

REPRESENTING HETEROGENEITY IN STRUCTURAL RELATIONSHIPS AMONG MULTIPLE CHOICE VARIABLES USING A LATENT SEGMENTATION APPROACH

Sebastian Astroza

The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: +56-41-220-3618; Email: sastroza@utexas.edu
and

Departamento de Ingeniería Industrial, Universidad de Concepción,
Edmundo Larenas 219, Concepción, Chile

Venu M. Garikapati

National Renewable Energy Laboratory
Systems Analysis & Integration Section
15013 Denver West Parkway, Golden, CO 80401
Tel: 303-275-4784; Email: venu.garikapati@nrel.gov

Ram M. Pendyala

Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Tel: 480-727-4587; Email: ram.pendyala@asu.edu

Chandra R. Bhat (corresponding author)

The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Patricia L. Mokhtarian

Georgia Institute of Technology
School of Civil and Environmental Engineering
790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-385-1443; Email: patmikh@gatech.edu

ABSTRACT

Travel model systems often adopt a single decision structure that links several activity-travel choices together. The single decision structure is then used to predict activity-travel choices, with those downstream in the decision-making chain influenced by those upstream in the sequence. The adoption of a singular sequential causal structure to depict relationships among activity-travel choices in travel demand model systems ignores the possibility that some choices are made jointly as a bundle as well as the possible presence of structural heterogeneity in the population with respect to decision-making processes. As different segments in the population may adopt and follow different causal decision-making mechanisms when making selected choices jointly, it would be of value to develop simultaneous equations model systems relating multiple endogenous choice variables that are able to identify population subgroups following alternative causal decision structures. Because the segments are not known a priori, they are considered latent and determined endogenously within a joint modeling framework proposed in this paper. The methodology is applied to a national mobility survey data set to identify population segments that follow different causal structures relating residential location choice, vehicle ownership, and car-share and mobility service usage. It is found that the model revealing three distinct latent segments best describes the data, confirming the efficacy of the modeling approach and the existence of structural heterogeneity in decision-making in the population. Future versions of activity-travel model systems should strive to incorporate such structural heterogeneity to better reflect varying decision processes across population subgroups.

Keywords: causal relationships, structural heterogeneity, simultaneous equations models, latent segmentation, joint estimation, vehicle ownership, residential location choice, mobility service usage.

1. INTRODUCTION

The presence of heterogeneity in people's activity-travel choice behavior is widely recognized in travel demand modeling systems (Krizek and Waddell, 2002; Johansson et al., 2006; Lavieri et al., 2017). As people differ with respect to their tastes, preferences, priorities, perceptions, and attitudes/values, it is not surprising that travel model systems have increasingly moved towards explicitly capturing such heterogeneity. In particular, model systems attempt to account for observed heterogeneity by defining market segments using a number of observed socio-economic and demographic variables. Variables such as income, vehicle ownership, household structure and composition, employment status, and age are often used to define segments; separate choice models are then estimated for different segments to reflect the fact that sensitivity to different variables (say, travel time in a mode choice model) may vary across segments (Yarlagadda and Srinivasan, 2008; Silva et al., 2012). More recently, advanced discrete choice modeling methods have incorporated randomness in coefficient effects, thus capturing population variance in sensitivity (taste variations) to selected variables (e.g., Bhat et al., 2016a). Further advances in econometric approaches to travel behavior analysis have provided the ability to also account for *unobserved* heterogeneity that may arise due to differences among people with respect to unmeasured attributes, such as attitudes, values, and lifestyle preferences (Bhat, 2015). These methods often employ flexible, heteroskedastic, and correlated error structures to account for unobserved heterogeneity in behavior.

While these advances have greatly aided in reflecting behavioral heterogeneity in travel demand model systems, there is a more fundamental structural heterogeneity in activity-travel behavior that continues to be ignored in operational activity-travel model systems despite the recognition of its existence (Pendyala, 1998; Waddell et al., 2007). While current methods reflect heterogeneity in tastes and preferences in relation to specific variables, they do not account for heterogeneity or jointness in structural relationships among multiple dependent (choice) variables of interest. In other words, current model systems assume that the same sequential causal structure among choice variables applies to all, only allowing model parameters on specific variables to vary across population segments. The structural or causal relationships embedded within the activity-travel model systems are assumed to hold true for everybody, and are often established on the basis of intuition, empirical observation, or computational feasibility considerations. Even in the context of traditional four-step travel demand models, the structure in which destination choice (trip distribution) precedes mode choice is widely used in practice and applied to an entire region – although there may be segments of the population for whom mode choice influences destination choice or the two choices are made jointly as a bundle.

Thus, the motivation for this study stems from the desire to examine the recursive causal effects in a joint model with multiple limited-dependent endogenous variables. The modeling effort in this paper recognizes jointness in multiple choice variables due to unobserved factors that influence multiple choice dimensions at once, while also considering population heterogeneity in the directionality of the recursive causal effects among a multitude of endogenous variables of interest. This paper aims to make a significant contribution to the field by considering jointness in decision-making and recognizing that different segments of the population may be choosing a lifestyle bundle (of mobility options) using different underlying recursive causal structures.

The importance of recognizing and accounting for the presence of structural or causal heterogeneity in the population is becoming increasingly apparent as emerging transportation technologies, the sharing economy, and mobility-on-demand services gain footholds in the transportation landscape. In addition, some suggest that there is a generational shift (particularly

among millennials) in preferences and attitudes that is fundamentally re-shaping how different segments make choices related to where they live and work, how – and how much – they travel, and how and with whom they spend time engaging in different activities (Garikapati et al., 2016). If travel demand model systems are to accurately forecast adoption and usage of autonomous vehicles, mobility-on-demand services, and shared vehicular systems (besides all of the other activity-travel choices), then alternative structural relationships that govern choice behaviors and jointness in decision-making need to be appropriately reflected for different segments in the population.

This paper offers a methodology for identifying and reflecting different causal structures among multiple endogenous variables of interest that may be prevalent in a population. Because the choice variables of interest are discrete in nature, two-way causal relationships cannot be estimated (due to identification and logical inconsistency issues). Yet, the choice between alternative one-way relationships is rather arbitrary, thus calling for the adoption of a latent segmentation approach that can endogenously identify subgroups in the population that follow different decision structures while also accommodating jointness that may exist in the decision-making process. Although the alternative causal structures that represent different decision-making mechanisms can be specified a priori, group membership is unknown. The analyst does not know the decision-making structure that applies to various agents in the population, thus rendering segments *latent* or *unobserved*. In this paper, a latent segmentation modeling approach is proposed to probabilistically associate behavioral units (agents) to the causal structure that best fits or describes their behavior. The methodology offers the ability to identify and define segments in the population that follow alternative decision structures, and subsequently model their behavior appropriately in travel demand forecasting systems. In this paper, latent segments that follow different decision structures are identified in the context of three key choice variables of interest: residential location (area type) choice, vehicle ownership, and extent of usage of car-share and mobility-on-demand services. Data from a national mobility attitudes survey conducted in 2014 is used to estimate the model system and demonstrate the efficacy of the modeling methodology.

The remainder of this paper is organized as follows. The next section provides a discussion of structural heterogeneity in activity-travel behaviors. The third section presents a data description, the fourth section offers an overview of the methodology, and the fifth section describes model estimation results. The sixth and final section provides a discussion of the disparate latent segments together with concluding thoughts.

2. STRUCTURAL HETEROGENEITY IN RELATIONSHIPS AMONG CHOICES

A number of studies have attempted to unravel causal relationships among multiple endogenous variables (e.g., Bagley and Mokhtarian, 2002; Ye et al., 2007; Bhat et al., 2016b; Mishra et al., 2017). In most, if not all, of these studies, empirical data sets are used to estimate alternative causal structures and the specification that offers the best statistical fit is considered the dominant decision-making pattern prevailing in the population, although the existence of plausible alternative structures is readily acknowledged (e.g., Ye et al., 2007). In some instances, a two-way or bi-directional causal relationship may exist among dependent variables of interest. This has led to the estimation of structural equations models that are capable of capturing such multi-way relationships (e.g., Cao et al., 2007; Lu and Pas, 1999; Golob, 2000); appropriate restrictions may be imposed to ensure that the parameters in the model system are identified (i.e., can be estimated).

Simultaneous equations models often include error covariance structures that account for the presence of correlated unobserved attributes that affect multiple endogenous variables.

However, model systems depicting bi-directional causal relationships between dependent variables can only be estimated when all endogenous variables are continuous (Pendyala and Bhat, 2004). In situations where one or more of the endogenous variables is a discrete nominal, ordinal, or count variable, simultaneous equations model systems reflecting bi-directional causal relations are not logically consistent. To be logically consistent, restrictions that render the model systems recursive in structure must be imposed. In other words, simultaneous equation model systems that include non-continuous endogenous variables must be specified such that there is a recursive decision structure implied by the model system (Maddala, 1983). This has inevitably led to comparing alternative model structures with respect to goodness-of-fit measures and then using the one with the best fit as the single structure driving travel demand forecasts (Ye et al., 2007). As mentioned earlier, this ignores the existence of multiple plausible decision-making mechanisms that may be driving travel choices in the population.

In this paper, a latent segmentation-based approach is adopted to accommodate the possibility that multiple decision structures may be prevalent in a population and to identify the population subgroups that follow each of the structures. Earlier work (Waddell et al., 2007) utilizing this approach examined the relationship between residential location choice and work location choice. It was posited that some may choose work location based on residential location while others may choose residential location based on work location. A latent segmentation approach was adopted to identify the prevalence of these causal structures in the population and identify the subgroups associated with each structure. However, in that study, the choice variable that appeared first in the decision hierarchy was treated as an exogenous or independent variable rather than an endogenous variable. In other words, each causal structure was depicted, for simplicity, as a single equation with one of the two choices appearing as an exogenous variable in the equation predicting the other choice. In this study, all choice variables are treated as endogenous variables, including the one that is not influenced by any of the other endogenous variables. Thus, the proposed model in this paper constitutes a true simultaneous equations model system that accounts for error covariances and reveals latent segments in the population following alternative decision-making structures.

More recently, Angueira et al. (2017) adopted a latent segmentation-based modeling approach to study the two-way relationship between vehicle type choice and daily distance traveled using a discrete-continuous joint modeling framework. They report that the structure where vehicle type choice influences daily distance traveled explains the vehicle usage patterns of 89 percent of the individuals in the data set used for model estimation, leaving a minor but non-negligible portion (11 percent) of the sample's vehicle usage behavior explained by the alternate causal structure (where distance traveled influences vehicle type choice).

The empirical context considered in this paper includes three choice variables of interest: residential location (area type) choice, vehicle ownership choice, and frequency of mobility service usage. All three variables are treated as multinomial discrete choice variables, thus necessitating the use of recursive structures to depict relationships among them. In recent work, Mishra et al. (2017) analyzed the two-way relationship between vehicle ownership and frequency of car share usage using a structural equations modeling approach. Their model specification treated both endogenous variables as continuous variables, thus enabling the estimation of the bi-directional causal relationship. As a significant extension to that work, this paper adds a third choice dimension (residential location choice) and treats all endogenous variables as multinomial discrete choices. The latent segmentation based approach allows the identification of population subgroups that follow each of the different decision structures.

Our study attempts to unravel the presence of alternative causal relationships among the three endogenous choice variables of interest using a cross-sectional survey data set that provides measurements only at one point in time. However, the notion of causality is often associated with observations of change in behavior over time, followed by an attribution of those changes to temporal changes in exogenous variables that may have contributed to the temporal behavior changes. Longitudinal data that tracks behaviors and exogenous factors over a period of time is needed to identify such cause-and-effect relationships – where a behavior change precedes or occurs subsequent to a change in exogenous (or other endogenous) variables. In using a regular cross-sectional data set to identify alternative plausible causal structures prevalent in the population, this study focuses on the notion of contemporaneous causation. In other words, the modeling effort is applicable to situations where individuals are making choices at a specific point in time based on the context, situation, and circumstances that prevail at that instant in time. In our empirical context, while it may seem that the time-scales of residential location choice and vehicle ownership choice are different from the time-scale corresponding to the choice of the daily or weekly use of shared mobility services, we posit that these time-scales correspond more to the window in which a prior joint choice made for these three decisions is exercised, but that the three choices themselves define a mobility lifestyle choice bundle determined by the individual jointly and contemporaneously. In other words, an individual is determining the kind of lifestyle that he or she would like to lead, and that entails making choices about these three aspects simultaneously. There is no temporal lag in the choice of these three entities, and hence longitudinal data is not needed to identify the causal relationships that exist among these choice dimensions. For example, an individual may choose to live a car-free lifestyle in a dense multimodal urban environment and use ride-hailing services when needed, as a lifestyle (mobility bundle) choice. The choice dimensions are therefore modeled simultaneously while recognizing endogeneity that arises from the presence of error covariances due to common unobserved attributes that simultaneously affect all three dependent variables of interest. The relationships depicted in the causal structures do not constitute a sequential decision-making process; rather they depict a contemporaneous decision-making process of a lifestyle bundle, albeit with distinct relationships among the choice dimensions. In other words, even within a contemporaneous causal context, it is entirely plausible to recognize the presence of different causal structures that lead to the chosen lifestyle bundle.

In summary, this study treats mobility choices as the elements of a lifestyle, with individuals exercising choices about residential location, vehicle ownership, and ride-hailing service use as a joint and collective bundle. However, within this contemporaneous causation framework, there may be alternative decision structures in terms of how the different dimensions influence one another. Recognizing that there may be alternative decision structures adopted by different segments in the population, this study aims to utilize the latent segmentation approach to tease out and identify the latent market segments and the different causal decision structures that they follow.

3. DATA AND SAMPLE DESCRIPTION

The data for this study is derived from the 2014 Who's On Board Mobility Attitudes Survey, an online survey conducted by Resource Systems Group, Inc. (Transit Center, 2017) in 46 metropolitan areas across the country. A total of 11,842 individuals responded to the survey; after removing observations with extensive missing data, there were 11,428 observations remaining in the data set. Table 1 shows the socio-economic and demographic characteristics of the sample. In general, the sample shows a reasonable distribution of various characteristics. Among person

characteristics, the sample has a slightly larger proportion of females and a mix of different generations represented -- with millennials and baby boomers exhibiting larger shares, consistent with their prevalence in the general population.

TABLE 1 Socio-Economic and Demographic Characteristics of the Sample

Person Characteristics (N=11,275)		Household Characteristics (N=11,275)	
Variable	Value	Variable	Value
Gender		Annual Household Income	
% Female	53.6%	< \$25K	10.7%
Age Group		\$25K to < \$35K	9.8%
Millennials (1980-1996)	36.9%	\$35K to < \$50K	14.2%
Generation X (1965-1979)	13.8%	\$50K to < \$75K	23.1%
Baby Boomers (1946-1964)	38.8%	\$75K to <\$100K	18.4%
Silent Generation (before 1946)	10.5%	\$100K or more	23.8%
Educational Attainment		Presence of Children	
High School or Less	17.2%	Child 0-4 years old	7.6%
Some College	31.5%	Child 5-15 years old	10.4%
College Graduate	33.7%	Geographical Region	
Any Graduate School	17.6%	Northeast	15.4%
Race/Ethnicity		South	19.9%
White	84.7%	West/Southwest	19.7%
Hispanic	7.8%	West Coast	21.0%
Other race/ethnicity	7.5%	Midwest	24.0%
Employment Status		Transit Richness	
Full-time Employed	38.7%	Transit Deficient City	39.8%
Part-time Employed	12.3%	Transit Progressive City	60.2%
Not Employed	51.0%	Residential Location Type	
Transit Use Frequency		Urban Mix/Residential	24.0%
Frequent: ≥ Once per Week	13.6%	Suburban/Small Town Mix	32.3%
Infrequent: < Once per Week	26.9%	Suburban/Small Town Residential	43.7%
Never	59.5%	Vehicle Availability	
Time Spent Online		Zero-vehicle or Vehicle-Deficient	13.4%
Always Online (>8 hours/day)	17.2%	Vehicle Sufficient	86.6%
Often Online (4-8 hours/day)	33.0%	<i>Note: 153 individuals residing in households with no drivers and no workers were removed from the sample. Shared Mobility Services refer to car-share service and taxi-car (e.g., Uber/Lyft) services.</i>	
Sometimes Online (<4 hours/day)	49.8%		
Use of Shared Mobility Services			
Frequent (Multiple Times per Week)	13.7%		
Rare (Sometimes per Year)	25.3%		
Never	61.0%		

About one-half of the sample is not employed and a large majority are white and non-Hispanic. Nearly 60 percent of the sample indicated that they never use transit services and nearly one-half of the sample spends less than four hours per day online. Among household characteristics, nearly 24 percent earn \$100,000 or more per year, indicating a high level of income among respondent households. Only 10 percent of respondents indicated that they resided in a household with a child between the ages of 5 and 15 years. Households are drawn from across the country and most respondents (60 percent) indicate that they reside in transit-progressive cities.

Three dependent variables are considered in this study. Residential location is categorized based on area type, with 43.7 percent of respondents indicating that they reside in *suburban or small town residential neighborhoods*. Another 32.3 percent of respondents reside in *mixed neighborhoods in suburban or small town areas*, while the remaining 24 percent are *urban dwellers (mixed or purely residential neighborhoods)*. Vehicle availability is defined as the ratio of number of vehicles to number of workers (if number of workers is equal to zero, then number of drivers is used instead; 153 individuals residing in households with no workers and no drivers were excluded from the analysis). Due to the extremely small number of individuals in zero-vehicle households, they had to be combined with individuals in *vehicle-deficient* households (less than one vehicle per worker or driver). The vast majority (86.6 percent) reside in households that are *vehicle sufficient*, i.e., there is at least one vehicle for every worker or driver. Finally, frequency of usage of shared mobility services depicts a distribution in which a majority (60 percent) *never use* car-share or taxi-car services. About one-quarter of respondents are *occasional users* while 13.7 percent are reasonably *frequent users* of such services.

In summary, the three dependent variables comprise a three-category residential location type choice (RLC), a binary vehicle availability choice (VEH), and a three-category shared mobility usage choice (SVC). These three variables may be related in six possible different recursive causal structures. While all six causal structures are plausible, the two structures in which shared mobility usage affects the other two endogenous choice variables are excluded from consideration in this paper. Although the causal decision processes in which ride-hailing service usage drives residential location and vehicle ownership choices are not completely unreasonable (and there is anecdotal evidence that the availability of and proclivity to use ride-hailing services is motivating certain residential location and vehicle ownership choices), it is rather unlikely that these types of decision structures are prevalent in the population to any substantial degree. In general, the mode share for and frequency of use of ride-hailing services is very small when compared with other more traditional modes of transportation, whereas if ride-hailing service usage were a driver of other choice dimensions (for a substantial portion of the population), then the mode share for this service would presumably be higher. With the passage of time and the maturation of these services, it is plausible that these decision processes will see greater prevalence – in which case the analysis would need to be revised to accommodate these two additional segments. However, another reason for excluding these causal structures from the analysis is the desire to reduce the computational complexity and burden associated with estimating latent segmentation models. As the number of segments increases, the computational complexity and burden increases as well. Inclusion of these two causal decision structures would have increased the number of possible segments to six, requiring additional step-wise estimation procedures. In the absence of any compelling evidence (in the literature) that these causal decision structures are prevalent to any significant degree, it was felt that these two causal decision structures could be eliminated from among the candidate in the interest of computational efficiency.

The four structures considered in this paper may then be depicted as shown below.

Structure 1:	RLC	Structure 3:	VEH
	VEH \leftarrow RLC		RLC \leftarrow VEH
	SVC \leftarrow RLC + VEH		SVC \leftarrow RLC + VEH
Structure 2:	RLC	Structure 4:	VEH
	SVC \leftarrow RLC		SVC \leftarrow VEH
	VEH \leftarrow RLC + SVC		RLC \leftarrow VEH + SVC

The four structures have reasonably intuitive interpretations. Note that all structures represent joint model systems in that the three dependent variables are modeled as a choice bundle by appropriately considering utility error covariances among alternatives within each dimension as well as across dimensions. Thus, an environmentally conscious (say, an unobserved variable) individual may choose to live in an urban neighborhood, be car-free, and use shared mobility services frequently. These kinds of effects are considered in every structure, and hence all structures represent joint model systems in which choices are made as a package or bundle. The directionality of effects being tested correspond to those within the joint package, with all three variables being considered as endogenous variables. Thus, the first structure should not necessarily be misconstrued as a system in which residential choice (RLC) is determined first, then vehicle availability (VEH) is determined second, and finally shared mobility usage (SVC) is determined last. Rather, the system should be viewed as representing a choice bundle in which, in addition to (i.e., after controlling for) the error covariations across the many dimensions, RLC impacts VEH, and both impact SVC. The same holds for all other structures. In this paper, the terms “recursive” and “causal” are used to refer to these directional effects within a joint model, and should not be confused with the term “sequential” that has a connotation of one choice completely preceding the other. Models representing sequential decision-making structures do not account for error covariances that reflect the presence of common unobserved attributes that simultaneously affect multiple choice dimensions. Moreover, such sequential model systems are estimated one equation or choice dimension at a time. Thus, sequential model structures may be considered a special case of the more general simultaneous equations modeling approach adopted in this paper. Regardless of the desire to model multiple behavioral choices simultaneously or sequentially, the latent segmentation approach presented in this paper offers insights on the prevalence of multiple decision structures in the population of interest.

In this paper, the behavioral unit of analysis is the person. The survey data set comprises cases where only one (random) individual in a household participated in and responded to the survey. So, the data set does not include responses from multiple individuals in the same household. As such, the unit of analysis is the person, and the model may be considered a person-level model. While vehicle ownership and residential location largely constitute household-level choices, and ride-hailing service usage is largely an individual person-level choice, these choices are all considered as elements of an individual’s mobility lifestyle bundle. This is consistent with the notion that person-level choices are made within a larger and broader household context, with persons in the household interacting and negotiating with one another to make household level choices. In other words, household level choices are nothing but a reflection of the collective desires (and compromises and trade-offs) of individuals that reside in the household. So, if a household has chosen to reside in a dense urban built environment, then it may be surmised that all of the persons (adults) residing in the household have made that choice. In summary, the modeling exercise presented in this paper uses the individual as the behavioral unit of analysis, with the explicit recognition that household level choices are nothing but manifestations of the collective decisions of individuals within the household.

4. MODELING METHODOLOGY

This section presents a detailed description of the model formulation adopted in this paper. Consider an individual q ($q=1, 2, 3, \dots, Q$) facing a multi-dimensional array of nominal (unordered-response) choices. Let the cardinality of the multi-dimensional array be G (that is, there are G

nominal (unordered-response) variables in the choice space), and let g be the index for the nominal variables ($g = 1, 2, 3, \dots, G$). In the empirical context of the current paper, $G=3$ (the nominal variables are residential location, vehicle availability, and extent of use of car-share service and taxi-car service). Also, let I_g ($I_g \geq 2$) be the number of alternatives corresponding to the g^{th} nominal variable and let i_g be the corresponding index ($i_g = 1, 2, 3, \dots, I_g$). Using a typical utility maximizing framework for the nominal variables, the utility for alternative i_g for the g^{th} nominal variable, given that individual q belongs to segment h , can be written as:

$$U_{qhg i_g} = U_{qg i_g} \mid (\text{individual } q \in \text{segment } h) = \beta'_{hg} \mathbf{x}_{qg i_g} + \varepsilon_{qhg i_g}, \quad (1)$$

where $\mathbf{x}_{qg i_g}$ is a $(K_g \times 1)$ -column vector of exogenous attributes as well as possibly the observed values of other endogenous nominal variables (introduced as dummy variables). β_{hg} is a $(K_g \times 1)$ -column vector of corresponding coefficients. Note that some elements of β_{hg} can be zero for some of the exogenous variables, indicating that the corresponding exogenous variables do not impact choice-making in segment h . Further, because latent segmentation is used as a way to introduce, across the segments, heterogeneity in the recursive effects among the endogenous variables, β_{hg} will necessarily be zero on some of the endogenous variables within each segment.

For example, assume that the first latent segment corresponds to the situation where vehicle availability impacts the extent of use of car-share services, but not the other way around. And let the second latent segment correspond to the reverse situation where the extent of use of car-share services impacts vehicle availability (ownership). Then, in the first segment, the elements of β_{hg} corresponding to the vehicle availability dummy variables in the car-share alternative utilities will be non-zero, while, in the second segment, these same elements will be zero. $\varepsilon_{qhg i_g}$ in Equation 1 is a segment-specific normal scalar error term. Let the variance-covariance matrix of the vertically stacked vector of errors $\boldsymbol{\varepsilon}_{qhg} = [(\varepsilon_{qhg 1}, \varepsilon_{qhg 2}, \dots, \varepsilon_{qhg I_g})']$ be Λ_{hg} . The size of $\boldsymbol{\varepsilon}_{qhg}$ is $(I_g \times 1)$, and the size of Λ_{hg} is $(I_g \times I_g)$. The model above may be written in a more compact form by defining the following vectors and matrices: $\mathbf{U}_{qhg} = (U_{qhg 1}, U_{qhg 2}, \dots, U_{qhg I_g})'$ ($I_g \times 1$ vector), $\mathbf{x}_{qg} = (\mathbf{x}_{qg 1}, \mathbf{x}_{qg 2}, \mathbf{x}_{qg 3}, \dots, \mathbf{x}_{qg I_g})'$ ($I_g \times K_g$ matrix), and $\mathbf{V}_{qhg} = \mathbf{x}_{qg} \beta_{hg}$ ($I_g \times 1$ vector). Then, $\mathbf{U}_{qhg} \sim MVN_{I_g}(\mathbf{V}_{qhg}, \Lambda_{hg})$, where $MVN_{I_g}(\mathbf{V}_{qhg}, \Lambda_{hg})$ is the multivariate normal distribution with mean vector \mathbf{V}_{qhg} and covariance Λ_{hg} . For later use, define the stacked $\vec{G} \times 1$ -vectors:

$$\begin{aligned} \mathbf{U}_{qh} &= (\mathbf{U}'_{qh1}, \mathbf{U}'_{qh2}, \dots, \mathbf{U}'_{qhG}), \\ \mathbf{V}_{qh} &= (\mathbf{V}'_{qh1}, \mathbf{V}'_{qh2}, \dots, \mathbf{V}'_{qhG}), \text{ and} \\ \boldsymbol{\varepsilon}_{qh} &= (\boldsymbol{\varepsilon}'_{qh1}, \boldsymbol{\varepsilon}'_{qh2}, \dots, \boldsymbol{\varepsilon}'_{qhG}), \end{aligned}$$

where $\vec{G} = \sum_{g=1}^G I_g$. Consider now that the individual q chooses alternative m_{qg} for the g^{th} nominal variable. Under the utility maximization paradigm, given that $q \in \text{segment } h$, $(U_{qhg i_g} - U_{qhg m_{qg}})$ must be less than zero for all $i_g \neq m_{qg}$, since the individual chose alternative m_{qg} . Let

$\mathbf{u}_{qhg_i m_{qg}}^* = (U_{qhg_i} - U_{qhg m_{qg}})$ ($i_g \neq m_{qg}$), and stack the latent utility differentials into an $[(I_g - 1) \times 1]$ vector $\mathbf{u}_{qhg}^* = \left(\mathbf{u}_{qhg_1 m_{qg}}^*, \mathbf{u}_{qhg_2 m_{qg}}^*, \dots, \mathbf{u}_{qhg_{I_g} m_{qg}}^* \right)'$, $i_g \neq m_{qg}$.

Define $\tilde{G} = \sum_{g=1}^G (I_g - 1)$. Further, let

$$\tilde{\mathbf{u}}_{qhg}^* = \left(U_{qhg_2} - U_{qhg_1}, U_{qhg_3} - U_{qhg_1}, \dots, U_{qhg_{I_g}} - U_{qhg_1} \right)' \quad [(I_g - 1) \times 1 \text{ vector}],$$

$$\tilde{\mathbf{u}}_{qh}^* = \left([\tilde{\mathbf{u}}_{qh1}^*]', [\tilde{\mathbf{u}}_{qh2}^*]', \dots, [\tilde{\mathbf{u}}_{qhG}^*]' \right)' \quad [(\tilde{G} \times 1) \text{ vector}], \text{ and}$$

$$\mathbf{u}_{qh}^* = \left([\mathbf{u}_{qh1}^*]', [\mathbf{u}_{qh2}^*]', \dots, [\mathbf{u}_{qhG}^*]' \right)' \quad [(\tilde{G} \times 1) \text{ vector}]$$

(so $\tilde{\mathbf{u}}_{qh}^*$ is the vector of utility differences taken with respect to the first alternative for each nominal variable, while \mathbf{u}_{qh}^* is the vector of utility differences taken with respect to the chosen alternative for each nominal variable). Now, for any nominal variable, the full covariance matrix $\mathbf{\Lambda}_{hg}$ of the original error terms in the utilities is not identifiable.

One approach to estimation, typically used in univariate multinomial probit models, is to take the difference of the utilities with respect to the first alternative, and estimate the covariance matrix of these differenced utilities after scaling the first diagonal term to zero. However, in the proposed estimation procedure, what is needed is the covariance matrix of the differenced utilities with respect to the chosen alternative. And it should be guaranteed that there is consistency between the above two differenced covariance matrices. The approach to do so is discussed in detail in Bhat (2015). Basically, it is necessary to ensure that the estimable covariance matrix of $\tilde{\mathbf{u}}_{qh}^*$ is consistent with the covariance matrix of \mathbf{u}_{qh}^* used in estimation. This needs some additional notation and discussion, which are omitted in the interest of brevity. Essentially, it is possible to construct an estimable covariance matrix $\mathbf{\Omega}_h^*$ associated with the undifferenced utilities U_{qh} , but

in which only $\sum_{g=1}^G \left(\frac{I_g * (I_g - 1)}{2} - 1 \right)$ elements are identified (see Bhat, 2015).

Let $\boldsymbol{\theta}_h$ be the collection of parameters to be estimated within segment h : $\boldsymbol{\theta}_h = [\boldsymbol{\beta}'_{h1}, \dots, \boldsymbol{\beta}'_{hG}, \text{Vech}(\mathbf{\Omega}_h^*)']'$, where $\text{Vech}(\mathbf{\Omega}_h^*)$ represents the vector of upper triangle elements of the non-zero and non-fixed elements of $\mathbf{\Omega}_h^*$. Construct a contrast matrix \mathbf{M}_q that allows the translation of the mean and covariance of U_{qh} into the mean and covariance of \mathbf{u}_{qh}^* , as discussed in Bhat (2015). Then the likelihood function for the individual q given that he/she belongs to segment h may be written as:

$$L_q(\boldsymbol{\theta}_h) | (q \in \text{segment } h) = F_{\tilde{G}} \left[\mathbf{M}_q \mathbf{V}_{qh}, \mathbf{M}_q \mathbf{\Omega}_h^* \mathbf{M}_q' \right], \quad (2)$$

where $F_{\tilde{G}}(\cdot, \cdot)$ is the $\tilde{G} = \left(\sum_{g=1}^G (I_g - 1) \right)$ -dimensional normal cumulative distribution function.

Of course, the actual assignment of individual q to a specific segment is not observed, but it is possible to attribute a probability π_{qh} ($h = 1, 2, \dots, H$) to individual q belonging to segment h .

The conditions that $0 \leq \pi_{qh} \leq 1$ and $\sum_{h=1}^H \pi_{qh} = 1$ must be met. To enforce these restrictions, following Bhat (1997), the following logit link function is used:

$$\pi_{qh} = \frac{\exp(\boldsymbol{\mu}'_h \mathbf{w}_q)}{\sum_{j=1}^H \exp(\boldsymbol{\mu}'_j \mathbf{w}_q)}, \quad (3)$$

where \mathbf{w}_q is a vector of individual exogenous variables, and $\boldsymbol{\mu}_1 = \mathbf{0}$ serves as a vector identification condition ($\boldsymbol{\mu}_h$ for $h = 2, 3, \dots, H$ is a vector of coefficients on \mathbf{w}_q determining the probabilistic assignment of individuals to segments).

Defining $\boldsymbol{\theta} = [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_H; \boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_H]'$, then the likelihood function for individual q is:

$$L_q(\boldsymbol{\theta}) = \sum_{h=1}^H \pi_{qh} [L_q(\boldsymbol{\theta}_h) | q \in \text{segment } h], \quad (4)$$

and the likelihood function is then given as:

$$L(\boldsymbol{\theta}) = \prod_q L_q(\boldsymbol{\theta}). \quad (5)$$

Typical simulation-based methods to approximate the multivariate normal cumulative distribution function in Equation 2 can become inaccurate and time-consuming. As an alternative, the Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat, 2011) is used, in which the multiple integrals are evaluated using a fast analytic approximation method. The MACML estimator is based solely on univariate and bivariate cumulative normal distribution evaluations, regardless of the dimensionality of integration, which considerably reduces computation time compared to other simulation techniques used to evaluate multidimensional integrals.

5. MODEL ESTIMATION RESULTS

The estimation of model systems that incorporate latent segments can prove to be challenging. To estimate models in this study, starting values for parameter estimates were first obtained by estimating the four different causal structures separately and independently on the sample data set. In other words, four different model systems – each corresponding to a different recursive structure – were first estimated with no latent segmentation. The model development and estimation process then proceeds sequentially, starting with a two-segment model. The starting values for each of the two segments are those derived from the separate model estimations. After the two-segment model is estimated, the three-segment and four-segment models are estimated. In each instance, parameter estimates from the separate model estimations are used as starting values. By estimating models with increasing numbers of segments, it is possible to perform a comparative evaluation of model fit and determine the number of segments that best describes behavioral patterns in the data. The evaluation of model fit was done using the Bayesian Information Criterion (BIC):

$$\text{BIC} = -LL(\boldsymbol{\theta}) + 0.5 \times R \times \ln(Q) \quad (6)$$

The first term on the right side is the negative of the log-likelihood value at convergence; R is the number of parameters estimated and Q is the number of observations (see Allenby, 1990;

Bhat, 1997). As the number of segments, H , increases, the BIC value keeps declining until a point is reached where an increase in H results in an increase in the BIC value. Estimation is terminated at this point and the number of segments corresponding to the lowest value of BIC is considered the appropriate number for H that best describes prevalent causal structures in the data. Based on the Bayesian Information Criterion (BIC), the three-segment model is found to offer the best fit. For this model, the log-likelihood value at convergence is -73,275.5 and, with 261 parameters, the BIC is 74,494.9; the corresponding values for the model with one segment, two segments, and four segments are 74,781.8, 74,585.4, and 74,512.1, respectively, indicating that the three-segment model provides a statistically significant improvement over model specifications with alternative numbers of latent segments. Even though four structures were considered plausible, inclusion of the fourth segment reduced statistical fit; moreover, the expected number of cases in this segment comprised only a small fraction (less than five percent) of the sample. For these reasons, the three-segment model was considered appropriate and the remainder of this section is devoted to a presentation of this model. The estimation results for the three-segment model are presented in Tables 2 through 6. Table 2 presents the results of the logit link function characterizing the probabilistic assignment of individuals to segments (see discussion in Section 5.1). Tables 3 through 5 present the estimates of the independent variable effects and endogenous effects for the multi-dimensional array of endogenous choices, each table corresponding to the results for a specific segment (see discussion in Section 5.2). Table 6 presents the estimates of the variance-covariance matrix of the error differences for each segment (see discussion in Section 5.3)

5.1 Assignment of individuals to latent segments

The model estimation results presented in Table 2 correspond to Equation (3) presented in the modeling methodology section. The key finding is that the first segment has a majority (expected) share of the cases (53.7 percent), as shown at the bottom of the table, which corresponds to a causal structure in which residential location choice affects vehicle availability, and both affect shared mobility service use. This is clearly the predominant pattern among the causal relationships that exist in the sample, and is actually quite consistent with travel demand model structures currently in use (although it should be noted that typical travel demand model structures represent this causal structure in a sequential paradigm, as opposed to within a joint system). In many travel models, land use choices (such as residential and workplace location choice) are treated as long term choices; these location choices influence more medium-term choices such as vehicle ownership and availability; and finally, location choices (built environment attributes) and vehicle availability influence short-term mode use choices (shared mobility service may be considered a mode).

At the same time, the results show that there is considerable structural heterogeneity in the sample. It is found that the second segment accounts for an expected share of 12.3 percent of the sample, while a full third (34 percent) of the sample is expected to fall in the third segment. In other words, 46.3 percent of the sample does not follow the predominant causal structure. Assuming that the same causal structure applies to all individuals in the sample is clearly invalid and would lead to erroneous travel forecasts and inferences of policy impacts. The segmentation model shows that baby boomers are more likely to belong to the third segment, while the silent generation is more likely to belong to the first segment. Frequent transit users are more likely to belong to the second segment. A few geographic differences are discernible, with those in the northeast and south more likely to associate with the second and third segments. The bottom half of Table 2 provides additional insights into the composition of each of the three segments. It can be seen from the right half of the table that the majority of individuals in every demographic group

follow the conventional (from a travel demand modeling standpoint) causal structure of residential location choice affecting vehicle ownership, and both impacting shared mobility service use.

TABLE 2 The Latent Segmentation Model and Characterization of the Three Segments

Segmentation Variable		Segment 1 (base)	Segment 2		Segment 3			
			Coeff.	t-stat	Coeff.	t-stat		
Segment specific constant		-	-1.158	-21.27	-0.619	-17.34		
Age group	Millennial/GenX (base)	-	-	-	-	-		
	Baby boomer	-	-1.178	-10.46	0.292	7.66		
	Silent generation	-	-1.159	-5.87	-0.091	-3.21		
Transit use frequency	Infrequent/non-user (base)	-	-	-	-	-		
	Frequent transit user	-	0.304	1.87	-0.789	-1.90		
Geographic region	West Coast/Midwest (base)	-	-	-	-	-		
	Northeast	-	0.237	1.99	0.113	1.97		
	South	-	0.203	2.05	0.416	1.99		
	West/Southwest	-	-0.214	-1.95	0.146	2.04		
Characteristics of the Three Segments								
Attribute	Category	Percent Within Segment Falling into Subgroup			Percent Within Attribute Belonging to Segment			Overall Sample
		Seg 1	Seg 2	Seg 3	Seg 1	Seg 2	Seg 3	
Age group	Millennial/Gen X	50.0	77.5	42.1	52.9	18.9	28.2	50.7%
	Baby boomer	38.2	17.2	47.6	52.8	5.5	41.7	38.8%
	Silent generation	11.8	5.3	10.3	60.3	6.3	33.4	10.5%
Transit use frequency	Infrequent or non-user	85.0	73.9	93.1	52.8	10.5	36.7	86.4%
	Frequent transit user	15.0	26.1	6.9	59.2	23.6	17.2	13.6%
Geographic region	West Coast/Midwest	47.2	44.8	41.6	56.2	12.3	31.5	45.0%
	Northeast	15.3	19.3	14.0	53.4	15.5	31.1	15.4%
	South	17.5	20.6	23.6	47.1	12.8	40.1	19.9%
	West/Southwest	20.0	15.3	20.8	54.6	9.5	35.9	19.7%
Residential location choice	Suburban/small town residential	43.3	37.3	46.6	53.2	10.5	36.3	43.7%
	Suburban/small town mix	32.4	32.0	32.4	53.7	12.2	34.1	32.3%
	Urban mix or residential	24.3	30.8	21.0	54.4	15.8	29.8	24.0%
Vehicle availability	Vehicle sufficient hhld	86.3	83.5	88.3	53.4	11.9	34.7	86.6%
	Zero/veh deficient hhld	13.7	16.5	11.7	55.1	15.2	29.7	13.4%
Shared mobility service use	Non-user	60.4	53.0	64.7	53.2	10.7	36.1	61.0%
	Rare user	25.4	25.0	25.3	53.8	12.2	34.0	25.3%
	Frequent user	14.2	22.0	10.0	55.3	19.8	24.9	13.7%
Segment size					53.7%	12.3%	34.0%	100.0%
					(6,131)	(1,410)	(3,887)	(11,428)
Notes:								
<ul style="list-style-type: none"> • Segment 1 Causal Structure: RLC → VEH; RLC + VEH → SVC • Segment 2 Causal Structure: RLC → SVC; RLC + SVC → VEH • Segment 3 Causal Structure: VEH → RLC; RLC + VEH → SVC 								

However, some interesting differences emerge. Millennials and GenX individuals, for example, show a greater inclination to be associated with the second segment than other generations do. Specifically, 18.9 percent of millennials belong to the second segment, while only 5.5 percent of baby boomers and 6.3 percent of the silent generation do. In the context of a cross-

sectional data set, generational cohorts may be considered synonymous with corresponding age groups defined by range of birth year (for each generation). However, rather than use age as the descriptors or labels for the different groups, generational cohort names were used to reflect the possibility that there may be structural differences in choice behaviors and preferences between the generations. As millennials age, for example, they would not adopt the coefficient of baby boomers; rather they would continue to retain the coefficient that is specific to their cohort – recognizing that their causal decision structures may be fundamentally different from those of other generations (McDonald, 2015).

Similarly, 23.6 percent of frequent transit users are expected to belong to the second segment, but only 10.5 percent of infrequent transit users are expected to belong to this segment. The same can be said for other choice variables that capture more active and alternative mode lifestyles. For example, 15.2 percent of individuals in zero-vehicle or vehicle-deficient households are expected to fall into the second segment, as opposed to 11.9 percent of individuals in vehicle-sufficient households; and 19.8 percent of frequent shared mobility service users are expected to belong to the second segment, which is considerably higher than the percent of infrequent users or non-users who are expected to fall into this segment.

It is encouraging that the predominant causal structure revealed by the model is that which is commonly adopted in travel demand forecasting models, although systems in practice are based on specifications that estimate the structure sequentially and ignore error covariances. Sequential estimation procedures often lead to incorrect (inconsistent) estimates of the impacts of one endogenous variable on another, as discussed later. The second most predominant structure (third segment) is that corresponding to a pattern where vehicle ownership impacts residential location choice, and both of these affect shared mobility service use. Individuals in this segment are likely to be choosing residential locations that are consistent with (and dependent on) their vehicle ownership proclivities. In particular, those who eschew vehicles in favor of a car-free lifestyle are likely to choose residential locations based on that preference, and use shared mobility services as needed to meet daily activity-travel needs. Thus, the two predominant causal structures (collectively accounting for 87.7 percent of the sample) correspond to those where longer term choices consistent with lifestyle preferences and household needs influence a (daily) mobility choice.

The second segment, which is smaller but certainly not negligible, is one in which residential location choice impacts shared mobility service use, and both of these affect vehicle ownership or availability. This segment has a higher proportion of millennials, transit users, and shared mobility users than other segments. It appears that these individuals choose a residential location type based on their socio-economic and demographic characteristics (similar to the first segment), but the level of vehicle ownership is mediated by their ability to use alternative transportation modes in the location that they choose to live. The findings suggest that these individuals are more inclined to use alternative modes of transportation (than those in other segments), and vehicle ownership levels are based on the extent to which they are able to do so. If interest in urban living and use of shared mobility services continues to grow over time, it is possible that this segment will grow in size in the future – presenting greater levels of uncertainty in travel forecasts.

5.2 Exogenous and endogenous effects within each segment

Tables 3 through 5 present model estimation results corresponding to exogenous as well as endogenous effects embedded within each causal structure. The estimation results presented in

these tables correspond to the formulation presented in Equation (1) in the modeling methodology section. In general, it is found that a number of socio-economic and demographic variables affect all three choice variables of interest considered in this paper, namely, residential location choice, vehicle availability choice, and shared mobility use. However, the effects of exogenous variables on the three choice variables generally do not differ across causal structures (and there is no reason that they should). Exogenous variable effects across all three causal structures show that older individuals are less likely to reside in urban and mixed land use neighborhoods, those with lower levels of education are more likely to be in zero-car or vehicle deficient households, and those belonging to minority ethnic groups are more likely to reside in urban and mixed land use neighborhoods and in zero-car and vehicle deficient households. However, non-white individuals are more likely to report shared mobility service use. Income is a significant determinant of all three choice variables. Clear patterns are seen in the impacts of income; as income increases, individuals are more likely to reside in suburban and small town residential areas as opposed to urban and mixed use neighborhoods, reside in households that are vehicle-sufficient, and exhibit higher levels of shared mobility use. Individuals in households with children are more likely to be in suburban and small town residential neighborhoods, and those residing in transit progressive cities (as defined in the study design) are more likely to exhibit lower levels of vehicle ownership. All of these findings are consistent with expectations.

As expected, the frequency of transit use is strongly associated with residential location, vehicle ownership, and ride-hailing service use. Individuals who are frequent transit users are more likely to reside in urban neighborhoods and own fewer vehicles than infrequent transit users, who themselves are more likely to reside in urban neighborhoods and own fewer vehicles than transit non-users. These results may be attributable to the good transit accessibility in urban areas relative to non-urban areas. Interestingly, our results also show a positive effect of the frequency of transit use and the frequency of shared mobility service use, perhaps because of complementary effects in the usage of these two modes (for example, using shared mobility service to more easily access fixed transit routes) or because both these modes are used as alternative options to increase access to activities by urban dwellers with few vehicles. In the latter context, the treatment of transit use frequency as an exogenous variable is not ideal when modeling a bundle of mobility choices. Transit use frequency is a mobility choice variable and should therefore be treated as an endogenous variable, similar to residential location choice, vehicle ownership choice, and ride-hailing service usage. There may be cyclic or two-way causal relationships between transit use frequency and other choice dimensions of the model, including vehicle ownership choice and residential location choice. For this study, however, transit use frequency has been treated as an exogenous variable to capture the association between transit use and the endogenous choice dimensions considered in this paper. If transit use frequency were treated as an endogenous variable, the study would have entailed modeling causal structures that involve four endogenous variables (as opposed to three). With four endogenous variables, the number of plausible causal structures increases quite substantially; this means that the number of possible latent segments also increases dramatically, resulting in considerable computational complexity and burden. To keep the problem size manageable, the authors limited the model to three endogenous variables. Residential location choice and vehicle ownership choice are classic measures of behavior of great interest to the profession, while the third choice (ride-hail service usage) is of considerable interest in the current context of emerging mobility services. Hence, these three variables were treated as endogenous variables and transit use frequency was considered exogenous. Moreover, transit use frequency may be considered to some degree as an (exogenous) indicator of transit proclivity –

measuring the extent to which an individual is willing to use transit, which in turn affects the endogenous choice dimensions of interest.

TABLE 3 Model Estimation Results – Segment 1 (RLC → VEH; RLC + VEH → SVC)

Variables	Residential location choice (base: suburban and small town residential, or rural area)				Vehicle availability (base: vehicle sufficient HH)		Use of car-share service and taxi-car service (base: non-user)			
	Suburban and small town mix		Urban mix or residential		Zero car or vehicle deficient HH		Rare user		Frequent user	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constant	-0.435	-10.52	-0.547	-5.01	-1.645	-29.18	-0.883	-35.34	-1.501	-37.10
Individual characteristics										
<i>Gender (base: male)</i>										
Female	0.074	5.71	--	--	--	--	--	--	--	--
<i>Age group (base: millennial)</i>										
Generation X	-0.093	-4.06	-0.173	-6.67	-0.086	-4.77	--	--	-0.331	-12.23
Baby boomer	-0.148	-6.07	-0.348	-8.72	-0.086	-4.77	-0.039	-2.35	-0.610	-26.79
Silent generation	-0.148	-6.07	-0.376	-8.15	--	--	--	--	-0.733	-16.09
<i>Educational Attainment (base: any degree)</i>										
High school or less	--	--	--	--	0.176	8.95	-0.208	-9.62	0.150	6.33
Some college	--	--	--	--	--	--	-0.108	-6.76	--	--
<i>Race/ethnicity (base: white and non-Hispanic)</i>										
Non-white	0.115	4.58	0.243	8.09	0.146	7.09	0.046	2.06	0.265	12.02
Hispanic	--	--	0.188	5.52	--	--	--	--	0.191	6.71
<i>Work status (base: unemployed)</i>										
Part-time worker	--	--	--	--	--	--	--	--	0.196	7.05
Full-time worker	--	--	--	--	-0.243	-13.43	--	--	0.171	7.98
<i>Transit use frequency (base: transit non-user)</i>										
Infrequent transit user	0.259	11.09	0.421	10.09	0.183	9.63	1.068	29.34	0.786	17.12
Frequent transit user	0.350	6.43	0.770	9.49	0.715	21.22	0.708	13.19	1.644	27.18
<i>Time spent online (base: sometimes/often online)</i>										
Always online	0.097	5.22	0.148	7.26	--	--	--	--	0.108	5.06
Household demographics										
<i>Household income (base: \$100,000 or more)</i>										
Less than \$25,000	0.434	14.25	0.509	11.99	0.493	19.05	-0.349	-12.61	-0.304	-12.61
\$25,000-\$34,999	0.302	10.27	0.414	10.71	0.285	10.52	-0.276	-9.85	-0.229	-7.03
\$35,000-\$49,999	0.328	12.84	0.402	11.48	0.274	11.43	-0.223	-9.36	-0.229	-7.03
\$50,000-\$74,999	0.222	11.71	0.222	11.71	0.200	9.43	-0.248	-12.20	-0.119	-4.71
\$75,000-\$99,999	0.135	7.69	0.135	7.69	--	--	-0.096	-4.63	--	--
<i>Presence of children (base: no children 15 or younger)</i>										
Presence of children 0-4 y/o	--	--	-0.072	-3.23	--	--	--	--	--	--
Presence of children 5-15 y/o	-0.113	-5.02	-0.126	-5.45	--	--	--	--	--	--
Location characteristics										
<i>Transit richness (base: transit-deficient city)</i>										
Transit-progressive city	--	--	--	--	0.116	6.93	--	--	--	--
Endogenous effects										
<i>Residential location choice (base: suburban and small town residential, or rural area)</i>										
Suburban and small town mix					0.177	9.27	0.177	10.74	0.197	7.03
Urban mix or residential					0.333	15.80	0.200	9.67	0.148	5.56
<i>Vehicle availability (base: vehicle sufficient HH)</i>										
Car-free or vehicle deficient HH							0.142	6.64	0.134	5.42

TABLE 4 Model Estimation Results –Segment 2 (RLC → SVC; RLC + SVC → VEH)

Variables	Residential location choice (base: suburban and small town residential, or rural area)				Vehicle availability (base: vehicle sufficient HH)		Use of car-share service and taxi-car service (base: non-user)			
	Suburban and small town mix		Urban mix or residential		Zero-car or vehicle deficient HH		Rare user		Frequent user	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constant	-0.440	-10.05	-0.677	-4.70	-1.677	29.20	-0.864	-23.94	-1.462	-26.12
Individual characteristics										
<i>Gender (base: male)</i>										
Female	0.073	5.57	--	--	--	--	--	--	--	--
<i>Age group (base: millennial)</i>										
Generation X	-0.095	-4.02	-0.198	-6.26	-0.082	-3.25	--	--	-0.332	-12.09
Baby boomer	-0.151	-5.97	-0.431	-7.74	-0.082	-3.25	-0.040	-2.42	-0.603	-24.45
Silent generation	-0.151	-5.97	-0.431	-7.74	--	--	--	--	-0.715	-15.12
<i>Educational Attainment (base: any degree)</i>										
High school or less	--	--	--	--	0.177	8.97	-0.190	-8.75	0.153	6.28
Some college	--	--	--	--	--	--	-0.104	-6.58	--	--
<i>Race/ethnicity (base: white and non-Hispanic)</i>										
Non-white	0.119	4.64	0.273	7.31	0.137	6.63	0.060	2.69	0.265	11.99
Hispanic	--	--	0.222	5.08	--	--	--	--	0.189	6.62
<i>Work status (base: unemployed)</i>										
Part-time worker	--	--	--	--	--	--	--	--	0.201	7.26
Full-time worker	--	--	--	--	-0.247	-13.62	--	--	0.167	7.83
<i>Transit use frequency (base: transit non-user)</i>										
Infrequent transit user	0.265	10.81	0.477	8.83	0.127	6.19	1.090	22.17	0.839	20.57
Frequent transit user	0.357	6.26	0.882	8.21	0.641	24.95	0.797	15.00	1.688	22.29
<i>Time spent online (base: sometimes/often online)</i>										
Always online	0.099	5.19	0.165	6.79	--	--	--	--	0.104	4.90
Household demographics										
<i>Household income (base: \$100,000 or more)</i>										
Less than \$25,000	0.445	14.28	0.558	11.07	0.508	19.54	-0.338	-12.43	-0.300	-9.31
\$25,000-\$34,999	0.310	10.35	0.442	9.84	0.298	10.97	-0.274	-9.84	-0.245	-7.30
\$35,000-\$49,999	0.337	12.94	0.442	9.84	0.298	10.97	-0.221	-9.36	-0.221	-9.36
\$50,000-\$74,999	0.234	12.14	0.234	12.14	0.209	9.83	-0.242	-12.00	-0.123	-4.98
\$75,000-\$99,999	0.141	7.78	0.141	7.78	--	--	-0.095	-4.62	--	--
<i>Presence of children (base: no children 15 or younger)</i>										
Presence of children 0-4 y/o	--	--	-0.085	-3.14	--	--	--	--	--	--
Presence of children 5-15 y/o	-0.115	-5.02	-0.142	-5.23	--	--	--	--	--	--
Location characteristics										
<i>Transit richness (base: transit-deficient city)</i>										
Transit-progressive city	--	--	--	--	0.115	6.86	--	--	--	--
Endogenous effects										
<i>Residential location choice (base: suburban/small town residential or rural area)</i>										
Suburban and small town mix					0.169	8.81	0.168	10.30	0.196	7.06
Urban mix or residential					0.321	15.17	0.203	9.94	0.174	4.57
<i>Use of car-share service and taxi-car service (base: non-user)</i>										
Rare user					0.137	6.56				
Frequent user					0.137	6.56				

TABLE 5 Model Estimation Results – Segment 3 (VEH → RLC; RLC + VEH → SVC)

Variables	Residential location choice (base: suburban and small town residential, or rural area)				Vehicle availability (base: vehicle sufficient HH)		Use of car-share service and taxi-car service (base: non-user)			
	Suburban and small town mix		Urban mix or residential		Zero-car or vehicle deficient HH		Rare user		Frequent user	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Constant	-0.440	-10.50	-0.614	-4.85	-1.550	-29.83	-0.874	-25.36	-1.530	-23.70
Individual characteristics										
<i>Gender (base: male)</i>										
Female	0.076	5.74	--	--	--	--	--	--	--	--
<i>Age group (base: millennial)</i>										
Generation X	-0.096	-4.13	-0.184	-6.52	-0.102	-3.93	--	--	-0.353	-12.79
Baby boomer	-0.155	-6.02	-0.376	-8.27	-0.102	-3.93	-0.042	-2.54	-0.622	-26.47
Silent generation	-0.155	-6.02	-0.423	-7.97	--	--	--	--	-0.792	-17.22
<i>Educational Attainment (base: any degree)</i>										
High school or less	--	--	--	--	0.178	9.02	-0.213	-9.89	0.164	6.79
Some college	--	--	--	--	--	--	-0.110	-6.82	--	--
<i>Race/ethnicity (base: white and non-Hispanic)</i>										
Non-white	0.108	4.33	0.245	7.52	0.168	8.16	--	--	0.251	11.19
Hispanic	--	--	0.198	5.21	0.091	3.36	--	--	0.186	6.37
<i>Work status (base: unemployed)</i>										
Part-time worker	--	--	--	--	--	--	--	--	0.184	6.88
Full-time worker	--	--	--	--	--	--	--	--	0.184	6.88
<i>Transit use frequency (base: transit non-user)</i>										
Infrequent transit user	0.258	11.08	0.436	9.51	0.227	12.11	1.050	26.21	0.728	16.76
Frequent transit user	0.328	6.29	0.766	8.78	0.793	26.15	0.634	11.85	1.604	21.25
<i>Time spent online (base: sometimes/often online)</i>										
Always online	0.096	5.10	0.151	6.94	--	--	--	--	0.100	4.59
Household demographics										
<i>Household income (base: \$100,000 or more)</i>										
Less than \$25,000	0.419	14.13	0.494	11.65	0.540	20.96	-0.337	-12.12	-0.276	-8.15
\$25,000-\$34,999	0.297	10.23	0.416	10.29	0.313	11.98	-0.269	-9.57	-0.210	-6.29
\$35,000-\$49,999	0.324	12.83	0.406	11.11	0.313	11.98	-0.217	-9.07	-0.210	-6.29
\$50,000-\$74,999	0.221	11.71	0.221	11.71	0.213	10.13	-0.248	-12.17	-0.094	-3.69
\$75,000-\$99,999	0.139	7.74	0.139	7.74	--	--	-0.104	-4.94	--	--
<i>Presence of children (base: no children 15 or younger)</i>										
Presence of children 0-4 y/o	--	--	-0.074	-3.07	--	--	--	--	--	--
Presence of children 5-15 y/o	-0.113	-4.99	-0.130	-5.27	--	--	--	--	--	--
Location characteristics										
<i>Transit richness (base: transit-deficient city)</i>										
Transit-progressive city	--	--	--	--	0.123	7.37	--	--	--	--
Endogenous effects										
<i>Residential location choice (base: suburban and small town residential, or rural area)</i>										
Suburban and small town mix							0.163	9.88	0.189	6.88
Urban mix or residential							0.174	8.46	0.159	4.36
<i>Vehicle availability (base: vehicle sufficient HH)</i>										
Car-free or vehicle deficient HH	0.177	7.24	0.292	8.51			0.145	6.74	0.145	6.74
Goodness-of-Fit Statistics for the Three-Segment Model:										
Log likelihood at convergence, L(θ) = -73,275.50 (261 parameters)										
Log likelihood with constants, L(c) = -93,691.20										
Log likelihood with no constants, L(0) = -101,529.66										
Adjusted ρ ² (c) = 0.215524										
Adjusted ρ ² (0) = 0.275857										

With respect to endogenous effects (presented towards the end of Tables 3 through 5), it can be seen that model estimation results offer behaviorally intuitive interpretations. In the first segment (RLC \rightarrow VEH; RLC + VEH \rightarrow SVC), residing in denser, mixed, and urban neighborhoods contributes to lower levels of vehicle ownership and higher levels of shared mobility use. Lower levels of vehicle ownership contribute to higher levels of shared mobility use. In the second segment (RLC \rightarrow SVC; RLC + SVC \rightarrow VEH), residential location effects are similar to those in the first segment; and shared mobility usage is associated with residing in zero-car or vehicle-deficient households. In the third segment (VEH \rightarrow RLC; RLC + VEH \rightarrow SVC), a zero-car or vehicle-deficient lifestyle choice is associated with living in urban and mixed neighborhoods and greater usage of shared mobility services. Residing in urban and mixed neighborhoods, in turn, contributes to higher level of shared mobility usage as well.

An important note is that the endogenous effects in Tables 3 through 5 control for jointness through error covariations. For example, while the first latent segment may seem to be consistent with traditional travel demand modeling procedures, this is not actually so because traditional procedures consider the structure to be a sequential hierarchy of choices as opposed to a choice bundle with correlated unobserved attributes that influence all choice dimensions simultaneously. In many survey data sets that do not include information about individual attitudes or lifestyle preferences, there are common individual-specific unobserved factors (for example, being environmentally conscious) that simultaneously and intrinsically increase the likelihood of residing in mixed urban environments as well as being car-free or vehicle deficient and using shared-ride services frequently. If these unobserved error covariances are ignored, the positive effect that an urban mixed environment has on an individual being car-free gets over-estimated as does the positive effect on using car-sharing and taxi services frequently. This is the classic case of self-selection in residential choice, as discussed extensively in the literature (see, for example, Bhat and Guo, 2007).

5.3 Error covariance matrix

A number of error covariances are found to be statistically significant, suggesting that there are correlated unobserved attributes that simultaneously affect choice behaviors. The error covariance matrix, depicting statistically significant error correlations, is presented in Table 6. Significant error covariances are found between residential locations of urban mix/residential and suburban/small town mix, suggesting that there are correlated unobserved attributes (such as innate preferences for an active lifestyle) that simultaneously affect both of these choice alternatives. Similarly, a significant error covariance exists between rare and frequent usage of shared mobility services, potentially suggesting that there are other correlated factors, such as being comfortable with digital and mobile technology (Lavieri et al., 2017), that influence usage of such services. Another significant error covariance suggests that individuals with active lifestyle preferences (unobserved attributes) may be simultaneously inclined to reside in mixed and urban neighborhoods and use shared mobility services, reducing dependence on vehicles. An interesting finding is that the error covariance between residing in a zero-car or vehicle-deficient household and using shared mobility services is significant only for the third segment (VEH \rightarrow RLC; RLC + VEH \rightarrow SVC). This is consistent with the finding that individuals in this segment exhibit the lowest levels of transit use and highest levels of vehicle ownership. The unobserved attributes that contribute to their highly auto-oriented lifestyle are also likely to diminish their use of shared mobility services.

TABLE 6 Variance-Covariance Matrix of the Error Differences for Each Segment

Error differences		Residential location choice (base: suburban and small town residential, or rural area)		Vehicle availability (base: vehicle sufficient HH)	Use of car-share service and taxi-car service (base: non-user)	
		Suburban and small town mix	Urban mix or residential	Zero-car or vehicle deficient HH	Rare user	Frequent user
First segment	Residential location choice (base: suburban and small town residential, or rural area) Suburban and small town mix Urban mix or residential	1.000*	0.496 0.747	0.000* 0.000*	0.000* 0.000*	0.000* 0.241
	Vehicle availability (base: vehicle sufficient HH) Car-free or vehicle deficient HH			1.000*	0.000*	0.000*
	Use of car-share service and taxi-car service (base: Non-user) Rare user Frequent user				1.000*	0.152 1.000*
Second segment	Residential location choice (base: suburban and small town residential, or rural area) Suburban and small town mix Urban mix or residential	1.000*	0.494 0.926	0.000* 0.000*	0.000* 0.000*	0.000* 0.245
	Vehicle availability (base: vehicle sufficient HH) Zero-car or vehicle deficient			1.000*	0.000*	0.129
	Use of car-share service and taxi-car service (base: Non-user) Rare user Frequent user				1.000*	0.213 1.000*
Third segment	Residential location choice (base: suburban and small town residential, or rural area) Suburban and small town mix Urban mix or residential	1.000*	0.508 0.808	0.000* 0.000*	0.000* 0.000*	0.000* 0.252
	Vehicle availability (base: vehicle sufficient HH) Zero-car or vehicle deficient			1.000*	0.000*	0.000*
	Use of car-share service and taxi-car service (base: Non-user) Rare user Frequent user				1.000*	-0.104 1.000*

*matrix element was fixed during the estimation because it was not statistically significantly different from the fixed value at even the 30% confidence level, or it was fixed because of identification issues.

Note: All matrix entries that were estimated are significant at the 5% level or better.

6. DISCUSSION AND CONCLUSIONS

This study is motivated by the increasing recognition that a single decision structure representing the nature of relationships among multiple activity-travel choice variables may not be appropriate for modeling and forecasting travel demand. The potential presence of structural heterogeneity, where a multiplicity of causal decision structures drive activity-travel choices for different subgroups of the population, calls for the development of models capable of accounting for such heterogeneity. Although it is possible to potentially enumerate plausible causal decision structures linking multiple endogenous variables a priori, the identity of population subgroups is not known

a priori. Moreover, two-way bi-directional relationships cannot be estimated (because of identification and logical consistency issues) when the dependent variables of interest are not continuous.

This paper addresses this need by developing a simultaneous equations model system with latent segmentation such that distinct population subgroups following different causal structures can be explicitly identified. Using a large national mobility survey data set collected in the United States in 2014, the study develops a model system linking residential location (area type) choice, vehicle ownership choice, and shared mobility service (car-share and taxi-car services) use. All three variables are treated as discrete multinomial choice variables. Four distinct causal structures are considered within this study; two structures in which residential location choice impacts the other two choices, and two structures in which vehicle ownership impacts the other two choices, all within a joint modeling system. Model estimation results show that the data is best described by three latent segments. It is found that just over one-half of the sample are expected to fall into the segment where the causal structure implies that residential location choice affects vehicle ownership, both of which then impact shared mobility service use. With just under one-half of the sample classified into two other structures, it is clear that the sample depicts significant structural heterogeneity in activity-travel decision-making. In addition, significant error covariances were found, suggesting that there are correlated unobserved attributes (such as attitudes) that simultaneously affect multiple choice behaviors.

The causal decision structure that did not show substantial presence in the sample is the one where vehicle ownership impacts shared mobility service use, both of which then influence residential location choice. This is quite reasonable; residential location choice is generally a long term choice, entails considerable expense, and is influenced by a number of other considerations (such as pricing, building stock, parks and recreation, crime, school quality, and distance to work). Thus it is unlikely that a causal structure in which residential location choice does not influence either vehicle ownership or shared mobility usage would be present (to any substantial degree) in the population. The findings also reveal that residential self-selection effects are present and significant for a substantial fraction of the population; for one of the latent segments, vehicle ownership choice drives residential location choice, suggesting that these individuals choose a residential area type that is consistent with their auto ownership/use proclivities.

The study results suggest that activity-travel demand model systems would benefit from the ability to reflect the presence of structural heterogeneity in causal decision structures prevalent in the population. The model system developed in this study offers a latent segmentation approach that allows the probabilistic allocation of individuals or households into distinct groups adopting alternative causal decision structures. The choice behaviors of agents can then be simulated using appropriate model structures depending on the group to which they belong. Through the incorporation of methods such as those offered in this paper, activity-based travel model systems can be enhanced to account for population heterogeneity in behavioral decision-making processes.

Future research efforts could focus on exploring heterogeneity in relationships among different sets of choice variables and testing the stability of findings across geographic contexts. In addition, although this study is motivated by the identifiability restrictions entailed by non-continuous endogenous variables, and the specific method adopted in this paper relies on this case, it should be noted that the need to address structural heterogeneity is also pertinent even when all endogenous variables are continuous and it is feasible to model two-way relationships. Future research should investigate approaches for dealing with structural heterogeneity in those cases as well. Another area of inquiry that merits additional research involves determining the extent to

which heterogeneity in decision-making structures is reduced when heterogeneity in behavior is accommodated through the use of random parameters or latent class approaches. Also, from an estimation perspective, latent segmentation models need a systematic way of building the model using converged parameters from simpler estimations as the starting point. Otherwise, there could be convergence problems because of computational instability in these mixture models (see, for example, a discussion of this issue by Bhat, 1997). Whether using converged values of simpler specifications to kick-start estimations in latent segmentation models puts this model at more or less risk of ending up with a local optimum remains an open question.

ACKNOWLEDGEMENTS

This research was partially supported by the Center for Teaching Old Models New Tricks (TOMNET) as well as the Data-Supported Transportation Operations and Planning (D-STOP) Center, both of which are Tier 1 University Transportation Centers sponsored by the US Department of Transportation. The authors are grateful to Lisa Macias for her help in formatting this document. The authors thank three anonymous reviewers for their valuable comments and input that greatly improved the paper.

Conflict of interest statement: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Authors' contributions:

S Astroza: Literature review, variable specification development, model estimation and coding

VM Garikapati: Literature review, manuscript writing, editing

RM Pendyala: Conceptual development, manuscript writing, variable specification development

CR Bhat: Conceptual development, methodology development, manuscript writing

PL Mokhtarian: Literature synthesis, editing, variable specification development.

REFERENCES

- Angueira, J., K.C. Konduri, V. Chakour, and N. Eluru (2017) Exploring the Relationship Between Vehicle Type Choice and Distance Traveled: A Latent Segmentation Approach. *Transportation Letters*, 1-12.
- Allenby, G. (1990) Hypothesis Testing with Scanner Data: The Advantage of Bayesian Methods. *Journal of Marketing Research*, 27, 379-389.
- Bagley, M.N. and P.L. Mokhtarian (2002) The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. *The Annals of Regional Science*, 36(2), 279-297.
- Bhat, C.R. (1997) An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, 31(1), 34-48.
- Bhat, C.R. (2011) The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. *Transportation Research Part B*, 45(7), 923-939.
- Bhat, C.R. (2015) A New Generalized Heterogeneous Data Model (GHDM) to Jointly Model Mixed Types of Dependent Variables. *Transportation Research Part B*, 79, 50-77.

- Bhat, C.R. and J.Y. Guo (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., S. Astroza, and A.C. Bhat (2016a) On Allowing a General Form for Unobserved Heterogeneity in the Multiple Discrete-Continuous Probit Model: Formulation and Application to Tourism Travel. *Transportation Research Part B*, 86, 223-249.
- Bhat, C.R., S. Astroza, A.C. Bhat, and K. Nagel (2016b) Incorporating a Multiple Discrete-Continuous Outcome in the Generalized Heterogeneous Data Model: Application to Residential Self-Selection Effects Analysis in an Activity Time-use Behavior Model. *Transportation Research Part B*, 91, 52-76.
- Cao, X., P.L. Mokhtarian, and S. Handy (2007) Do Changes in Neighborhood Characteristics Lead to Changes in Travel Behavior? A Structural Equations Modeling Approach. *Transportation*, 34(5), 535-556.
- Garikapati, V.M., R.M. Pendyala, E.A. Morris, P.L. Mokhtarian, and N. McDonald (2016) Activity Patterns, Time Use, and Travel of Millennials: A Generation in Transition? *Transport Reviews*, 36(5), 558-584.
- Golob, T.F. (2000) A Simultaneous Model of Household Activity Participation and Trip Chain Generation. *Transportation Research Part B*, 34(5), 355-376.
- Johansson, M.V., T. Heldt, and P. Johansson (2006) The Effects of Attitudes and Personality Traits on Mode Choice. *Transportation Research Part A*, 40(6), 507-525.
- Krizek, K. and P. Waddell (2002) Analysis of Lifestyle Choices: Neighborhood Type, Travel Patterns, and Activity Participation. *Transportation Research Record: Journal of the Transportation Research Board*, 1807, 119-128.
- Lavieri, P., V.M. Garikapati, C.R. Bhat, and R.M. Pendyala (2017) Investigation of Heterogeneity in Vehicle Ownership and Usage for the Millennial Generation. *Transportation Research Record: Journal of the Transportation Research Board*, 2664, 91-99.
- Lu, X. and E.I. Pas (1999) Socio-demographics, Activity Participation, and Travel Behavior. *Transportation Research Part A*, 33(1), 1-18.
- Maddala, G.S. (1983) *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, MA.
- McDonald, N.C. (2015) Are Millennials Really the “GoNowhere” Generation? *Journal of the American Planning Association*, 81(2), 90-103.
- Mishra, G.S., P.L. Mokhtarian, R.R. Clewlow, and K.F. Widaman (2017) Addressing the Joint Occurrence of Self-Selection and Simultaneity Biases in the Estimation of Program Effects Based on Cross-Sectional Observational Surveys: Case Study of Travel Behavior Effects in Carsharing. *Transportation (in press)*, <https://doi.org/10.1007/s11116-017-9791-1>.
- Pendyala, R.M. (1998) Causal Analysis in Travel Behaviour Research: A Cautionary Note. In J.D. Ortuzar, D. Hensher, and S. Jara-Diaz (eds.) *Travel Behaviour Research: Updating the State of Play*, Elsevier Science Publishers, B.V., The Netherlands, 35-48.
- Pendyala, R.M. and C.R. Bhat (2004) An Exploration of the Relationship between Timing and Duration of Maintenance Activities. *Transportation*, 31(4), 429-456.

- Silva, J.D.A., C. Morency, and K.G. Goulias (2012) Using Structural Equations Modeling to Unravel the Influence of Land Use Patterns on Travel Behavior of Workers in Montreal. *Transportation Research Part A*, 46(8), 1252-1264.
- Transit Center. *Who's On Board: 2014 Mobility Attitudes Survey*. Transit Center, New York, NY. <http://transitcenter.org/wp-content/uploads/2014/08/WhosOnBoard2014-ForWeb.pdf>. Accessed July 15, 2017.
- Waddell, P., C.R. Bhat, N. Eluru, L. Wang, and R.M. Pendyala (2007) Modeling Interdependence in Household Residence and Workplace Choices. *Transportation Research Record: Journal of the Transportation Research Board*, 2003, 84-92.
- Yarlagadda, A.K. and S. Srinivasan (2008) Modeling Children's School Travel Mode and Parental Escort Decisions. *Transportation*, 35(2), 201-218.
- Ye, X., R.M. Pendyala, and G. Gottardi (2007) An Exploration of the Relationship between Mode Choice and Complexity of Trip Chaining Patterns. *Transportation Research Part B*, 41(1), 96-113.

Author Bios

Sebastian Astroza is finishing his PhD in Civil Engineering (Transportation Engineering) at The University of Texas at Austin and just joined the Departamento de Ingeniería Industrial at Universidad de Concepción in Concepcion, Chile. His primary research interests include discrete choice models and travel behavior, spatial econometrics, and discrete-continuous econometric systems. He received his MS degree in Transportation Engineering and his BS degree in Civil Engineering from Universidad de Chile in Santiago, Chile.

Venu Garikapati is a project leader in the Mobility, Behavior and Advanced Powertrains group at the National Renewable Energy Laboratory (NREL). At NREL, he oversees the day-to-day operations of the Transportation Secure Data Center and conducts research on the travel, and energy related impacts of emerging transportation technologies such as connected and automated vehicles, and transportation networking companies. Venu received his Ph.D. in Civil Engineering with a specialization in transportation systems from Arizona State University.

Ram M. Pendyala is a professor of transportation and Director of a University Transportation Center in the School of Sustainable Engineering and the Built Environment at Arizona State University. He specializes in activity-based travel behavior analysis, transportation demand forecasting, and development of agent-based integrated land use-transport microsimulation models. He is past chair of the International Association for Travel Behaviour Research and the Transportation Research Board's Planning and Environment Group, Travel Analysis Methods Section, and Traveler Behavior and Values Committee. He has his PhD in Civil Engineering (Transportation) from the University of California at Davis.

Chandra R. Bhat is the Director of the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 USDOT University Transportation Center and the Joe J. King Chair in Engineering at The University of Texas at Austin. He has contributed toward the development of advanced econometric techniques for travel behavior analysis, in recognition of which he received the 2015 Frank M. Masters Transportation Engineering Award from the American Society of Civil

Engineers (ASCE), and the 2017 Lifetime Achievement in Transportation Research and Education Award (Academic) from the Council of University Transportation Centers (CUTC). He is the current chair of the Transportation Research Board Travel Analysis Methods Section.

Patricia L. Mokhtarian is the Susan G. and Christopher D. Pappas Professor of Civil and Environmental Engineering at the Georgia Institute of Technology. She has specialized in the study of travel behavior for more than 35 years, emphasizing ICT impacts on travel, attitudes toward travel itself, residential self-selection, and multitasking. She is the immediate past chair of the International Association for Travel Behaviour Research.