Travel Mode Choice Model for Mega-Regions in an Autonomous Driving World

Chandra Bhat, Aupal Mondal
September 2024

Technical Report Documentation Page

1. Report No. CM2-67	2. Government Accession No.	3. Recipient's Catalog No. ORCID: 0000-0002-0715-8121
4. Title and Subtitle Travel Mode Choice Model for Mega-Regions in an Autonomous Driving World		5. Report Date September 2024
		6. Performing Organization Code
7. Author(s) Chandra Bhat and Aupal Mondal		8. Performing Organization Report No. CM2-67
9. Performing Organization Name and Address The University of Texas at Austin School of Architecture 310 Inner Campus Drive, B7500 Austin, TX 78712		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. USDOT 69A3551747135
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Transit Administration Office of the Assistant Secretary for Research and Technology, UTC Program 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Technical Report conducted October 2021-September 2024
		14. Sponsoring Agency Code
15. Supplementary Notes Project performed under a grant from the U.S. Department of Transportation's University Transportation Center's Program.		
16. Abstract Studies in residential self-selection literature, which use attitudinal factors to model the jointness of residential and travel-related choices, assume a unidirectional impact from attitudes to behavior; however, such an assumption may be violated under several circumstances. In this study, we allow the error terms of the attitudinal factors to be correlated with the main outcomes as we jointly model residential choice, auto-ownership level and ranked-based modal preferences. In our joint model, we use Green Lifestyle Propensity (GLP) and Luxury Lifestyle Propensity (LLP) as the two stochastic latent constructs. The empirical data for this study is drawn from the 2019 multi-city Transformative Technologies in Transportation (T4) Survey for the city of Austin that elicited information regarding individuals' residential location, auto-ownership, and modal preferences through a stated preference experiment in a futuristic AV environment. Results indicate significant unobserved correlations between the latent constructs and the main outcomes; ignoring such endogeneity leads to underestimations of the "true" causal effect of high-density neighborhood (HDN) living on travel-related choices, which can have consequences for policy-making. In our analysis, the "true" causal effect of HDN living on auto-ownership suggests that, on average, the auto-ownership level would reduce by about 29% when an individual is shifted from a non-HDN to an HDN. Furthermore, the probability of using the bicycle mode for non-work pursuits is estimated to increase by 8%, and that of using a private vehicle is estimated to decrease by 3.1%, when individuals are moved from a non-HDN to an HDN.		
17. Key Words Residential self-selection; latent constructs; GHDM; auto-ownership	18. Distribution Statement No restrictions.	
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 48	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgements

Table of Contents

Executive Summary

Studies in residential self-selection literature, which use attitudinal factors to model the jointness of residential and travel-related choices, assume a unidirectional impact from attitudes to behavior; however, such an assumption may be violated under several circumstances. In this study, we allow the error terms of the attitudinal factors to be correlated with the main outcomes as we jointly model residential choice, auto-ownership level and ranked-based modal preferences. In our joint model, we use Green Lifestyle Propensity (GLP) and Luxury Lifestyle Propensity (LLP) as the two stochastic latent constructs. The empirical data for this study is drawn from the 2019 multi-city Transformative Technologies in Transportation (T4) Survey for the city of Austin that elicited information regarding individuals' residential location, auto-ownership, and modal preferences through a stated preference experiment in a futuristic AV environment. Results indicate significant unobserved correlations between the latent constructs and the main outcomes; ignoring such endogeneity leads to underestimations of the "true" causal effect of high-density neighborhood (HDN) living on travel-related choices, which can have consequences for policy-making. In our analysis, the "true" causal effect of HDN living on auto-ownership suggests that, on average, the auto-ownership level would reduce by about 29% when an individual is shifted from a non-HDN to an HDN. Furthermore, the probability of using the bicycle mode for non-work pursuits is estimated to increase by 8%, and that of using a private vehicle is estimated to decrease by 3.1%, when individuals are moved from a non-HDN to an HDN.

Chapter 1. Introduction

There is substantial interest in the land use-transportation literature on disentangling associative effects from causal effects in the impacts of residential built environment on traveler behavior. In this direction of research, motivated by the potential of influencing individual's travel-related choices through built environment (BE) configurations and policies, it is critical to understand whether the co-movement of BE characteristics and the travel-related variables is "truly" a reflection of causality or simply a spurious correlation due to intrinsic attitudinal factors leading individuals to live in specific built environments and also pursuing distinctive activity-travel patterns (see Bhat and Guo, 2007, for an extended discussion of residential self-selection considerations and alternative ways to address this concern; more recently, Guan et al., 2020 provide an extensive review of studies that accommodate residential self-selection when investigating the effect of the built environment (BE) on travel-related variables such as trip frequency, active travel, transit use, private vehicle use, travel duration, auto-ownership, and travel mode choice). Thus, it is important to explicitly model the jointness in (that is, correlation in unobserved factors impacting) residential living choice and traveler behavior choices.

In our current effort, we contribute to this thread of land use-transportation relationship by investigating residential location effects on auto-ownership levels (number of motorized four-wheelers owned by a household) and rank-based travel mode preferences of individuals, within a hypothetical futuristic autonomous vehicle (AV) landscape. Rank-based preference surveys can be exploited to achieve a certain desired precision in choice model estimation with a much smaller sample size, making ranked data surveys more cost-effective than traditional first-choice surveys (see Nair et al., 2019). Accordingly, and, to our knowledge, for the first time in the literature, we jointly model ranked modal preferences with residential choice and auto-ownership choices to understand BE effects. To do so in a parsimonious manner, we consider two stochastic latent constructs as determinants of residential choice, auto-ownership, and modal preferences within Bhat's (2015a) Generalized Heterogeneous Data Model (GHDM) framework. The latent constructs (or social lifestyle factors) are Green Lifestyle Propensity (GLP) and Luxury Lifestyle Propensity (LLP). The first latent construct relates to a general environmental-consciousness and a pro-environmental lifestyle; the second LLP construct is characterized by a penchant for consuming more, marked by a desire for privacy, spaciousness, and signaling exclusivity.

The GHDM model is based on using the latent constructs as unobserved stochastic factors that impact the underlying propensities/utilities (determining non-continuous outcomes of interest) or directly influencing a continuous outcome of interest. In this report, we will use the label "latent construct" strictly for unobserved stochastic lifestyle factors, and the label "latent variable" strictly for propensities/utilities underlying non-continuous outcomes of interest. Then, through the latent construct mediating factors, which are generally fewer than the latent variables underlying the main outcomes of interest, one can engender jointness (correlations due to unobserved factors) in a parsimonious fashion among the main outcomes of interest (see Bhat, 2015a).

One important assumption in the GHDM model (and the other factorization-based parsimonious models that are subsumed within the broad GHDM framework) is that the latent constructs are completely independent of the main outcomes of interest (high-density neighborhood or HDN living, auto-ownership, and mode choice in our empirical context). This is because the stochastic latent constructs are used in a strictly forward-facing manner to impact the latent variables underlying the main outcomes to generate correlation among the main outcomes. Such a forward-facing assumption is also theoretically embedded in the popular Theory of Planned Behavior or TPB (Ajzen, 1991), which proposes a unidirectional impact of attitudes on behavior. The underlying notion is that, together with subjective norms and behavioral control, our attitudes mold our behavioral intentions. Yet, such a unidirectional relationship may not always hold true. For example, imagine an individual with low green lifestyle propensity (GLP) who decides to reside in an HDN (characterized by higher use of sustainable modes) because of employer-provided housing incentives. Over a period of time, such individuals may begin to appreciate the 'lower-carbon-footprint' lifestyle stemming from the greater usage of sustainable modes, which may elevate a sense of self-pride for contributing positively toward environmental conservation. This, in turn, may increase the levels of GLP in such an individual. In such a case, one might reasonably argue that HDN living and the use of sustainable travel modes (the main behavioral outcomes) themselves shape the GLP latent construct (the attitudinal precedent of the behavioral outcomes, in the context of chronological time). That is, the GLP construct and the underlying latent variables determining HDN living and sustainable mode use, when measured at a particular cross-sectional point in time, would be positively correlated. Bhat and Mondal (2022) label such potential correlations between latent constructs and the main outcomes as the "latent construct endogeneity effect (or LC endogeneity effect)". Of course, this bi-directional relationship between

attitudes and behavioral choices has a temporal component to it, where attitudes affect behavior and then the behavioral actions influence attitudes over time. However, many publicly collected and available multivariate data sets on attitudes and behaviors are cross-sectional, thereby comingling attitudes and behaviors in a sense of an equilibrium state at the time of data collection. Thus, with such cross-sectional data, the analyst needs to consider the attitude-behavior data as a package observation. Ignoring the package nature of the attitude-behavior data can lead to biased estimates of the "true" causal effect of HDN living on travel-related behavior, as discussed in detail in Bhat and Mondal (2022).[1]

In the current report, we follow the flexible GHDM approach (accommodating LC endogeneity effects) proposed by Bhat and Mondal (2022), but extend their approach to the consideration of a ranked variable among the main outcomes. To our knowledge, this is the first study to employ a latent construct-based approach to model a rank-ordered travel mode variable jointly with other main outcome variable types (residential location and auto-ownership), while also recognizing, in a cross-sectional analysis context, the potential endogeneity of latent constructs in the modeling of the main outcomes. From a substantive standpoint, we contribute to the study of land use-transportation relationships by investigating residential location effects on auto-ownership levels and travel mode preferences of individuals, within a hypothetical futuristic autonomous vehicle (AV) landscape. Further, unlike the vast number of studies that focus on commute mode choice, our focus is on mode choice for non-work travel.

---

[1] After accommodating the correlations underlying the latent constructs and the manifested behaviors (in our empirical case, HDN living, auto-ownership level, and mode choice), the residual "true" causal direction of effect may be from the latent constructs to behavior, or from behavior to latent constructs (but cannot be in both directions; see Maddala, 1983, Bhat, 2015a, and Lavieri and Bhat, 2019 for a full discussion of this point). Similarly, while one observed endogenous outcome can affect another dependent outcome, in limited-dependent outcome model systems (that is, model systems that include non-continuous dependent outcomes), these endogenous outcome effects can only be recursive. In our empirical context, we consistently achieved better data fit measures when considering the causal structure direction in which (a) the latent constructs impact HDN living, auto-ownership, and mode choice, (b) HDN living impacts auto-ownership level and mode choice, and (c) auto-ownership level affects mode choice. All further discussion of the methodology will be based on these causal direction effects (after accommodating for correlations between the latent constructs and the outcomes).

Chapter 2. Methodology

For ease of presentation, we will suppress the index for decision-makers in our exposition below and assume that all error terms are independent and identically distributed across decision-makers. Following Bhat's (2015a) GHDM formulation, let l be an index for latent variables ($l=1, 2, …, L$). In our case, $L=2$, corresponding to the two latent constructs. Consider the latent construct z_l^* and write it 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. 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 $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, …, \eta_L)'$. We allow a multivariate normal (MVN) correlation structure for $\boldsymbol{\eta}$ to accommodate interactions among the unobserved latent variables. $\boldsymbol{\eta} \sim MVN_L[\boldsymbol{0}_L, \boldsymbol{\Gamma}]$, where $\boldsymbol{0}_L$ 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:

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

Now consider N ordinal outcomes (indicator variables for the latent constructs as well as main outcomes) and let n be the index for the ordinal outcomes $(n = 1, 2, ..., N)$. In our empirical context, $N=8$, corresponding to a total of six indicators of the two latent constructs (three indicators for each of the two latent constructs, as discussed later in Section 3.2) and the two ordinal main outcomes (HDN living and auto-ownership; the HDN variable is characterized as a binary variable in our empirical analysis, which can be treated as a special case of an ordinal variable with two categories). Also, let J_n be the number of categories for the nth ordinal outcome $(J_n \geq 2)$ and let the corresponding index be j_n $(j_n = 1, 2, ..., J_n)$. Let $\tilde{y}_n^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the nth 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:

$$\tilde{y}_n^* = \tilde{\boldsymbol{\gamma}}_n' \boldsymbol{x} + \tilde{\boldsymbol{d}}_n' \boldsymbol{z}^* + \tilde{\varepsilon}_n, \text{ and } \tilde{\psi}_{n,a_n-1} < \tilde{y}_n^* < \tilde{\psi}_{n,a_n} \tag{3}$$

where x is an $(A \times 1)$ vector of exogenous variables (including a constant) as well as possibly the observed values of other endogenous variables, $\tilde{\gamma}_n$ is a corresponding vector of coefficients to be estimated, $\tilde{d}_n$ is an $(L \times 1)$ vector of latent variable loadings on the nth ordinal outcome, the $\tilde{\psi}$ terms represent thresholds, and $\tilde{\varepsilon}_n$ is the standard normal random error for the nth ordinal outcome (note, however, that for the indicators (but not the main outcomes), the x vector will not appear on the right side of Equation (3); also, there are specific identification conditions for the number of non-zero elements of $\tilde{d}_n$ that can be present in each indicator equation and across all indicator equations; please see Bhat (2015a) for additional details). For each ordinal outcome, $\tilde{\psi}_{n,0} < \tilde{\psi}_{n,1} < \tilde{\psi}_{n,2} ... < \tilde{\psi}_{n,J_n-1} < \tilde{\psi}_{n,J_n}$; $\tilde{\psi}_{n,0} = -\infty$, $\tilde{\psi}_{n,1} = 0$, and $\tilde{\psi}_{n,J_n} = +\infty$. For later use, let $\tilde{\psi}_n = (\tilde{\psi}_{n,2}, \tilde{\psi}_{n,3}, ..., \tilde{\psi}_{n,J_n-1})'$ and $\tilde{\psi} = (\tilde{\psi}_1', \tilde{\psi}_2', ..., \tilde{\psi}_N')'$. Stack the N underlying continuous variables $\tilde{y}_n^*$ into an $(N \times 1)$ vector $\tilde{y}^*$, and the N error terms $\tilde{\varepsilon}_n$ into another $(N \times 1)$ vector $\tilde{\varepsilon}$. Define $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, ..., \tilde{\gamma}_N)'$ [$(N \times A)$ matrix] and $\tilde{d} = (\tilde{d}_1, \tilde{d}_2, ..., \tilde{d}_N)$ [$(N \times L)$ matrix], and let $\mathbf{IDEN}_N$ be the identity matrix of dimension N representing the correlation matrix of $\tilde{\varepsilon}$. Finally, stack the lower thresholds for the decision-maker $\tilde{\psi}_{n,a_n-1}(n = 1, 2, ..., N)$ into an $(N \times 1)$ vector $\tilde{\psi}_{low}$ and the upper thresholds $\tilde{\psi}_{n,a_n}(n = 1, 2, ..., N)$ into another vector $\tilde{\psi}_{up}$. Then, in matrix form, the measurement equation for the ordinal outcomes for the decision-maker may be written as:

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

Now let there be G ranked outcome variables for an individual, and let g be the index for the ranked variables $(g = 1, 2, 3, ..., G)$. Also, let I_g be the number of alternatives being ranked for the gth ranked variable $(I_g \geq 3)$ and let i_g be the corresponding index $(i_g = 1, 2, 3, ..., I_g)$. In our case, $G=1$ and $I_1 = 7$; however, we present the framework for any number of ranked outcomes. Consider the gth ranked variable and assume the usual random utility structure for each alternative i_g.

$$U_{gi_g} = b'_{gi_g}x + \vartheta'_{gi_g}(\beta_{gi_g}z^*) + \varsigma_{gi_g} \tag{5}$$

where x is an $(A \times 1)$ vector of exogenous variables (including a constant) as well as possibly the observed values of other endogenous variables (introduced in a recursive fashion), b_{gi_g} is an $(A \times 1)$ column vector of corresponding coefficients, and ς_{gi_g} is a normal error term. β_{gi_g} is an

$(N_{gi_g} \times L)$ -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 their interaction effects with other exogenous variables. Let $\boldsymbol{\varsigma}_g = (\varsigma_{g1}, \varsigma_{g2}, ... \varsigma_{gI_g})'$ $[(I_g \times 1)$ vector], and $\boldsymbol{\varsigma}_g \sim MVN_{I_g}(\mathbf{0}_{I_g}, \mathbf{IDEN}_{I_g})$. The error terms in the alternatives are assumed to be independent in order to achieve a parsimonious specification; however, the utilities of the alternatives are correlated (because of unobserved factors) through the stochastic latent constructs. Moreover, for our modeling, we assume the scales of all the alternatives to be the same and fixed to one (the latter is needed for identification). In addition, the usual identification restriction is imposed such that one of the alternatives serves as the base when introducing alternative-specific constants and variables that do not vary across alternatives. To proceed, define $U_g = (U_{g1}, U_{g2}, ..., U_{gI_g})'$ $(I_g \times 1$ vector), $\boldsymbol{b}_g = (\boldsymbol{b}_{g1}, \boldsymbol{b}_{g2}, \boldsymbol{b}_{g3}, ..., \boldsymbol{b}_{gI_g})'$ $(I_g \times A$ matrix), and

$\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, ..., \boldsymbol{\beta}'_{gI_g})'$ $\left(\left(\sum_{i_g=1}^{I_g} N_{gi_g}\right) \times L \text{ matrix}\right)$. Also, define the $\left(I_g \times \left(\sum_{i_g=1}^{I_g} N_{gi_g}\right) \text{matrix}\right) \boldsymbol{\vartheta}_g$,

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), $\ddot{G} = \sum_{g=1}^{G} I_g$, $\tilde{G} = \sum_{g=1}^{G} (I_g - 1)$,

$U = (U'_1, U'_2, ... , U'_G)'$ $(\ddot{G} \times 1 \text{ vector})$, $\boldsymbol{\varsigma} = (\boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_2, ... \boldsymbol{\varsigma}_G)'$ $(\ddot{G} \times 1 \text{ vector})$, $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_G)'$ $(\ddot{G} \times A$ matrix), $\boldsymbol{\varpi} = (\boldsymbol{\varpi}'_1, \boldsymbol{\varpi}'_2, ..., \boldsymbol{\varpi}'_G)'$ $(\ddot{G} \times L \text{ matrix})$, and $\boldsymbol{\vartheta} = \text{Vech}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G)$ (that is, $\boldsymbol{\vartheta}$ 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 (5) as:

$$U = \boldsymbol{b}x + \boldsymbol{\varpi} z^* + \boldsymbol{\varsigma} \tag{6}$$

where $\boldsymbol{\varsigma} \sim MVN_{\ddot{G}}(\mathbf{0}_{\ddot{G}}, \mathbf{IDEN}_{\ddot{G}})$. \hfill (7)

With the matrix definitions above, the continuous components of the model system may be written compactly as:

$$z^* = \boldsymbol{\alpha}w + \boldsymbol{\eta} \tag{8}$$

11

$$\tilde{y}^* = \tilde{\gamma}x + \tilde{d}z^* + \tilde{\varepsilon}, \text{ with } \text{Var}(\tilde{\varepsilon}) = \mathbf{IDEN}_N \ (N \times N \text{ matrix}) \tag{9}$$

$$U = bx + \varpi z^* + \varsigma \tag{10}$$

The vector equation for the latent constructs (represented by Equation (8)) constitutes the structural equation system. The vector equation for the ordinal indicators and ordinal outcomes (represented by Equation (9)) and the vector equation for the ranked outcomes (represented by Equation (10)) constitute the measurement equation system in our framework.

Next, we explicitly consider that the latent constructs and the main outcomes are correlated; i.e., we consider η to be correlated with $\tilde{\varepsilon}$ as well as ς (while still maintaining the independence assumption between $\tilde{\varepsilon}$ and ς). Let the matrix $\mathbf{\Omega}_{\text{LN}}(L \times N \text{ matrix})$ contain the correlation elements between each of the latent constructs and the ordinal outcomes, and let the matrix $\mathbf{\Omega}_{\text{LG}}(L \times \ddot{G} \text{ matrix})$ contain the correlation terms between each of the latent constructs and the alternatives of the rank-ordered multinomial outcomes in differenced form. Earlier studies in residential self-selection explicitly consider $\mathbf{\Omega}_{\text{LN}}$ and $\mathbf{\Omega}_{\text{LG}}$ to be zero matrices. Let $\boldsymbol{\delta}$ be the collection of parameters to be estimated: $\boldsymbol{\delta} = [\text{Vech}(\boldsymbol{\alpha}), \text{Vechup}(\boldsymbol{\Gamma}), \text{Vech}(\tilde{\gamma}), \text{Vech}(\tilde{d}), \tilde{\psi}, \text{Vech}(\boldsymbol{b}), \boldsymbol{\vartheta}, \text{Vechup}(\mathbf{\Omega}_{\text{LN}}), \text{Vechup}(\mathbf{\Omega}_{\text{LG}})]$, where the operator "Vech(.)" vectorizes all the non-zero elements of the matrix/vector on which it operates and "Vechup(.)" indicates strictly upper diagonal elements.

To develop the reduced form equations, replace the right side of Equation (8) for z^* in Equations (9) and (10) to obtain the following system:

$$\tilde{y}^* = \tilde{\gamma}x + \tilde{d}z^* + \tilde{\varepsilon} = \tilde{\gamma}x + \tilde{d}(\alpha w + \eta) + \tilde{\varepsilon} = \tilde{\gamma}x + \tilde{d}\alpha w + \tilde{d}\eta + \tilde{\varepsilon} \tag{11}$$

$$U = bx + \varpi z^* + \varsigma = bx + \varpi(\alpha w + \eta) + \varsigma = bx + \varpi\alpha w + \varpi\eta + \varsigma$$

Now, consider the $[(N + \ddot{G}) \times 1)]$ vector $\boldsymbol{yU} = \left[[\tilde{y}^*]', U' \right]'$. Define

$$\boldsymbol{B} = \begin{bmatrix} \tilde{\gamma}x + \tilde{d}\alpha w \\ bx + \varpi \alpha w \end{bmatrix} \text{ and } \boldsymbol{\Omega} = \begin{bmatrix} \tilde{d}\Gamma\tilde{d}' + \mathbf{IDEN}_N + \tilde{d}\mathbf{\Omega}_{\text{LN}} + \mathbf{\Omega}'_{\text{LN}}\tilde{d}' & \tilde{d}\Gamma\varpi' + \tilde{d}\mathbf{\Omega}_{\text{LG}} + \mathbf{\Omega}'_{\text{LN}}\varpi' \\ \varpi\Gamma\tilde{d}' + \varpi\mathbf{\Omega}_{\text{LN}} + \mathbf{\Omega}'_{\text{LG}}\tilde{d}' & \varpi\Gamma\varpi' + \mathbf{IDEN}_{\ddot{G}} + \varpi\mathbf{\Omega}_{\text{LG}} + \mathbf{\Omega}'_{\text{LG}}\varpi' \end{bmatrix} \tag{12}$$

Then $\boldsymbol{yU} \sim \boldsymbol{MVN}_{N+\ddot{G}}(\boldsymbol{B}, \boldsymbol{\Omega})$.

For the estimation of the model, define a matrix $\mathbf{M}$ of size $\left[N + \tilde{G} \right] \times \left[N + \ddot{G} \right]$ which is similar to the mask matrix used in the single choice discrete model except that now this is defined with

respect to the respondent-provided ranking order of the alternatives (see Appendix A for the definition of this matrix $\mathbf{M}$). With this matrix $\mathbf{M}$, we can write $\boldsymbol{yu} \sim MVN_{N+\tilde{G}}(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\boldsymbol{B}} = \mathbf{M}\boldsymbol{B}$ and $\tilde{\boldsymbol{\Omega}} = \mathbf{M}\boldsymbol{\Omega}\mathbf{M}'$.

Next, define threshold vectors as follows:

$$\ddot{\boldsymbol{\psi}}_{low} = \left[\tilde{\boldsymbol{\psi}}'_{low}, \left(-\infty_{\tilde{G}}\right)'\right]' \quad ([(N+\tilde{G})\times 1] \text{ vector}) \text{ and } \ddot{\boldsymbol{\psi}}_{up} = \left[\tilde{\boldsymbol{\psi}}'_{up}, \left(\mathbf{0}_{\tilde{G}}\right)'\right]' \quad ([(N+\tilde{G})\times 1] \text{ vector}), \text{ where}$$

$-\infty_{\tilde{G}}$ is a $\tilde{G}\times 1$-column vector of negative infinities, and $\mathbf{0}_{\tilde{G}}$ is another $\tilde{G}\times 1$-column vector of zeros. Then the likelihood function may be written as:

$$
\begin{aligned}
L(\boldsymbol{\delta}) &= \Pr\left[\ddot{\boldsymbol{\psi}}_{low} \leq \boldsymbol{yu} \leq \ddot{\boldsymbol{\psi}}_{up}\right] \\
&= \int_{D_r} f_{N+\tilde{G}}(\boldsymbol{r} \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}}) dr
\end{aligned}
\tag{13}
$$

where the integration domain $D_r = \{\boldsymbol{r} : \ddot{\boldsymbol{\psi}}_{low} \leq \boldsymbol{r} \leq \ddot{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\boldsymbol{yu}$ vector determined by the observed ordinal outcomes, and the range $(-\infty_{\tilde{G}}, \mathbf{0}_{\tilde{G}})$ for the utility differences taken with respect to the utility of the ranked preference for the rank-ordered outcome. The likelihood function for a sample of Q decision-makers is obtained as the product of the individual-level likelihood functions. Since a closed-form expression does not exist for this integral and evaluation using simulation techniques can be time consuming, we used the analytical methods proposed by Bhat (2018) for approximating this integral.

The overall methodology, in our empirical context, may be visualized as shown in Figure 1. The exogenous individual and household socio-demographics (left side of the figure and represented by the vector $\boldsymbol{w}$ in the notation of Equation (8)) affect the latent constructs (in the middle of Figure 1 and represented by the vector $\boldsymbol{z}^*$) in the structural equation model (SEM) component of the GHDM (these effects are captured by the elements of the vector $\boldsymbol{\alpha}$). As part of the SEM component, the correlation matrix $\boldsymbol{\Gamma}$ among the elements of the error term vector $\boldsymbol{\eta}$ of the latent constructs is also estimated (indicated by the double-headed arrow between the two latent constructs). This SEM system is estimated through the loadings of the latent constructs on the latent construct indicators. These loading pathways of the constructs on the indicators, which constitute one part of the measurement equation model (MEM) component, are not shown in the figure to avoid clutter. The second part of the MEM component, shown in the figure, corresponds

13

to two effects: (a) the effects (as captured by the matrices $\tilde{\gamma}$ and $\boldsymbol{b}$) of individual/household characteristics (represented by the vector $\boldsymbol{x}$ in Equations (9) and (10)); the trip-level attributes, shown toward the right bottom of the figure and relevant only for the mode choice model, and the observed values of endogenous outcome variables affecting other endogenous variables, are also a part of the vector $\boldsymbol{x}$, and (b) the effects (as captured by the matrices $\tilde{\boldsymbol{d}}$ and $\boldsymbol{\varpi}$) of the latent constructs on the main outcome variables, while also recognizing the potential endogeneity of the latent constructs, as identified by the correlation matrices $\boldsymbol{\Omega}_{\mathrm{LN}}$ and $\boldsymbol{\Omega}_{\mathrm{LG}}$ in Figure 1. Finally, part of the $\tilde{\gamma}$ and $\boldsymbol{b}$ matrices also correspond to the effects of observed endogenous outcomes on other endogenous outcomes, as shown within the far right box entitled "Main Outcomes".
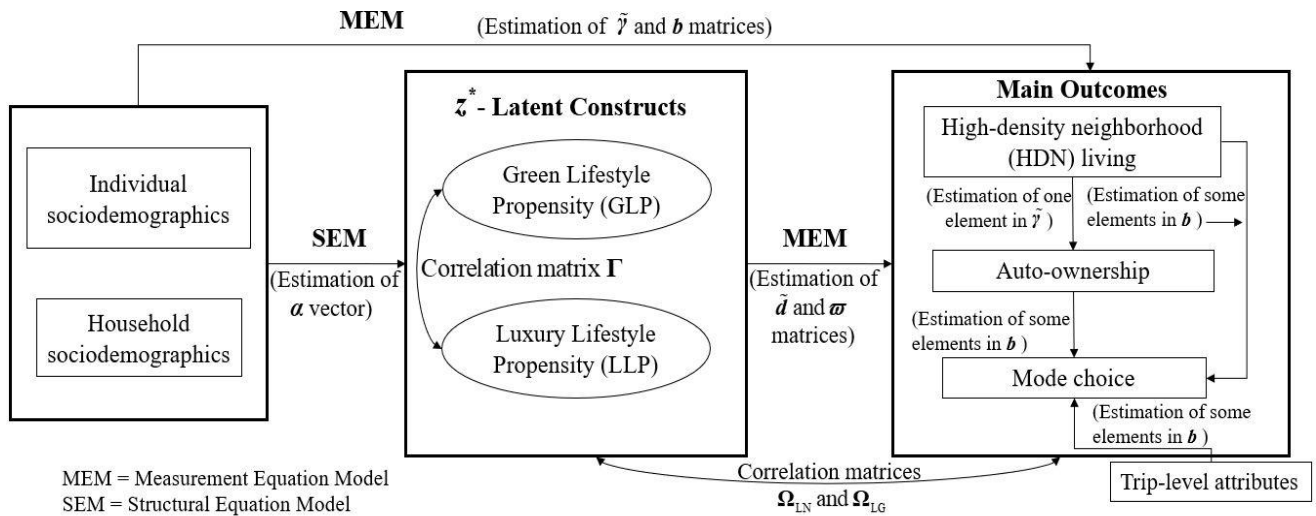

Figure 1. Analytical Framework

Chapter 3. Data and Sample Description

3.1. Survey and Sample

The data used in this study is drawn from the 2019 multi-city "Transformative Technologies in Transportation (T4)" Survey (Asmussen et al., 2020). The T4 survey was conducted in Phoenix (Arizona), Atlanta (Georgia), Tampa (Florida), and Austin (Texas). This survey was a general-purpose survey that collected information on a wide variety of emerging mobility options (including e-scooters, e-bikes, sharing arrangements, autonomous vehicle adoption, and intended usage), general lifestyle attitudes, and current travel behaviors. Full details of the survey design and administration procedures, and the survey instrument itself, are available at https://tomnet-utc.engineering.asu.edu/data/t4-survey/.

For the analysis in the current report, we focus on the data collected exclusively from Austin. This was because, though the survey was conducted across the four cities, there was quite a bit of variation in the details of the survey design across the cities to accommodate for the current availability and extent of use of different travel modes. For example, two of the cities (Phoenix and Tampa) did not have a pooled ride-hailing service (one of the model alternatives in our empirical context; see Section 3.3.3) at the time of the survey, and the cities also varied considerably in the current provision of transit services and travel modal shares. Also, the Stated Preference (SP) travel attribute values were generated specific to the travel patterns for each city. For all these reasons, we do not expect travel-related behaviors and lifestyles across the four cities to be related in the same fashion to exogenous variables.

For the Austin survey, a financial incentive was provided in the form of $10 Amazon gift cards for the first 250 respondents, while the remaining respondents were entered into a drawing to win one of the remaining one hundred $10 Amazon gift cards. The ensuing survey distribution effort resulted in a convenience sample of 1,127 respondents (for the city of Austin), which was reduced to a final sample of 928 individuals after removing 199 individuals who provided clearly inappropriate or incomplete responses (such as inconsistent socio-demographic information, incomplete modal ranks, and missing information on residential location).

The survey elicited several user characteristics and choices, which we discuss in three categories below: (a) those related to transportation/technology attitudinal indicators and lifestyle preferences related to transportation, environment, and residential living space (these form the

indicators for the latent constructs, as discussed in Section 3.2 below), (b) current residential neighborhood, auto-ownership levels, and stated preferences for emerging mobility options of ride-hailing and autonomous vehicles (AV) (these constitute the endogenous outcomes of interest, as discussed in Section 3.3 below; as part of this section, we also discuss the trip-level attributes presented in the experimental design, which constitute exogenous variables), and (c) individual and household socio-demographics (which, along with the trip-level attributes discussed in Section 3.3, form the exogenous variables, and are discussed in Section 3.4 below).

3.2. Latent Constructs

In our empirical analysis, we consider two latent constructs: Green Lifestyle Propensity (GLP) and Luxury Lifestyle Propensity (LLP). [Note that additional latent constructs, including those associated with variety-seeking lifestyle, security concern, time sensitivity, and tech-savviness were also constructed and tested in the model, but did not turn out to be statistically significant in explaining any of the main outcomes; this is because of correlation between these other constructs and the constructs already considered in this report]. The stochastic latent constructs are not observed directly from the sample, but are estimated based on attitudinal questions (indicators) capturing user preferences. A traditional confirmatory factor analysis determined the most suitable indicators for each of the selected two latent constructs, as discussed further below.

The first latent construct, GLP (sometimes also termed Environmental Consciousness), refers to a general consciousness about the degrading quality of the environment and concerns about the personal carbon footprint on environmental quality. Several studies in the land use-travel behavior literature have established GLP (or Environmental Consciousness) as an important attitudinal factor impacting individuals' travel decisions (see, for example, Zhu et al., 2020 and Blazanin et al., 2022). For this construct, we use the following three ordinal indicators (all collected on a five-point Likert scale from "Strongly Disagree" to "Strongly Agree"):

(a) The government should raise the gas tax to help reduce the negative impacts of transportation on the environment.

(b) I am committed to an environmentally-friendly lifestyle.

(c) I am committed to using a less polluting means of transportation (e.g., walking, biking, and public transit) as much as possible.

The second latent construct, Luxury Lifestyle Propensity or LLP, is characterized by a penchant for consuming more, marked by a desire for privacy, spaciousness, and signaling exclusivity. Again, within the body of the land use-travel behavior literature, several studies have used luxury lifestyle propensity or similar attitudes indicative of a lavish lifestyle to understand their role in travel decisions (see, for example, Lavieri and Bhat, 2019 and Kim et al., 2020). The LLP stochastic latent construct is based on three attitudinal indicators:

(a) I like to be among the first to have the latest technology.
(b) I prefer to live in a spacious home, even if it is farther from public transportation or many places I go.
(c) I definitely like the idea of owning my own car.

3.3. Endogenous Outcomes

The three endogenous outcomes for our study are discussed in turn in the sections below.

3.3.1. High-Density Neighborhood (HDN) Living

This endogenous outcome is a binary variable indicating whether an individual resides in a dense neighborhood or not. Each individual's residence address (closest cross-streets) is mapped to a Census Block Group (CBG), and the population density of the CBG (as extracted from the U.S. EPA Smart Location Database; see Ramsey and Bell, 2014) is attributed to the individual's residence. Next, the individual is designated as residing in an HDN if the residential CBG has a population density of more than 10 individuals/acre (the average population density of Austin is close to 5 individuals/acre (see City of Austin, 2021)). The most densely populated CBGs are found around the University of Texas (Hyde Park, North University, West Campus) and in the Downtown region. Other pockets of high-density neighborhoods are located around the areas of the Domain, South Congress, North Lamar, East Riverside-Oltorf, and Round Rock, rounding out the multi-centric nature of the Austin metro area. 149 of the 967 CBGs in the Austin-Round Rock Metropolitan Statistical Area (MSA) are HDN. In the sample used in this study, 228 individuals (24.5% of the sample) reside in an HDN, while the remaining individuals reside in non-HDNs.

3.3.2. Auto-Ownership

The auto-ownership endogenous outcome corresponds to the number of motorized four-wheelers available to the household. This outcome is considered as an ordered variable (with levels of 0,1,2,3, and 4+). The highest value for the variable was considered as 4+ because only 34 households (less than 3.7% of the sample) owned more than four vehicles. For modeling such capped non-negative integer counts, the ordered-response is a particularly suitable framework (see Haddad et al., 2022 for further discussion). Our sample statistics reveal the following distribution of household auto-ownership level among the respondents: 8% of individuals reside in zero-auto households; about 25% reside in single-auto households; a little under 34% reside in two-auto households, about 20% live in three-auto households, and the rest in 4+-auto households.

3.3.3. Ranked Mode Preference

The mode choice outcome is in the form of a rank-ordered multinomial choice variable. The survey elicited users' modal preferences through the use of a stated preference (SP) question that asked respondents to rank, in the context of a future autonomous vehicle environment, their mode choice preferences (from most preferred to least preferred) for non-work/non-mandatory trips. The mode choice alternatives were: private vehicle (human-driven or autonomous), bicycle, public transport (bus/rail), human-driven private ride-hailing (ride-hailing alone with a human driver), human-driven pooled ride-hailing (ride-hailing with others with a human driver), autonomous vehicle (AV) private ride-hailing (same as private ride-hailing, except the vehicle is autonomous without a human driver), and AV pooled ride-hailing (same as pooled ride-hailing, except the vehicle is autonomous). In the rest of this report, we will use the acronym "HD" for human-driven and "RH" for ride-hailing. Therefore, the seven modes will be referred to as private vehicle, bicycle, public transport, HD private RH, HD pooled RH, AV private RH, and AV pooled RH.

Three trip attributes were used to characterize the SP experiment – wait time, in-vehicle travel time (IVTT), and total trip cost. The attribute levels for each attribute varied by travel mode, with 3 to 9 levels for wait time, 5 to 31 levels for IVTT, and 3 to 44 levels for total trip cost (see Appendix B for additional details).

A note is in order here. The reader may have observed the conspicuous absence of the walk mode, and the lack of distinction between the human-driven and AV variants for the private vehicle mode, in the list of modal alternatives identified earlier. The survey team initially had included the

walk mode and also separated out the human-driven and private AV modes, but was concerned (and also received feedback from pilots) that ranking nine alternatives was a little much. So, in thinking through alternatives to cut down, we decided to drop the walk alternative, supported by the difficulty in engendering sufficient variation in travel times within a compact walk distance threshold in the SP experiment, as well as the irrelevant nature of both wait time and cost for the walk mode. Additionally, the team decided not to separate out the private vehicle mode by human-driven versus AV, with the view that the private vehicle mode choice would be less impacted by the distinction, while earlier studies have strongly pointed to this distinction being important in the context of ride-hailing (see, for example, Lavieri and Bhat, 2019, Menon et al., 2019 and Mo et al., 2022). With these intentional (even if admittedly semi-executive) decisions, the number of alternatives reduced to seven in the SP ranking exercise.

The SP experimental procedure entailed an orthogonal fractional factorial design with the seven alternatives, three trip attributes of wait time, IVTT, and cost, and the multiple levels for each trip attribute. Supplemented by a random blocking approach, a total of 36 scenarios were developed. A single scenario was randomly assigned for presentation to each respondent. Additionally, a non-mandatory trip purpose was also randomly assigned to this question (which varied across individuals); the trip purposes were "Shopping", Airport-access", "Socializing" and "Eating-out"; these were later used as exogenous variables in our model to recognize that mode choice may be purpose-specific. Figure 2 provides a sample of the actual question presented to respondents.

The descriptive statistics of the mode choice ranking indicate a strong preference for the use of private vehicles among the individuals in the dataset, with more than 70% assigning this mode the top rank. Interestingly, beyond the top-most rank, HD private RH turned out to be the second-most preferred mode, which suggests that a large fraction of the individuals who choose private vehicle as the top rank are also inclined toward the use of HD private RH. Additionally, the rank-preference distribution also suggests that private RH modes (HD or AV) are preferred to pooled RH modes, and human-driven RH modes are generally preferred over AV RH modes. These results hint at the notion that in addition to the trip attributes presented in the SP experiment, and other sociodemographic and built environment contexts, modal familiarity (that is, the familiarity with human-driven private modes) appears to also play a role in modal rankings.

Bicycling is the least preferred mode, with only about 36% of the survey respondents ranking it within the top five.

Suppose you are going to a mall to do some shopping (e.g., to purchase clothes, books, etc.). You have the following seven options for your transportation. Rank the alternatives listed from most preferred (Rank 1) to least preferred (Rank 7). Please do not give the same rank to multiple alternatives.

Rank	Alternative	Wait time	In-vehicle travel time	Cost for entire trip
	Private vehicle: Use your own private vehicle (human-driven or AV)	No wait	10 min	$0.40
	Bicycle	No wait	35 min	$0.00
	Public transit: Use bus or rail	10 min	23 min	$1.25
	Private ride-hailing: Get a ride with a human-driven ride-hailing service (*e.g.*, Uber, Lyft)	3 min	10 min	$12.50
	Pooled ride-hailing: Get a human-driven ride in a vehicle in which other passengers may be added.	6 min	20 min	$9.40
	AV private ride-hailing: Same as ride-hailing, except that the vehicle will be autonomous.	3 min	10 min	$8.75
	AV pooled ride-hailing: Same as shared ride-hailing, except that the vehicle will be autonomous.	6 min	20 min	$6.60

Figure 2. Example of the Ranking Question Presented to Users (Shopping Purpose)

3.4. Individual and Household Demographics

Table 1 provides a brief description of the individual and household characteristics of the 928 sample respondents. The table also presents statistics corresponding to the Austin metropolitan statistical area (MSA) population demographics, which are obtained from the U.S. Census Bureau (2018). In cases where the Austin MSA values are not readily available, there is a "-" in the table. As may be observed from the table, our sample shows an overrepresentation of women, young adults (in the age group of 18-29 years), individuals who completed some college or a higher education level, unemployed individuals, individuals from multi-adult and zero-children households, and low-income households.

Table 1. Sample Description

Variable	Sample		Census
	Count	Percentage	Percentage
Gender			
Male	331	35.7%	50.0%
Female	597	64.3%	50.0%
Age			
18 to 29	565	60.9%	28.1%*
30 to 50	197	21.2%	37.7%*
50+	166	17.9%	34.2%*
Possession of Driving License			
No	113	12.0%	-
Yes	815	88.0%	-
Student			
No	462	49.8%	-
Yes	466	50.2%	-
Education Qualification			
Completed high-school or less	120	12.9%	26.0%
Completed some college or technical school	323	34.8%	24.0%
Completed undergraduate degree	316	34.1%	32.0%
Completed graduate degree	169	18.2%	18.0%
Employment Status			
Employed	359	38.7%	68.5%
Not Employed	569	61.3%	31.5%
No. of Adults			
1	234	25.2%	32.8%
2+	694	74.8%	67.2%
No. of Kids			
0	788	84.9%	73.6%
1	81	8.8%	-
2	43	4.6%	-
3+	16	1.7%	-
Household Annual Income			
Less than $50,000	374	40.3%	29.0%
$50,000 - $99,999	281	30.3%	28.0%
≥ $100,000	273	29.4%	43.0%

* Percentages are normalized (to add up to 100%) for the population above 18 years of age.

The sociodemographic differences between our sample and the Austin area population are not surprising, given that our survey was administered online and disseminated, in part, through social media outlets. Such an administration approach would naturally draw in individuals with advanced degrees and those who are technology-savvy. In fact, the very topic of autonomous vehicles would likely be of more interest to such individuals, contributing further to the sample bias. Besides, the financial incentives to participate in the survey would tend to attract students,

unemployed individuals and those from low-income households. However, these skews in the sample should not affect our investigation of individual-level causal relationships (that is, how changes in exogenous variables impact the endogenous variables of interest). This is because the basis for teasing out causal relationships does not hinge on having a representative sample, but only on good variation in the range of each exogenous variable (and, of course, good variation in the indicators and the outcomes in our GHDM framework). As our descriptive statistics above illustrate, we do have good variation within the sample in the exogenous sociodemographic variables. Also, because our sampling process is itself not based on any of the endogenous outcomes of the residential living environment, auto-ownership level, or travel mode preferences (that is, our data collection is based on exogenous sampling), an unweighted approach provides consistent estimates as well as yields more efficient estimates relative to a weighted procedure (see Wooldridge, 1995 and Solon et al., 2015 for an extensive discussion of this point).

Chapter 4. Model Results

4.1. Latent Construct Results

Table 2 provides the results for the latent construct indicator loadings (these are elements of the $\tilde{d}$ matrix in Section 2). All the indicator variables have the expected direction of loadings on each of the two latent constructs. This forms one component of the measurement equation in the GHDM framework.

Table 2. Latent Construct Indicator Loadings (elements of the $\tilde{d}$ matrix)

Indicators	Green Lifestyle Propensity (GLP)		Luxury Lifestyle Propensity (LLP)	
	Coef.	t-stat	Coef.	t-stat
The government should raise the gas tax to help reduce the negative impacts of transportation on the environment.	0.594	25.40		
I am committed to an environmentally-friendly lifestyle.	0.476	17.97		
I am committed to using a less polluting means of transportation (e.g., walking, biking, and public transit) as much as possible.	1.349	19.77		
I like to be among the first to have the latest technology.			0.125	7.85
I prefer to live in a spacious home, even if it is farther from public transportation or many places I go.			0.658	13.42
I definitely like the idea of owning my own car.			1.050	18.62

Table 3 presents the structural equation model results that relate the latent constructs to observed demographic variables (these effects represent the elements of the matrix α). The results indicate that there is no statistically significant gender difference in green lifestyle propensity (GLP), consistent with the studies of Xiao and McCright (2014) and Blazanin et al. (2022) (although contrary to the studies of Astroza et al., 2017 and Strapko et al., 2016 that find women having a higher GLP). However, there is a significant gender difference in Luxury Lifestyle Propensity (LLP); in particular, women have higher levels of LLP compared to men. Indeed, Stokburger-Sauer and Teichman (2013) found a generally higher proclivity of women toward luxury and exclusive items. Further, women have been known to desire large spacious homes (an

indicator of LLP in our model system) as a signaling mechanism of (effectively a stage play to project) privileged motherhood to the wider social world (see Mulder and Lauster, 2010 and Bhat, 2015b).

Table 3. **Determinants of Latent Constructs** (elements of the matrix α)

Variables (base category)		Green Lifestyle Propensity (GLP)		Luxury Lifestyle Propensity (LLP)	
		Coef.	t-stat	Coef.	t-stat
Individual-Level Characteristics					
Female (base: Male)		-	-	0.148	2.54
Age 50 years or greater (base: below 50 years)		-0.336	-7.56	0.268	4.33
Student (base: not student)		-	-	-0.151	-2.04
Graduate or higher (base: less than Graduate)		0.349	3.45	-	-
Household-Level Characteristics					
Presence of child below 18 years (base: not present)		-0.202	-4.54	0.326	4.13
Annual household income (base: less than $50,000)					
$50,000-$100,000		-0.214	-5.43	-	-
Greater than $100,000		-0.214	-5.43	0.181	2.33
Correlation Γ among latent constructs	**GLP**	1.000	NA	-0.489	-4.04
	LLP	-0.489	-4.04	1.000	NA

NOTE: '-' indicates that the variable was not found to be statistically significant.

Age appears to be a key determinant of both GLP and LLP. In particular, older individuals (age 50 or above) exhibit lower levels of "pro-environmental" inclination as compared to their younger counterparts, an observation which is corroborated by several earlier studies in the field of gerontology, environmental science, and transportation-land use (for example, see Liu et al., 2014 and Clements, 2012). The younger generation has grown up in an era of an information-rich environment that appears to have made them more knowledgeable about environmental and climate change issues, thereby encouraging them to adopt a more environmental-friendly lifestyle (Hassim, 2021). Further, and not inconsistent with the finding that older individuals have a lower GLP, older individuals have a higher LLP compared to their younger peers. This may be attributed

to a generally higher level of financial security and a lower need to "save for the future" as one ages, while younger individuals show more restraint in spending during their formative years of asset-building (see Kahn, 2018, and Henager and Cude, 2016). For similar reasons, it is not surprising to find in our results that students are low on LLP.

Higher education levels (Graduate or higher) are positively associated with GLP, a result that has also been found in many earlier studies (see, for example, Fisher et al., 2012; Franzen and Vogl, 2013; and Blazanin et al., 2022). Educated individuals are more likely to appreciate the human impact on natural resource degradation and hence more likely to be environment-friendly (Philippssen et al., 2017).

The results also suggest that individuals with children in their household have lower levels of GLP and higher levels of LLP than other individuals. As opposed to the legacy hypothesis which suggests that parents are likely to be more concerned about the future quality of the environment that they leave behind for their children, Thomas et al. (2018) found that the immediate short-term well-being and comfort of their children is of far greater importance to parents than any future environmental threats. Add to that the time-pressure faced by parents, it leaves little time for them to adopt time-consuming sustainable practices, leading to an under-emphasis on environmental issues to manage the resulting dissonance (see for example, Strazdins and Loughrey, 2007). The higher LLP among parents is similarly not surprising, given the strong need to provide a comfortable living environment for the children (especially given that the LLP indicators include a preference for a large home and to own a car).

As per our results in Table 3, individuals from non-low-income households (annual income higher than $50,000) are found to score low on GLP, while those from high-income households (annual income higher than $100,000) have a high LLP. Although a few earlier studies associate higher income with greater levels of environmental awareness (Awan and Abbasi, 2013, Bülbül et al., 2020), many recent studies have found that lower-income individuals are more likely to have witnessed first-hand the negative impacts of environmental degradation and climate change issues, resulting in a higher GLP (see, for example, Pearson et al., 2018, Wenz, 2015 and Banzhaf et al., 2019).

Finally, as one might expect, there is a significant negative correlation (-0.489) between the GLP and LLP constructs (see the bottom row panel of Table 3); unobserved factors that increase GLP, decrease LLP. After all, while a green lifestyle is associated with careful and

conservative consumption of resources, a luxury lifestyle correlates with extravagant living and indulgence beyond an "indispensable minimum".

4.2. Main Outcome Results

Table 4 provides the results for the HDN/auto-ownership outcome (Section 4.2.1), and Table 5 presents the results for the mode choice outcome (Section 4.2.2). For each table, we discuss, in turn, the latent construct effects, the individual and household effects, the trip variable effects (only for Table 5), and the endogenous outcome effects. Subsequently, Section 4.2.3 discusses the correlation effects between the endogenous outcomes and the latent constructs.

4.2.1. High-Density Neighborhood (HDN) Living and Auto-Ownership Level

The results discussed below correspond to the impact of variables on the underlying propensities of the HDN and the auto-ownership levels (Table 4). The thresholds toward the bottom of Table 4 do not have any substantive interpretation, but provide the mapping of the underlying latent propensities to the actual observed outcomes (these are elements of the $\tilde{\psi}$ vector in Section 2).

Latent construct effects (elements of the $\tilde{d}$ matrix)

Green Lifestyle Propensity (GLP) positively impacts HDN living, while Luxury Lifestyle Propensity (LLP) negatively impacts HDN living propensity. Not surprisingly, the signs switch for auto-ownership level propensity. High-density or urban neighborhoods are typically associated with shorter trip distances and lower private-vehicle use, and provide greater opportunities for the use of sustainable modes of transportation such as walking, bicycling, and public transit; therefore, environmentally conscious individuals, who are likely to be more concerned about their carbon footprint, have a higher tendency to reside in such HDNs as well as have a lower propensity for auto-ownership (see also Etezady et al., 2021). On the contrary, individuals with higher LLP are likely to have a penchant for spacious living with greater desires for privacy and exclusivity, as well as may feel more of a need for power and opulence signaling, both of which can explain the lower propensity to live in an HDN and higher propensity of auto-ownership.

Table 4. Main Outcome Results: HDN and Auto-Ownership

Variables	High-density neighborhood living (HDN)		Auto-ownership	
	Coef.	t-stat	Coef.	t-stat
Latent construct effects				
Green Lifestyle Propensity (GLP)	0.122	2.11	-0.634	-5.66
Luxury Lifestyle Propensity (LLP)	-0.441	-4.52	0.241	3.91
Individual characteristics				
Age (base: below 30 years)				
30 to 50	-0.435	-5.34	-	-
50+	-0.532	-4.44	-0.336	-2.33
Student (base: non-student)	0.482	2.89	-	-
Household characteristics				
Household size 3 or more (Base: Less than 3)	-0.249	-4.55	-	-
Number of adults	-	-	0.376	3.56
Number of children (below 18 years)	-	-	0.136	2.33
Annual household income (base: <$50,000)				
$50,000 - $99,999	-0.377	-4.05	0.144	2.13
Greater $100,000	-0.492	-3.03	0.579	4.55
Endogenous outcome effect				
High-density neighborhood (HDN) living	NA	NA	-0.116	-3.66
Thresholds				
Threshold 1	0.356	6.22	-0.411	-7.43
Threshold 2	NA	NA	0.845	11.55
Threshold 3	NA	NA	1.964	10.92
Threshold 4	NA	NA	2.813	16.02

NOTE: '-' indicates that the variable was not found to be statistically significant.

Individual and household sociodemographic effects (elements of the $\tilde{\gamma}$ matrix)

The individual and household effects in Table 4 are direct effects after accommodating any moderating effects of sociodemographic variables through the GLP and LLP latent construct effects. The results suggest that older adults (age 30 or more) tend to have a lower proclivity for HDN living compared to their younger peers; younger individuals (less than 30 years of age) are more likely to have a high desire for a fast-paced, entertainment-accessible, socially-oriented and physically-active urban lifestyle, which gets reflected in their HDN preference (De Vos and Alemi, 2020). The negative effect of those over the age of 50 years on auto-ownership propensity is

interesting and finds support in earlier literature (see Clark et al., 2016a). In particular, panel studies have revealed that there is a household life-cycle effect associated with auto-ownership in which the number of vehicles owned by a household increases as the head of the household reaches 50 years, and declines thereafter. This effect is different from the age-effect on the actual modal preference of these individuals; that is, these individuals may still be highly inclined toward private vehicle use (as we note later), but from a sheer number of auto-ownership standpoint, there is a negative relationship. Students (of any age) are found to have a higher HDN living propensity relative to non-students, presumably because of their need to stay in proximity to academic institutions and social opportunities.

In terms of household effects, individuals from households with three or more adults and progressively higher annual incomes, are less likely to opt for HDN living relative to their peer groups. On the other hand, the effects of these variables are diametrically flipped in terms of auto-ownership propensity. The earlier literature (see, for example, Moos, 2016, and Clark et al., 2016b) has established that larger households (especially with many children) tend to gravitate toward lower-density living, accompanied by a higher auto-ownership level, because of residential space considerations and movement flexibility considerations to accommodate the travel needs of all household members.

Endogenous outcome effect (elements of the $\tilde{\gamma}$ matrix)

The direct endogenous outcome effect of HDN living on auto-ownership level suggests that the built environment in high-density neighborhoods has a negative causal effect on the propensity to own private vehicles. This is consistent with the earlier literature, as HDNs are often characterized by better opportunities to walk, bicycle, and use public transit; moreover, high-density and congested neighborhoods are often known to present parking-related challenges, which may further discourage urban residents to own private cars (Prieto et al., 2017).

4.2.2. Mode Choice Model

Latent construct effects (elements of the ϖ matrix)

Our results from Table 5 indicate that individuals with higher levels of GLP have a higher preference for the bicycle mode, but consistently lower preferences for private vehicles and all forms of RH services (relative to the public transit mode, which is treated as the base mode for our

analysis). On the contrary, individuals with high LLP have significantly higher preferences for private vehicles and all forms of RH services (relative to the public transit mode and bicycling).

*Individual and household sociodemographic effects (elements of the **b** matrix)*

There is a significant gender difference in mode preferences. First, women are found to have a higher preference for private vehicles. This result is consistent with the finding from Asmussen et al. (2020) that women exhibit a higher need for driving control, which may be, among other things, attributed to the time-poor nature of women who have to juggle work and non-work household chores/child-care responsibilities (see also, Giurge et al., 2020, and Shirgaokar and Lanyi-Bennett, 2020). Asmussen et al. also suggest, based on the social-psychological studies of Skuladottir and Halldorsdottir (2008) and Leung et al. (2018), that, in a rather asymmetric, male-dominated world in which women feel a lower sense of general life control, women are not willing to relinquish the feelings of free-spiritedness and empowerment they derive from the ability to drive by themselves. Further, the preference of women for private vehicles may be associated with the heightened personal security considerations felt by women when traveling with strangers (see Gardner et al., 2017). Second, and in line with several earlier studies, our results indicate that women have a lower preference for the bicycling mode, perhaps due to a combination of the difficulty riding with the outfits worn by women (such as skirts, dresses, and high-heeled shoes; see, for example, Kaplan, 2015 and Singleton and Goddard, 2016), the complex activity-chaining patterns typically undertaken by women (Singleton and Goddard, 2016), and the heightened concerns of women regarding bicycling-related crash risk (Akar et al., 2013). Third, women also have a lower propensity for AV RH mode use; this is presumably because of elevated feelings of skepticism among women toward the reliability and safety of newer technologies (Othman, 2021).

Older adults (age 50 or greater) are found to have a stronger inclination toward the use of private vehicles compared to their younger peers, which may be attributed to their greater need for spatiotemporal or mobility control. Presumably because of their gradual slow-down of reflexes and difficulty in physiological maneuverability with age, older adults are also found to be significantly disinclined toward the use of the bicycle mode. In addition, older individuals have much lower preferences toward AV ride-hailing modes (private and pooled), which may be attributed to their general distrust toward newer technology, heightened safety concerns, and lower technology-savviness (Asmussen et al., 2020, Faber and Lierop, 2020, and Siegfried et al., 2021)

As would be expected, the possession of a driving license increases the propensity for the use of privately owned vehicles, perhaps due to these individuals' "pro-driving" preference and strong affective emotions of empowerment when driving. Interestingly, our results suggest a lower preference for the bicycle mode, but consistently high preferences for all the RH alternatives, among those with an undergraduate/graduate degree relative to those who have not obtained a college degree.

Among the exogenous household variables, income and household size play an important role in travel mode decisions. Individuals from high-income households are predisposed toward privately owned vehicles and private ride-hailing (human-driven and AV) modes, primarily because such individuals assign a high premium for comfort and privacy. Individuals from households with more than 3 members appear to prefer the use of private vehicles, presumably because it is convenient and offers them movement control to travel together as a family.

*Trip variable effects (elements of the **b** matrix)*

Table 5 highlights the need to analyze purpose-specific modal preferences even for non-work trip purposes. Private vehicles are strongly preferred for socializing, shopping, and airport-access purposes perhaps because of the social convenience, convenient baggage carrying capacity, and high time control that this mode provides (Gilibert et al., 2019). For similar reasons, the impacts of shopping trips and airport-access trips indicate that individuals have a significantly higher preference for private RH (human-driven and AV) for these purposes. Interactions of these purpose-specific trips with demographics were also attempted, but the trip purpose effects seemed to be uniform across all sociodemographic groups.

As expected, the travel cost and travel time coefficients are negative and estimated to be -0.06 and -0.11 respectively. We use a generic coefficient for the travel time and cost variables since these coefficients did not show much variability across the alternatives. From the estimated coefficients, the value of travel time (VTT) comes out to be $32/hour, which is a reasonable estimate relative to what has been found in the literature (Zhong et al., 2020). Interestingly, the wait time coefficient did not turn out to be statistically significant.

Table 5. Main Outcome Results: Mode Choice Dimension [coefficient estimates (t-stats)]

Variables	Public Transport (base)	Private vehicle	Bicycle	HD private RH	HD pooled RH	AV private RH	AV pooled RH
Latent construct effects							
Green Lifestyle Propensity (GLP)		-0.212	0.112	-0.120	-0.098	-0.078	-0.089
		(-3.44)	(2.33)	(-2.34)	(-1.98)	(-1.87)	(-2.05)
Luxury Lifestyle Propensity (LLP)		0.596	-	0.336	0.160	0.185	0.122
		(6.44)		(3.92)	(3.04)	(3.44)	(2.33)
Individual characteristics							
Female (Base: Male)		0.231	-0.133	-	-	-0.092	-0.081
		(3.60)	(-2.86)			(-4.12)	(-2.29)
Age greater than 50 years (Base: Less than 50 years)		0.179	-0.192	-	-	-0.096	-0.078
		(1.98)	(-2.04)			(-3.55)	(-1.91)
Possession of a driver's license (Base: No possession)		0.332	-	-	-	-	-
		(3.76)					
Education level (Base: lower than undergraduate)							
Completed undergraduate degree		-	-	0.075	0.089	0.142	0.132
				(2.21)	(2.86)	(3.44)	(2.33)
Completed graduate degree		-	-0.090	0.097	0.076	0.103	0.132
			(-1.89)	(2.56)	(1.79)	(2.34)	(2.33)
Household characteristics							
Household size 3 or more (Base: Less than 3)		0.121	-	-	-	-	-
		(3.18)					
Household income (Base: <$50,000)							
Income ≥ $100,000		0.098	-	0.068	-	0.093	-
		(1.92)		(1.89)		(1.98)	
Trip level attributes							
Trip purpose (Base: Eat-out)							
Shopping purpose		0.193	-	0.080	-	0.011	-
		(1.97)		(2.34)		(2.01)	
Airport-access purpose		0.459	-	0.101	-	0.083	-
		(4.17)		(2.00)		(1.97)	
Social purpose		0.224	-	-	-		
		(2.84)					
Travel time (minutes)		-0.060 (-3.52)					
Travel cost ($)		-0.112 (-7.52)					
Endogenous outcome effects							
High-density neighborhood (HDN) living		-0.092	0.167	-	-	-	-
		(-2.12)	(3.33)				
Auto-ownership level		0.110	-	-	-0.067	-	-0.067
		(3.45)			(-1.82)		(-1.82)
Constant		-0.224	-0.266	-0.074	0.072	-0.044	-0.040
		(-4.50)	(-6.07)	(-1.67)	(2.24)	(-2.26)	(-1.99)

NOTE: '-' indicates that the variable was not found to be statistically significant.

*Endogenous outcome effect (elements of the **b** matrix)*

There are two sets of endogenous outcome effects for the mode choice component of the joint model. First, the high-density neighborhood or HDN impact suggests that individuals residing in HDNs have a lower preference for the private vehicle mode, but higher preferences for the bike and HD RH modes (private and pooled). The second set of endogenous outcome effects relate to auto-ownership effects on mode choice. As one would anticipate, high auto-ownership is associated with a higher preference for private vehicle use. Interestingly, our results also suggest that individuals in higher auto-ownership households have lower preferences for pooled RH (HD or AV), which is presumably because such individuals are more used to traveling in private and in the comfort of their own car.

4.2.2. Correlations between Main Outcomes and Latent Constructs (elements of the Ω_{LN} and Ω_{LG} matrices)

Our proposed model allows the error terms of the latent constructs to be correlated with the error terms of the main outcomes. Among all the possible correlations, five correlation terms turned out to be significant. These estimated correlations (and t-stats) are as follows: GLP and HDN living with a correlation of +0.121 (t-stat = 2.22), GLP and the Bicycle mode with a correlation of +0.142* (t-stat = 2.85), LLP and HDN living with a correlation of -0.090 (t-stat = -1.89), LLP and auto-ownership with a correlation of +0.105 (t-stat = 3.00), and LLP and the Private vehicle mode with a correlation of +0.114* (t-stat = 2.12) [the * indicates that the correlations with the alternatives in the mode choice model are with respect to the error differenced terms with Public Transport as the base]. Analyzing the signs of these correlations and the signs of the direct effects of the latent constructs on the main outcomes reveals that ignoring these endogeneity effects will consistently overestimate the self-selection effects and underestimate the "true" causal HDN effect on the main outcomes for this particular empirical context, as we further illustrate in Section 5.

4.3. Data Fit Comparison

We compare the data fit measures of our proposed model, which considers the endogeneity of the latent constructs to the main outcomes (we call our proposed model FLEX-GHDM), with the traditional GHDM model (TRAD-GHDM). The TRAD-GHDM model is estimated by ignoring the endogeneity of the latent constructs; that is, by considering Ω_{LN} and Ω_{LG} to be zero matrices.

Table 6 provides the log-likelihood values at convergence for both models (first row), and related statistics. The BIC measure (fourth row) and the adjusted likelihood ratio index (fifth row) both favor the FLEX-GHDM model. Further, a likelihood ratio test (sixth row) indicates that the superior fit of the FLEX-GHDM model is statistically significant at the 0.05 level of significance (and, in fact, at any reasonable level of significance). To further understand the gains in prediction for our proposed model over the TRAD-GHDM model, we compute the predictive log-likelihood at convergence for only the main outcome variables (that is, the HDN, auto-ownership, and modal choices) in the GHDM framework (see the penultimate row of Table 6). This again favors the FLEX-GHDM model. We can also use an informal likelihood ratio index test to compare the two models on predicting only the main outcomes. Such an informal test confirms the statistical superiority of the FLEX-GHDM over the TRAD-GHDM model (see the last row of Table 6).

Table 6. Disaggregate Data Fit Measures

Summary Statistics	Model	
	TRAD-GHDM	**FLEX-GHDM**
Log-likelihood at convergence	-172,273.83	-172,115.23
Number of non-constant parameters	88	93
Constants-only log-likelihood	-179,670.41	
Bayesian Information Criterion (BIC)	172,724.81	172,583.29
Adjusted likelihood ratio index	0.0407	0.0415
Likelihood ratio (LR) test	$LR = 317.2 > \chi^2_{(5,0.05)} = 11.070$	
Predictive log-likelihood of only main outcomes	-9684.33	-9620.90
Informal predictive likelihood ratio test of only main outcomes	$LR = 126.86 > \chi^2_{(5,0.05)} = 11.070$	

Chapter 5. Estimating the "True" Effects of Residential Location on Auto-Ownership and Mode Choice

For our study, there are three effects associated with the relationship between HDN living propensity on the one hand and auto-ownership/mode choice on the other: (1) the *"true" causal* effect of HDN living on auto-ownership/mode choice, (2) an associative *sample selection effect,* and (3) the *latent construct* (*LC*) *endogeneity effect*. To tease out these components of the total effect of HDN living on travel behavior choices, we use the "average treatment effects" metric (see Appendix C for details on the computation process).

Table 7 presents the magnitude effect as well as the contributing percentages (in parentheses) of each of the three effects for the TRAD-GHDM and FLEX-GHDM. The top row panel provides the three HDN-associated living effects on auto-ownership levels, while the bottom two row panels provide the HDN-associated living effects on the choice of the private vehicle and bicycle modes (we focus on the HDN effects only on these two alternatives to keep the discussion focused, and also because these two alternatives are directly impacted by the HDN outcome; see Table 5). Beginning with the top row panel of Table 7, the "true" causal effect of HDN living on auto-ownership is estimated as -0.24 in the TRAD-GHDM model and -0.29 in the FLEX-GHDM model. The way to interpret this is as follows. If a random individual is supplanted from a non-HDN to an HDN, auto-ownership will, on average, reduce by 0.24 based on the TRAD-GHDM and by 0.29 based on the FLEX-GHDM; equivalently, if 100 random individuals are transplanted from a non-HDN environment to an HDN environment, there would be a reduction (among these 100 individuals) of 24 vehicles according to the TRAD-GHDM and 29 vehicles according to the FLEX-GHDM. Essentially, by ignoring the endogeneity of the two latent constructs to the main outcomes, the TRAD-GHDM increases the correlative associative effect between HDN living and auto-ownership (that is, incorrectly elevates the sample selection effect), thereby underestimating the true HDN causal effect on auto-ownership. This is also clearly noticeable in the relative percentage of the three effects (shown in parenthesis) in Table 7. A similar result is evident from the bottom two rows in which the TRAD-GHDM underestimates the "true" negative causal effect of HDN living on private vehicle mode choice and also underestimates the "true" positive causal effect of HDN living on the bicycling mode. Overall, in our particular empirical context, the results suggest that efforts at neo-urbanist designs that focus on neighborhood densification and

investments in improved bicycling infrastructures may get undervalued in planning and policy decisions if the endogeneity of latent constructs is ignored. Of course, whether this result would be universally applicable (that is, transferable) to other metropolitan areas is an open question; most likely, there will be specific local elements related to infrastructure investment levels, public environmental-consciousness levels, and overall activity-travel behaviors that will either elevate this undervaluation of the effectiveness of neo-urbanist designs or temper this undervaluation. In any case, the extent of difference in the HDN effect between ignoring and not ignoring the latent construct endogeneity is likely to be local context-specific and will require estimating the FLEX-GHDM in the local context.

Overall, our study continues to emphasize the value of neo-urbanist design in reducing traffic through decreased auto-ownership levels and more non-motorized mode use. This may be achieved, among other ways, through transit-oriented development (TOD) and bicycling infrastructure improvements. In fact, such efforts can have a mutually reinforcing and snowballing effect with, for example, improved bicycling infrastructure enhancing transit stop accessibility and increasing transit use. However, from a policy standpoint, affordability becomes a critical equity issue for TOD housing in particular and neo-urbanist housing in general. For instance, transit-proximal settlements have attractive property values and provide convenient opportunities for activity engagement, thereby creating high demand and leading to surges in housing prices. This translates to an affordability paradox wherein high-income households who tend to use private vehicles more often live nearer to transit, while low-income households who use transit more are priced out of such housing (Dong, 2017). Indeed, due to this affordability and housing price imbalance, low-income households sometimes find it more reasonable to own a private vehicle (even if that causes financial stress) than to locate close to transit areas, which in turn, increases private auto dependency. Thus, to reap the benefits of neo-urbanist forms on reduced motorized travel, there needs to be a synergistic approach among the several elements of housing affordability, transit operations, and integrated bicycling/walking infrastructure.

TABLE 7 Quantifying HDN Living Effect on Auto-Ownership and Modal Preferences

Metric	TRAD-GHDM	FLEX-GHDM
HDN effect auto-ownership: magnitude change (% contribution)		
"True" causal HDN effect %	-0.24 (69.3%)	-0.29 (85.4%)
Estimated self-selection effect %	-0.10 (30.7%)	-0.02 (5.5%)
Estimated EC endogeneity effect %	-	-0.03 (9.1%)
HDN effect on the private vehicle mode in the mode choice model: % share change (% contribution)		
"True" causal HDN effect %	-2.3% (55.3%)	-3.1% (73.2%)
Estimated self-selection effect %	-1.9% (44.7%)	-0.6% (14.7%)
Estimated EC endogeneity effect %	-	-0.5% (12.1%)
HDN effect on the bicycle mode in the mode choice model: % share change (% contribution)		
"True" causal HDN effect %	6.7% (68.2%)	8.0% (81.2%)
Estimated self-selection effect %	3.2% (31.8%)	1.1% (10.7%)
Estimated EC endogeneity effect %	-	0.8% (8.1%)

Chapter 6. Conclusions

In our current effort, we propose a methodological framework based on Bhat's (2015a) GHDM framework that accounts for the endogeneity of the latent constructs while investigating the relationship between residential location and two markers of travel behavior (auto-ownership levels and rank-based travel mode preferences of individuals) within a hypothetical futuristic autonomous vehicle (AV) landscape. The data for this study is drawn from the 2019 multi-city Transformative Technologies in Transportation (T4) Survey for the city of Austin. Our results reveal the presence of significant endogeneity effects, i.e., significant unobserved correlations between the main outcomes and the latent constructs used. Specifically, the unobserved correlations between Green Lifestyle Propensity (GLP) and HDN living, GLP and the Bicycle mode, LLP and HDN living, LLP and auto-ownership, and LLP and the private vehicle mode are found to be statistically significant. Moreover, ignoring such endogeneity effects, as done in earlier studies, can underestimate the "true" causal impact of HDN on the main outcomes. Additionally, by evaluating the average treatment effect of residential density, we are able to quantify the contribution of the "endogeneity effects", the "spurious" self-selection effects, and the "true" BE effects on auto-ownership and mode choice behavior. The "true" causal effects of HDN living on auto-ownership suggests that on average, auto-ownership level would reduce by about 29% when a random individual is shifted from a non-HDN to an HDN. Furthermore, such a shift would increase non-work bicycle mode share by 8% and decrease non-work private vehicle share by 3.1%. Overall, our results highlight the need to recognize dependency structures between attitudinal factors and travel-based outcomes of interest.

Of course, as with all research efforts, our work is not without limitations. First, there is a temporal component in the interactions between attitude and behavior, which cross-sectional data would not be able to capture. However, panel data are expensive and rare to come by. Besides, panel data are themselves subject to other problematic issues, such as unobserved changes in the environment (between one time period and another) getting comingled with "true" causal effects. In any case, panel data analysis should provide additional insights into cross-temporal behavior-attitude interactions, and would be a fruitful avenue for further research. Second, in our study, we achieved better model fit using the causal structure in which attitudes (latent constructs) impact behavior (travel-related choices) after accommodating unobserved error correlations. However, it

is possible that this causal structure itself is different across individuals, which could be accounted for through the use of a latent class approach as in Sharda et al. (2019). Third, from a substantive standpoint, a multi-dimensional characterization of residential living rather than a simple binary classification of HDN versus non-HDN living would be a better representation of the living environment and would provide richer insights into the relationship between the living environment and travel outcomes. Finally, the analysis could be extended to include the walk mode as well as to distinguish between autonomous and human-driven variants of the private vehicle mode.

Appendix A

Define a matrix $\mathbf{M}$ of size $\left[N+\tilde{G}\right]\times\left[N+\tilde{G}\right]$ (N, $\tilde{G}$, and $\ddot{G}$ are as defined in the main text). Fill all the elements of the matrix with zeros. Then, insert an identity matrix of size N into the first N rows and N columns of the matrix $\mathbf{M}$. Next, consider the rows from $N+1$ to $N+I_1-1$, and columns from $N+1$ to $N+I_1$ (these rows and columns correspond to the first ranked variable), and do the following: in the first row, place an entry of '1' in the column corresponding to the second-ranked alternative, and a '-1' in the column corresponding to the first-ranked alternative; in the second row, place an entry of '1' in the column corresponding to the third-ranked alternative, and a '-1' in the column corresponding to the second-ranked alternative; and so on until placing an entry of '-1' in the column corresponding to the penultimate-ranked alternative, and a '1' in the column corresponding to the last-ranked alternative. Repeat this entire step for the second-ranked variable (if present) in the next I_2-1 rows and I_2 columns. Continue this procedure for all G-ranked variables. Thus, in the case of N=2, G=1 with I_1=5, if the first individual's ranking (from the top choice to the last choice) is 4>1>2>3>5, then the $\mathbf{M}$ matrix for this individual is as below:

$$
\mathbf{M} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & -1 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1
\end{bmatrix}
$$

Appendix B

The travel attribute values (levels) for each mode were carefully designed to simulate realistic scenarios according to our region of interest (Austin, Texas). Several logical checks for scenario building were in place, such as ensuring that the travel times of pooled ride-hailing are always larger than that of private ride-hailing and private vehicle modes; travel costs of pooled ride-hailing are always less than that of private ride-hailing; and so on. A total of 36 scenarios were developed for the survey, and each user was presented with one of the scenarios drawn at random.

Table B.1. SP Attribute Levels

Modes	Travel time (in minutes)	Travel cost (in $)	Wait time (in minutes)
Public Transport (base)	9 to 108 (31 levels)	0.75, 1.25, 1.75 (3 levels)	5, 10, 15 (3 levels)
Private vehicle (HD or AV)	5, 17, 24, 36, 48 (5 levels)	0.25 to 19 (31 levels)	0
Bicycle	9 to 108 (31 levels)	0	0
HD private RH	5, 17, 24, 36, 48 (5 levels)	4.50 to 60 (12 levels)	3, 6, 9 (3 levels)
HD pooled RH	10 to 63 (15 levels)	1.5 to 54 (31 levels)	4 to 14 (9 levels)
AV private RH	5, 17, 24, 36, 48 (5 levels)	2 to 78 (31 levels)	3, 6, 9 (3 levels)
AV pooled RH	10 to 63 (15 levels)	1.5 to 58.5 (44 levels)	4 to 14 (9 levels)

Appendix C

To tease out the components of the total effect of HDN living on travel behavior choices, we use the "average treatment effects" metric, which computes the impact on a downstream variable of interest due to a treatment that changes the state of an antecedent variable from one state to another. In our case, the downstream variables are auto-ownership and travel modal preferences, while the antecedent variable is the HDN living binary variable (note that the downstream and the antecedent variables are correlated in our proposed model). For ease of presentation, interpretability, and understanding, we use the travel-related endogenous outcome variables in specific forms for our analysis. For the auto-ownership outcome, since the ordered levels themselves indicate the cardinal values of the number of private vehicles owned, we use the ordered values as they are except for the highest ordered level (which groups the ownership levels of 4 vehicles or greater) for which we use a cardinal value of 4.5 for the number of vehicles owned. Therefore, with j as the number of ordered levels in the auto-ownership model (j=0,1,2,3,4) and c_j as the cardinality of these ordered levels (0, 1, 2, 3, and 4.5, respectively), we can write the expected household auto-ownership $\bar{y}_q$ for an individual q (where $\tilde{y}_q$ is the ordered level) as follows:

$$\bar{y}_q = \left[\sum_{j=0}^{4} c_j \left(P(\tilde{y}_q = j) \right) \right]$$

To simplify presentability and interpretability along the mode choice model component, we compute the ATEs as a percentage change in the first-choice shares (and not a rank-based choice) of the modes between the case where all individuals in the dataset are assumed to be in a non-high density neighborhood and the case where all individuals are assumed to be in a high-density neighborhood.

As discussed in the main text, there are three effects associated with HDN living propensity on auto-ownership and mode choice dimensions. To disentangle these three effects, we first estimate an independent heterogeneous data model (IHDM). This IHDM ignores the correlations among the three main outcomes and does not consider the stochastic latent constructs in the framework; thus, the second and the third effects (the self-selection and the endogeneity effects) are completely ignored, which then get lumped up as the "true" causal effect within the coefficient (direct effect) on the HDN variable in the travel-behavior outcome equations. Therefore, the ATEs computed from the IHDM model (with the "treatment" being an individual shifting from a non-

HDN living to an HDN-living scenario) may be considered to cumulatively include all three effects. Moreover, in the traditional GHDM framework (TRAD-GHDM), which does consider the endogeneity effects, only the first and the second effects are present (and the third effect gets clubbed with the self-selection effect). Using the steps discussed below, we also quantify the "true" causal and the self-selection effects of the TRAD-GHDM framework and compare them to our proposed GHDM framework (or FLEX-GHDM, which considers the latent construct endogeneity).

To compute and compare the contributions of the various effects, we undertake the following steps:

Step 1: Using the estimates from the IHDM model, compute the ATEs for the main outcomes with respect to the HDN-living variable; say this value is X_{IHDM} (we will use this value as the base since this value may be considered to include the total sum of all the three effects lumped into the direct "true" causal effect for the IHDM framework, i.e., the total effect of HDN on the travel-related outcomes across all the models discussed below are considered to be the same as this value, but differing in the contribution-split of the three effects discussed earlier).

Step 2: For the TRAD-GHDM model, compute the ATE values with the coefficient estimates from the estimated TRAD-GHDM model but ignore the correlations among the main outcome equations (i.e., consider the corresponding $\boldsymbol{\Omega}$ matrix in the TRAD-GHDM model to be an identity matrix while computing the ATE values); say this value is $X_{TRAD-TRUE}$. The $X_{TRAD-TRUE}$ denotes the "true" causal effect for the TRAD-GHDM model, while the self-selection effect is obtained as $X_{IHDM} - X_{TRAD-TRUE}$. Express these values as percentages of the X_{IHDM} quantity in absolute terms along with their actual magnitudes (as reported in Table 7).

Step 3: For our proposed FLEX-GHDM model, compute the ATE values from the coefficient estimates (as reported in Table 4) but consider the $\boldsymbol{\Omega}$ matrix (shown in Equation 12) to be an identity matrix, i.e., completely ignore the correlations between the main outcome equations (as well as the latent-construct endogeneity effects); say this value is $X_{FLEX-TRUE}$. The $X_{FLEX-TRUE}$ denotes the "true" causal effect for our proposed framework. Next, compute another ATE value but this time consider the $\boldsymbol{\Omega}$ matrix as described in Equation 12 but assume $\boldsymbol{\Omega_{LN}}$ and $\boldsymbol{\Omega_{LG}}$ to be

zero matrices, i.e., ignore the endogeneity effects; call this value $X_{FLEX-TRUE-SS}$. The $X_{FLEX-TRUE-SS}$ value cumulatively denotes the "true" causal effect and the self-selection effect. Therefore, the quantity $X_{FLEX-TRUE-SS} - X_{FLEX-TRUE}$ is the self-selection effect for our proposed model. Finally, the endogeneity effect may be obtained as $X_{IHDM} - X_{FLEX-TRUE-SS}$. Express all these values as a percentage of the X_{IHDM} quantity in absolute terms along with their actual magnitudes (as reported in Table 7).

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.

Akar, G., Fischer, N., and Namgung, M. (2013). Bicycling choice and gender case study: The Ohio State University. *International Journal of Sustainable Transportation*, 7(5), 347-365.

Asmussen, K.E., Mondal, A., and Bhat, C.R. (2020). A socio-technical model of autonomous vehicle adoption using ranked choice stated preference data. *Transportation Research Part C*, 121, 102835.

Astroza, S., Garikapati, V. M., Bhat, C. R., Pendyala, R. M., Lavieri, P. S., and Dias, F. F. (2017). Analysis of the impact of technology use on multimodality and activity travel characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2666(1), 19-28.

Awan, U., and Abbasi, A. S. (2013). Environmental sustainability through determinism the level of environmental awareness, knowledge and behavior among business graduates. *Research Journal of Environmental and Earth Science*, 5(9), 505-515.

Banzhaf, S., Ma, L., and Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, 33(1), 185-208.

Bhat, C. R. (2015a). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50-77.

Bhat, C. R. (2015b). A comprehensive dwelling unit choice model accommodating psychological constructs within a search strategy for consideration set formation. *Transportation Research Part B*, 79, 161-188.

Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B*, 109, 238-256.

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., and Mondal, A. (2022). A new flexible generalized heterogeneous data model (GHDM) with an application to examine the effect of high density neighborhood living on bicycling frequency. *Transportation Research Part B*, 164, 244-266.

Blazanin, G., Mondal, A., Asmussen, K. E., and Bhat, C. R. (2022). E-scooter sharing and bikesharing systems: An individual-level analysis of factors affecting first-use and use frequency. *Transportation Research Part C*, 135, 103515.

Bülbül, H., Büyükkeklik, A., Topal, A., and Özoğlu, B. (2020). The relationship between environmental awareness, environmental behaviors, and carbon footprint in Turkish households. *Environmental Science and Pollution Research*, 27(20), 25009-25028.

City of Austin (2021). Austin By the Numbers: City Releases Tract-Level Analysis of 2020 Census Data for Austin Metro Service Area. https://www.austintexas.gov/news/austin-numbers-city-releases-tract-level-analysis-2020-census-data-austin-metro-service-area

Clark, B., Chatterjee, K., and Melia, S. (2016a). Changes in level of household car ownership: The role of life events and spatial context. *Transportation*, 43(4), 565-599.

Clark, B., Lyons, G. and Chatterjee, K. (2016b). Understanding the process that gives rise to household car ownership level changes. *Journal of Transport Geography*, 55, 110-120.

Clements, B. (2012). Exploring public opinion on the issue of climate change in Britain. *British Politics,* 7(2), 183-202.

De Vos, J., and Alemi, F. (2020). Are young adults car-loving urbanites? Comparing young and older adults' residential location choice, travel behavior and attitudes. *Transportation Research Part A*, 132, 986-998.

Dong, H. (2017). Rail-transit-induced gentrification and the affordability paradox of TOD. *Journal of Transport Geography*, 63, 1-10.

Etezady, A., Shaw, F. A., Mokhtarian, P. L., and Circella, G. (2021). What drives the gap? Applying the Blinder–Oaxaca decomposition method to examine generational differences in transportation-related attitudes. *Transportation*, 48(2), 857-883.

Faber, K., and van Lierop, D. (2020). How will older adults use automated vehicles? Assessing the role of AVs in overcoming perceived mobility barriers. *Transportation Research Part A*, 133, 353-363.

Fisher, C., Bashyal, S., and Bachman, B. (2012). Demographic impacts on environmentally friendly purchase behaviors. *Journal of Targeting, Measurement and Analysis for Marketing*, 20(3), 172-184.

Franzen, A., and Vogl, D. (2013). Two decades of measuring environmental attitudes: A comparative analysis of 33 countries. *Global Environmental Change*, 23(5), 1001-1008.

Gardner, N., Cui, J., and Coiacetto, E. (2017). Harassment on public transport and its impacts on women's travel behaviour. *Australian Planner*, 54(1), 8-15.

Gilibert, M., I. Ribas, N. Maslekar, C. Rosen, and A. Siebeneich (2019). Mapping of service deployment use cases and user requirements for an on-demand shared ride-hailing service: MOIA test service case study. *Case Studies on Transport Policy*, 7(3), 598-606.

Giurge, L. M., Whillans, A. V., and West, C. (2020). Why time poverty matters for individuals, organisations and nations. *Nature Human Behaviour*, 4(10), 993-1003.

Guan, X., Wang, D., and Cao, X. J. (2020). The role of residential self-selection in land use-travel research: A review of recent findings. *Transport Reviews*, 40(3), 267-287.

Haddad, A., Mondal, A., and Bhat, C.R. (2022). Eat-in or Eat-out? A joint model to analyze the new landscape of dinner meal preferences. *Transportation Research Part C*, 147, 104016.

Hassim, A. (2021). Why younger generations are more willing to change in the name of sustainability. GreenBiz.com. Available at: https://www.greenbiz.com/article/why-younger-generations-are-more-willing-change-name-sustainability

Henager, R., and Cude, B. J. (2016). Financial Literacy and Long-and Short-Term Financial Behavior in Different Age Groups. *Journal of Financial Counseling and Planning*, 27(1), 3-19.

Kahn, K. (2018). Work and wealth. *A Report from the 2017 Aspen Institute Economic Security*.

Kaplan, D. H. (2015). Transportation sustainability on a university campus. *International Journal of Sustainability in Higher Education*, 16(2), 173-186.

Kim, S. H., Mokhtarian, P. L., and Circella, G. (2020). Will autonomous vehicles change residential location and vehicle ownership? Glimpses from Georgia. *Transportation Research Part D*, 82, 102291.

Lavieri, P. S., and Bhat, C. R. (2019). Investigating objective and subjective factors influencing the adoption, frequency, and characteristics of ride-hailing trips. *Transportation Research Part C*, 105, 100-125.

Leung, E., Paolacci, G., and Puntoni, S., 2018. Man versus machine: Resisting automation in identity-based consumer behavior. *Journal of Marketing Research*, 55(6), 818-831.

Liu, X., Vedlitz, A., and Shi, L. (2014). Examining the determinants of public environmental concern: Evidence from national public surveys. *Environmental Science and Policy*, 39, 77-94.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics* (No. 3). Cambridge University Press, Cambridge, MA

Menon, N., Barbour, N., Zhang, Y., Pinjari, A. R., and Mannering, F. (2019). Shared autonomous vehicles and their potential impacts on household vehicle ownership: An exploratory empirical assessment. *International Journal of Sustainable Transportation*, 13(2), 111-122.

Mo, D., Chen, X. M., and Zhang, J. (2022). Modeling and managing mixed on-demand ride services of human-driven vehicles and autonomous vehicles. *Transportation Research Part B*, 157, 80-119.

Moos, M. (2016). From gentrification to youthification? The increasing importance of young age in delineating high-density living. *Urban Studies*, 53(14), 2903-2920.

Mulder, C. H., and Lauster, N. T. (2010). Housing and family: An introduction. *Housing Studies*, 25(4), 433-440.

Nair, G. S., Bhat, C. R., Pendyala, R. M., Loo, B. P. Y., and Lam, W. H. K. (2019). On the use of probit-based models for ranking data analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(4), 229-240.

Othman, K. (2021). Public acceptance and perception of autonomous vehicles: A comprehensive review. *AI and Ethics*, 1(3), 355-387.

Pearson, A., Schuldt, J., Romero-Canyas, R., Ballew, M., and Larson-Konar, D. (2018). Diverse segments of the US public underestimate the environmental concerns of minority and low-income Americans. *Proceedings of the National Academy of Sciences*, 115(49), 12429-12434.

Philippssen, J. S., Soares Angeoletto, F. H., and Santana, R. G. (2017). Education level and income are important for good environmental awareness: A case study from south Brazil. *Ecología Austral*, 27(1), 39-44.

Prieto, M., Baltas, G., and Stan, V. (2017). Car sharing adoption intention in urban areas: What are the key sociodemographic drivers?. *Transportation Research Part A*, 101, 218-227.

Ramsey, K., and Bell, A. (2014). Smart Location Database, Version 2.0 User Guide. Available at: https://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf [Accessed by: 7/26/21]

Sharda, S., Astroza, S., Khoeini, S., Batur, I., Pendyala, R. M., and Bhat, C. R. (2019). Do attitudes affect behavioral choices or vice-versa: Uncovering latent segments within a heterogeneous population. Transportation Research Board 98th Annual Meeting Compendium of Papers, Washington D.C., January.

Shirgaokar, M., and Lanyi-Bennett, K. (2020). I'll have to drive there: How daily time constraints impact women's car use differently than men's. *Transportation*, 47(3), 1365-1392.

Siegfried, A. L., Bayne, A., Beck, L. F., and Freund, K. (2021). Older adult willingness to use fully autonomous vehicle (FAV) ride sharing. *Geriatrics*, 6(2), 47.

Singleton, P. A., and Goddard, T. (2016). Cycling by choice or necessity?: Exploring the gender gap in bicycling in Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, 2598(1), 110-118.

Skuladottir, H., and Halldorsdottir, S. (2008). Women in chronic pain: Sense of control and encounters with health professionals. *Qualitative Health Research*, 18(7), 891-901.

Solon, G., Haider, S.J., and Wooldridge, J.M. (2015). What are we weighting for? *Journal of Human Resources*, 50(2), 301-316.

Stokburger-Sauer, N. E., and Teichmann, K. (2013). Is luxury just a female thing? The role of gender in luxury brand consumption. *Journal of Business Research*, 66(7), 889-896.

Strapko, N., Hempel, L., MacIlroy, K., and Smith, K. (2016). Gender differences in environmental concern: Reevaluating gender socialization. *Society and Natural Resources*, 29(9), 1015-1031.

Strazdins, L., and Loughrey, B. (2007). Too busy: why time is a health and environmental problem. *New South Wales Public Health Bulletin*, 18(12), 219-221.

Thomas, G. O., Fisher, R., Whitmarsh, L., Milfont, T. L., and Poortinga, W. (2018). The impact of parenthood on environmental attitudes and behaviour: A longitudinal investigation of the legacy hypothesis. *Population and Environment*, 39(3), 261-276.

U.S. Census Bureau(2018). American Community Survey 1-year estimates. Census Reporter Profile page for Austin-Round Rock, TX Metro Area. https://censusreporter.org/profiles/31000US12420-austin-round-rock-tx-metro-area/.

Wenz, P. (2015). Environmental justice. In *Thinking About the Environment* (pp. 214-220). Routledge.

Wooldridge, J.M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1), 115-132.

Xiao, C., and McCright, A. M. (2014). A test of the biographical availability argument for gender differences in environmental behaviors. *Environment and Behavior*, 46(2), 241-263.

Zhong, H., W. Li, M. W. Burris, A. Talebpour, and K. C. Sinha (2020). Will autonomous vehicles change auto commuters' value of travel time?. *Transportation Research Part D*, 83, 102303.

Zhu, M., X. Hu, Z. Lin, J. Li, S. Wang, and C. Wang (2020). Intention to adopt bicycle-sharing in China: introducing environmental concern into the theory of planned behavior model. *Environmental Science And Pollution Research*, 7(33), 41740-41750.