). These findings point to the important and significant role played by latent constructs in home ownership and residential choice.

Based on the above results, it can be concluded that the bundle of residential choices, corresponding to location, neighborhood land use pattern, availability of public transportation, housing unit type, and dwelling ownership status, and the latent variables corresponding to household preferences and attitudes are closely related.

4.4. Data Fit

The performance of the GHDM structure used here may be compared to a more restrictive model that does not consider latent constructs, but includes the determinants of the latent constructs as explanatory variables. Essentially, this is an independent model in that the error term correlations across the dimensions are ignored, but the best specification of the explanatory variables (including those used in the GHDM model in the structural equation system to explain the latent constructs) is considered to explain the residential choice dimensions. We will refer to this as the independent heterogeneous data model (or IHDM model). The GHDM and the IHDM models are not nested, but they may be compared using the composite likelihood information criterion (CLIC), which takes the following form:

$$\log L^{*}_{CML}(\hat{\boldsymbol{\theta}}) = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \tag{8}$$

The model that provides a higher value of CLIC is preferred. Another way to examine the performance of the two models is to compute the equivalent GHDM predictive household-level likelihood value and computing the log-likelihood value across all households at convergence $\mathscr{L}(\hat{\boldsymbol{\theta}})$. The corresponding IHDM predictive log-likelihood value may also be computed. Then, one can compute the adjusted likelihood ratio index of each model with respect to the log-likelihood with only the constants for each dimension in the IHDM model:

$$\bar{\rho}^2 = 1 - \frac{\mathscr{L}(\hat{\boldsymbol{\theta}}) - M}{\mathscr{L}(c)}, \tag{9}$$

where $\mathscr{L}(\hat{\boldsymbol{\theta}})$ and $\mathscr{L}(c)$ are the log-likelihood functions at convergence and at constants, respectively, and M is the number of parameters (not including the constant(s) for each dimension) estimated in the model. To test the performance of the two non-nested models (*i.e.* the GHDM and IHDM models) statistically, the non-nested adjusted likelihood ratio test may be used. This test determines if the adjusted likelihood ratio indices of two non-nested models are significantly different. In particular, if the difference in the indices is $(\bar{\rho}_2^2 - \bar{\rho}_1^2) = \tau$, then the probability that this difference could have occurred by chance is no larger than $\Phi\left\{-[-2\tau\mathscr{L}(c) + (M_2 - M_1)]^{0.5}\right\}$ in the asymptotic limit. A small value of the probability of chance occurrence indicates that the difference is statistically significant and that the model with the higher value of adjusted likelihood ratio index is to be preferred.

The results of our data fit evaluation are provided in Table 6. The CLIC values in Table 6 clearly favor the GHDM model over the IHDM model. The same result is obtained when comparing the predictive likelihood values, the predictive adjusted likelihood ratio indices, and computing the non-nested likelihood ratio statistic; the probability that the adjusted likelihood ratio index difference between the GHDM and the IHDM models could have occurred by chance is literally zero.

5. CONCLUSIONS

Modeling of residential location choice has been an important facet of travel demand forecasting due to the important role that residential location attributes play in shaping daily activity-travel patterns. Residential location models have hitherto focused largely on predicting the spatial unit (such as a traffic analysis zone) chosen by households for their residence. However, there is a growing recognition that residential choice involves a bundle of multiple dimensions that are interrelated. In order to more comprehensively model residential choice processes of households, this paper employs the Generalized Heterogeneous Data Model framework proposed by Bhat (2015) to jointly model five residential choice dimensions including location type, neighborhood land-use pattern, public transportation availability in the neighborhood, housing type, and dwelling ownership. Four latent variables that describe individual/household travel attitudes and residential preferences are considered to help account for self-selection effects in explaining and modeling household residential location decisions.

The current study utilizes data from the 2013 Housing, Transportation and Community Survey conducted nationwide by the Urban Land Institute (ULI). The study offers several important findings. The latent constructs depicting individual attitudes towards travel modes and lifestyle preferences were found to play an important role in the multiple dimensions of residential choice. In general, individuals (or households) making residential choice decisions seek the residence that best satisfies their array of lifestyle and modal preferences. This is an important consideration in the context of implementing policies that aim to modify travel behavior through changes in the built environment and land use, as the effects of such policies cannot be accurately estimated without considering the self-selection effects that derive from attitudes and preferences. As expected, individual and household socio-demographic characteristics are found to be strongly associated with residential choice. For example, retirees have a greater propensity to own their homes, and are less likely to live in urban neighborhoods with mixed land-use patterns. Education, race, employment status, marital status, and age are other socio-economic variables that play a significant role in shaping residential choices.

The results confirm the key role played by latent attitudinal and lifestyle variables in shaping the multiple dimensions of residential choice. Specifically, residential location, neighborhood types (including land-use pattern and public transportation availability), housing type, and dwelling ownership are all endogenous variables depicting residential choice, emphasizing the need for multidimensional modeling of residential attributes. The construction of a latent variable structural equations model offers the ability to estimate latent attitudinal and lifestyle constructs as a function of observed socio-economic and demographic variables, and include such constructs in models of residential choice. Thus the model system presented in this paper overcomes the challenge associated with including attitudinal and lifestyle variables that are not typically observed in travel surveys, in residential choice model specifications. The model system presented in this paper should be extended to include built environment and level of service variables, along with additional choice dimensions such as housing cost and housing unit configuration (square feet, number of rooms, year of construction) to develop a comprehensive model system of residential choice.

ACKNOWLEDGEMENTS

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The corresponding author would also like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany. Finally, the authors are grateful to Lisa Macias for her help in formatting this document and to Subodh Dubey for help with coding.

REFERENCES

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

Bagley, M.N. and Mokhtarian, P.L. 1999. The role of lifestyle and attitudinal characteristics in residential neighborhood choice. *Transportation and Traffic Theory*, 735-758.

Barrios García, J.A. and Rodríguez Hernández, J.E. 2008. Housing demand in Spain according to dwelling type: Microeconometric evidence. *Regional Science and Urban Economics* 38(4), 363-377.

Bhat, C.R. 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45(7), 923-939.

Bhat, C.R. 2015. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research, Part B*, forthcoming.

Bhat, C.R. and Dubey, S.K. 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, 67(9), 68-85.

Bhat, C.R. and Eluru, N. 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, 43(7), 749-765.

Bhat, C.R. and Guo, J.Y. 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B*, 41(5), 506-526.

Bhat, C.R., Astroza, S., Sidharthan, R., Jobair Bin Alam, M., Khushefati, W.H. 2014. A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B*, 68, 31-51.

Böheim, R. and Taylor, M., 1999. Residential mobility, housing tenure and the labour market in Britain. Institute for Social and Economic Research, University of Essex.

Bohte, W., Maat, K., and van Wee, B. 2009. Measuring attitudes in research on residential self-selection and travel behavior: A review of theories and empirical research. *Transport Reviews*, 29(3), 325-357.

Bolduc, D., Boucher, N., and Alvarez-Daziano, R. 2008. Hybrid choice modeling of new technologies for car choice in Canada. *Transportation Research Record: Journal of the Transportation Research Board*, 2082, 63-71.

Chen, J. and Jin, M., 2014. Income elasticity of housing demand in China: micro-data evidence from Shanghai. *Journal of Contemporary China*, 23(85), 68-84.

Chen, R.B., Gehrke, S.R., Liu, J.H., Jang, Y. and Clifton, K.J. 2013. Exploring engagement in household activities and decisions on residential tenure and household type. *Transportation Research Record: Journal of the Transportation Research Board*, 2344, 68-78.

Coulombel, N. 2010. Residential choice and household behavior: State of the art. SustainCity Working Paper, 2.2 a, ENS Cachan.

Daly, A., Hess, S., Patruni, B., Potoglou, D., and Rohr, C. 2012. Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behavior. *Transportation*, 39(2), 267-297.

de Abreu e Silva, J., Sottile, E., and Cherchi, E. 2014. Effects of land use patterns on tour type choice: Application of a hybrid choice model. *Transportation Research Record: Journal of the Transportation Research Board*, 2453, 100-108.

Dieleman, F.M. and Mulder, C.H. 2002. The geography of residential choice, in *Residential Environments: Choice, Satisfaction, and Behavior*, (eds. J.I. Aragonés , G. Francescato and T. Gärling), 35-54. Bergin and Garvey , Westport, CT.

Dubin, J.A. and McFadden, D.L. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 52(2), 345-362.

Flavin, M. and Nakagawa, S. 2008. A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence. *The American Economic Review*, 98(1), 474-495.

Fleischer, F. 2007. To choose a house means to choose a lifestyle. The consumption of housing and class-structuration in urban China. *City & Society*, 19(2), 287-311.

Frenkel, A. and Kaplan, S. 2014. The joint choice of tenure, dwelling type, size and location: the effect of home-oriented versus culture-oriented lifestyle. *Letters in Spatial and Resource Sciences*, 1-19.

Guo, J.Y. and Bhat, C.R. 2007. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 15(1), 31-45.

Handy, S.L. and Clifton, K.J. 2001. Local shopping as a strategy for reducing automobile travel. *Transportation*, 28(4), 317-346.

Harold, E.W. and Leonard, Z.V. 1991. Tenure choice, housing demand and residential location. *Journal of Real Estate Research*, 6(3), 341-356.

Jansen, S.J. 2012. What is the worth of values in guiding residential preferences and choices? *Journal of Housing and the built Environment*, 27(3), 273-300.

La Paix, L., Bierlaire, M., Cherchi, E., and Monzón, A. 2013. How urban environment affects travel behaviour: integrated choice and latent variable model for travel schedules. *Choice Modelling: The State of the Art and the State of Practice*, (eds. S. Hess and A. Daly), 211-228, Edward Elgar Publishing.

Lee, L.F. and Trost, R.P. 1978. Estimation of some limited dependent variable models with application to housing demand. *Journal of Econometrics*, 8(3), 357-382.

Lee, B.H. and Waddell, P. 2010. Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, 37(4), 587-601.

Lerman, S.R. 1977. Location, housing, automobile, ownership and mode to work: a joint choice model. *Transportation Research A*, 13, 1-19

Mulder, C.H. 2007. The family context and residential choice: A challenge for new research. *Population, Space and Place*, 13(4), 265-278.

Ory, D.T. and Mokhtarian, P.L. 2005. When is getting there half the fun? Modeling the liking for travel. *Transportation Research Part A*, 39(2), 97-123.

Pinjari, A.R., Bhat, C.R., and Hensher, D.A. 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B*, 43(7), 729-748.

Quigley, J.M. 1976. Housing demand in the short run: An analysis of polytomous choice. *In Explorations in Economic Research*, 3(1), 76-102.

Rashidi, T.H., Auld, J., Mohammadian, A.K. 2012. A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A*, 46(7), 1097-1107.

Rosen, H.S. 1979. Housing decisions and the US income tax: An econometric analysis. *Journal of Public Economics*, 11(1), 1-23.

Rouwendal, J. and Meijer, E. 2001. Preferences for housing, jobs, and commuting: a mixed logit analysis. *Journal of Regional Science*, 41(3), 475-505.

Schwanen, T. and Mokhtarian, P.L. 2005. What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D*, 10(2), 127-151.

Schwanen, T. and Mokhtarian, P.L. 2007. Attitudes toward travel and land use and choice of residential neighborhood type: Evidence from the San Francisco bay area. *Housing Policy Debate*, 18(1), 171-207.

Srinivasan, S. and Ferreira, J. 2002. Travel behavior at the household level: Understanding linkages with residential choice. *Transportation Research Part D*, 7(3), 225-242.

Stapleton, D.C. 1978. Analyzing political participation data with a MIMIC Model. *Sociological Methodology*, 52-74.

Temme, D., Paulssen, M., and Dannewald, T. 2008. Incorporating latent variables into discrete choice models-A simultaneous estimation approach using SEM software. *BuR-Business Research*, 1(2), 220-237.

Van Acker, V., Mokhtarian, P.L., and Witlox, F. 2011. Going soft: on how subjective variables explain modal choices for leisure travel. *European Journal of Transport and Infrastructure Research*, 11(2), 115-146.

Van Wee, B. 2009. Self-selection: A key to a better understanding of location choices, travel behaviour and transport externalities? *Transport Reviews*, 29(3), 279-292.

Waddell, P. 2001. Towards a behavioral integration of land use and transportation modeling. Presented at the 9th International Association for Travel Behavior Research Conference, Queensland, Australia.

Waddell, P., Ulfarsson, G.F., Franklin, J.P. and Lobb, J. 2007. Incorporating land use in metropolitan transportation planning. *Transportation Research Part A*, 41(5), 382-410.

Zolfaghari, A., Sivakumar, A., and Polak, J. 2013. Simplified probabilistic choice set formation models in a residential location choice context. *Journal of Choice Modelling*, 9, 3-13.

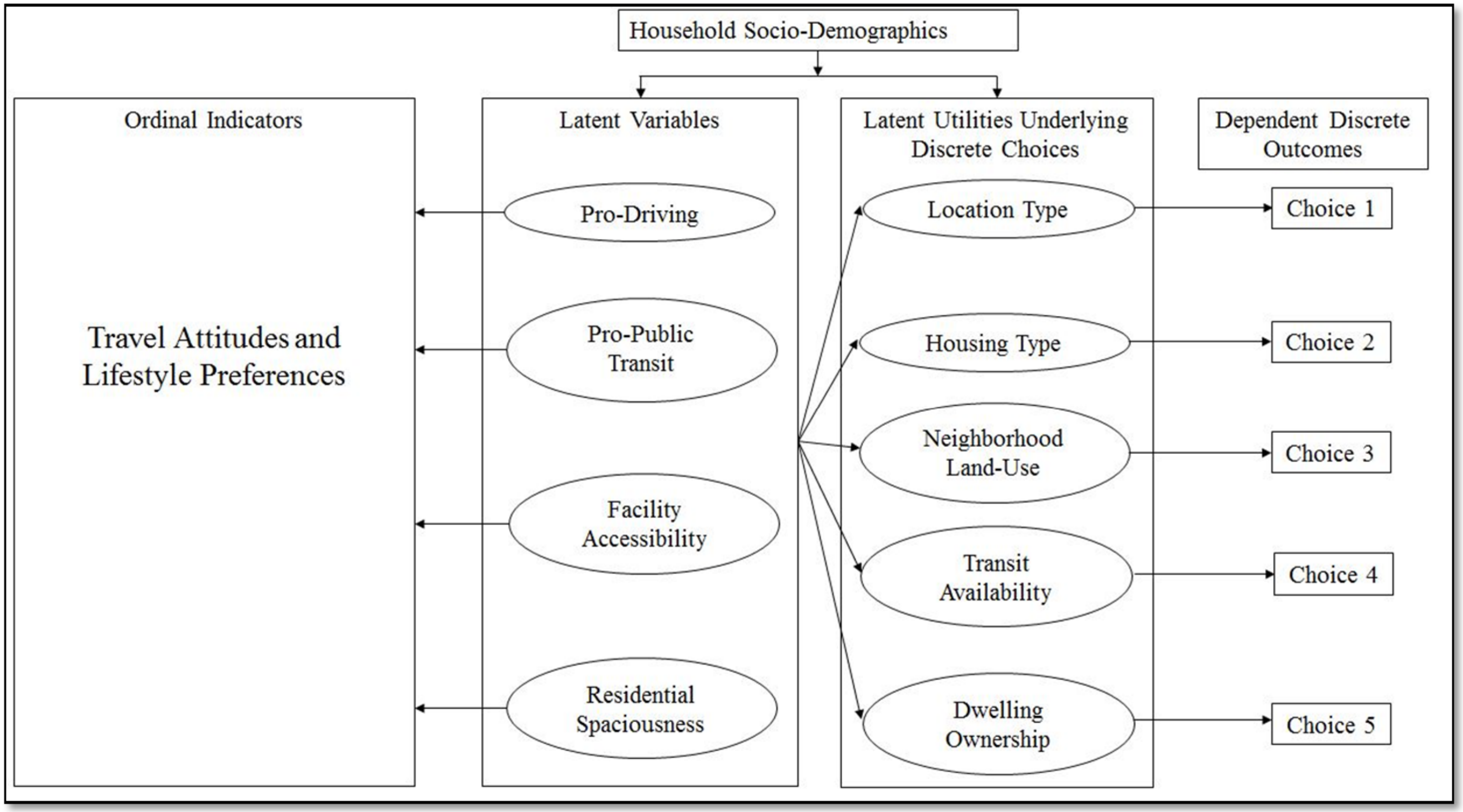

Figure 1. Behavioral Framework of Residential Choice Model

Table 1. Descriptive Statistics of Dependent Variables

Choice dimension	Alternatives	Proportion (%)
Location	Rural	21.46
	Suburban	43.46
	Urban	35.08
Housing type	Single-family detached house	63.38
	Apartment/Condominium	24.08
	Single-family attached house	12.54
Neighborhood land-use pattern	Residential	55.38
	Mixed land-use	44.62
Public transportation availability	No	33.46
	Yes	66.54
Dwelling ownership	Rent	36.69
	Own	63.31

Table 2. Indicators of the Latent Variables

Indicators	Likert Scale	Mean	Std. Dev.
Importance of “Expanding highways”	5*	3.42	1.13
Importance of “Maintaining local streets and roads”	5	4.32	0.81
Willing to pay taxes for “New roads”	3+	2.14	0.54
Importance of “Expanding local bus services”	5	3.46	1.28
Willing to pay taxes for “Expansion of public transportation, like bus or rail”	3+	2.08	0.59
Willing to pay taxes for “Better quality and service from existing public transportation, like bus or rail”	3+	2.13	0.65
Shops and restaurants within walking distance of home	3#	2.35	0.61
Parks and playgrounds in the vicinity of home	3#	2.24	0.55
Places to walk and exercise in the vicinity of home	3#	2.32	0.56
Large discount stores or warehouse stores in the vicinity of home	3#	2.20	0.60
Importance of “having a large house”	5	2.53	0.97
Importance of “privacy from neighbors”	5	3.33	0.74
Importance of “being able to buy as large a house as you can”	5	2.56	0.99

* 5-point Likert Scale: 1-Not at all important,…, 5-Very important
\+ 3-point Likert Scale: 1-Neither willing nor unwilling, 2-Fairly willing, 3-Very willing
\# 3-point Likert Scale: 1-Somewhat important, 2-Important, 3-Very important

Table 3. Latent Variable Structural Equation Model

Variable	Coefficient (t-stat)
Pro-Driving	
Household Income (base is more than $100,000)	
Less than $25,000 (Yes=1, No=0)	-1.262 (-29.01)
$25,000–$50,000 (Yes=1, No=0)	-0.652 (-28.10)
$50,001–$75,000 (Yes=1, No=0)	-0.230 (-16.08)
$75,001–$100,000 (Yes=1, No=0)	-0.020 (-1.38)
Gender (Male=1, Female=0)	0.114 (11.18)
Pro-Public Transportation	
Household Income (base is more than $100,000)	
Less than $25,000 (Yes=1, No=0)	0.168 (16.8)
$25,000–$50,000 (Yes=1, No=0)	0.044 (6.88)
Education Status (base is some college)	
Bachelor degree (Yes=1, No=0)	0.187 (30.66)
Post-graduate degree (Yes=1, No=0)	0.203 (24.17)
Age (base is 18–29 years old)	
50–64 years old (Yes=1, No=0)	-0.283 (-37.73)
Older than 64 years old (Yes=1, No=0)	-0.453 (-37.75)
Facility Availability	
Household Income (base is more than $100,000)	
Less than $25,000 (Yes=1, No=0)	0.173 (1.95)
$25,000–$50,000 (Yes=1, No=0)	0.066 (1.82)
Gender (Male=1, Female=0)	-0.200 (-2.04)
Age (base is 18–29 years old)	
30–49 years old	*-0.039 (-1.75)*
Luxury Living	
Presence of children in the household (Yes=1, No=0)	0.497 (7.41)
Household Income (base is more than $100,000)	
$50,001–$75,000	0.082 (5.19)
$75,001–$100,000	0.138 (6.00)
Gender (Male=1, Female=0)	0.135 (6.49)
Age (base is 18–29 years old)	
Older than 64 years old (Yes=1, No=0)	-0.553 (-7.48)

Table 4. Latent Variable Measurement Equation Model for Ordinal Indicators

Latent variables	Indicators	Constant (t-stat)	Factor loading (t-stat)
Pro-driving	Importance of "Expanding highways"	1.443 (44.13)	0.394 (11.42)
	Importance of "Maintaining local streets and roads"	2.173 (72.92)	0.11 (4.7)
	Willing to pay taxes for "New roads"	0.188 (8.91)	0.32 (8.31)
Pro-public transportation	Importance of "Expanding local bus services"	1.19 (21.79)	0.783 (8.23)
	Willing to pay taxes for "Expansion of public transportation, like bus or rail"	-	3.372 (1.04)
	Willing to pay taxes for "Better quality and service from existing public transportation, like bus or rail"	-	2.394 (1.94)
Facility accessibility	Attitudes: "Shops or restaurants within an easy walk of your house"	1.54 (3.86)	0.691 (1.26)
	Attitudes: "Parks and playgrounds"	1.944 (1.86)	-
	Attitudes: "Places to walk or exercise for fun"	2.342 (1.23)	-
	Attitudes: "Large discount or warehouse stores"	1.23 (5.67)	0.568 (1.41)
Luxury living	Importance of "Having a large house"	1.386 (1.43)	2.208 (1.17)
	Importance of "Privacy from neighbors"	1.87 (38.09)	0.398 (5.36)
	Importance of "Being able to buy as large a house as you can"	1.082 (2.41)	1.716 (1.75)

Table 5. Latent Variable Measurement Equation Model for the Discrete Choices

Variable	Coef	t-stat	Coef	t-stat	Coef	t-stat
Residential location (base: rural area)	**Rural**		**Suburban**		**Urban**	
Constant			0.502	15.79	-0.826	-10.35
Socio-demographic attributes						
Age (base is 18–29 years old)						
30–49 years old (Yes=1, No=0)			0.043	4.22		
Education Status (base is some college)						
High school degree (Yes=1, No=0)			-0.183	-17.10		
Bachelor degree (Yes=1, No=0)			0.086	7.75		
Post-graduate degree (Yes=1, No=0)			0.143	9.17		
Race (base is white)						
Black (Yes=1, No=0)					0.863	13.99
Asian (Yes=1, No=0)					0.652	13.15
Marriage status (base is unmarried and living with partners)						
Married (Yes=1, No=0)			-0.220	-19.47		
Single (Yes=1, No=0)					0.520	14.02
Employment status (base is others, including students, homemakers, and unemployed)						
Employed (Yes=1, No=0)					0.426	14.01
Retired (Yes=1, No=0)					-0.394	-11.55
Moving in the last three years (Yes=1, No=0)					0.588	13.71
Latent variables						
Pro-driving			0.539	16.23		
Pro-public transportation	-0.459	-9.72			0.282	6.18
Facility availability						
Luxury living	0.098	5.36				
House type (base is single-family detached house)	**Single-family detached house**		**Apartment**		**Single-family attached house**	
Constant			-1.850	-27.17	-1.679	-2.38
Socio-demographic attributes						
Education Status (base is some college)						
High school degree (Yes=1, No=0)					0.057	2.01
Bachelor degree (Yes=1, No=0)					-0.208	-2.37
Post-graduate degree (Yes=1, No=0)					-0.482	-2.37
Household Income (base is more than $100,000)						
$50,001–$75,000 (Yes=1, No=0)			0.036	1.67		
$75,001–$100,000 (Yes=1, No=0)			-0.004	-0.15		
Race (base is white)						
Black (Yes=1, No=0)			0.330	16.34		
Asian (Yes=1, No=0)			0.555	15.72		
Marriage status (base is unmarried and living with partners)						
Married (Yes=1, No=0)					-0.222	-2.36
Single (Yes=1, No=0)			0.323	18.56		
Separated (Yes=1, No=0)			0.514	20.16		
Moving in the last three years (Yes=1, No=0)			0.755	24.59		
Latent variables						
Pro-driving	0.776	13.59			0.405	3.04
Pro-public transportation			0.647	13.12		
Facility availability						
Luxury living	0.197	6.06			0.350	4.26

Table 5 (Cont.) Latent Variable Measurement Equation Model for the Discrete Choices

Neighborhood land-use pattern (base is only houses)	**Only houses**		**Mixed land-use**	
Constant			-0.302	-29.90
Socio-demographic attributes				
Race (base is white)				
Black (Yes=1, No=0)			0.292	27.29
Hispanic (Yes=1, No=0)			0.247	25.73
Asian (Yes=1, No=0)			0.222	14.32
Marriage status (base is unmarried and living with partners)				
Married (Yes=1, No=0)			-0.014	-1.77
Employment status (base is others, including students, homemakers, and unemployed)				
Employed (Yes=1, No=0)			0.205	23.56
Retired (Yes=1, No=0)			-0.030	-2.50
Presence of children (Yes=1, No=0)			-0.133	-9.05
Moving in the last three years (Yes=1, No=0)			0.210	26.58
Latent variables				
Pro-driving				
Pro-public transportation			0.595	18.89
Facility availability				
Luxury living	0.067	2.89		
Public transportation availability in the neighborhood (base is no)	**No**		**Yes**	
Constant			0.171	13.05
Socio-demographic attributes				
Race (base is white)				
Black (Yes=1, No=0)			0.253	25.56
Hispanic (Yes=1, No=0)			0.495	53.80
Asian (Yes=1, No=0)			0.684	41.45
Marriage status (base is unmarried and living with partners)				
Married (Yes=1, No=0)			-0.210	-18.10
Single (Yes=1, No=0)			0.118	9.37
Separated (Yes=1, No=0)			0.191	14.04
Employment status (base is others, including students, homemakers, and unemployed)				
Employed (Yes=1, No=0)			0.301	43.62
Moving in the last three years (Yes=1, No=0)			0.224	30.27
Latent variables				
Pro-driving				
Pro-public transportation			0.331	16.39
Facility availability				
Residential spaciousness				

Table 5 (Cont.) Latent Variable Measurement Equation Model for the Discrete Choices

Dwelling ownership (base is renting)	Renting		Owning	
Constant			1.872	8.59
Socio-demographic attributes				
Age (base is 18–29 years old)				
30–49 years old (Yes=1, No=0)			-0.351	-8.34
50–64 years old (Yes=1, No=0)			0.445	8.32
Education Status (base is some college)				
High school degree (Yes=1, No=0)			-0.519	-8.58
Race (base is white)				
Black (Yes=1, No=0)			-0.867	-8.61
Hispanic (Yes=1, No=0)			-0.544	-8.51
Asian (Yes=1, No=0)			-0.428	-6.90
Marriage status (base is unmarried and living with partners)				
Married (Yes=1, No=0)			0.919	8.93
Employment status (base is others, including students, homemakers, and unemployed)				
Employed (Yes=1, No=0)			0.166	6.86
Retired (Yes=1, No=0)			0.647	8.80
Moving in the last three years (Yes=1, No=0)			-1.844	-8.89
Latent variables				
Pro-driving			1.739	6.77
Pro-public transportation	0.535	6.14		
Facility availability	0.189	1.94		
Luxury living				

Table 6. Disaggregate Data Fit Measures

Summary Statistics	Model	
	GHDM	IHDM
Composite marginal log-likelihood value at convergence	-331492.20	-341029.00
Composite Likelihood Information Criterion (CLIC)	-331848.92	-341307.56
Log-likelihood at constants	-3954.25	
Predictive log-likelihood at convergence	-3376.29	-3515.74
Number of parameters	97	125
Number of observations	1300	1300
Predictive adjusted likelihood ratio index	0.122	0.079
Non-nested adjusted likelihood ratio test between the GHDM and IHDM	$\Phi[-18.58] << 0.0001$	